

# Recent Novel High-Tech Researches in Molecular Biology

Guest Editors: Calvin Yu-Chian Chen, Kuo-Chen Chou, James David Adams,  
Tai-Ping Fan, and Gerhard Litscher





---

# **Recent Novel High-Tech Researches in Molecular Biology**

BioMed Research International

---

## **Recent Novel High-Tech Researches in Molecular Biology**

Guest Editors: Calvin Yu-Chian Chen, Kuo-Chen Chou,  
James David Adams, Tai-Ping Fan, and Gerhard Litscher



---

Copyright © 2015 Hindawi Publishing Corporation. All rights reserved.

This is a special issue published in “BioMed Research International.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

# Contents

**Recent Novel High-Tech Researches in Molecular Biology**, Calvin Yu-Chian Chen, Kuo-Chen Chou, James David Adams, Tai-Ping Fan, and Gerhard Litscher  
Volume 2015, Article ID 749160, 3 pages

**Transgenic Sugarcane Resistant to *Sorghum mosaic virus* Based on Coat Protein Gene Silencing by RNA Interference**, Jinlong Guo, Shiwu Gao, Qinliang Lin, Hengbo Wang, Youxiong Que, and Liping Xu  
Volume 2015, Article ID 861907, 9 pages

**Next-Generation Technologies for Multiomics Approaches Including Interactome Sequencing**, Hiroyuki Ohashi, Mai Hasegawa, Kentaro Wakimoto, and Etsuko Miyamoto-Sato  
Volume 2015, Article ID 104209, 9 pages

**Targeted Next Generation Sequencing Identifies Novel Mutations in *RPI* as a Relatively Common Cause of Autosomal Recessive Rod-Cone Dystrophy**, Said El Shamieh, Elise Boulanger-Scemama, Marie-Elise Lancelot, Aline Antonio, Vanessa Démontant, Christel Condroyer, Mélanie Letexier, Jean-Paul Saraiva, Saddek Mohand-Saïd, José-Alain Sahel, Isabelle Audo, and Christina Zeitz  
Volume 2015, Article ID 485624, 11 pages

**Interaction Analysis through Proteomic Phage Display**, Gustav N. Sundell and Ylva Ivarsson  
Volume 2014, Article ID 176172, 9 pages

**Applications of Flow Cytometry to Characterize Bacterial Physiological Responses**, Verónica Ambriz-Aviña, Jorge A. Contreras-Garduño, and Mario Pedraza-Reyes  
Volume 2014, Article ID 461941, 14 pages

**Female Aging Alters Expression of Human Cumulus Cells Genes that Are Essential for Oocyte Quality**, Tamadir Al-Edani, Said Assou, Alice Ferrières, Sophie Bringer Deutsch, Anna Gala, Charles-Henri Lecellier, Ounissa Aït-Ahmed, and Samir Hamamah  
Volume 2014, Article ID 964614, 10 pages

**Quantitative Analysis of Pork and Chicken Products by Droplet Digital PCR**, Yicun Cai, Xiang Li, Rong Lv, Jieli Yang, Jian Li, Yuping He, and Liangwen Pan  
Volume 2014, Article ID 810209, 6 pages

**Methylation-Associated Gene Silencing of *RARB* in Areca Carcinogens Induced Mouse Oral Squamous Cell Carcinoma**, Zi-Lun Lai, Yung-An Tsou, Shin-Ru Fan, Ming-Hsui Tsai, Hsiao-Ling Chen, Nai-Wen Chang, Ju-Chien Cheng, and Chuan-Mu Chen  
Volume 2014, Article ID 378358, 13 pages

**Osteoponin Promoter Controlled by DNA Methylation: Aberrant Methylation in Cloned Porcine Genome**, Chih-Jie Shen, Yung-An Tsou, Hsiao-Ling Chen, Hung-Jin Huang, Shinn-Chih Wu, Winston T. K. Cheng, Calvin Yu-Chian Chen, and Chuan-Mu Chen  
Volume 2014, Article ID 327538, 16 pages

**Muscle-Type Specific Autophosphorylation of CaMKII Isoforms after Paced Contractions**, Wouter Eilers, Wouter Gevers, Daniëlle van Overbeek, Arnold de Haan, Richard T. Jaspers, Peter A. Hilbers, Natal van Riel, and Martin Flück  
Volume 2014, Article ID 943806, 20 pages

**CELLCOUNTER: Novel Open-Source Software for Counting Cell Migration and Invasion *In Vitro*,**

Xiaoni Li, Hongshun Yang, Hailiang Huang, and Tao Zhu

Volume 2014, Article ID 863564, 6 pages

**Treatment of Cardiovascular Disease by Traditional Chinese Medicine against Pregnane X Receptor,**

Kuen-Bao Chen, Hsin-Yi Chen, Kuan-Chung Chen, and Calvin Yu-Chian Chen

Volume 2014, Article ID 950191, 17 pages

**Evaluation of Correlation of Cell Cycle Proteins and Ki-67 Interaction in Paranasal Sinus Inverted**

**Papilloma Prognosis and Squamous Cell Carcinoma Transformation,** Yung-An Tsou, Hung-Jin Huang,

Tang-Chuan Wang, Chih-Jaan Tai, Chuan-Mu Chen, and Calvin Yu-Chian Chen

Volume 2014, Article ID 634945, 16 pages

## Editorial

# Recent Novel High-Tech Researches in Molecular Biology

**Calvin Yu-Chian Chen,<sup>1,2,3,4</sup> Kuo-Chen Chou,<sup>5,6</sup> James David Adams,<sup>7</sup>  
Tai-Ping Fan,<sup>8</sup> and Gerhard Litscher<sup>3,9</sup>**

<sup>1</sup>Human Genetic Center, Department of Medical Research, China Medical University Hospital, Taichung 40447, Taiwan

<sup>2</sup>Research Center for Chinese Medicine & Acupuncture, China Medical University, Taichung 40402, Taiwan

<sup>3</sup>School of Medicine, College of Medicine, China Medical University, Taichung 40402, Taiwan

<sup>4</sup>Department of Biomedical Informatics, Asia University, Taichung 41354, Taiwan

<sup>5</sup>Gordon Life Science Institute, Boston, MA 02478, USA

<sup>6</sup>Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University, Jeddah 21589, Saudi Arabia

<sup>7</sup>School of Pharmacy, University of Southern California, Los Angeles, CA 90033, USA

<sup>8</sup>Department of Pharmacology, University of Cambridge, Cambridge, UK

<sup>9</sup>Research Unit of Biomedical Engineering in Anesthesia and Intensive Care Medicine, Research Unit for Complementary and Integrative Laser Medicine, and TCM Research Center Graz, Medical University of Graz, 8036 Graz, Austria

Correspondence should be addressed to Calvin Yu-Chian Chen; [ycc929@mit.edu](mailto:ycc929@mit.edu)

Received 7 October 2014; Accepted 7 October 2014

Copyright © 2015 Calvin Yu-Chian Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this special issue recent novel high-tech research in molecular biology is discussed. Since the fast deployment in physics, chemistry, mathematics, and computer science, the recent molecular biology research becomes an evidence-based study even to a single molecular-level study. Thus we call this special issue for novel modern technology including next-generation sequencing methods, proteomics, bioinformatics, genomics, and computational systems biology.

S. El Shamieh et al. reported ophthalmic and genetic findings in families with autosomal recessive rod-cone dystrophy (arRCD) and RPI mutations. Genomic DNA was investigated using a customized next-generation sequencing panel targeting up to 123 genes implicated in inherited retinal disorders. Sequencing identified 9 RPI mutations in 7 index cases. Eight of the mutations were novel. Among these mutations, 4 belong to a region previously associated with arRCD and 5 others in a region previously associated with adRCD. Interestingly, a prevalence of  $\approx 2.5\%$  points out the necessity of sequencing RPI in sporadic and recessive cases of RCD. The authors pointed out that further functional studies would strengthen our knowledge in the physiology of retinal photoreceptors.

The topic of H. Ohashi et al. was “*Next-generation technologies for multiomics approaches including interactome sequencing.*” They outlined a variety of new innovative techniques and discussed their use in omics research (e.g., genomics, transcriptomics, metabolomics, proteomics, and interactomics). The possible applications of these methods in future medical and life science research were also discussed, including an interactome-sequencing technology, developed by the authors.

J. Guo et al. investigated a virus-resistant transgenic sugarcane involving coat protein gene silencing by RNA interference (RNAi), which is a novel strategy for producing viral resistant plants. It can lead to target gene silencing, thus suppressing target gene expression. In this study, the conserved region of coat protein (CP) genes was selected as the target gene. The genetically modified sorghum mosaic virus-resistant lines of cultivar ROC22 provide resistant germplasm for breeding lines and can also serve as resistant lines having the same genetic background for study of resistance mechanisms.

The paper by T. Al-Edani et al. deals with the influence of female aging on human cumulus cells (CCs) genes. There

was a need for an extensive analysis of age impact on transcriptome profile to link oocyte quality and developmental potential with patient's age. CCs from patients of three age groups were analyzed individually using whole genome U133 Plus 2.0 GeneChip Affymetrix microarrays. The authors focused on pathways affected by aging in CCs that may explain the decline of oocyte quality with age. Specific molecular signatures were characterized for the three age categories. It was revealed that the pathways impacted by age were potential targets of specific microRNAs previously identified in CCs small RNAs sequencing.

V. Ambriz-Aviña et al. applied flow cytometry (FCM) to characterize bacterial physiological responses. They reviewed how FCM has been applied to characterize distinct physiological conditions in bacteria including responses to antibiotics and other cytotoxic chemicals and physical factors. Since FCM is suitable for performing studies at the single-cell level, the authors were able to describe how this powerful technique has yielded invaluable information about the heterogeneous distribution of differently and even specialized responding cells and how it may help to provide insights about how cell interaction takes place in complex structures, such as those that prevail in bacterial biofilms.

In the project introduced by Y. Cai et al., a highly precise quantitative method based on the digital polymerase chain reaction (dPCR) technique was developed to determine the weight of pork and chicken in meat products. Currently, real-time quantitative polymerase chain reaction (qPCR) is used for quantitative molecular analysis of the presence of species-specific DNAs in meat products, but it is limited in several aspects. By using the dPCR method, the authors found that the relationships between the raw meat weight and DNA weight and between the DNA weight and DNA copy number were both close to linear. This enabled them to establish formulae to calculate the raw meat weight based on the DNA copy number. The accuracy was verified using samples of pork and chicken powder mixed in known proportions. Quantitative analysis indicated that dPCR is highly precise and therefore has the potential to be used in routine analysis by government regulators and quality control departments, once some technical flaws have been resolved.

G. N. Sundell and Y. Ivarsson investigated interaction analysis through proteomic phage display, which is a powerful technique for profiling specificities of peptide-binding domains. Using highly diverse combinatorial peptide phage libraries, the method is suited for the identification of high-affinity ligands with inhibitor potential. A complementary but considerably less explored approach is to display expression products from genomic DNA, cDNA, open reading frames (ORFs) or from microarray oligonucleotide libraries designed to encode for defined regions of a target proteome are displayed on phage particles. This review focused on the use of proteomic phage display to uncover protein-protein interactions of potential relevance for cellular function. The method is particularly suited for the discovery of interactions between peptide-binding domains and their targets. The authors discussed the largely unexplored potential of this method in the discovery of domain-motif interactions of potential biological relevance.

Z.-L. Lai et al. contributed a paper entitled "*Methylation-associated gene silencing of RARB in areca carcinogens induced mouse oral squamous cell carcinoma.*" DNA methylation is a major epigenetic alternation of genome that regulates this crucial aspect of its function without changes in the DNA sequence. It is also thought to play an important role in carcinogenesis. Regarding oral squamous cell carcinoma (OSCC) development, chewing areca is known to be a strong risk factor in many Asian cultures. Therefore, the authors established an OSCC induced mouse model by 4-nitroquinoline-1-oxide (4-NQO), or arecoline, or both treatments, respectively. These are the main two components of the areca nut that could increase the occurrence of OSCC. The effects were examined with the noncommercial MCGI (mouse CpG islands) microarray for genome wide screening of the DNA methylation aberrant in induced OSCC mice. The results showed that retinoic acid receptor b (RARb) was indicated in hypermethylation at the promoter region and the loss of expression during cancer development. According to the results of real-time PCR, it was shown that de novo DNA methyltransferases were involved in gene epigenetic alternations of OSCC. Collectively, the results showed that RARB hypermethylation was involved in the areca-associated oral carcinogenesis.

W. Eilers et al. explored to which extent isoforms of the regulator of excitation-contraction and excitation-transcription coupling, calcium/calmodulin protein kinase II (CaMKII), contribute to the specificity of myocellular calcium sensing between muscle types and whether concentration transients in its autophosphorylation can be simulated. Qualitative differences existed between fast (gastrocnemius medialis) and slow type (muscle soleus) for the expression pattern of CaMKII isoforms. In silico assessment emphasized the importance of mitochondrial calcium buffer capacity for excitation-induced CaMKII autophosphorylation but did not predict its isoform specificity. The findings exposed that CaMKII autophosphorylation with paced contractions is regulated in an isoform and muscle type-specific fashion and highlight properties emerging for phenotype-specific regulation of CaMKII.

X. Li et al. introduced a novel open-source software (CELLCOUNTER) for counting cell migration and invasion in vitro. In contrast to the usually performed manual counting of cells in Transwell Boyden chamber based migration/invasion assays, this application is reported to be capable of recognizing and counting the total number of cells through an intuitive graphical user interface. The counting can be performed in batch, and the counting results can be visualized and further curated manually. The authors therefore conclude that the new software will be helpful in streamlining the experimental process and improving the reliability of the data acquisition.

In their paper, C.-J. Shen et al. investigated aberrant methylation in cloned porcine genome. Cloned animals usually exhibit defects in physical characteristics or aberrant epigenetic reprogramming, especially in some important organ development, such as heart valve and bone retardation. Osteopontin (OPN) is an extracellular-matrix protein that is involved with heart and bone development and

diseases. The authors investigated the correlation between OPN mRNA and its promoter methylation changes by the 5-aza-dc treatment in fibroblast cell and promoter assay. Data revealed that four methylated CpG sites presenting in the -2615 to -2239 bp region cause significant downregulation (approximately 75%) of OPN promoter activity. They also perform the protein-protein docking by software named Z-dock. Besides, the protein-protein complex also performed molecular dynamics simulation for validation. From all the evidences, they propose a novel mechanism and suggest that methylation in the OPN promoter plays a crucial role in the regulation of OPN expression that was found in cloned pigs genome.

The topic of Y.-A. Tsou et al.'s paper was "*Evaluation of correlation of cell cycle proteins and Ki-67 interaction in paranasal sinus inverted papilloma prognosis and squamous cell carcinoma transformation.*" The authors used protein expression patterns by immunohistochemical methods to see that the expression of p53, p16, p21, and p27 belongs to cell-cycle-regulators and PCNA, Ki-67 the proliferation markers in 60 inverted papilloma and 10 of them with squamous cell carcinoma transformation of the sinonasal tract. Significantly elevated levels of Ki67 and PCNA in IP with squamous cell carcinoma transformation of sinonasal tract compared with inverted papilloma were revealed. No variation of p16, p21, p27, and p53 expression was correlated to the IP malignant transformation. In conclusion, this is a first study that showed the correlation of Ki67 interacted with CDK1 and leads to malignant transformation and the elevated PLUNC expression in the sinonasal IPs with multiple recurrences in humans. They also employ the Z-Dock for protein-protein docking of related target proteins, CDK1 with target proteins (a) Ki-67, (b) p27, and (c) PCNA. From dihedrals angle of key binding residues and cluster analyses of all CDK1 and the binding proteins, they propose a novel mechanism of Ki-67, p27, and PCNA in cell cycle.

K.-B. Chen et al. compared the use of traditional Chinese medicine (TCM) against pregnane X receptor in the treatment of cardiovascular disease. The human pregnane X receptor, PXR, plays a crucial role in exogenous and endobiotic metabolism for rabbits, rats, mice, and humans. PXR activation can protect the blood vessels from damage caused by hazardous substances. The authors aimed to investigate the potent lead compounds as PXR receptor agonist against cardiovascular disease. To improve drug development of TCM compounds, they also aimed to investigate the potent lead compounds as PXR agonists from the TCM compounds in TCM Database@Taiwan. The top three TCM compounds, BEMG, Ixeriside, and Tangshenoside II, displayed higher potent binding affinities than the positive control, PNU-142721, in the docking simulation. They also perform the very time-consuming molecular dynamics simulation for validation of the stability of these potent compounds binding with PXR protein. Hence, the authors propose BEMG and Tangshenoside II, as potential lead compounds for further study in drug development process with the PXR protein.

## Acknowledgments

We express our sincere thanks and gratitude to the editorial board for their approval of this concept and continuous help in the successful publication of this special issue. We would also like to thank contributors to this special issue for their scientifically sound papers. With great pleasure and respect we extend our thanks to the reviewers for critical assessment of each paper, their constructive criticisms, and timely responses that made this special issue possible.

Calvin Yu-Chian Chen  
Kuo-Chen Chou  
James David Adams  
Tai-Ping Fan  
Gerhard Litscher

## Research Article

# Transgenic Sugarcane Resistant to *Sorghum mosaic virus* Based on Coat Protein Gene Silencing by RNA Interference

Jinlong Guo,<sup>1,2</sup> Shiwu Gao,<sup>1,2</sup> Qinliang Lin,<sup>1,2</sup> Hengbo Wang,<sup>1,2</sup>  
Youxiong Que,<sup>1,2</sup> and Liping Xu<sup>1,2</sup>

<sup>1</sup> Key Laboratory of Sugarcane Biology and Genetic Breeding, Ministry of Agriculture, Fujian Agriculture and Forestry University, Fuzhou 350002, China

<sup>2</sup> National Research and Development Center for Sugarcane Industry Technology, Fujian Agriculture and Forestry University, Fuzhou 350002, China

Correspondence should be addressed to Youxiong Que; queyouxiong@hotmail.com and Liping Xu; xlpmail@126.com

Received 25 June 2014; Revised 30 July 2014; Accepted 21 August 2014

Academic Editor: Calvin Yu-Chian Chen

Copyright © 2015 Jinlong Guo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As one of the critical diseases of sugarcane, sugarcane mosaic disease can lead to serious decline in stalk yield and sucrose content. It is mainly caused by *Potyvirus sugarcane mosaic virus* (SCMV) and/or *Sorghum mosaic virus* (SrMV), with additional differences in viral strains. RNA interference (RNAi) is a novel strategy for producing viral resistant plants. In this study, based on multiple sequence alignment conducted on genomic sequences of different strains and isolates of SrMV, the conserved region of coat protein (CP) genes was selected as the target gene and the interference sequence with size of 423 bp in length was obtained through PCR amplification. The RNAi vector pGII00-HACP with an expression cassette containing both hairpin interference sequence and *cp4-epsps* herbicide-tolerant gene was transferred to sugarcane cultivar ROC22 via *Agrobacterium*-mediated transformation. After herbicide screening, PCR molecular identification, and artificial inoculation challenge, anti-SrMV positive transgenic lines were successfully obtained. SrMV resistance rate of the transgenic lines with the interference sequence was 87.5% based on SrMV challenge by artificial inoculation. The genetically modified SrMV-resistant lines of cultivar ROC22 provide resistant germplasm for breeding lines and can also serve as resistant lines having the same genetic background for study of resistance mechanisms.

## 1. Introduction

Sugarcane (*Saccharum* spp. L.), a major sucrose accumulator and biomass producer, is one of the most important field crops grown in the tropics and subtropics [1]. It accounts for 92% of all sugar produced in China [2] and 80% of that in the world. Sugarcane mosaic disease is one of the most serious sugarcane diseases. It primarily damages chloroplasts, blocks photosynthesis, and decreases photosynthetic products, thus resulting in a decline in yield and sugar content [2]. Sugarcane mosaic disease is caused by the sugarcane mosaic virus subgroup of *Potyvirus sugarcane mosaic virus* (SCMV) and/or *Sorghum mosaic virus* (SrMV) [3]. *Potyvirus* is a single-stranded RNA virus, with simple genome structure encoding 10 mature proteins, named from N-terminal to C-terminal: the first protein (P1), helper component proteinase (HC-pro),

the third protein (P3), the first 6K protein (6K1), cylindrical inclusion protein (CI), the second 6K protein (6K2), viral protein genome-linked (VPg), nuclear inclusion a protein (NIa), nuclear inclusion b protein (NIb), and coat protein (CP) [4].

The virus strain differentiation is complex, with both of the virus members that cause sugarcane mosaic disease having several different virus strains [5]. At least eight strains have been reported [6], including five from SCMV and three from SrMV. The mixed infection of different virus strains also occurs [3, 6, 7], and dominant virus strains are variable [8]. In the 1980s there were at least three strains of SCMV including strains A, D, and E in mainland China [9]. However, the dominant pathogen has become strain H of SrMV in the last ten years [10]. The simplicity of pathogenic virus genome quickens the change of dominant strains. Coupled with

the complexity of the genetic background of sugarcane, the difficulty in the crossbreeding of virus-resistant varieties is obvious, especially for breeding sugarcane varieties resistant to multiple virus strains.

Improving plant antiviral resistance by gene silencing has proven to be effective in several plant-virus biosystems. Abel et al. first transferred CP genes of *tobacco mosaic virus* (TMV) into tobacco and successfully obtained anti-TMV tobacco plants [11]. Subsequently, different genes in *Potyvirus* genome were introduced into various plants to obtain corresponding resistant plants [12]. Joyce et al. introduced the CP gene of SCMV into sugarcane, and the CP-transformed plants displayed various phenotypes after SCMV challenge [13]. Ingelbrecht et al. introduced the CP gene of SrMV-H strain into sugarcane and obtained a range of different resistance types [14]. Yao et al. transferred the CP gene of SCMV-E strains into *S. officinarum* Badila and obtained SCMV-resistant transgenic lines but, after field experiments, found that some of them showed symptoms of mosaic disease, which were shown to be infected with SrMV-H and SCMV by RT-PCR [15]. Therefore, resistance performance of transgenic offspring obtained by the introduction of complete CP genes is complex, and resistance loss to the same or different virus strains in transgenic plants suggests that an improved method is necessary.

Gene silencing through RNA interference (RNAi) appears to be present in most eukaryotic organisms. Homologous RNA is degraded with the introduction of double-stranded RNA (dsRNA), which can lead to target gene silencing [16]. The target gene to be silenced can include a single gene or part sequence of a single gene that is targeted for suppression or can include multiple consecutive segments of a target gene, multiple nonconsecutive segments of a target gene, multiple alleles of a target gene, or multiple target genes from one or more species. RNAi-based antiviral breeding appears to be a promising strategy for development of virus resistance transgenic plants. There have been many successful examples of RNAi-mediated virus resistance improvement in crops, such as soybean [17], tobacco [18, 19], potato [20], barley [21], tomato [22], maize [23], and rice [24]. In sugarcane, the application of RNAi technology to suppress lignin biosynthesis was reported [25, 26], but no study on the application of RNAi technology in improving disease resistance has been reported. SrMV, the pathogen of sugarcane mosaic disease, is a single-stranded RNA virus, which replicates using a viral RNA polymerase. Viral genes in the form of dsRNA generate during replication, which is the basis of using RNAi technology for its control.

In this study, we have used RNAi technology, taking highly conserved sequences of SrMV CP gene as a silencing target, and RNAi expression vector with hairpin structures and introduced them into sugarcane via *Agrobacterium*-mediated transformation. We then performed screening and biological identification to obtain anti-*Sorghum mosaic virus* transgenic sugarcane plants. This study provides sugarcane transgenic lines with different resistances in the same genetic background for study of resistance mechanisms and for breeding of multiresistance to various SrMV strains.

## 2. Materials and Methods

**2.1. Bacterial Strains and Plasmids.** *Escherichia coli* strain DH5A, *Agrobacterium tumefaciens* strain EHA105, and intermediate vector pHANNIBAL were provided by the Key Laboratory of Sugarcane Biology and Genetic Breeding, Ministry of Agriculture (Fuzhou, China). The glyphosate tolerance gene *cp4-epsps* was obtained from roundup ready soybean by PCR and verified by sequencing, and the intermediate vector pGIIHA containing 35S promoter-*cp4-epsps*-CaMV polyA cassette was constructed subsequently in previous study.

**2.2. Reagents and Plant Materials.** Reverse transcriptase (AMV), restriction endonucleases, T4 DNA ligase, and PCR kits were purchased from Fermentas (USA); dephosphorylation (BAP) kit was purchased from Takara (Dalian, China); Wizard DNA clean-up kit gel extraction kit was purchased from Promega Corporation (USA); plant genomic DNA extraction kit was purchased from TIANGEN (Beijing, China); components in MS medium were purchased from Sangon (Shanghai, China); Trizol reagents were purchased from Invitrogen (USA); Timentin disodium salt and 2,4-dichlorophenoxyacetic acid (2-4-D) were purchased from Sigma (USA); and herbicide (47% isopropylamine salt of N-glycine) applicable by foliar spraying was purchased from Sannong Co., Ltd. (Fujian, China). ROC22 was the most popular cultivar in China, which was provided by the Key Laboratory of Sugarcane Biology and Genetic Breeding, Ministry of Agriculture (Fuzhou, China).

**2.3. Medium.** Infection medium M1 is 1/2 MS + 100  $\mu$ mol/L acetosyringone + 20 g/L sucrose, pH 5.8; cocultivation medium M2 is 1/2 MS + 3.0 mg/L 2-4-D + 100  $\mu$ mol/L acetosyringone + 20 g/L sucrose + 5 g/L agar powder, pH 5.8; subculture medium M3 is MS + 3.0 mg/L 2-4-D + 300 mg/L Timentin + 8.0 mg/L herbicide + 30 g/L sucrose + 6 g/L agar powder, pH 5.8; differential medium M4 is MS + 2.0 mg/L BA + 0.5 mg/L KT + 0.2 mg/L NAA + 300 mg/L Timentin + 6.0 mg/L herbicide + 30 g/L sucrose + 6 g/L agar powder, pH 5.8; rooting medium M5 is 1/2 MS + 0.2 mg/L 6-BA + 3 mg/L NAA + 60 g/L sucrose + 6 g/L agar powder, pH 5.8.

**2.4. RNAi Target Sequence Selection.** Genome sequences of the SrMV strains (H, I, and M) isolated from sugarcane were collected from GenBank. Using DNAMAN 5.22 software (<http://www.lynnon.com/>), multiple sequence alignment was performed to determine the most conservative nucleic acid segment as RNAi target sequence. The fast alignment was generated using DNAMAN 5.22 with default parameters (Gap penalty was set at 7, K-tuple at 3, and number of Top at 5). The accession numbers of the chosen sequences in alignment were EU189035, EU189036, EU189037, EU189041, EU189042, EU189038, EU189043, EU189044, EU189045, EU189046, EU189039, EU189040, U07219, AJ310198, NC004035, and SMU57358 and SrMV FZ strain was kept in our lab (a SrMV strain isolated from Fuzhou, China, unsubmitted).

**2.5. Interference Fragment Preparation and Hairpin Intermediate Vector Construction.** According to multiple sequence alignment results, a pair of specific primers targeting the most conservative segment were designed, with extra *Xba* I and *Xho* I endonuclease restriction sites on the 5' end of the forward primer CPS and *Cla* I and *Kpn* I on the 5' end of the forward primer CPA. The primer sequences are as follows:

CPS:

5'-GATCTAGA CTCGAGTGT TGGACAATGATG-3'  
*Xba* I                      *Xho* I

CPA:

5'-CTATCGAT GGTACCGCACATCAGTGGTTCT-3'  
*Cla* I                      *Kpn* I

Target sequence for RNAi was amplified by PCR using SrMV FZ as template. The 50  $\mu$ L PCR reaction mix contained 5.0  $\mu$ L 10  $\times$  PCR buffer, 4.0  $\mu$ L deoxynucleotide triphosphates (dNTPs) (2.5 mM), 2.0  $\mu$ L each of forward and reverse primers (10  $\mu$ M), 2.0  $\mu$ L template (100 ng), and 0.25  $\mu$ L Ex-Taq enzyme (5 U/ $\mu$ L). The ddH<sub>2</sub>O was added as supplement. The PCR amplification program consisted of predenaturation for 5 min at 94°C, denaturation for 30 s at 94°C, annealing for 30 s at 60°C, and extension for 30 s at 72°C for 30 cycles; final extension was for 10 min at 72°C. The PCR product was purified by gel extraction kit to prepare for digestion. The PCR product was separated in 2% agarose gel. The target DNA fragments were excised and purified using an agarose gel purification kit. Using the two sets of restriction enzymes, *Cla* I/*Xba* I and *Xho* I/*Kpn* I, successively, the target sequence was inserted into the two sides of the intronic region of pHANNIABL vector. A clone with a recombinant plasmid was validated by PCR, double digestion, and sequencing and was termed as pHANNIABL-CP.

**2.6. Construction of the RNAi Expression Vector.** The pGIIHA intermediate vector was digested with *Not* I and then purified by gel extraction kit. The purified products were dephosphorylated according to manual of the dephosphorylation (BAP) kit. *Not* I-digested hairpin interference cassette fragments from pHANNIABL-CP were inserted into the *Not* I site of pGIIHA. A clone with a recombinant plasmid was validated by PCR, double digestion, and sequencing and was termed as pGII00-HACP.

**2.7. Preparation of the Engineering Bacteria.** According to the freeze-thaw method reported by Holsters et al. [27], the RNAi vector pGII00-HACP was transformed into *A. tumefaciens* EHA105. The positive clone identified by PCR was inoculated into the LB medium containing kanamycin (50  $\mu$ g·mL<sup>-1</sup>) and rifampicin (35  $\mu$ g·mL<sup>-1</sup>) for shake culture at 150 rpm at 37°C. When OD<sub>600</sub> reached 1.0 to 1.2, the culture was centrifuged at 5,000 rpm for 5 min at room temperature to discard the supernatant. The pellet was collected and resuspended with M1 medium and then centrifuged at 5,000 rpm for 5 min at room temperature again to discard the supernatant. The pellet thus obtained was resuspended and diluted with M1 medium to OD<sub>600</sub> = 1.0.

**2.8. Agrobacterium-Mediated Transformation and Screening.** Leaf explants from sugarcane ROC22 were cultured on MS

medium supplemented with 3.0 mg/L 2,4-D for one week and then cocultivated with recombinant *A. tumefaciens* EHA105 for 30 min. The plant tissue was picked out and sucked dry with filter paper and cultured on M2 medium for 2-3 d; then the plant tissue was transferred to M3 culture medium and screened for 2-3 generations. After that, the plant tissue was transferred to M4 differential medium, followed by a period of culture on M5 rooting medium when the tissue culture seedlings grew to 4~5 cm in length. The seedlings were then transferred to 72-well nutrition plate when their roots reached about 2.5 cm. After the seedlings were transplanted, the plants were sprayed with liquid herbicide solutions at the concentrations of 3.0‰ and survivors from the herbicide treatments were selected.

**2.9. PCR for Positive Identification.** Genome DNA extracted from herbicide-resistant and control plants was diluted to a concentration of 50 ng/ $\mu$ L and was used as PCR template. 35S promoter and *cp4-epsps* gene were selected as the target genes for identification. Genome DNA of the herbicide-resistant plants was isolated and tested by PCR, using pGII00-HACP as a positive control, genome DNA of untransformed plants as a negative control, and ddH<sub>2</sub>O as a blank control.

The primer sequences for the 35S promoter were 35SPro F: 5'-TCTAACAGA AACTCGCCGTGAA-3' and 35Spro R: 5'-AAGGGTCTTGCGAAGGATAGT-3', and the primers sequences for the *cp4-epsps* gene were cp4-epsps F: 5'-GTCCTTCATGTTTCGGCGGTCTC-3' and cp4-epsps R: 5'-ACGTCGATGACTTGGCTGGTGA-3'. The 50  $\mu$ L PCR reaction mix contained 5.0  $\mu$ L 10  $\times$  PCR buffer; 4.0  $\mu$ L deoxynucleotide triphosphates (dNTPs) (2.5 mM); 2.0  $\mu$ L each of forward and reverse primers (10  $\mu$ M); 2.0  $\mu$ L plasmid DNA (100 ng); and 0.25  $\mu$ L Ex-Taq enzyme (5 U/ $\mu$ L). The sterile ddH<sub>2</sub>O was added as supplement. The PCR amplification program consisted of predenaturation for 5 min at 94°C, denaturation for 30 s at 94°C, annealing for 30 s at 57°C, and extension for 30 s at 72°C for 30 cycles, and final extension was for 10 min at 72°C. The PCR products were separated by 2.0% agarose gel electrophoresis, and the results were analyzed by gel imaging and analysis system.

**2.10. Identification of SrMV Resistance by Artificial Inoculation.** According to Gómez et al. [28], sugarcane leaves with typical mosaic symptoms were collected and diagnosed with SrMV infection. The artificial inoculation method according to Uzarowska et al. [29] used the SrMV-positive leaf samples as the virus infection source.

### 3. Results and Analysis

**3.1. RNAi Target Sequence Selection.** Thirteen sequences of the SrMV CP gene and four sequences of the SrMV whole genome in NCBI were selected for multiple sequence alignment analysis. Figure 1 showed that CP genes with 87.39% homology were the most conservative fragment in SrMV genome sequence. Therefore, a conserved region of 423 bp, from 573 bp to 995 bp in CP genes, was identified as the interference fragment sequence.

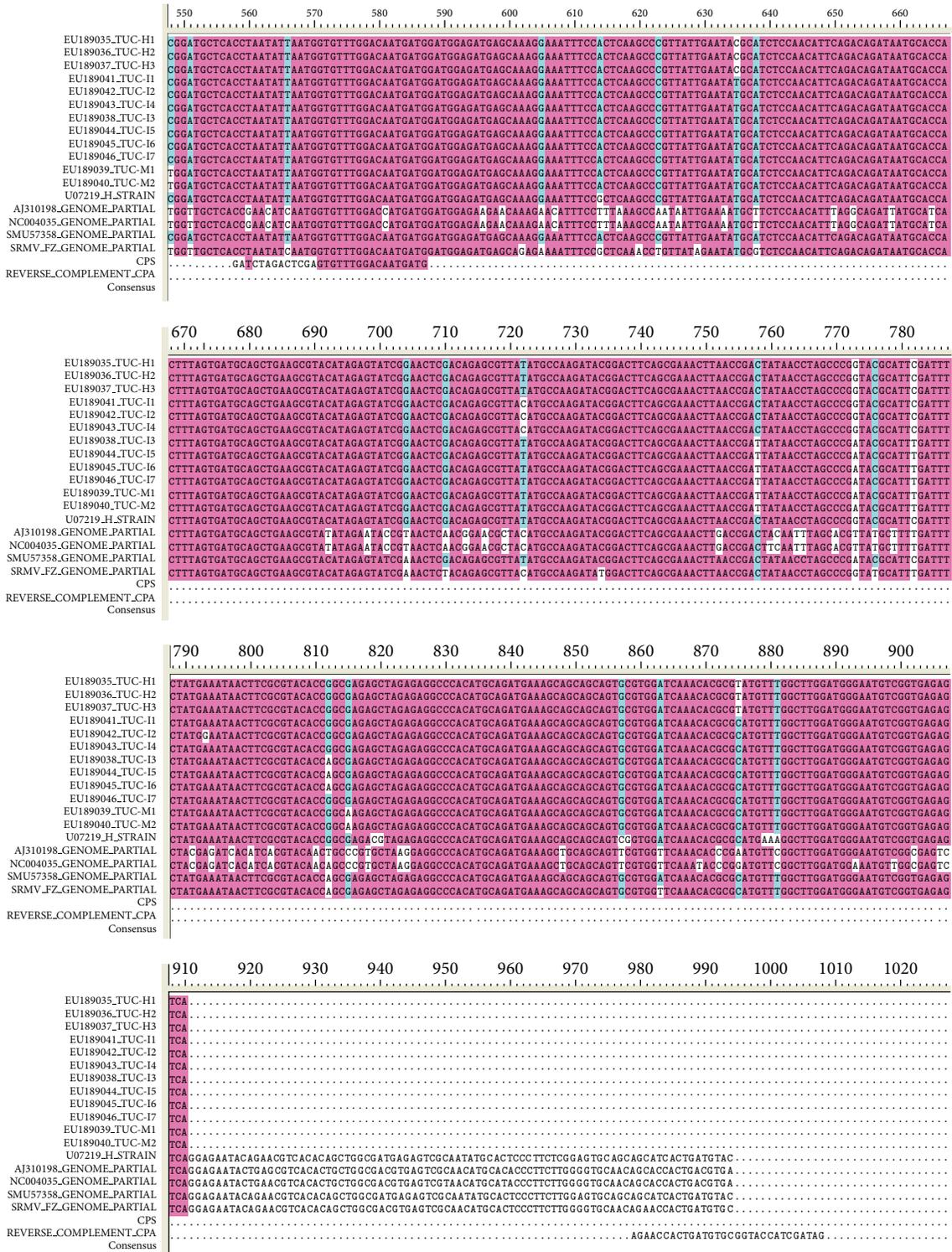


FIGURE 1: Multiple sequence alignment of SrMV CP genes.

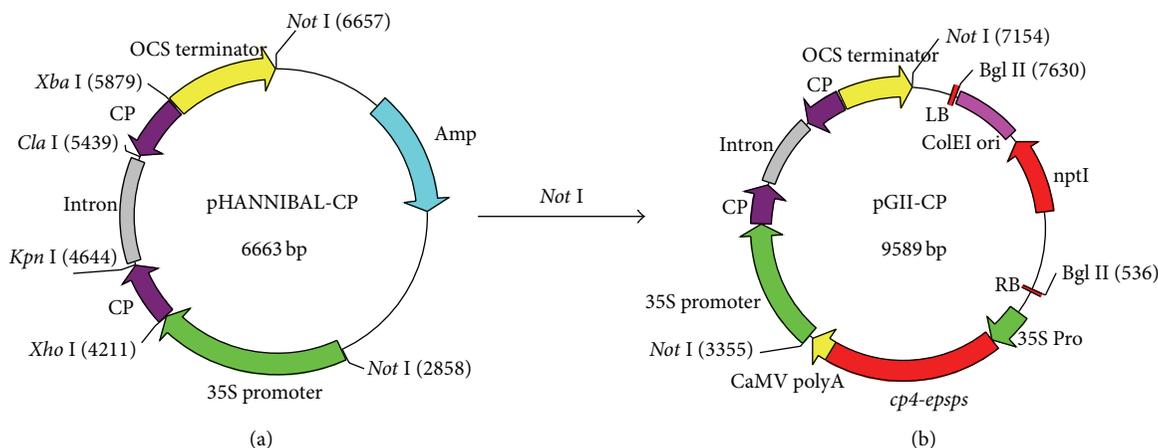


FIGURE 2: A simplified flowchart showing the construction of binary vectors. (a) Vector diagram of pHANNIBAL-CP; (b) diagram of RNAi expression vector pGII00-HACP.

**3.2. Interference Fragment Preparation and Hairpin Intermediate Vector Construction.** The target interference fragment with the expected length of 423 bp was obtained by PCR and verified by sequencing. Using the two sets of restriction endonuclease—*Kpn* I/*Xho* I or *Xba* I/*Cla* I, the 423 bp of partial CP gene and its reverse complement fragment were successively inserted into each intron site contained in pHANNIBAL and enabled the hairpin intermediate vector pHANNIBAL-CP to make a hairpin loop (Figure 2(a)).

**3.3. Construction of RNAi Expression Vector.** Recombinant RNAi expression vector was identified by *Not* I restriction analysis, PCR, and sequencing (data not shown), and the positive hairpin RNAi expression vector was termed as pGII00-HACP (Figure 2(b)).

**3.4. Agrobacterium-Mediated Transformation and Screening.** The constructed RNAi expression vector pGII00-HACP was transformed into *A. tumefaciens* EHA105 and used to infect sugarcane calli. After coculture, selective subculture and differentiation culture under herbicide stress, and rooting culture (Figure 3), about five hundred regenerated seedlings were obtained.

**3.5. Herbicide Resistance Screening and PCR Detection of Resistant Regenerated Plants.** A portion of the regenerated putative recombinants survived herbicide treatment (Figure 4). Among these, 16 plants from 50 survivors were further identified as positive by PCR, exhibiting existence of 463 bp specific band in the 35S promoter detection and 623 bp specific band in the *cp4-epsps* gene detection. In order to get more putative resistant transgenic plants, 0.3% herbicide, which was not a complete lethal concentration for sugarcane, was used in this study, although it led to higher false-positive rate. Figure 5 showed part of PCR products identified by gel electrophoresis.

**3.6. Disease Incidence of Artificially Inoculated Transgenic Lines.** After artificial inoculation with SrMV, 14 transgenic

plants showed no symptoms and no virus in RT-PCR detection and were judged to be uninfected after SrMV challenge; two transgenic plants and nontransgenic control plants showed symptoms and SrMV in RT-PCR detection, diagnosed as infected after SrMV challenge (Figures 6 and 7). Therefore, it could be concluded that hairpin RNAi expression vector pGII00-HACP, which resulted in production of resistance against SrMV, was successfully introduced into sugarcane and according to 87.5% transgenic plants showed improved resistance to SrMV.

## 4. Discussion

Mosaic virus-resistant transgenic sugarcane plants have been obtained via particle gun bombardment [14, 15, 30–32], and some transgenic plants were significantly improved in mosaic virus resistance. However, most events produced by gene gun bombardment tend to show high copy numbers of recombinant inserts [33]. Modern sugarcane varieties are a complex allopolyploid and aneuploid genetic background of *S. officinarum* (chromosome number 80) and *S. spontaneum* (chromosome number from 40 to 128), with even *Erianthus arundinaceus* included in sugarcane clones bred during last five years in China [34]. Hence, it was hard to prove clearly characters such as copy numbers, insertion sites, and border sequences in genetically modified (GM) sugarcane via gun bombardment. However, such information is necessary for any GM organisms including GM sugarcane before the application of transgenic field trials. Also high copy numbers of exogenous genes, such as the selective marker, in GM organism can even cause cosuppression [35].

It has been reported that *Agrobacterium*-mediated transformation leads to clean, discrete, low copy, well-defined, unarranged DNA insertions into the plant genome [36, 37]. However, *Agrobacterium*-mediated transformation is not as successful as gene gun bombardment in sugarcane. In the present study, hairpin RNAi expression vector pGII00-HACP was transferred to sugarcane cultivar ROC22 via *Agrobacterium*-mediated method. The 423 bp interference

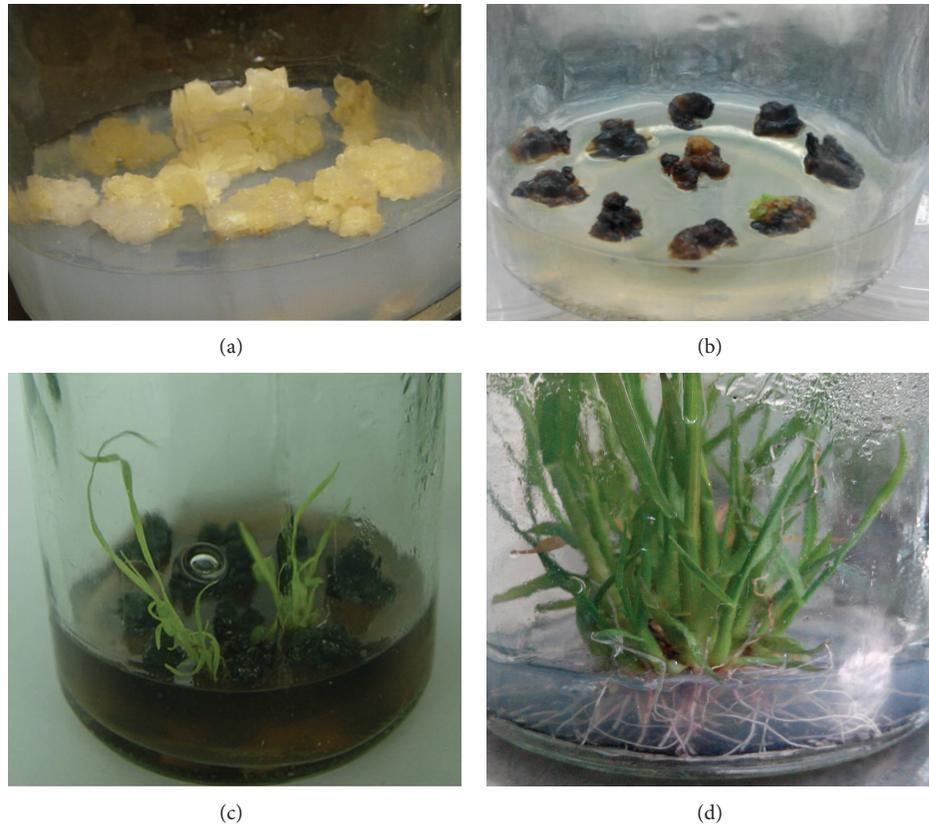


FIGURE 3: Putative recombinant screening. (a) Wild-type calli, (b) transformed calli screening by herbicide, (c) regenerated seedlings at the stage of differentiation selection culture, and (d) regenerated seedlings at the stage of rooting culture.

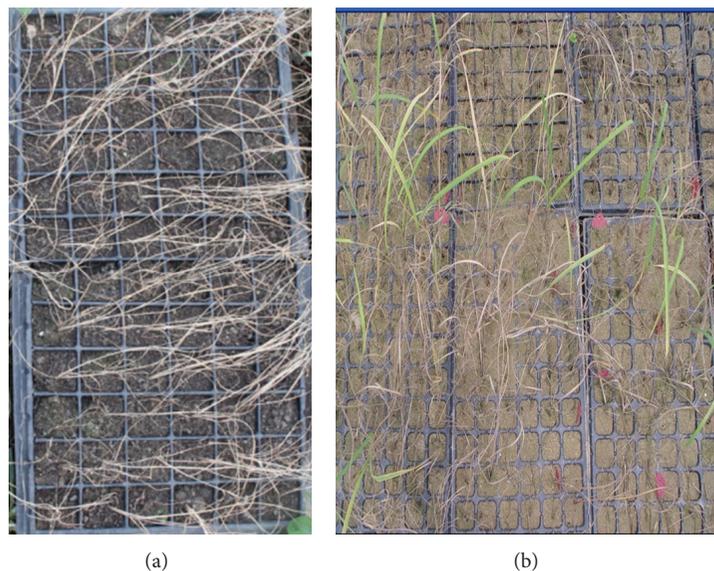


FIGURE 4: Spraying screening by 0.3% herbicide. (a) Wild-type plants and (b) putative transformants.

fragment derived from the most conservative region of the CP gene of SrMV based on multiple alignment analysis of all the three SrMV strains (H, I, and M). The purpose is to obtain multistrains resistant sugarcane plants. In addition, *cp4-epsps* gene contained in pGII00-HACP can be used as

a high-efficiency selective marker and also endows sugarcane with a herbicide-tolerant trait. This enables farmers to make the process of weed control more efficient and flexible.

RNAi is a highly conserved dsRNA-guided mechanism that mediates sequence-specific posttranscriptional gene

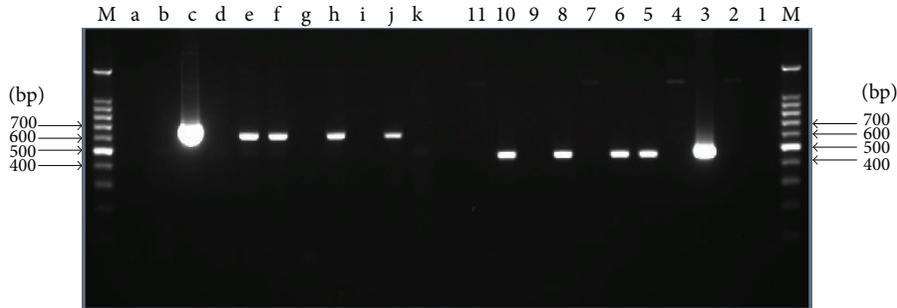


FIGURE 5: PCR amplification products of *cp4-epsps* gene and 35S promoter. M: DNA marker; d~k: 35S promoter detection; 4~11: *cp4-epsps* gene detection; c and 3: positive control; b and 2: negative control; a and 1: ddH<sub>2</sub>O blank control.

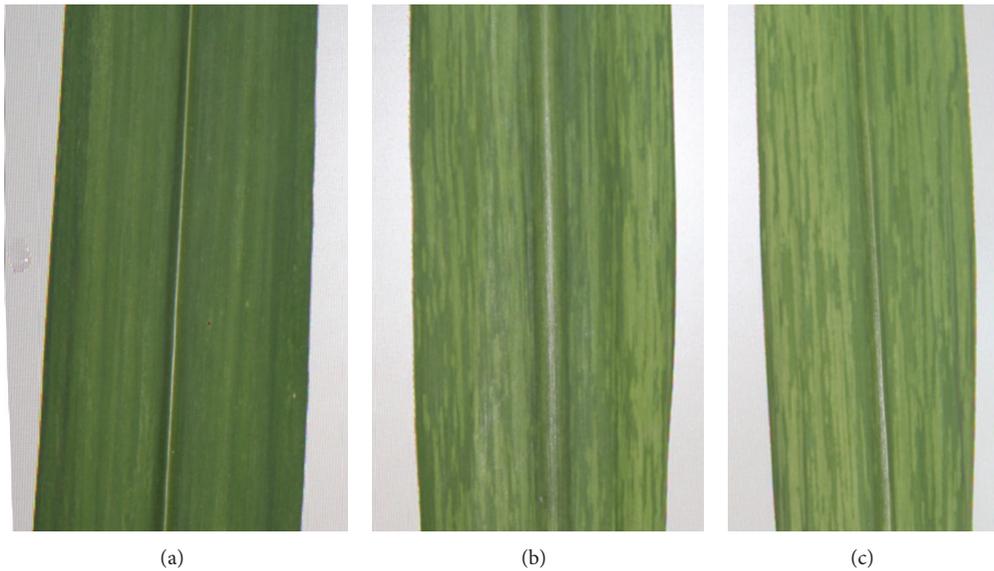


FIGURE 6: The SrMV-resistant and sensitive symptoms. (a) Transgenic plants had no symptoms; (b) transgenic plants displayed symptoms; (c) nontransgenic control plants displayed appeared symptoms.

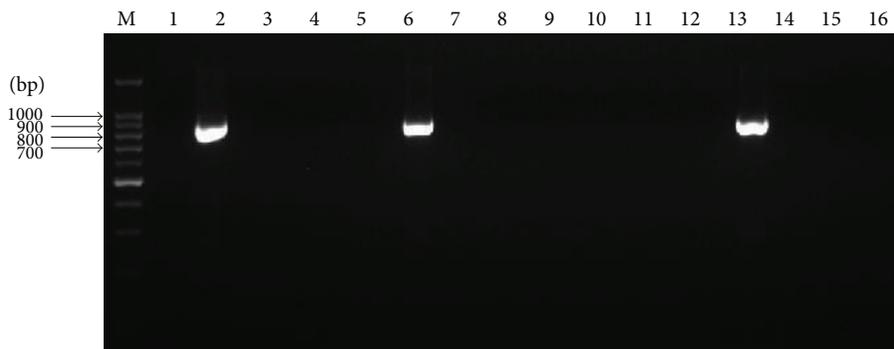


FIGURE 7: RT-PCR detection of SrMV in experimental plants. M: DNA marker; 1: ddH<sub>2</sub>O blank control; 2: nontransgenic control showed symptoms; 3~6, 8~14, and 16: transgenic plants without symptoms; 7, 15: transgenic plants showed symptoms.

silencing [38]. As a source of dsRNA, plasmid-expressed short hairpin RNA (shRNA) has been demonstrated to be able to trigger RNAi silencing [21–24]. The study of Varsha Wesley et al. [39] showed that intron-containing constructs (ihpRNA) can generally enable 90–100% of independent

transgenic plants to show silencing. The average percentages of ihpRNA, hpRNA, cosuppression, and antisense constructs at silencing were 90%, 58%, 13%, and 12%, respectively [39]. pHANNIBAL, an intermediate generic vector used in this study, allows a simple, single PCR product from CP gene to

be easily converted into a highly effective ihpRNA silencing construct. Similar to Varsha Wesley et al. [39], the ihpRNA silencing construct targeting SrMV CP gene in this study exhibited 87.5% high silencing effect.

In summary, a 423 bp highly conserved region from the CP gene of SrMV was selected as the interference sequence based on multiple alignment analysis of all the three SrMV strains (H, I, and M) and several other isolates. The hairpin RNAi expression vector pGII00-HACP was transferred to sugarcane cultivar ROC22 via *Agrobacterium*-mediated transformation. After herbicide screening, PCR molecular identification, and artificial inoculation challenge, anti-SrMV positive transgenic lines were successfully obtained. This study provides the foundation for a further study on silencing mechanism of SrMV-CP gene expression based on RNA interference and provides novel materials to evaluate the silencing effect connected with exogenous gene copy numbers and insertion sites. It also provides new material for broad-spectrum antiviral sugarcane breeding.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Authors' Contribution

Jinlong Guo and Shiwu Gao equally contributed to this paper.

### Acknowledgments

This work was supported by the National Natural Science Foundation of China (30871581), the Earmarked Fund for the Modern Agroindustry Technology Research System (CARS-20), and the 948 Program on the Introduction of International Advanced Agricultural Science and Techniques of the Department of Agriculture (2014-S18).

### References

- [1] J. L. Guo, L. P. Xu, Y. C. Su et al., "ScMT2-1-3, a metallothionein gene of sugarcane, plays an important role in the regulation of heavy metal tolerance/accumulation," *BioMed Research International*, vol. 2013, Article ID 904769, 12 pages, 2013.
- [2] R. K. Chen, L. P. Xu, Y. Q. Lin et al., *Chinese Monograph: Modern Sugarcane Genetic Breeding*, China Agriculture Press, 2011.
- [3] Y. Xie, M. Wang, D. Xu, R. Li, and G. Zhou, "Simultaneous detection and identification of four sugarcane viruses by one-step RT-PCR," *Journal of Virological Methods*, vol. 162, no. 1-2, pp. 64-68, 2009.
- [4] F. Revers, O. L. Gall, T. Candresse, and A. J. Maule, "New advances in understanding the molecular biology of plant/potyvirus interactions," *Molecular Plant-Microbe Interactions*, vol. 12, no. 5, pp. 367-376, 1999.
- [5] Z. N. Yang and T. E. Mirkov, "Sequence and relationships of sugarcane mosaic and sorghum mosaic virus strains and development of RT-PCR-based RFLPs for strain discrimination," *Phytopathology*, vol. 87, no. 9, pp. 932-939, 1997.
- [6] V. R. Mali and R. P. Thakur, "Natural infection of sugarcane by an immunity breaking strain of sorghum mosaic potyvirus (SRMV-IBS) in peninsular India," *Sugar Tech*, vol. 2, no. 3, pp. 20-25, 2000.
- [7] C. G. Marcos, P. R. Luciana, C. S. Silvana, and G. A. L. Marcos, "Virus diseases of sugarcane A constant challenge to sugarcane breeding in Brazil," *Functional Plant Science and Biotechnology*, vol. 6, pp. 108-112, 2012.
- [8] M. P. Grisham and Y.-B. Pan, "A genetic shift in the virus strains that cause mosaic in Louisiana sugarcane," *Plant Disease*, vol. 91, no. 4, pp. 453-458, 2007.
- [9] Y. H. Chen, Z. J. Zhou, Q. Y. Lin, and L. H. Xie, "A preliminary report on strains of sugarcane mosaic virus," *Chinese Journal: Journal of Fujian Agriculture College*, vol. 17, no. 1, pp. 44-48, 1988.
- [10] Y. Guo, "Identification of the viral pathogen of sugarcane mosaic disease in 4 sugarcane planting states of China and gene analysis of virus protein," in *Chinese Degree: Fujian Agriculture and Forestry University*, Supervisor: M. Q. Zhang, pp. 26-27, 2008.
- [11] P. P. Abel, R. S. Nelson, B. de et al., "Delay of disease development in transgenic plants that express the tobacco mosaic virus coat protein gene," *Science*, vol. 232, no. 4751, pp. 738-743, 1986.
- [12] E. Gammelgård, M. Mohan, and J. P. T. Valkonen, "Potyvirus-induced gene silencing: the dynamic process of systemic silencing and silencing suppression," *Journal of General Virology*, vol. 88, no. 8, pp. 2337-2346, 2007.
- [13] P. A. Joyce, R. B. McQualter, J. A. Handley, J. L. Dale, R. M. Harding, and G. R. Smith, "Transgenic sugarcane resistant to sugarcane mosaic virus," in *Proceedings of the 20th Conference of the Australian Society of Sugar Cane Technologists*, Ballina, Australia, 1998.
- [14] I. L. Ingelbrecht, J. E. Irvine, and T. E. Mirkov, "Posttranscriptional gene silencing in transgenic sugarcane. Dissection of homology-dependent virus resistance in a monocot that has a complex polyploid genome," *Plant Physiology*, vol. 119, no. 4, pp. 1187-1198, 1999.
- [15] W. Yao, A. L. Yu, J. S. Xu, G. L. Geng, M. Q. Zhang, and R. K. Chen, "Analysis and identification for transgenic sugarcane of ScMV-CP gene," *Chinese Journal of Molecular Plant Breeding*, vol. 2, no. 1, pp. 13-18, 2004.
- [16] A. Aigner, "Gene silencing through RNA interference (RNAi) in vivo: strategies based on the direct application of siRNAs," *Journal of Biotechnology*, vol. 124, no. 1, pp. 12-25, 2006.
- [17] H. J. Kim, M.-J. Kim, J. H. Pak et al., "Characterization of SMV resistance of soybean produced by genetic transformation of SMV-CP gene in RNAi," *Plant Biotechnology Reports*, vol. 7, no. 4, pp. 425-433, 2013.
- [18] Y. B. Niu, D. F. Wang, M. Yao, Z. Yan, and W. X. You, "Transgenic tobacco plants resistant to two viruses via RNA silencing," *Acta Agronomica Sinica*, vol. 37, no. 3, pp. 484-488, 2011.
- [19] E. Bucher, D. Lohuis, P. M. J. A. van Poppel, C. Geerts-Dimitriadou, R. Goldbach, and M. Prins, "Multiple virus resistance at a high frequency using a single transgene construct," *Journal of General Virology*, vol. 87, no. 12, pp. 3697-3701, 2006.
- [20] V. O. Ntui, K. Kynet, P. Azadi et al., "Transgenic accumulation of a defective cucumber mosaic virus (CMV) replicase derived double stranded RNA modulates plant defence against CMV strains O and Y in potato," *Transgenic Research*, vol. 22, no. 6, pp. 1191-1205, 2013.
- [21] M.-B. Wang, D. C. Abbott, and P. M. Waterhouse, "A single copy of a virus-derived transgene encoding hairpin RNA

- gives immunity to barley yellow dwarf virus," *Molecular Plant Pathology*, vol. 1, no. 6, pp. 347–356, 2000.
- [22] C.-Y. Lin, H.-M. Ku, W.-S. Tsai, S. K. Green, and F.-J. Jan, "Resistance to a DNA and a RNA virus in transgenic plants by using a single chimeric transgene construct," *Transgenic Research*, vol. 20, no. 2, pp. 261–270, 2011.
- [23] Z.-Y. Zhang, L. Yang, S.-F. Zhou, H.-G. Wang, W.-C. Li, and F.-L. Fu, "Improvement of resistance to maize dwarf mosaic virus mediated by transgenic RNA interference," *Journal of Biotechnology*, vol. 153, no. 3-4, pp. 181–187, 2011.
- [24] T. Shimizu, E. Nakazono-Nagaoka, F. Akita et al., "Hairpin RNA derived from the gene for Pns9, a viroplasm matrix protein of Rice gall dwarf virus, confers strong resistance to virus infection in transgenic rice plants," *Journal of Biotechnology*, vol. 157, no. 3, pp. 421–427, 2012.
- [25] K. Osabe, S. R. Mudge, M. W. Graham, and R. G. Birch, "RNAi mediated down-regulation of PDS gene expression in sugarcane (*Saccharum*), a highly polyploid crop," *Tropical Plant Biology*, vol. 2, no. 3, pp. 143–148, 2009.
- [26] J. H. Jung, W. M. Fouad, W. Vermerris, M. Gallo, and F. Altpeter, "RNAi suppression of lignin biosynthesis in sugarcane reduces recalcitrance for biofuel production from lignocellulosic biomass," *Plant Biotechnology Journal*, vol. 10, no. 9, pp. 1067–1076, 2012.
- [27] M. Holsters, D. de Waele, A. Depicker, E. Messens, M. van Montagu, and J. Schell, "Transfection and transformation of *Agrobacterium tumefaciens*," *Molecular & General Genetics*, vol. 163, no. 2, pp. 181–187, 1978.
- [28] M. Gómez, A. M. Rago, and G. Serino, "Rapid identification of viruses causing sugarcane mosaic by direct sequencing of RT-PCR products from crude extracts: a method for large scale virus surveys," *Journal of Virological Methods*, vol. 157, no. 2, pp. 188–194, 2009.
- [29] A. Użarowska, G. Dionisio, B. Sarholz et al., "Validation of candidate genes putatively associated with resistance to SCMV and MDMV in maize (*Zea mays* L.) by expression profiling," *BMC Plant Biology*, vol. 9, article 15, 2009.
- [30] G. R. Smith, R. L. Gambley, and B. T. Egan, "Progress in development of a sugarcane meristem transformation system and production of SCMV-resistant transgenics," *Proceedings of the Australian Society of Sugar Cane Technologists*, vol. 15, pp. 237–250, 1994.
- [31] R. A. Gilbert, M. Gallo-Meagher, J. C. Comstock, J. D. Miller, M. Jain, and A. Abouzid, "Agronomic evaluation of sugarcane lines transformed for resistance to Sugarcane mosaic virus strain E," *Crop Science*, vol. 45, no. 5, pp. 2060–2067, 2005.
- [32] Y. Guo, M. H. Ruan, W. Yao, L. Chen, R. K. Chen, and M. Q. Zhang, "Difference of coat protein mediated resistance to sugarcane mosaic virus between Badila and Funong 91-4621," *Journal of Fujian Agriculture and Forestry University (Natural Science Edition)*, vol. 37, no. 1, pp. 7–12, 2008 (Chinese).
- [33] A. D. Arencibia, E. R. Carmona, P. Téllez et al., "An efficient protocol for sugarcane (*Saccharum spp.* L.) transformation mediated by *Agrobacterium tumefaciens*," *Transgenic Research*, vol. 7, no. 3, pp. 213–222, 1998.
- [34] B. Xue, J. Guo, Y. Que, Z. Fu, L. Wu, and L. Xu, "Selection of suitable endogenous reference genes for relative copy number detection in sugarcane," *International Journal of Molecular Sciences*, vol. 15, no. 5, pp. 8846–8862, 2014.
- [35] W. Tang, R. J. Newton, and D. A. Weidner, "Genetic transformation and gene silencing mediated by multiple copies of a transgene in eastern white pine," *Journal of Experimental Botany*, vol. 58, no. 3, pp. 545–554, 2007.
- [36] A. R. Wenck, M. Quinn, R. W. Whetten, G. Pullman, and R. Sederoff, "High-efficiency agrobacterium-mediated transformation of Norway spruce (*Picea abies*) and loblolly pine (*Pinus taeda*)," *Plant Molecular Biology*, vol. 39, no. 3, pp. 407–416, 1999.
- [37] M. N. Somleva, Z. Tomaszewski, and B. V. Conger, "Agrobacterium-mediated genetic transformation of switchgrass," *Crop Science*, vol. 42, no. 6, pp. 2080–2087, 2002.
- [38] G. Meister and T. Tuschl, "Mechanisms of gene silencing by double-stranded RNA," *Nature*, vol. 431, no. 7006, pp. 343–349, 2004.
- [39] S. Varsha Wesley, C. A. Helliwell, N. A. Smith et al., "Construct design for efficient, effective and high-throughput gene silencing in plants," *Plant Journal*, vol. 27, no. 6, pp. 581–590, 2001.

## Review Article

# Next-Generation Technologies for Multiomics Approaches Including Interactome Sequencing

**Hiroyuki Ohashi,<sup>1</sup> Mai Hasegawa,<sup>1,2</sup> Kentaro Wakimoto,<sup>1,2</sup> and Etsuko Miyamoto-Sato<sup>1,2</sup>**

<sup>1</sup>*Division of Interactome Medical Sciences, The Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan*

<sup>2</sup>*Division of Molecular Biology, Tokyo University of Science Research Institute for Biomedical Science, 2669 Yamazaki, Noda, Chiba 278-0022, Japan*

Correspondence should be addressed to Etsuko Miyamoto-Sato; nekoneko@rs.tus.ac.jp

Received 27 June 2014; Revised 30 August 2014; Accepted 31 August 2014

Academic Editor: Calvin Yu-Chian Chen

Copyright © 2015 Hiroyuki Ohashi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The development of high-speed analytical techniques such as next-generation sequencing and microarrays allows high-throughput analysis of biological information at a low cost. These techniques contribute to medical and bioscience advancements and provide new avenues for scientific research. Here, we outline a variety of new innovative techniques and discuss their use in omics research (e.g., genomics, transcriptomics, metabolomics, proteomics, and interactomics). We also discuss the possible applications of these methods, including an interactome sequencing technology that we developed, in future medical and life science research.

## 1. Introduction

More than a decade has passed since the human genome sequence was decoded. Subsequent advancements in and integration of personal genome analysis, post-genome functional analysis, and multiomics analyses have facilitated the development of personalized medicine, which is emerging as the optimal therapeutic direction for the future of medical science (Figure 1). The advent of next-generation sequencing (NGS) and its use in clinical practice will enable the adaptation of multiomics data to personal medical care. However, the costs of these methods and the amount of data generated using multiomics approaches have emerged as challenges that must be tackled. Interactome analysis is considered to be a crucial integrator of multiomics analysis. Currently, the “integrome” is being investigated to determine how the large amounts of data generated using multiomics approaches can be integrated most advantageously. In this review, we discuss these innovative new approaches used in genomics, transcriptomics, metabolomics, and proteomics. We also present an overview of the new insights into complex biological systems that are provided by the use of these technologies.

## 2. Genomics

NGS has contributed substantially to recent advances in omics research. In NGS, a technology used in genome sequencing, sequences containing millions of DNA fragments are read by performing numerous reactions in parallel [1]. The use of this technology has dramatically reduced the time and cost required for sequencing and has facilitated analysis of the human genome, epigenome, and transcriptome. Several NGS platforms have been released by various companies, and a few representative platforms are HiSeq and MiSeq (Illumina), 454 GS FLX (Roche), and PacBio (Pacific Biotechnology) (Table 1). In these platforms, distinct methods of template preparation and signal detection are used [2, 3].

NGS can be used for performing genomic and epigenomic analyses (Table 2). In genomic analysis, somatic mutations are detected using whole-genome sequencing or whole-exome sequencing. In whole-genome sequencing, somatic mutations (e.g., single nucleotide polymorphisms or insertion-deletion mutations) are identified by sequencing the entire genome, and this approach has been used to

TABLE 1: Comparison of representative NGS platforms.

Platform	Company	Detection	Run time	Read length (bp)
454 GS FLX Titanium XL+	Roche	Pyrosequencing	23 hours	700
454 GS Junior System	Roche	Pyrosequencing	10 hours	400
HiSeq 2000/2500	Illumina	Fluorescence	12 days	$2 \times 100$
MiSeq	Illumina	Fluorescence	65 hours	$2 \times 300$
Ion torrent	Life Technologies	Proton release	3 hours	35–400
Ion Proton	Life Technologies	Proton release	4 hours	125
Abi/solid	Life Technologies	Fluorescence	10 days	50
PacBio RSII	Pacific Bioscience	Fluorescence	2 days	–8500

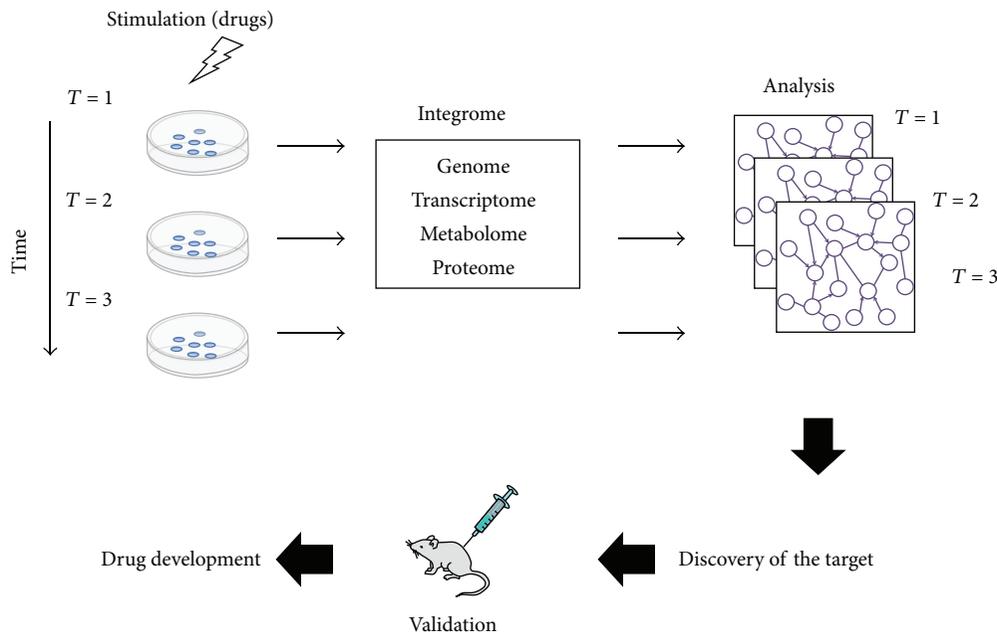


FIGURE 1: The dynamics of pharmacological response mechanisms can be examined by analyzing integrated multiomics data. First, the time series of the multiomics data are integrated. Second, an efficient module-detecting algorithm is applied to the composite maps. The maps can then be used for comparing cancer cells and normal cells and for assessing the effects of medicines. Lastly, the identified targets can be validated in animal experiments designed for the purpose of subsequent drug development.

identify several somatic mutations in various cancers [4, 5].

The use of whole-exome sequencing, which is employed for analyzing exon regions, has identified numerous mutations that occur in disease, such as *BRAF* mutations in papillary craniopharyngiomas [6] and somatic mutations of *BCOR* in myeloid leukemia [7]. Furthermore, this approach has been used for analyzing tumor borders and for detecting the *BRAF* mutation characteristic to borderline tumors. By this approach, 15 novel somatic mutations were detected in serous borderline tumors of the ovary [8]. Thus, genomic analysis performed using NGS provides extensive information about somatic mutations. In addition to genomic analysis, epigenetic analyses are performed using NGS. DNA methylation is involved in transcriptional regulation and it potently affects disease progression. One of the methods used for analyzing the methylation status of DNA (in particular, the methylation of cytosine residues) is bisulfite sequencing. This application

was developed based on exploiting the feature that bisulfite treatment converts all residues except methylated cytosine into uracil. The use of bisulfite sequencing has yielded key information regarding the epigenome in the context of cancer and other diseases [9, 10]. Thus, analyzing DNA methylation is critical in the field of epigenetics.

Chromatin immunoprecipitation sequencing (ChIP-seq) can be used for detecting protein binding to target DNA sequences and histone modifications. The method enables analyses of transcription factor binding to gene promoters and epigenetic modifications (e.g., histone modifications) [11, 12]. Moreover, the chromosome conformation capture (3C) method [13, 14] is used for detecting protein-DNA interaction-mediated spatial chromosome proximity, which is involved in transcriptional regulation and co-expression. Studies in which genome-wide 3C methods were employed together with NGS, such as chromosome conformation capture-on-chip (4C) [15], Hi-C [16], and tethered

TABLE 2: Types and features of next-generation sequencing technologies.

Type of analysis	Type of sequencing	Feature
Genome	Whole-genome sequencing	Used to detect somatic mutations by sequencing the whole genome
	Whole-exome sequencing	Used to detect somatic mutations by sequencing the whole exon region
Epigenome	Bisulfite sequencing	Used for analyzing methylation by sequencing genome exhaustively
	ChIP-seq	Used to detect the targets of transcription factors or analysis of histone modifications
	DNase-seq	
	FAIRE-seq	Used for analysis of chromatin architecture
	Hi-C	
	ChIA-PET	Used to characterize chromatin interactions that are mediated by nuclear protein of interest
Transcriptome	RNA sequencing	Used for analysis of gene expression or detection of fusion genes and splice variants
Interactome	IVV-HiTSeq	Used to detect reliable protein (domain) interactome without cloning including interactions of protein-protein/DNA/RNA/metabolic compounds/small molecules/drugs and so forth, suitable for high-throughput application, acquisition of high-reliability datasets, and analysis of cytotoxic proteins
	Y2H-seq	Used to detect interacting proteins or protein-domain pairs, but mating and the following diploid culture become the rate-limiting steps when applied in high-throughput technologies

ChIP-seq: chromatin immunoprecipitation sequencing; FAIRE-seq: formaldehyde-assisted isolation of regulatory elements sequencing; ChIA-PET: chromatin interaction analysis by means of paired-end tag sequencing; IVV-HiTSeq: IVV high-throughput sequencing; Y2H-seq: yeast two-hybrid interaction screening approach involving short-read second-generation sequencing.

conformation capture (TCC) [17], have shown that the spatial architecture of interphase chromosomes is closely related to DNA-replication timing, activity of genes, and cell differentiation (reviewed in [18]). Chromatin interaction analysis by means of paired-end tag sequencing (ChIA-PET), which is regarded as a combination of ChIP-seq and 3C, has been used for detecting the chromatin organization that is caused by a specific transcription factor [19, 20]. Recently, ChIA-PET studies performed on RNA polymerase II, which is present in the transcription preinitiation complex, comprehensively revealed active promoters, the transcription factors involved in their activation, and the spatial relationships among them [21, 22]. Thus, DNA sequencing has facilitated advances in both genomic and epigenomic analyses.

### 3. Transcriptomics

Similar to the manner in which DNA sequencing analysis has contributed to genomics and epigenomics, RNA sequencing (RNA-seq) has contributed to transcriptome analysis (Table 2). RNA-seq is an RNA-sequencing technology that is mainly used for sequencing mRNAs or long noncoding RNAs (lncRNAs). The mRNA in cells is analyzed in order to quantify gene expression or to detect fusion genes and splice variants in various cancers [23–25]. RNA-seq has also been used for studying gene expression patterns unique to certain cancers, including lung and renal carcinomas, and this has enabled researchers to identify novel biomarkers for

specific types of cancer [24, 26, 27]. Microarrays are also used for analyzing gene expression, but RNA-seq differs from that approach in the following manner: using RNA-seq, absolute quantification of expression is performed, whereas, using microarrays, relative expression is calculated. Moreover, an additional advantage of RNA-seq is that it can be used for detecting unknown transcription products, whereas microarrays cannot be used for this purpose. Certain transcription products have been detected using RNA-seq, and, in a few recent studies, RNA-seq was used for analyzing lncRNAs.

Thus, advances in this field of study have been made possible by the use of NGS-mediated analysis of the transcriptome. NGS has also already been successfully used for detecting mutations in cancer genes, and future research is expected to identify more of these mutations, which might be of therapeutic value.

### 4. Metabolomics

Metabolomics differs from nucleic acid-based-omics methods. Using metabolomics approaches, metabolites contained in a sample can be detected and their concentrations can be determined. This strategy is based on the premise that differences in metabolites reflect differences in biological processes.

Shifts in metabolite composition and changes at the genetic level enable the screening of potential biomarker candidates or therapeutic targets. For instance, high levels of

TABLE 3: Comparison of major metabolomics methods.

Method	Benefit	Drawback
NMR	(1) Results obtained in a single experiment (2) Fast (3) Available for studying various nuclei (4) Consistent with liquid and solid matrices (5) Quantitative (6) Sample can be recovered after analysis	(1) Requires highly skilled technicians and statisticians (2) Limited sensitivity (1–10 $\mu\text{mol/L}$ ) (3) Coresonant metabolites can be challenging to quantify
GC-MS	(1) High accuracy and repeatability of results (2) Small amounts of samples are required (3) High discrimination between molecules exhibiting very similar structures (4) High sensitivity (1 pmol/L)	(1) Sample preparation (including derivatization) can be time-consuming (2) Not all compounds are suitable for gas chromatography
LC-MS	(1) Wide application (2) High sensitivity (1 pmol/L)	(1) Requires extensive sample preparation, including derivatization (2) Long analytical time (20–60 min per sample) (3) Limited to volatile compounds (4) Suffers from ion suppression
CE-MS	(1) Suitable for polar molecules (2) Large separation capacity (3) High sensitivity (1 pmol/L)	(1) Low repeatability

reactive carbonyl compounds and low levels of vitamin B<sub>6</sub> are observed in the plasma of patients with certain subtypes of schizophrenia [28], suggesting that the use of the carbonyl-scavenger pyridoxamine might provide therapeutic benefits for these patients [29]. The relationship between cancer and changes in metabolites is widely recognized. For example, the Warburg effect describes the process whereby cancer cells preferably use the glycolytic pathway to produce ATP, even when sufficient oxygen is present [30]. The recent accumulation of knowledge based on metabolomics could enable advances in early cancer detection. For example, metabolomics studies have revealed that the profile of free amino acids in plasma is altered in the presence of cancer [31]. This information might lead to the development of novel metabolomics-based screening for early detection of a malignancy.

Most methods used in metabolomics involve separation and detection processes (Table 3) [32–35]. Researchers have typically relied on chromatography—gas chromatography (GC) and high-performance liquid chromatography (HPLC)—and capillary electrophoresis (CE) for separation, whereas they have used nuclear magnetic resonance (NMR) or mass spectrometry (MS) for detection [36, 37]. However, the drawbacks of these approaches have led researchers to combine two or more methods (e.g., liquid chromatography and MS (LC-MS/MS) plus NMR) in metabolomics studies [38].

Metabolomics is divided roughly into two categories based on the experimental methods used: nontargeted and target-defined metabolomics. Nontargeted metabolome analysis is extremely attractive because this method can be used to identify an unknown metabolite and to concurrently determine its relative amount; thus, this method is suitable for nonbiased metabolite fingerprinting and diagnostic-marker

exploration. However, nontargeted analysis performed using a single routine method remains challenging. This is because the metabolome includes compounds that differ considerably in molecular weight, electric charge, and concentration. Furthermore, although reference mass-spectrum databases [39–43] have grown rapidly and NMR microassays [44] have been improved, the molecular structures of unknown compounds present in trace amounts cannot be readily determined. Conversely, in several cases, identifying and determining the concentrations of all metabolites is not necessary. For example, only approximately 3,000 types of compounds are currently recognized in relation to human disease [45]. Because MS/MS can be used for detecting hundreds of compounds present in a single extract [46], targeted analysis is more suitable than nontargeted analysis for certain types of application, such as when the research is focused on a specific metabolic pathway. Therefore, future analyses are likely to involve the use of wide-targeted approaches in which several targeted experiments are combined when target compounds are available.

## 5. Proteomics

Comprehensive proteome analysis includes expression proteomics and interaction proteomics. Expression proteomics reveals protein expression patterns in cells, and this approach has been typically used for analyzing the expression status of various proteins by using two-dimensional electrophoresis (differential display) and MS [47]. However, proteins typically do not act alone and must interact with other proteins in order to perform functions. Therefore, interactions between proteins should be analyzed comprehensively. Comprehensive analysis of protein-protein interactions (PPIs) is critical in the fields of proteomics, functional genomics, and

TABLE 4: Comparison of comprehensive protein-protein interaction analysis methods.

Method	Selection condition of PPIs	Library size	Cell cloning required	Next-generation sequencing
Y2H	<i>In vivo</i>	10 <sup>6</sup>	Yes	Applicable but limited
TAP-MS	<i>In vitro</i>	Living body sample	Yes	Inapplicable
Protein microarray	<i>In vitro</i>	Living body sample	No	Inapplicable
Shotgun proteomics	<i>In vitro</i>	Living body sample	No	Inapplicable
IVV	<i>In vitro</i>	10 <sup>12</sup>	No	Applicable and effective

See also Table 1.

Y2H: yeast two-hybrid; TAP-MS: tandem affinity purification-mass spectrometry; IVV: *in vitro* virus.

systems biology. PPIs are detected using methods that can be divided into *in vivo* and *in vitro* techniques (Table 4). Among these methods, *in vitro* virus (IVV) and yeast two-hybrid (Y2H) tests allow the application of interactome sequencing (Table 2).

**5.1. In Vivo Methods.** The Y2H assay is an *in vivo* approach used to detect PPIs [48]. The assay requires two protein domains: a DNA-binding domain and an activation domain that is involved in the activation of DNA transcription. These domains are necessary for the transcription of reporter genes [49, 50]. The Y2H assay allows PPIs to be directly recognized, although in the analysis performed using this method false-positive interactions can appear. The data generated using *in vivo* techniques contain extremely high levels of false positives and false negatives. For example, although the exact rate of false-positive results in Y2H experiments is unknown, early estimates were that these are as high as 70% [51]. False-positive rates in AP experiments could be as high as 77% [52].

**5.2. In Vitro Methods.** Traditionally, PPIs *in vitro* have been detected using the tandem affinity purification (TAP) method [53, 54]. In this method, a bait containing tags allows protein complexes to be purified. The purified complexes are identified using a separate approach, such as MS [53, 54]. TAP tags have been developed which can be used for studying complex *in vivo* PPIs without the requirement of prior knowledge [53]. The TAP tagging technology has been used for analyzing the interactome in yeast [55]. The method is based on the attachment of two tags to a protein of interest, which is followed by a two-step purification process [56]. Proteins complexed with the target protein can then be identified using MS [57] and sodium dodecyl sulfate-polyacrylamide gel electrophoresis. The main advantage of TAP-MS is that it can be used for analyzing PPIs comprehensively in the form of protein complexes (which the target protein would be a part of *in vivo*) through the identification of a wide variety of protein complexes [56].

The protein microarray method is currently being established as a powerful tool for detecting proteins, observing their expression levels, and probing their interactions and functions. A protein microarray is a glass plate on which single proteins are bound at distinct positions according to a defined procedure [58]. Protein microarrays have been

developed in order to allow an operator to process multiple samples in parallel by using an automated process; this enables efficient and sensitive high-throughput protein analysis.

Another method that allows high-throughput identification of PPIs in original extracts is shotgun proteomics [59]. In this technique, proteins are digested with a protease immediately after extraction and then the resulting peptides are separated using LC; subsequently, the amino acid sequences of the peptides are determined using MS/MS.

The development of the techniques described thus far has facilitated the high-throughput identification of protein interactions. However, the number of proteins that cannot be detected or identified even using these methods, such as the proteins that are expressed at extremely low levels or are highly insoluble, is considerably greater than the number of proteins that can be detected or identified (e.g., abundant, soluble proteins). Therefore, new technologies are required for performing highly efficient high-throughput analysis of numerous proteins.

The IVV-high-throughput sequencing (IVV-HiTSeq, Figure 2) method [60], which is a combination of NGS and IVV, has been developed with the aim of overcoming the aforementioned challenges. In the IVV method, protein interactions are selected under cell-free conditions [61–65], and the subsequent sequencing by means of NGS is not limited by cloning steps performed using any specific type of cell. Thus, using the IVV-HiTSeq method, large amounts of accurate protein-interaction data can potentially be generated. The cell-free aspect of the experimental procedure is one of the main advantages: it allows highly efficient production of interaction data. When IVV and HiTSeq are combined, no host cells are required for the purpose of DNA cloning, a step that previously limited the efficiency of screening and the number of interactions that could be examined. Furthermore, the IVV method can be used to select proteins from a cDNA library consisting of 10<sup>12</sup> molecules, which is beyond the capacity of conventional high-throughput protein-selection methods [66, 67]. The coverage of the interactome is expected to increase in line with the further increases expected in the NGS throughput. Importantly, this completely cell-free procedure will also allow cytotoxic proteins to be analyzed, which will make interactome analysis more comprehensive than it currently is. With respect to the accuracy of IVV-HiTSeq data, the use of library-specific barcoded primers and

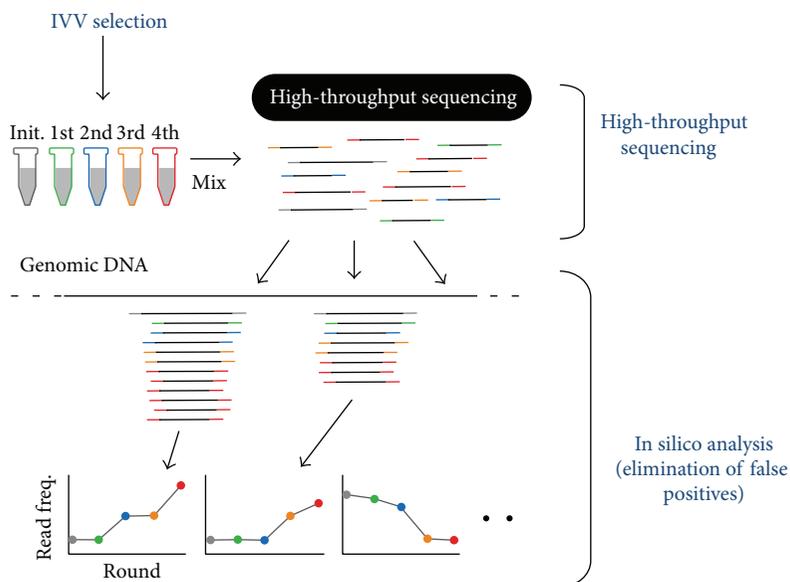


FIGURE 2: The type of primer used contains a barcoded region (indicated in grey, green, blue, yellow, and red), with four selection-round-specific bases. The reads generated using high-throughput sequencing are sorted according to their barcoded parts and mapped to known genomic sequences. Read frequencies of each genomic position are calculated for each selection round and used for determining the enriched regions. Using barcoded primers can reduce the risk of cross-contamination between libraries. Moreover, in a PPI analysis, increasing the sequencing depth can help detect contamination between samples. In the experiment shown, the Roche 454 Sequencer was used. Statistical significance was calculated by comparing the read frequencies with the frequencies of the initial library and the negative control.

*in silico* analysis reduces the number of false-positive interactions contained in the initial raw data [60]. IVV-HiTSeq was compared with conventional IVV performed using Sanger sequencing for the same prey library and bait. Whereas 640 sequences (87%) determined using Sanger sequencing were also obtained using IVV-HiTSeq, most of the sequences (99.7%) obtained using IVV-HiTSeq were new and were not detected using Sanger sequencing. Moreover, 88% of the real-time polymerase chain reaction (PCR) assays performed and followed up with the use of IVV-HiTSeq, which included *in silico* analyses, were positive. IVV-HiTSeq can potentially be applied to several cell-free display technologies, such as mRNA display, DNA display, and ribosome display. Moreover, the IVV test can be applied not only to *in vitro* selection of PPIs but also to the detection of protein-DNA, protein-RNA, and protein-chemical compound interactions [68]; this suggests that IVV-HiTSeq could become a universal tool for exploring protein sequences and interaction networks [47].

## 6. Discussion

To further our understanding of living cells, we must collect data by using multiomics analysis. Multiomics includes gene-, transcription-, metabolite-, and protein-specific information. By contrast, the interactome includes network data that are obtained based on direct interactions between molecules. Therefore, an integrated approach, which includes both interactome and multiomics data, is required for comparing the identities of individual cells. This approach is referred to as “integrome” analysis [69]. The integrome is a network map

of the interactome together with a list of multiomics data that allows analyses of differences between cancer cells and normal cells [70], the effects of treatments, and key factors such as biomarkers. Because PPIs lie at the core of biomolecular networks, an IVV system designed for detecting PPIs has been developed in an attempt to work toward personal genomics [71], and, using this system, noteworthy results of interactome analysis have been obtained. An advantage of IVV is the large library size (up to  $10^{12}$ ) that can be analyzed. This large size of the library increases the probability of selecting extremely rare sequences, and it also enhances the diversity of the selected sequences.

The use of NGS will allow the potential of IVV to be maximized. The latest Roche 454 Sequencer can sequence approximately  $10^6$  reads, at a rate of approximately 1000 bp per run. This capacity is sufficient for covering both of the linked variable regions. Although  $10^6$  reads are not adequate for covering the entire selected IVV library, this advance in NGS might make it possible to obtain unique, high-affinity binders.

Improvements in NGS might also facilitate the development of additional applications, and particularly notable is the speed with which NGS technology is being improved. When we have the capability to sequence the entire selected IVV library, we should be able to select low-affinity ligands that are commonly lost in a typical selection of repetitious rounds. IVV libraries will be subjected to high-throughput sequencing performed using NGS in order to generate interactome information, which will facilitate the archiving of the interactome map of a whole-cell library at a low cost.

The use of IVV systems could make key contributions to our understanding of the interactome networks in cells and thus help in the development of pharmaceutical agents for treating currently intractable diseases [71].

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## References

- [1] E. R. Mardis, "The impact of next-generation sequencing technology on genetics," *Trends in Genetics*, vol. 24, no. 3, pp. 133–141, 2008.
- [2] M. L. Metzker, "Sequencing technologies the next generation," *Nature Reviews Genetics*, vol. 11, no. 1, pp. 31–46, 2010.
- [3] H. P. Buermans and J. T. den Dunnen, "Next generation sequencing technology: advances and applications," *Biochimica et Biophysica Acta—Molecular Basis of Disease*, vol. 1842, no. 10, pp. 1932–1941, 2014.
- [4] J.-B. Cazier, S. R. Rao, C. M. McLean et al., "Whole-genome sequencing of bladder cancers reveals somatic *CDKN1A* mutations and clinicopathological associations with mutation burden," *Nature Communications*, vol. 5, article 3756, 2014.
- [5] Y. Zhao, J. Yang, Z. Chen et al., "Identification of somatic alterations in stage I lung adenocarcinomas by next-generation sequencing," *Genes Chromosomes and Cancer*, vol. 53, no. 4, pp. 289–298, 2014.
- [6] P. K. Brastianos, A. Taylor-Weiner, P. E. Manley et al., "Exome sequencing identifies BRAF mutations in papillary craniopharyngiomas," *Nature Genetics*, vol. 46, no. 2, pp. 161–165, 2014.
- [7] V. Grossmann, E. Tiacci, A. B. Holmes et al., "Whole-exome sequencing identifies somatic mutations of *BCOR* in acute myeloid leukemia with normal karyotype," *Blood*, vol. 118, no. 23, pp. 6153–6163, 2011.
- [8] J. Boyd, B. Luo, S. Peri et al., "Whole exome sequence analysis of serous borderline tumors of the ovary," *Gynecologic Oncology*, vol. 130, no. 3, pp. 560–564, 2013.
- [9] A. Adey and J. Shendure, "Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing," *Genome Research*, vol. 22, no. 6, pp. 1139–1143, 2012.
- [10] S. P. Kabekkodu, S. Bhat, R. Radhakrishnan et al., "DNA promoter methylation-dependent transcription of the double C2-like domain  $\beta$  (*DOC2B*) gene regulates tumor growth in human cervical cancer," *The Journal of Biological Chemistry*, vol. 289, no. 15, pp. 10637–10649, 2014.
- [11] S. Mahony, M. D. Edwards, E. O. Mazzoni et al., "An integrated model of multiple-condition ChIP-Seq data reveals predeterminants of Cdx2 binding," *PLoS Computational Biology*, vol. 10, no. 3, Article ID e1003501, 2014.
- [12] H. Pick, S. Kilic, and B. Fierz, "Engineering chromatin states: chemical and synthetic biology approaches to investigate histone modification function," *Biochimica et Biophysica Acta*, vol. 1839, no. 8, pp. 644–656, 2014.
- [13] J. Dekker, K. Rippe, M. Dekker, and N. Kleckner, "Capturing chromosome conformation," *Science*, vol. 295, no. 5558, pp. 1306–1311, 2002.
- [14] B. Tolhuis, R. J. Palstra, E. Splinter, F. Grosveld, and W. de Laat, "Looping and interaction between hypersensitive sites in the active  $\beta$ -globin locus," *Molecular Cell*, vol. 10, no. 6, pp. 1453–1465, 2002.
- [15] E. Splinter, E. de Wit, E. P. Nora et al., "The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA," *Genes & Development*, vol. 25, no. 13, pp. 1371–1383, 2011.
- [16] E. Lieberman-Aiden, N. L. van Berkum, L. Williams et al., "Comprehensive mapping of long-range interactions reveals folding principles of the human genome," *Science*, vol. 326, no. 5950, pp. 289–293, 2009.
- [17] R. Kalthor, H. Tjong, N. Jayathilaka, F. Alber, and L. Chen, "Genome architectures revealed by tethered chromosome conformation capture and population-based modeling," *Nature Biotechnology*, vol. 30, no. 1, pp. 90–98, 2012.
- [18] W. A. Bickmore and B. Van Steensel, "Genome architecture: domain organization of interphase chromosomes," *Cell*, vol. 152, no. 6, pp. 1270–1284, 2013.
- [19] M. J. Fullwood, M. H. Liu, Y. F. Pan et al., "An oestrogen-receptor- $\alpha$ -bound human chromatin interactome," *Nature*, vol. 462, no. 7269, pp. 58–64, 2009.
- [20] J. Zhang, H. M. Poh, S. Q. Peh et al., "ChIA-PET analysis of transcriptional chromatin interactions," *Methods*, vol. 58, no. 3, pp. 289–299, 2012.
- [21] Y. Zhang, C.-H. Wong, R. Y. Birnbaum et al., "Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations," *Nature*, vol. 504, no. 7479, pp. 306–310, 2013.
- [22] K. R. Kieffer-Kwon, Z. Tang, E. Mathe et al., "Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation," *Cell*, vol. 155, no. 7, pp. 1507–1520, 2013.
- [23] C. A. Maher, C. Kumar-Sinha, X. Cao et al., "Transcriptome sequencing to detect gene fusions in cancer," *Nature*, vol. 458, no. 7234, pp. 97–101, 2009.
- [24] Z.-H. Yang, R. Zheng, Y. Gao, Q. Zhang, and H. Zhang, "Abnormal gene expression and gene fusion in lung adenocarcinoma with high-throughput RNA sequencing," *Cancer Gene Therapy*, vol. 21, no. 2, pp. 74–82, 2014.
- [25] J. Eswaran, A. Horvath, S. Godbole et al., "RNA sequencing of cancer reveals novel splicing alterations," *Scientific Reports*, vol. 3, article 1689, 2013.
- [26] S. S. Han, W. J. Kim, Y. Hong et al., "RNA sequencing identifies novel markers of non-small cell lung cancer," *Lung Cancer*, vol. 84, no. 3, pp. 229–235, 2014.
- [27] Z. Xiong, H. Yu, Y. Ding et al., "RNA sequencing reveals upregulation of *RUNX1-RUNX1T1* gene signatures in clear cell renal cell carcinoma," *BioMed Research International*, vol. 2014, Article ID 450621, 8 pages, 2014.
- [28] M. Arai, H. Yuzawa, I. Nohara et al., "Enhanced carbonyl stress in a subpopulation of schizophrenia," *Archives of General Psychiatry*, vol. 67, no. 6, pp. 589–597, 2010.
- [29] M. Miyashita, M. Arai, A. Kobori et al., "Clinical features of schizophrenia with enhanced carbonyl stress," *Schizophrenia Bulletin*, vol. 40, no. 5, pp. 1040–1046, 2014.
- [30] O. Warburg, "On the origin of cancer cells," *Science*, vol. 123, no. 3191, pp. 309–314, 1956.
- [31] Y. Miyagi, M. Higashiyama, A. Gochi et al., "Plasma free amino acid profiling of five types of cancer patients and its application for early detection," *PLoS ONE*, vol. 6, no. 9, Article ID e24143, 2011.
- [32] M. Mussap, R. Antonucci, A. Noto, and V. Fanos, "The role of metabolomics in neonatal and pediatric laboratory medicine," *Clinica Chimica Acta*, vol. 426, pp. 127–138, 2013.

- [33] J. L. Griffin and J. P. Shockcor, "Metabolic profiles of cancer cells," *Nature Reviews Cancer*, vol. 4, no. 7, pp. 551–561, 2004.
- [34] S. P. Putri, S. Yamamoto, H. Tsugawa, and E. Fukusaki, "Current metabolomics: technological advances," *Journal of Bioscience and Bioengineering*, vol. 116, no. 1, pp. 9–16, 2013.
- [35] P. A. Belyakov, V. I. Kadentsev, A. O. Chizhov, N. G. Kolotyrykina, A. S. Shashkov, and V. P. Ananikov, "Mechanistic insight into organic and catalytic reactions by joint studies using mass spectrometry and NMR spectroscopy," *Mendeleev Communications*, vol. 20, no. 3, pp. 125–131, 2010.
- [36] E. G. Armitage and C. Barbas, "Metabolomics in cancer biomarker discovery: current trends and future perspectives," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 87, pp. 1–11, 2014.
- [37] S. P. Putri, Y. Nakayama, F. Matsuda et al., "Current metabolomics: practical applications," *Journal of Bioscience and Bioengineering*, vol. 115, no. 6, pp. 579–589, 2013.
- [38] P. J. Eugster, G. Glauser, and J.-L. Wolfender, "Strategies in biomarker discovery. peak annotation by MS and targeted LC-MS micro-fractionation for de novo structure identification by micro-NMR," *Methods in Molecular Biology*, vol. 1055, pp. 267–289, 2013.
- [39] J. Hummel, N. Strehmel, J. Selbig, D. Walther, and J. Kopka, "Decision tree supported substructure prediction of metabolites from GC-MS profiles," *Metabolomics*, vol. 6, no. 2, pp. 322–333, 2010.
- [40] C. A. Smith, G. O'Maille, E. J. Want et al., "METLIN: a metabolite mass spectral database," *Therapeutic Drug Monitoring*, vol. 27, no. 6, pp. 747–751, 2005.
- [41] Q. Cui, I. A. Lewis, A. D. Hegeman et al., "Metabolite identification via the Madison Metabolomics Consortium Database," *Nature Biotechnology*, vol. 26, no. 2, pp. 162–164, 2008.
- [42] D. S. Wishart, D. Tzur, C. Knox et al., "HMDB: the human metabolome database," *Nucleic Acids Research*, vol. 35, no. 1, pp. D521–D526, 2007.
- [43] H. Horai, M. Arita, S. Kanaya et al., "MassBank: a public repository for sharing mass spectral data for life sciences," *Journal of Mass Spectrometry*, vol. 45, no. 7, pp. 703–714, 2010.
- [44] S. S. Zalesskiy, E. Danieli, B. Blümich, and V. P. Ananikov, "Miniaturization of NMR Systems: desktop spectrometers, microcoil spectroscopy, and "NMR on a chip" for chemistry, biochemistry, and industry," *Chemical Reviews*, vol. 114, no. 11, pp. 5641–5694, 2014.
- [45] D. S. Wishart, T. Jewison, A. C. Guo et al., "HMDB 3.0—the human metabolome database in 2013," *Nucleic Acids Research*, vol. 41, no. 1, pp. D801–D807, 2013.
- [46] I. M. Abu-Reidah, D. Arráez-Román, A. Segura-Carretero, and A. Fernández-Gutiérrez, "Profiling of phenolic and other polar constituents from hydro-methanolic extract of watermelon (*Citrullus lanatus*) by means of accurate-mass spectrometry (HPLC-ESI-QTOF-MS)," *Food Research International*, vol. 51, no. 1, pp. 354–362, 2013.
- [47] S. Wittke, T. Kaiser, and H. Mischak, "Differential polypeptide display: the search for the elusive target," *Journal of Chromatography B*, vol. 803, no. 1, pp. 17–26, 2004.
- [48] P. Uetz, L. Glot, G. Cagney et al., "A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*," *Nature*, vol. 403, no. 6770, pp. 623–627, 2000.
- [49] T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki, "A comprehensive two-hybrid analysis to explore the yeast protein interactome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 8, pp. 4569–4574, 2001.
- [50] P. James, J. Halladay, and E. A. Craig, "Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast," *Genetics*, vol. 144, no. 4, pp. 1425–1436, 1996.
- [51] C. M. Deane, Ł. Salwiński, I. Xenarios, and D. Eisenberg, "Protein interactions: two methods for assessment of the reliability of high throughput observations," *Molecular and Cellular Proteomics*, vol. 1, no. 5, pp. 349–356, 2002.
- [52] A. M. Edwards, B. Kus, R. Jansen, D. Greenbaum, J. Greenblatt, and M. Gerstein, "Bridging structural biology and genomics: Assessing protein interaction data with known complexes," *Trends in Genetics*, vol. 18, no. 10, pp. 529–536, 2002.
- [53] G. Rigaut, A. Shevchenko, B. Rutz, M. Wilm, M. Mann, and B. Seraphin, "A generic protein purification method for protein complex characterization and proteome exploration," *Nature Biotechnology*, vol. 17, no. 10, pp. 1030–1032, 1999.
- [54] O. Puig, F. Caspary, G. Rigaut et al., "The tandem affinity purification (TAP) method: a general procedure of protein complex purification," *Methods*, vol. 24, no. 3, pp. 218–229, 2001.
- [55] A. C. Gavin, M. Bösch, R. Krause et al., "Functional organization of the yeast proteome by systematic analysis of protein complexes," *Nature*, vol. 415, no. 6868, pp. 141–147, 2002.
- [56] S. Pitre, M. Alamgir, J. R. Green, M. Dumontier, F. Dehne, and A. Golshani, "Computational methods for predicting protein-protein interactions," *Advances in Biochemical Engineering/Biotechnology*, vol. 110, pp. 247–267, 2008.
- [57] J. S. Rohila, M. Chen, R. Cerny, and M. E. Fromm, "Improved tandem affinity purification tag and methods for isolation of protein heterocomplexes from plants," *Plant Journal*, vol. 38, no. 1, pp. 172–181, 2004.
- [58] G. MacBeath and S. L. Schreiber, "Printing proteins as microarrays for high-throughput function determination," *Science*, vol. 289, no. 5485, pp. 1760–1763, 2000.
- [59] J. J. Moresco, P. C. Carvalho, and J. R. Yates, "Identifying components of protein complexes in *C. elegans* using co-immunoprecipitation and mass spectrometry," *Journal of Proteomics*, vol. 73, no. 11, pp. 2198–2204, 2010.
- [60] S. Fujimori, N. Hirai, H. Ohashi et al., "Next-generation sequencing coupled with a cell-free display technology for high-throughput production of reliable interactome data," *Scientific Reports*, vol. 2, article 691, 2012.
- [61] E. Miyamoto-Sato, S. Fujimori, M. Ishizaka et al., "A comprehensive resource of interacting protein regions for refining human transcription factor networks," *PLoS ONE*, vol. 5, no. 2, Article ID e9289, 2010.
- [62] E. Miyamoto-Sato, M. Ishizaka, K. Horisawa et al., "Cell-free cotranslation and selection using in vitro virus for high-throughput analysis of protein-protein interactions and complexes," *Genome Research*, vol. 15, no. 5, pp. 710–717, 2005.
- [63] E. Miyamoto-Sato and H. Yanagawa, "Toward functional analysis of protein interactome using "in vitro virus": in silico analyses of Fos/Jun interactors," *Journal of Drug Targeting*, vol. 14, no. 7, pp. 505–511, 2006.
- [64] E. Miyamoto-Sato, "Next-generation sequencing coupled with a cell-free display technology for reliable interactome of translational factors," *Methods in Molecular Biology*, vol. 1164, pp. 23–32, 2014.
- [65] H. Ohashi, S. Fujimori, N. Hirai, H. Yanagawa, and E. Miyamoto-Sato, "Analysis of transcription factor networks using IVV method," *Methods in Molecular Biology*, vol. 1164, pp. 15–22, 2014.

- [66] R. W. Roberts, "Totally in vitro protein selection using mRNA-protein fusions and ribosome display," *Current Opinion in Chemical Biology*, vol. 3, no. 3, pp. 268–273, 1999.
- [67] H. Wang and R. Liu, "Advantages of mRNA display selections over other selection techniques for investigation of protein-protein interactions," *Expert Review of Proteomics*, vol. 8, no. 3, pp. 335–346, 2011.
- [68] T. T. Takahashi, R. J. Austin, and R. W. Roberts, "mRNA display: ligand discovery, interaction analysis and beyond," *Trends in Biochemical Sciences*, vol. 28, no. 3, pp. 159–165, 2003.
- [69] M. Baker, "Big biology: the 'omes puzzle," *Nature*, vol. 494, no. 7438, pp. 416–419, 2013.
- [70] K. Dimitrakopoulou, G. N. Dimitrakopoulos, K. N. Sgarbas, and A. Bezerianos, "Tamoxifen integromics and personalized medicine: dynamic modular transformations underpinning response to tamoxifen in breast cancer treatment," *OMICS: A Journal of Integrative Biology*, vol. 18, no. 1, pp. 15–33, 2014.
- [71] H. Ohashi and E. Miyamoto-Sato, "Towards personalized medicine mediated by in vitro virus-based interactome approaches," *International Journal of Molecular Sciences*, vol. 15, no. 4, pp. 6717–6724, 2014.

## Research Article

# Targeted Next Generation Sequencing Identifies Novel Mutations in *RP1* as a Relatively Common Cause of Autosomal Recessive Rod-Cone Dystrophy

Said El Shamieh,<sup>1,2,3</sup> Elise Boulanger-Scemama,<sup>1,2,3</sup> Marie-Elise Lancelot,<sup>1,2,3</sup>  
Aline Antonio,<sup>1,2,3</sup> Vanessa Démontant,<sup>1,2,3</sup> Christel Condroyer,<sup>1,2,3</sup> Mélanie Letexier,<sup>4</sup>  
Jean-Paul Saraiva,<sup>4</sup> Saddek Mohand-Saïd,<sup>1,2,3,5</sup> José-Alain Sahel,<sup>1,2,3,5,6,7,8</sup>  
Isabelle Audo,<sup>1,2,3,5,8</sup> and Christina Zeitz<sup>1,2,3</sup>

<sup>1</sup>INSERM, U968, 75012 Paris, France

<sup>2</sup>Sorbonne Universités, UPMC University, Paris 06, UMR\_S 968, Institut de la Vision, 75012 Paris, France

<sup>3</sup>CNRS, UMR\_7210, 75012 Paris, France

<sup>4</sup>IntegraGen SA, Genopole Campus 1, Building G8, 91030 Evry, France

<sup>5</sup>Centre Hospitalier National d'Ophthalmologie des Quinze-Vingts, DHU ViewMaintain, INSERM-DHOS CIC 1423, 75012 Paris, France

<sup>6</sup>Fondation Ophthalmologique Adolphe de Rothschild, 75019 Paris, France

<sup>7</sup>Académie des Sciences-Institut de France, 75006 Paris, France

<sup>8</sup>University College London Institute of Ophthalmology, 11-43 Bath Street, London EC1V 9EL, UK

Correspondence should be addressed to Isabelle Audo; [isabelle.audo@inserm.fr](mailto:isabelle.audo@inserm.fr) and Christina Zeitz; [christina.zeitz@inserm.fr](mailto:christina.zeitz@inserm.fr)

Received 27 June 2014; Accepted 10 July 2014

Academic Editor: Calvin Yu-Chian Chen

Copyright © 2015 Said El Shamieh et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We report ophthalmic and genetic findings in families with autosomal recessive rod-cone dystrophy (arRCD) and *RP1* mutations. Detailed ophthalmic examination was performed in 242 sporadic and arRCD subjects. Genomic DNA was investigated using our customized next generation sequencing panel targeting up to 123 genes implicated in inherited retinal disorders. Stringent filtering coupled with Sanger sequencing and followed by cosegregation analysis was performed to confirm biallelism and the implication of the most likely disease causing variants. Sequencing identified 9 *RP1* mutations in 7 index cases. Eight of the mutations were novel, and all cosegregated with severe arRCD phenotype, found associated with additional macular changes. Among the identified mutations, 4 belong to a region, previously associated with arRCD, and 5 others in a region previously associated with adRCD. Our prevalence studies showed that *RP1* mutations account for up to 2.5% of arRCD. These results point out for the necessity of sequencing *RP1* when genetically investigating sporadic and arRCD. It further highlights the interest of unbiased sequencing technique, which allows investigating the implication of the same gene in different modes of inheritance. Finally, it reports that different regions of *RP1* can also lead to arRCD.

## 1. Introduction

Rod-cone dystrophy (RCD), also known as retinitis pigmentosa, is a heterogeneous group of inherited disorders affecting primary rod photoreceptors in the majority of cases with secondary cone degeneration [1, 2]. Population-based studies showed that 1 in 4,000 individuals is affected around the

world [1]. Patients diagnosed with RCD initially complain of night blindness due to rod dysfunction followed by progressive visual field constriction, abnormal color vision, and eventually loss of central vision due to cone photoreceptor involvement [1].

RCD is inherited as a Mendelian trait in most cases [3]. On the basis of its mode of inheritance and prevalence,

RCD can be divided into 3 groups: autosomal dominant (ad) (30–40%), autosomal recessive (ar) (50–60%), and X-linked (xl) (5–15%) [3]. To date, mutations in at least 53 genes were reported to cause nonsyndromic RCD (till 25 June 2014, <https://sph.uth.edu/retnet/>). Prevalence studies revealed *rhodopsin* (*RHO*), *retinitis pigmentosa GTPase regulator* (*RPGR*), and *usherin* (*USH2A*) as being the most frequently mutated genes in adRCD [4, 5], xlRCD [4], and arRCD, respectively [6]. Of note is that many other genes with lower prevalence are also implicated in the genetic etiology of RCD [7, 8]. Mutations in *RPI* were first shown to cause adRCD [9–11]; however, since 2005, articles have shed light on its implication in arRCD etiology [12–20]. *RPI* mutations were shown to account for  $\approx 5.5\%$  and  $\approx 1\%$  of adRCD and arRCD cases, respectively [8–20]. Interestingly, Avila-Fernandez et al. [12] reported that a founder nonsense mutation in the Spanish population p.Ser542\* is responsible for 4.5% of arRCD cases suggesting that *RPI* mutations are more prevalent in arRCD than previously thought [12].

Retinitis pigmentosa 1 (*RPI*) is a photoreceptor-specific gene encoding a protein regulated by oxygen [10]. *RPI* protein is required for correct orientation and higher-order stacking of outer segment disks [21] and was shown to be part of the photoreceptor axoneme [22]. *RPI* localizes to the connecting cilia of photoreceptors and may assist in maintenance of ciliary structure or transport down the photoreceptor [22]. Like many retinal degeneration genes, the mechanism by which mutations in *RPI* lead to photoreceptor cell death is still unclear.

We developed an unbiased and time-efficient retinal gene next generation sequencing array (NGS), which was further revised and improved to target more than 120 genes implicated in inherited retinal diseases (IRDs) (list available upon request) [23]. Using this NGS panel, we screened a total of 242 subjects with sporadic and recessive RCD in order to detect disease causing mutations and to report the prevalence of pathogenic mutations in *RPI* causing arRCD.

## 2. Methods

**2.1. Ethics Statement and Clinical Diagnosis of Rod-Cone Dystrophy.** The study protocol adhered to the tenets of the Declaration of Helsinki and was approved by the local Ethics Committee (CPP, Ile de France V). Informed written consent was obtained from each study participant. Index patients underwent full ophthalmic examination as previously described [23].

**2.2. Targeted Next Generation Sequencing.** A cohort of 242 subjects affected with sporadic and arRCD was investigated in the present study. Prior to NGS screening, molecular genetic analysis with microarray (Asper Ophthalmics, Tartu, Estonia), followed by direct Sanger sequencing of *EYS* and *C2orf71* (major and minor genes implicated in RCD, newly discovered at the beginning of our study), was performed in 201 index subjects (82%) [2, 24]. As *RPGR* exon *ORF15* (MIM 312610) is not targeted by existing NGS panels, we excluded mutations in this “hot spot” by Sanger sequencing.

Although our NGS panel was selected from the SureSelect Human All Exon Kits Version 4 (Agilent, Massy, France), this design was improved after analyzing the first 83 subjects with sporadic and arRCD. More precisely, a total of  $\approx 300$  Kb regions were added in order to cover all the previously nontargeted regions. Thus, whereas the first design covered the exons and the flanking intronic regions of 120 genes implicated in IRDs, the second covered 123 genes in total. The eArray web-based probe design tool was used for this purpose (<https://earray.chem.agilent.com/earray>). All probes were designed and synthesized by Agilent Technologies (Santa Clara, CA, USA). Sequence capture, enrichment, and elution were performed according to Agilent’s instructions. The complete details were described elsewhere [23].

**2.3. Assembly and Variant Calling.** Sequence reads were aligned to the reference human genome (UCSC hg19) using CASAVA1.7 software (Illumina) and the ELANDv2 alignment algorithm. Sequence variation annotation was performed using the IntegraGen in-house pipeline, which consisted of gene annotation (RefSeq), detection of known single nucleotide polymorphisms (dbSNP 135) followed by mutation characterization (missense, intronic, synonymous, nonsense, splice site, and insertions/deletions).

**2.4. Quality Control and Coverage Assessment.** The first NGS retinal panel harbored 120 IRDs genes, encompassing 321,240 kb length per sample. However, after improvement, the same panel contained  $\approx 600$  Kb and covered 123 IRD genes. The depth of coverage was calculated by counting the number of sequenced bases mapping to the target regions. Mean depth of coverage was calculated per base pair for all samples; however, only the results of subjects having *RPI* mutations were shown.

**2.5. Discrete Filtering of Annotated Variants.** In order to identify disease causing mutations among nonpathogenic single nucleotide polymorphisms, we used a filtering approach against a set of polymorphisms that are available in the public databases: dbSNP 137, 1000 Genomes Project [25], HapMap [26], and Exome Variant Server [27] with removal of variants with a minor allele frequency (MAF)  $\geq 0.005$  in case of presumed autosomal recessive mode of inheritance.

**2.6. Pathogenicity Assessment.** We stratified candidate mutations based on their functional class by giving a priority to frameshifts, stop codons, and disruptions of canonical splice sites variants [28]. For missense changes, amino acid conservation across 46 different species was studied using the UCSC Genome Browser [29]. If no amino acid change was found, then the residue was considered as “highly conserved.” If a different change was seen in less than four species and not in the primates, then it was considered as “moderately conserved” and if a change was present in 5–7, it was considered as “marginally conserved”; otherwise, the amino acid residue was considered as “not conserved.” Pathogenic prediction was performed using two software programs: Polyphen2 [30] and SIFT [31], based on species/homologue conservation,

putative structural domains, and 3D structures (if available). Analysis of potential splice site variant consequences when relevant was done using human splicing finder [32].

**2.7. Known Genotype-Phenotype Correlations.** The search for previous genotype-phenotype associations was done by searching numerous literature databases, including Online Mendelian Inheritance in Man (<http://omim.org/>), Human Gene Mutation Database [33], Leiden Open Variation Database [34], and RetNet (<https://sph.uth.edu/retnet/>).

**2.8. Validation of the Identified Variants and Cosegregation Analyses.** Sanger sequencing was performed to validate disease causing mutations in *RPI*. The respective primer information can be communicated on request. In addition, blood samples were collected from additional family members when possible and cosegregation analyses on extracted DNA were performed as previously described [35, 36].

### 3. Results

**3.1. Clinical Data.** Clinical data are summarized in Table 1. Among identified patients, 5 were females, 2 were male, and ages at time of examination ranged from 25 to 42. All patients were diagnosed before age 20 mostly based on night blindness from early childhood and secondary central vision loss. They all showed severe RCD with constricted visual fields, no detectable responses on full field electroretinogram, and both peripheral involvement and macular involvement (Figure 1 presents fundus pictures of patient II.1 (CIC01245) in family F752 as an example). Comparing visual acuity and visual fields for these arRCD patients with those of adRCD cases published by Audo and coworkers [8], we noticed a more severe phenotype in recessive cases. However, more cases with *RPI* mutations would be needed to draw statistical conclusion.

**3.2. Sequencing Statistics.** In index patients, the overall sequencing coverage of the target regions was  $\geq 88\%$  for a 25X depth of coverage in each of the chromosomes (Figure 2(a)), resulting a mean sequencing depth of 299 times per base. Mean sequencing results per base in each target chromosome gene regions were shown in Figure 2(b). It is of importance to mention that  $<1\%$  of target regions were not covered at all. These were fragments of 120 bp belonging in 66% of the cases only to a fraction of an exon. The remaining uncovered targets corresponded each to an entire exon in genes such as *CHM*, *PDZD7*, *RP9*, *CC2D2A*, *IMPDH1*, *CNGA1*, and *EYS*.

**3.3. Detection of Disease Causing Mutations in *RPI* Gene.** After data filtering, the total number of putative disease causing variants was reduced by 99.3%. Thus, in total, filtering enriched the percentage of putative disease causing mutations from 0.7% (25/3339 variants) to 33.3% (9/25 variants) in the 7 subjects presented here (Table 2). These subjects exhibit *RPI* mutations in the last exon 4 that are predicted to lead to a premature stop codon. We found 9 pathogenic mutations in *RPI* among which one (p.Ser542\* in CIC00445) was already

reported by Avila-Fernandez et al. [12] as a founder nonsense mutation in the Spanish population, responsible for 4.5% of arRCD. Although F303 is from French origin, we cannot exclude the possibility of a founder effect of p.Ser542\* in our cohort.

Patient family F303: II.1 (CIC00445) was found to carry compound heterozygous variants: a nonsense mutation c.1625C>G, leading to a predicted premature stop (p.Ser542\*) and a deletion c.4587\_4590delTAAG leading to a frameshift and a premature termination codon (p.Ser1529Argfs\*9) (Table 2, Figure 3). Patient family F752: II.1 (CIC01245) was also found to carry compound heterozygous variants: a 1 bp duplication c.2025dupA leading to p.Ser676Ilefs\*22 and a 1 bp deletion c.2377delA leading to p.Arg793Glufs\*55 (Table 2).

Patients from family F335: III.1 (CIC00491), family F674: III.6 (CIC01106), family F782: II.5 (CIC01300), family F1941: III.1 (CIC04130), and family F3110: III.5 (CIC05941) were found to carry homozygous deletions c.4089\_4092delAAGA leading to p.Arg1364Valfs\*8; c.1205delG leading to p.Gly402Alafs\*7; c.1719\_1723delCTCAA leading to p.Ser574Cysfs\*7; c.1329delG leading to p.Lys443Asnfs\*12; and c.2391\_2392delAA leading to p.Asp799\*, resp.) (Table 2 and Figure 3). It is important to note that consanguinity was reported in families F335, F674, F782 and F1941.

All *RPI* mutations detected by NGS were further validated by Sanger sequencing. All variants cosegregated with the phenotype in available family members. Based on the previous findings, the measured prevalence of *RPI*-associated arRCD in this cohort is  $\approx 2.5\%$ .

### 4. Discussion

The current study further demonstrates the usefulness of NGS as a comprehensive genetic diagnostic tool for IRDs with further impact on patients counseling and participation for potential therapeutic trials. Our study applied to a large cohort of sporadic and autosomal recessive cases of RCD identifies 8 novel mutations in a gene not classically screened in arRCD by other methods such as Sanger sequencing or microarray analysis, outlining the interest of this massive parallel sequencing method. Consequently, a prevalence of *RPI* mutation in 2.5% of sporadic or arRCD cases in the European population is herein reported.

*RPI* is a 15 kb single copy gene clustering the small arm of the chromosome 8 (8q12.1). It encodes a 2506 amino acid protein having a molecular weight of 241 kDa containing a *Drosophila melanogaster* bifocal (BIF) (amino acid 486–635) and two doublecortin domains. Whereas the BIF domain helps to maintain the photoreceptor morphogenesis, doublecortin domains bind microtubules and regulate their polymerization [22]. Along with RPIL1 (Retinitis Pigmentosa 1-like 1, another retinal-specific protein), *RPI* plays essential and synergistic roles in outer segment morphogenesis of rod photoreceptors [22].

To date, at least 50 mutations in *RPI* were identified in RCD [8, 12–20], the majority of which are located in its

TABLE 1: Clinical data of the 7 index patients with *RP1* recessive mutations.

Patient	Age at time of testing	Age of onset	Sex	Family history	Symptoms at time of diagnosis	BCVA OD/OS With refraction	Color vision (15 desaturated Hue)	Binocular kinetic visual field (III4e stimulus)	FF and mfERG	Fundus examination	FAF	Sd-OCT
F303: II.1 (CIC00445)	42	6	F	No other affected FM, from France.	Night blindness	Hand motion in both eyes	Impossible due to low vision	Reduced to peripheral islands of perception	Both undetectable	Pale optic disc narrowed blood vessels, macular atrophic changes, and optic nerve drusen	Hypoautofluorescence in the macular region	Thinning of outer retina in the macular region
F335: III.1 (CIC00491)	36	3	M	Two other brothers affected; parents first cousins	Night blindness and rapid decreased vision	LP in both eyes	Impossible due to low vision	Impossible due to low vision	Both undetectable	Widespread RPE changes and retinal atrophy in both the periphery and the macular area	Widespread loss of FAF	Widespread thinning of outer retina
F674: III.6 (CIC01106)	25	19	F	Parents first cousins from Turkey, one female and male cousins affected also from a consanguineous union	Night blindness and decreased vision	HM -3 (-1) 0° 20/160 -3 (-0.50) 0°	Dyschromatopsia with no specific axis	Reduced to 5 central degrees	Both undetectable	Well-colored optic disc and no narrowing of retinal vessels; RPE changes in the periphery and macular atrophic changes	Hypoautofluorescence in the macular region and outside the vascular arcades	Thinning of outer retina in the macular region
F752: II.1 (CIC01245)	31	Early teens	F	Two sisters affected	Night blindness	20/63 plano (-3) 180° 20/50 plano (-1.75) 180°	Deutan defect on both eyes	Reduced to 10° × 20°	Both undetectable	Pale optic disc head, narrowed retinal vessels, and RPE changes in the periphery with some macular atrophic changes	Hypoautofluorescence in the macular region and outside the vascular arcades	Thinning of outer retina in the macular region
F782 II.5 (CIC01300)	27	9	M	Parents from Algeria, first cousins	Night blindness and decreased vision	20/50 -9.25 (-2.50) 15° 20/50 -9 (-1.75) 100°	—	Reduced to the 10 central degrees	Both undetectable	Mild optic disc pallor, atrophic macular changes, and peripheral pigment deposits	Hypoautofluorescence in the macular region	Thinning of outer retina in the macular region

TABLE 1: Continued.

Patient	Age at time of testing	Age of onset	Sex	Family history	Symptoms at time of diagnosis	BCVA OD/OS With refraction	Color vision (15 desaturated Hue)	Binocular kinetic visual field (III4e stimulus)	FF and mfERG	Fundus examination	EAF	Sd-OCT
F1941: III.1 (CIC04130)	30	childhood	F	Parents from Algeria, first cousins	Night blindness	20/100 -4.25 (-1.25) 150° 20/80 -4.25 (-1.25) 150°	Normal at the saturated test	Reduced to the 10 central degrees	Both undetectable	Well-colored optic disc but narrowed retinal vessels; RPE changes in the periphery and macular atrophic changes	Hypoautofluorescence in the macular region and outside the vascular arcades	Thinning of outer retina in the macular region
F3110: III.5 (CIC05941)	27	5	F	One cousin on mother side may have RCD	Night blindness and decreased vision	20/125 +2 (-2) 95° 20/125 +1.75 (-2) 70°c	Dyschromatopsia with no specific axis	Reduced to the 10 central degrees	Both undetectable	Pale optic disc, narrowed retinal vessels, and RPE changes in the periphery with some macular atrophic changes	Hypoautofluorescence in the macular region and outside the vascular arcades	Thinning of outer retina in the macular region

F: female, FM: family member, M: male, BCVA: best corrected visual acuity; OD: oculo dextra (right eye); OS: oculo sinistra (left eye); FF and mfERG: full-field and multifocal ERG; EAF: fundus autofluorescence; Sd-OCT: spectral domain optical coherence tomography; RPE: retinal pigment epithelium; LP: light perception; HM: hand motion.

TABLE 2: List of mutations detected by next generation sequencing after applying relevant filters.

Patient	Gene	Exon	Allele state	Nucleotide exchange	Protein effect	rs ID	Conservation	Polyphen 2	SIFT	Pathogenicity	Note
F303:	<i>NPHP4</i>	12	HTZ	A>G	p.Ser481Asn	no	NC	B	T	Neutral	
II.1 (CIC00445)	<i>RPI</i>	4	HTZ	<b>c.1625C&gt;G</b>	<b>p.Ser542*</b>	—	—	—	—	<b>Disease causing</b>	<b>R M [12]</b>
	<i>RPI</i>	4	HTZ	<b>c.4587_4590delTAAG</b>	<b>p.Ser1529Argfs*9</b>	—	—	—	—	<b>Disease causing</b>	<b>N M</b>
F335:	<i>PROM1</i>	4	HTZ	T>C	p.Ile178Val	—	NC	B	T	Neutral	
III.1 (CIC00491)	<i>GPR98</i>	29	HTZ	G>A	p.Arg2128Gln	rs149390094	NC	B	T	Neutral	
	<i>RPI</i>	4	HMZ	<b>c.4089_4092delAAGA</b>	<b>p.Arg1364Valfs*8</b>	—	—	—	—	<b>Disease causing</b>	<b>N M</b>
F674:	<i>USH2A</i>	39	HTZ	T>G	p.Ser2450Arg	No	HC	P D	D	<b>Probably disease causing</b>	
III.6 (CIC01106)	<i>RPI</i>	4	HMZ	<b>c.1205delG</b>	<b>p.Gly402Alafs*7</b>	—	—	—	—	<b>Disease causing</b>	<b>N M</b>
F752:	<i>USH1C</i>	17	HTZ	G>A	p.Arg477Trp	TMP_ESP_IL17532053	HC	P D	D	Probably disease causing	
II.1 (CIC01245)	<i>PDE6B</i>	10	HTZ	T>C	p.Thr432Ile	—	HC	B	T	Uncertain pathogenicity	
	<i>RPI</i>	4	HTZ	<b>c.2025dupA</b>	<b>p.Ser676Ilefs*22</b>	—	—	—	—	<b>Disease causing</b>	<b>N M</b>
	<i>RPI</i>	4	HTZ	<b>c.2377delA</b>	<b>p.Arg793Gluifs*55</b>	—	—	—	—	<b>Disease causing</b>	<b>N M</b>
F782: II.5 (CIC01300)	<i>RPI</i>	4	HMZ	<b>c.1719_1723delCTCAA</b>	<b>p.Ser574Cysfs*7</b>	—	—	—	—	<b>Disease causing</b>	<b>N M</b>
	<i>TULP1</i>	5	HMZ	<b>c.395_418dup</b>	<b>p.Asp124_Glu131del</b>	rs63749128	—	—	—	Neutral	
	<i>PCDH15</i>	33	HTZ	C>T	p.Arg1889His	rs145851144	NC	B	T	Neutral	
F1941: III.1 (CIC04130)	<i>C2orf71</i>	1	HTZ	C>A	p.Arg656Ser	rs201980758	NC	B	T	Neutral	
	<i>CACNA2D4</i>	Intron 37	HTZ	C>T	—	rs80092457	NC	—	—	Neutral	
	<i>RPI</i>	4	HMZ	<b>c.1329delG</b>	<b>p.Lys443Asnfs*12</b>	—	—	—	—	<b>Disease causing</b>	<b>N M</b>
	<i>EYS</i>	6	HTZ	C>T	p.Ser326Asn	rs112822256	NC	B	T	Neutral	
	<i>MERTK</i>	8	HTZ	C>G	p.Arg421Trp	rs138908058	NC	B	D	Neutral	
F3110:	<i>PRPF6</i>	21	HTZ	A>G	p.Val915Met	rs139778757	MC	P D	D	Uncertain pathogenicity	
III.5 (CIC05941)	<i>TULP1</i>	14	HTZ	G>A	p.Alala496Thr	rs141980901	MC	B	D	Neutral	
	<i>EYS</i>	26	HTZ	G>A	p.Lys1365Glu	rs16895519	NC	B	D	Neutral	
	<i>MERTK</i>	18	HTZ	G>C	p.Glu823Gln	rs55924349	MC	B	D	Neutral	
	<i>RPI</i>	4	HMZ	<b>c.2391_2392delAA</b>	<b>p.Asp799*</b>	—	—	—	—	<b>Disease causing</b>	<b>N M</b>

Probably disease causing mutations are highlighted in bold.

B: benign, HMZ: homozygous, HTZ: heterozygous, M C: marginally conserved, N C: not conserved, N M: novel mutation, R M: recurrent mutation, T: tolerated, P,D: possibly damaging.



FIGURE 1: Ophthalmic features of family F752: II.1 (CIC01245): fundus color photographs ((a) and (d) for right and left eye resp.), autofluorescence ((b) and (e) for right and left eye resp.), and spectral domain optical coherence tomography horizontal macula scans ((c) and (f) for right and left eye resp.), showing severe rod-cone dystrophy signs with macular involvement.

last exon (exon 4) and shown to be transmitted in an autosomal dominant mode of inheritance. Most of *RPI* disease causing variants represent nonsense mutations, deletions, or insertions. In mammalian genes, nonsense mutations lead to unstable mRNAs that are degraded by nonsense-mediated

decay (NMD). However, exceptions might arise when premature stop codons occur in the last exon [37]. These variants are thought to abolish *RPI* function by resulting in a truncated protein lacking important functional domains although still able to interact with some of its protein partner(s) [21]. The

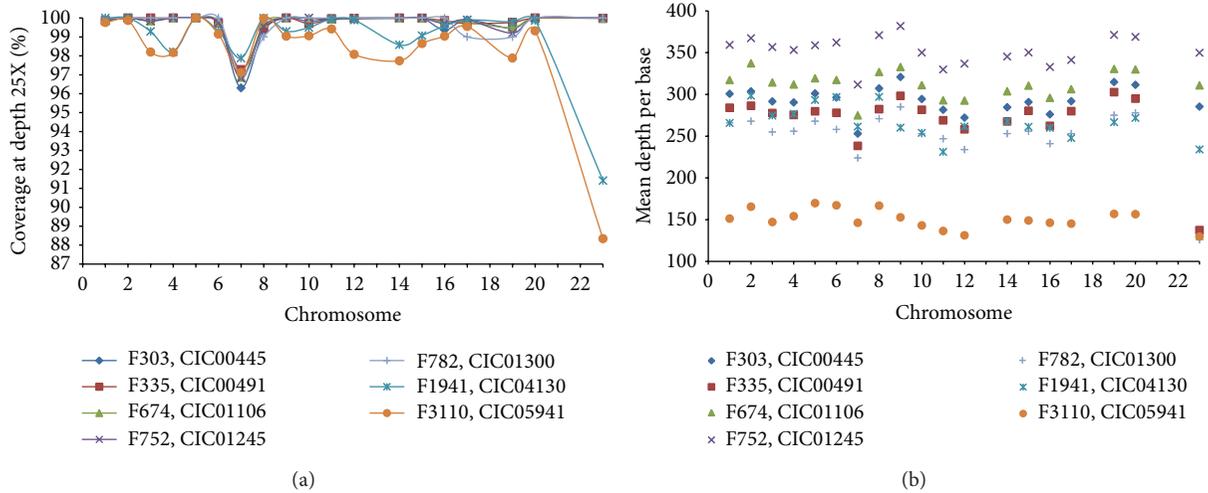


FIGURE 2: Sequencing statistics in index patients. (a) The overall sequencing coverage of the target regions at 25X depth of coverage is shown in each of the chromosomes. No values were indicated for chromosomes 13, 18, 21, and 22 as they were not targeted. The term chromosome 23 was used to designate the X chromosome. F1941: III.1 (CIC04130) and F3110: III.5 (CIC05941) showed the lowest coverage results. (b) The average mean depth per base pair is shown for each of the chromosomes. Most targets showed coverage around 300 times.

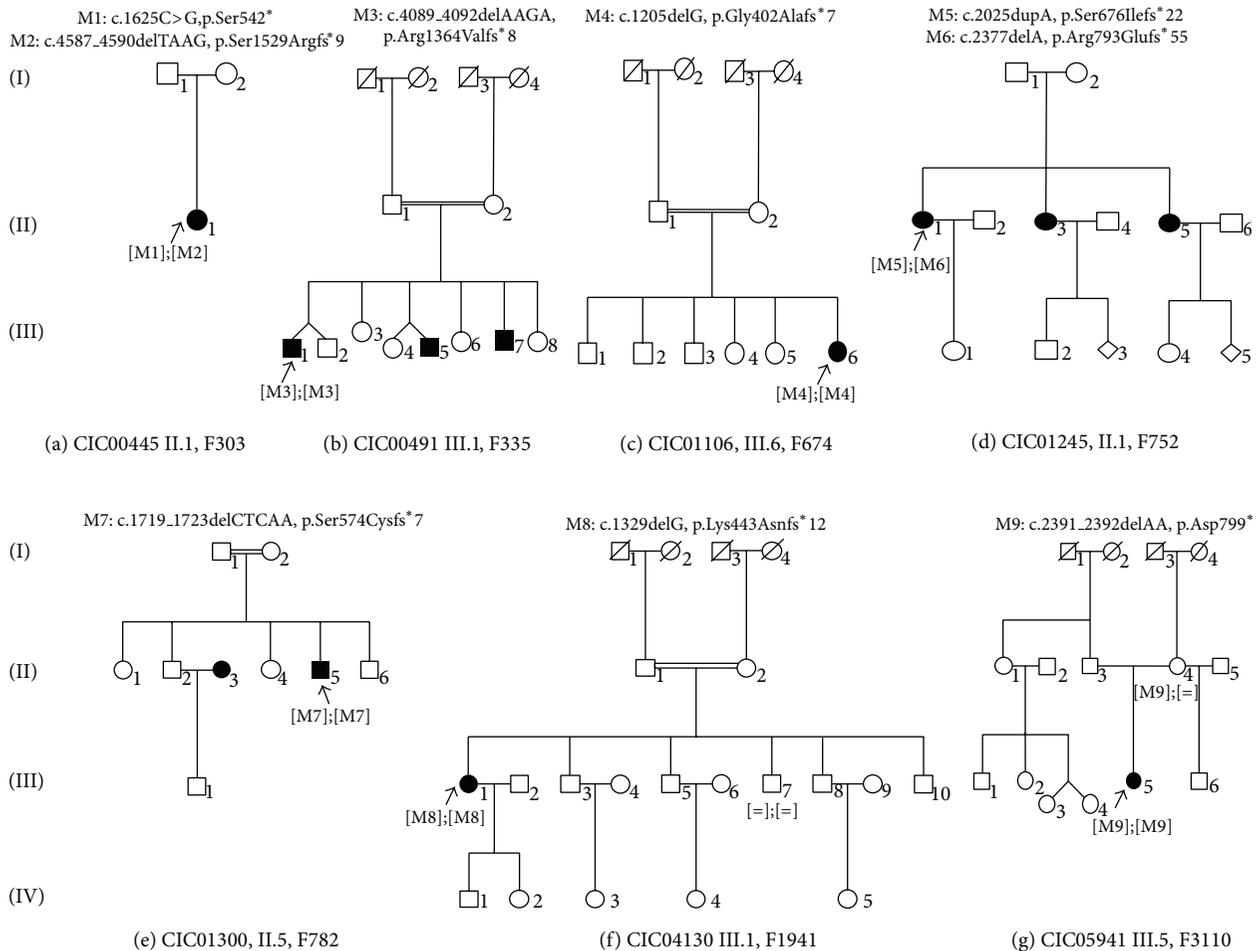


FIGURE 3: Pedigrees of seven families with *RPI* mutations underlying autosomal recessive rod-cone dystrophy. Affected and unaffected individuals are represented by shapes filled with black and white colors, respectively. Men and women are indicated by squares and circles, respectively. Index subjects are marked by  $\sphericalangle$ . Consanguinity is marked by a double horizontal line.

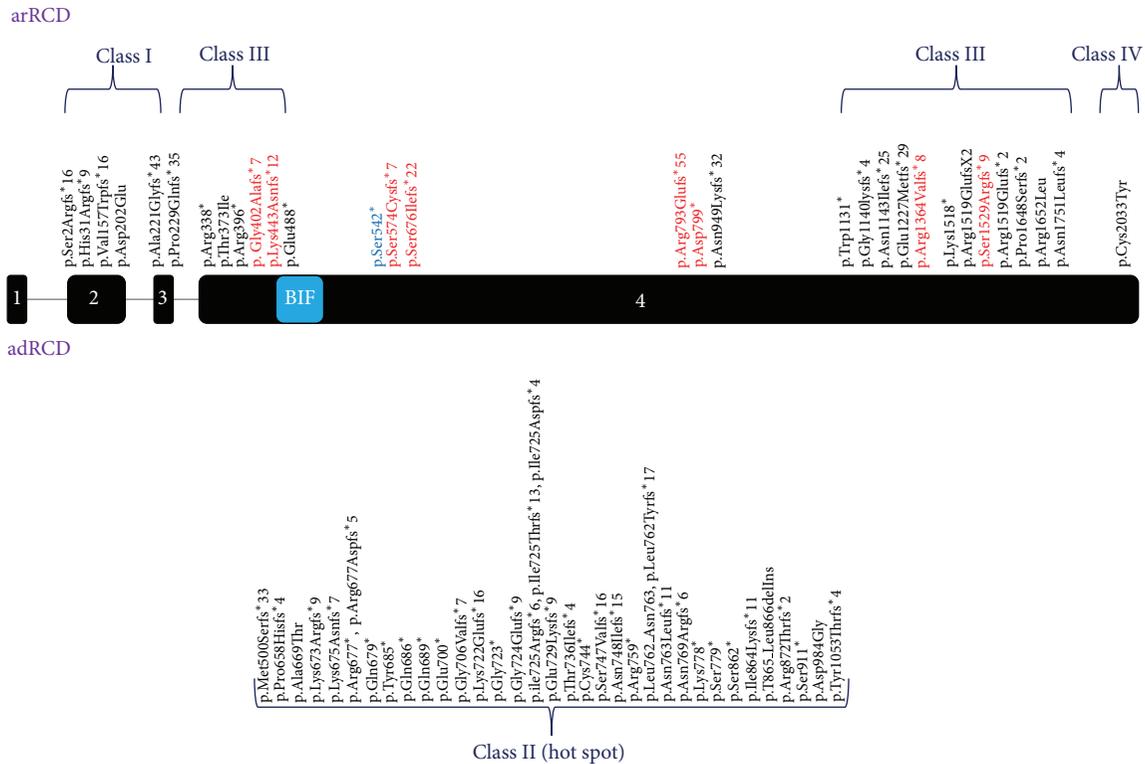


FIGURE 4: Schematic presentation of *RPI* disease causing mutations. Disease causing mutations were represented based on the classification by Chen and coworkers [13]. Mutations responsible for recessive arRCD were shown in the upper half, whereas mutations causing adRCD were shown in the lower half. p.Gly402Alafs\*7, p.Lys443Asnfs\*12, p.Arg1364Valfs\*8, and p.Ser1529Argfs\*9 belong to class III. Although p.Ser574Cysfs\*7, p.Ser676Ilefs\*22, p.Arg793Glufs\*55, and p.Asp799\* are class II mutations, these variants do not cause adRCD but arRCD instead. Amino acid modifications shown in red and blue represent novel frameshift or nonsense mutations and the recurrent p.Ser542\* mutation respectively. Protein localization of p.Ser542\* was highlighted in blue as it marked a recurrent mutation. adRCD: autosomal dominant rod-cone dystrophy, arRCD: autosomal recessive rod-cone dystrophy, BIF: drosophila melanogaster bifocal.

latter observation is supported by finding that *RPI* mutant mRNA is expressed in a human cell line carrying a homozygous p.Arg677\* mutation [21].

Based on Chen et al. [13], *RPI* truncating mutations leading to arRCD or adRCD can be divided into four distinct groups. Class I is composed of truncating mutations located in exons 2 and 3. These variants are sensitive to NMD and thus are considered as true loss-of-function alleles (Figure 4) [13]. Class II involves truncating mutations that are located in a spot between codons 500 and 1053 in exon 4 [13], the so called “*RPI* hot spot.” The “hot spot” variants tend to be insensitive to NMD process and thus result in a protein with a potential dominant negative effect leading to adRCD (Figure 4) [13]. Class III includes truncating mutations insensitive to NMD located between codons 264 and 499 and between codons 1054 to 1751 in exon 4. These truncating proteins result in a loss of function leading to arRCD (Figure 4) [13]. Finally, class IV includes protein-truncating mutations near the 3’ end of the fourth exon (Figure 4) [13]. Most likely, the resulting proteins display only a minor loss of their C-terminal portion, preserving the majority of functional domains and keeping a residual activity. According to the classification of Chen et al. [13], p.Gly402Alafs\*7, p.Lys443Asnfs\*12, p.Arg1364Valfs\*8, and p.Ser1529Argfs\*9 belong to class III (Figure 4).

The predicted physiopathology for p.Ser542\*, p.Ser574Cysfs\*7, p.Ser676Ilefs\*22, p.Arg793Glufs\*55, and p.Asp799\* is more complex. According to Chen’s classification, these frameshift deletions and nonsense mutations should belong to class II, previously only associated with adRCD. However, herein, they are causing presumably arRCD (Figure 4). To further confirm these findings, clinical and genetic testing of the reported unaffected parents should be done.

Based on the previous findings, we speculate that the classification by Chen and coworkers does not hold true for all mutations. Supporting this statement, Avila-Fernandez et al. [12] reported the same nonsense mutation (p.Ser542\*) found in (F303: II.1 (CIC00445)) and located at the 5’ end of the “hot spot” to cause arRP in the Spanish population [12]. These observations are of interest as they point out for an implication of hot spot region for adRCD-*RPI* mutation also in case of arRCD. Future studies will need to clarify why some class II mutations lead to adRCD and others to arRCD.

Patients with arRCD and *RPI* mutations show a more severe disease than adRCD-*RPI* mutant patients with macular atrophy in all our cases. This was first outlined by Lafont et al. [17]. When patients are presenting with late, severe disease, the diagnostic distinction between RCD, with

initial rod involvement, and cone-rod dystrophy (CRD) with initial cone involvement is difficult. Of note is that one of the patients (CIC01300) in the present study was initially classified as possibly having severe CRD and his diagnosis was actually revisited after NGS results. This also outlines the interest of unbiased massive parallel sequencing methods for a more precise clinical diagnostic in case of end stage disease. This point will most likely become even more critical with the perspective of therapeutic trials.

**4.1. Strength and Limitations.** We estimate that 1% of our target regions were not covered. Partially uncovered exons are a real common issue when capturing the DNA sequences using commercially available probes; this bias might imply a loss of some candidate variants. However, we found that rate of 1% is very reasonable when compared with other NGS panels. In addition, in order to exclude the possibility of finding other candidate variants, we have sequenced by Sanger method the majority of these regions. Five of our patients carried homozygous *RPI* mutations. For four of the subjects carrying homozygous variants, namely CIC00491, F335; CIC01106, F674; CIC01300, F782 and CIC04130, F1941; co-segregation analysis needs to be done to confirm autosomal recessive inheritance but we do not have access to parent's DNA. CIC05941 was the only one not to report clear consanguinity in the family, and we cannot exclude the possibility of a large deletion on the second allele of *RPI* gene. Again, DNA of the father, not available for us, would be helpful to prove autosomal recessive inheritance and the homozygous state of the mutation.

In conclusion, we have reported 9 mutations in *RPI* of which 8 were novel causing arRCD [8, 12–20]. Interestingly, a prevalence of  $\approx 2.5\%$  points out for the necessity of sequencing *RPI* in sporadic and recessive cases of RCD. Further functional studies are needed to understand the impact of *RPI* structure on its function at the molecular level; such a step would strengthen our knowledge in the physiology of retinal photoreceptors.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Authors' Contribution

Isabelle Audo and Christina Zeitz contributed equally to this work.

## Acknowledgments

The authors express their sincere gratitude to the families who participated in this study and to the clinical staff for their help in clinical data and DNA collection. This work was supported by Foundation Voir et Entendre (CZ), Prix Dalloz for "la recherche en ophtalmologie" (CZ), Foundation Fighting Blindness (FFB) [CD-CL-0808-0466-CHNO] (IA) and FFB center (FFB grantC-GE-0912-0601-INSERM02),

Prix de la Fondation de l'Œil (IA), Ville de Paris and Region Ile de France, and the French State program "Investissements d'Avenir" managed by the Agence Nationale de la Recherche [LIFESENSES: ANR-10-LABX-65].

## References

- [1] D. T. Hartong, E. L. Berson, and T. P. Dryja, "Retinitis pigmentosa," *The Lancet*, vol. 368, no. 9549, pp. 1795–1809, 2006.
- [2] I. Audo, M. Lancelot, S. Mohand-Saïd et al., "Novel C2orf71 mutations account for approximately  $\sim 1\%$  of cases in a large French arRP cohort," *Human Mutation*, vol. 32, no. 4, pp. E2091–E2103, 2011.
- [3] A. Anasagasti, C. Irigoyen, O. Barandika, A. López de Munain, and J. Ruiz-Ederra, "Current mutation discovery approaches in Retinitis Pigmentosa," *Vision Research*, vol. 75, pp. 117–129, 2012.
- [4] L. S. Sullivan, S. J. Bowne, D. G. Birch et al., "Prevalence of disease-causing mutations in families with autosomal dominant retinitis pigmentosa: a screen of known genes in 200 families," *Investigative Ophthalmology and Visual Science*, vol. 47, no. 7, pp. 3052–3064, 2006.
- [5] B. J. Seyedahmadi, C. Rivolta, J. A. Keene, E. L. Berson, and T. P. Dryja, "Comprehensive screening of the *USH2A* gene in Usher syndrome type II and non-syndromic recessive retinitis pigmentosa," *Experimental Eye Research*, vol. 79, no. 2, pp. 167–173, 2004.
- [6] A. Ávila-Fernández, D. Cantalapedra, E. Aller et al., "Mutation analysis of 272 Spanish families affected by autosomal recessive retinitis pigmentosa using a genotyping microarray," *Molecular Vision*, vol. 16, pp. 2550–2558, 2010.
- [7] S. El Shamieh, M. Neuille, A. Terray et al., "Whole-exome sequencing identifies KIZ as a ciliary gene associated with autosomal-recessive rod-cone dystrophy," *American Journal of Human Genetics*, vol. 94, no. 4, pp. 625–633, 2014.
- [8] I. Audo, S. Mohand-Saïd, C. M. Dhaenens et al., "RPI and autosomal dominant rod-cone dystrophy: novel mutations, a review of published variants, and genotype-phenotype correlation," *Human Mutation*, vol. 33, no. 1, pp. 73–80, 2012.
- [9] X. Guillonneau, N. I. Piriev, M. Danciger et al., "A nonsense mutation in a novel gene is associated with retinitis pigmentosa in a family linked to the RPI locus," *Human Molecular Genetics*, vol. 8, no. 8, pp. 1541–1546, 1999.
- [10] E. A. Pierce, T. Quinn, T. Meehan, T. L. McGee, E. L. Berson, and T. P. Dryja, "Mutations in a gene encoding a new oxygen-regulated photoreceptor protein cause dominant retinitis pigmentosa," *Nature Genetics*, vol. 22, no. 3, pp. 248–254, 1999.
- [11] L. S. Sullivan, J. R. Heckenlively, S. J. Bowne et al., "Mutations in a novel retina-specific gene cause autosomal dominant retinitis pigmentosa," *Nature Genetics*, vol. 22, no. 3, pp. 255–259, 1999.
- [12] A. Avila-Fernandez, M. Corton, K. M. Nishiguchi et al., "Identification of an RPI prevalent founder mutation and related phenotype in Spanish patients with early-onset autosomal recessive retinitis," *Ophthalmology*, vol. 119, no. 12, pp. 2616–2621, 2012.
- [13] L. J. Chen, T. Y. Y. Lai, P. O. S. Tam et al., "Compound heterozygosity of two novel truncation mutations in RPI causing autosomal recessive retinitis pigmentosa," *Investigative Ophthalmology and Visual Science*, vol. 51, no. 4, pp. 2236–2242, 2010.
- [14] S. Khaliq, A. Abid, M. Ismail et al., "Novel association of RPI gene mutations with autosomal recessive retinitis pigmentosa," *Journal of Medical Genetics*, vol. 42, no. 5, pp. 436–438, 2005.

- [15] A. M. Siemiakowska, G. D. N. Astuti, K. Arimadyo et al., "Identification of a novel nonsense mutation in RPI that causes autosomal recessive retinitis pigmentosa in an Indonesian family," *Molecular Vision*, vol. 18, pp. 2411–2419, 2012.
- [16] B. Bocquet, N. A. Marzouka, M. Hebrard et al., "Homozygosity mapping in autosomal recessive retinitis pigmentosa families detects novel mutations," *Molecular Vision*, vol. 19, pp. 2487–2500, 2013.
- [17] E. M. Lafont, G. Manes, G. Sénéchal et al., "Patients with retinitis pigmentosa due to RPI mutations show greater severity in recessive than in dominant cases," *Journal of Clinical & Experimental Ophthalmology*, vol. 2, article 194, 2011.
- [18] M. Al-Rashed, L. Abu Safieh, H. Alkuraya et al., "RPI and retinitis pigmentosa: report of novel mutations and insight into mutational mechanism," *British Journal of Ophthalmology*, vol. 96, no. 7, pp. 1018–1022, 2012.
- [19] S. A. Riazuddin, A. Shahzadi, C. Zeitz et al., "A mutation in SLC24A1 implicated in autosomal-recessive congenital stationary night blindness," *The American Journal of Human Genetics*, vol. 87, no. 4, pp. 523–531, 2010.
- [20] H. P. Singh, S. Jalali, R. Narayanan, and C. Kannabiran, "Genetic analysis of indian families with autosomal recessive retinitis pigmentosa by homozygosity screening," *Investigative Ophthalmology and Visual Science*, vol. 50, no. 9, pp. 4065–4071, 2009.
- [21] Q. Liu, A. Lyubarsky, J. H. Skalet, E. N. Pugh Jr., and E. A. Pierce, "RPI is required for the correct stacking of outer segment discs," *Investigative Ophthalmology & Visual Science*, vol. 44, no. 10, pp. 4171–4183, 2003.
- [22] Q. Liu, J. Zuo, and E. A. Pierce, "The retinitis pigmentosa 1 protein is a photoreceptor microtubule-associated protein," *Journal of Neuroscience*, vol. 24, no. 29, pp. 6427–6436, 2004.
- [23] I. Audo, K. M. Bujakowska, T. Léveillard et al., "Development and application of a next-generation-sequencing (NGS) approach to detect known and novel gene defects underlying retinal diseases," *Orphanet Journal of Rare Diseases*, vol. 7, no. 1, article 8, 2012.
- [24] I. Audo, J. A. Sahel, S. Mohand-Saïd et al., "EYS is a major gene for rod-cone dystrophies in France," *Human Mutation*, vol. 31, no. 5, pp. E1406–E1435, 2010.
- [25] T. G. P. Consortium, "A map of human genome variation from population-scale sequencing," *Nature*, vol. 467, no. 7319, pp. 1061–1073, 2010.
- [26] D. M. Altshuler, R. A. Gibbs, L. Peltonen et al., "Integrating common and rare genetic variation in diverse human populations," *Nature*, vol. 467, no. 7311, pp. 52–58, 2010.
- [27] J. A. Tennessen, A. W. Bigham, T. D. O'Connor et al., "Evolution and functional impact of rare coding variation from deep sequencing of human exomes," *Science*, vol. 337, no. 6090, pp. 64–69, 2012.
- [28] M. J. Bamshad, S. B. Ng, A. W. Bigham et al., "Exome sequencing as a tool for Mendelian disease gene discovery," *Nature Reviews Genetics*, vol. 12, no. 11, pp. 745–755, 2011.
- [29] W. J. Kent, C. W. Sugnet, T. S. Furey et al., "The human genome browser at UCSC," *Genome Research*, vol. 12, no. 6, pp. 996–1006, 2002.
- [30] I. A. Adzhubei, S. Schmidt, L. Peshkin et al., "A method and server for predicting damaging missense mutations," *Nature Methods*, vol. 7, no. 4, pp. 248–249, 2010.
- [31] P. Kumar, S. Henikoff, and P. C. Ng, "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm," *Nature Protocols*, vol. 4, no. 7, pp. 1073–1082, 2009.
- [32] F. O. Desmet, D. Hamroun, M. Lalande, G. Collod-Bèroud, M. Claustres, and C. Bèroud, "Human splicing finder: an online bioinformatics tool to predict splicing signals," *Nucleic Acids Research*, vol. 37, no. 9, article no. e67, 2009.
- [33] P. D. Stenson, E. V. Ball, M. Mort, A. D. Phillips, K. Shaw, and D. N. Cooper, "The Human Gene Mutation Database (HGMD) and its exploitation in the fields of personalized genomics and molecular evolution," in *Current Protocols in Bioinformatics*, chapter 1–13, 2012.
- [34] I. F. A. C. Fokkema, J. T. Den Dunnen, and P. E. M. Taschner, "LOVD: easy creation of a locus-specific sequence variation database using an "LSDB-in-a-Box" approach," *Human Mutation*, vol. 26, no. 2, pp. 63–68, 2005.
- [35] S. A. Miller, D. D. Dykes, and H. F. Polesky, "A simple salting out procedure for extracting DNA from human nucleated cells," *Nucleic Acids Research*, vol. 16, no. 3, p. 1215, 1988.
- [36] C. Zeitz, B. Kloeckener-Gruissem, U. Forster et al., "Mutations in *CABP4*, the gene encoding the  $Ca^{2+}$ -binding protein 4, cause autosomal recessive night blindness," *American Journal of Human Genetics*, vol. 79, no. 4, pp. 657–667, 2006.
- [37] S. A. Riazuddin, F. Zulfiqar, Q. Zhang et al., "Autosomal recessive retinitis pigmentosa is associated with mutations in RPI in three consanguineous Pakistani families," *Investigative Ophthalmology and Visual Science*, vol. 46, no. 7, pp. 2264–2270, 2005.

## Review Article

# Interaction Analysis through Proteomic Phage Display

**Gustav N. Sundell and Ylva Ivarsson**

*Department of Chemistry-BMC, Uppsala University, P.O. Box 576, 751 23 Uppsala, Sweden*

Correspondence should be addressed to Ylva Ivarsson; [ylva.ivarsson@kemi.uu.se](mailto:ylva.ivarsson@kemi.uu.se)

Received 25 June 2014; Revised 11 August 2014; Accepted 19 August 2014; Published 11 September 2014

Academic Editor: Calvin Yu-Chian Chen

Copyright © 2014 G. N. Sundell and Y. Ivarsson. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Phage display is a powerful technique for profiling specificities of peptide binding domains. The method is suited for the identification of high-affinity ligands with inhibitor potential when using highly diverse combinatorial peptide phage libraries. Such experiments further provide consensus motifs for genome-wide scanning of ligands of potential biological relevance. A complementary but considerably less explored approach is to display expression products of genomic DNA, cDNA, open reading frames (ORFs), or oligonucleotide libraries designed to encode defined regions of a target proteome on phage particles. One of the main applications of such proteomic libraries has been the elucidation of antibody epitopes. This review is focused on the use of proteomic phage display to uncover protein-protein interactions of potential relevance for cellular function. The method is particularly suited for the discovery of interactions between peptide binding domains and their targets. We discuss the largely unexplored potential of this method in the discovery of domain-motif interactions of potential biological relevance.

## 1. Introduction

The human interactome is estimated to contain about 130,000 binary protein-protein interactions (PPIs), of which the majority remains to be discovered [1]. PPIs are crucial for cellular function and dysfunction and large efforts are therefore invested in their identification and in constructing PPI based networks [2]. Different high-throughput methods render complementary data. For example, affinity purification coupled to mass spectrometry (AP-MS) [3, 4] and luminescence-based mammalian interactome mapping (LUMIER) [5] provide information on complexes, and yeast-two-hybrid (Y2H) experiments give insights into binary PPIs [1], as summarized in Table 1. Despite the significant advances being made the last decade, the human interactome is still largely uncharted and the accumulated knowledge is biased towards well-studied proteins [1, 6].

Particularly elusive to high-throughput methods are the interactions between peptide binding domains and their target motifs, which are typically less than ten residues in length [7, 8]. The peptide motifs are typically located in regions of intrinsic disorder, which can be found in about 35% of the human proteins [9]. Currently, there are more than 2,400 instances reported in the eukaryotic linear motif (ELM)

resource for functional sites in proteins [10], including binding motifs and posttranslational modification sites. This, however, covers only a fraction of the motifs expected to be present in the human proteome [8].

Among the most abundant peptide binding domains in the human proteome are the PSD-95/Discs-large/ZO-1 (PDZ) domains that typically interact with C-terminal sequences of target proteins [11]. Other domains, such as the Src Homology 2 (SH2), bind to phosphorylated target motifs [12]. Domain-motif interactions tend to be of rather low affinities and hence are easily lost in methods such as AP-MS. Although difficult to capture experimentally, transient protein-peptide interactions are crucial for cell function and may be perturbed by disease-causing genetic variations or by viral interferences [13, 14].

Phage display is a powerful approach for establishing binding preferences of peptide binding domains and in extension to discover novel motifs. In combinatorial peptide phage display, highly diverse libraries are used to identify high-affinity ligands with potential to serve as inhibitor [15]. Consensus motifs are derived based on the retained sequences and can be used for predictions of potential ligands in a target proteome [16]. These predictions, however, are not always accurate, which can lead to tedious experimental validations

TABLE 1: Summary of high-throughput methods for identification of PPIs, types of interactions identified, and major advantages and disadvantages of the respective method.

Method	Type of interaction	Advantage	Disadvantage
AP-MS	Binary and complexes	Physiological	Bias towards stable interactions, limited to specific condition (e.g., cell type)
LUMIER	Binary and complexes	Physiological	Bias towards stable interactions
Y2H	Binary	Low-tech	Bias towards stable interactions, bias towards soluble proteins that can translocate to the nucleus
Combinatorial peptide phage display	Binary	Large library size (up to $10^{10}$ ) Identification of consensus motifs	Need for bioinformatics, limited to natural amino acids, limited to protein-peptide interactions
Proteomic phage display	Binary	Identification of target proteins and consensus motif	Limited to natural amino acids

of putative targets. Luck and Travé demonstrated that predictions of human PDZ domain ligands based on results of combinatorial phage display may be hampered due to a bias towards overly hydrophobic (i.e., Trp containing) peptides [17].

A promising strategy to discover novel protein-motif interactions is to reduce the search space to comprise only sequences of a target proteome. In such proteomic phage display, expression products from genomic DNA, cDNA, open reading frames (ORFs), or from designed synthetic oligonucleotides are displayed on phage particles (Figure 1). Proteomic phage display has been used for the identification of allergens [18], antibody epitopes, tumor polypeptides producing immune response [19], and PPIs as well as for the identification of proteins binding to phospholipids and small chemical compounds [20, 21]. In this review, we survey the features, the development, and the applications of various phage display systems used for proteomic phage display, with a particular focus on the elucidation of cellular PPIs. For extensive reviews on cDNA/ORF display for antibody epitope mapping of antigen and pathogen research we refer the readers to dedicated reviews [22–26].

## 2. Phage Display Systems Used for Proteomic Phage Display

Phages that have been used for proteomic phage display include the filamentous M13 phage, the lytic T7 phage, and the temperate phage  $\lambda$ . The main advantage of the M13 phage display system is the ease of its manipulation and handling as detailed in the following section. The main drawback of the M13 phage is that the displayed proteins are secreted through the periplasmic space of the *Escherichia coli* membrane (Figure 2(a)), which may confer problems with unpaired cysteine residues that form undesired disulfide bridges and thereby compromise the successful display of designed sequences [27]. The T7 phage and the  $\lambda$  phage do not suffer from this issue, as they are not passing through the membrane but lysing the host cells (Figure 2(b)).

**2.1. M13 Filamentous Phage.** The M13 filamentous bacteriophage was the first developed phage display system [28].

The M13 phage consists of a circular single stranded DNA (ssDNA) that is covered by five different coat proteins (pIII, pVI, pVII, pVIII, and pIX). The 2,700 copies of the major coat protein pVIII cover the length of the phage [29]. The minor coat proteins pVII and pIX cover one end of the phage particle and pIII and pVI the other end (Figure 2(a)). The minor coat protein pIII is crucial for infection as it initiates the interaction with the F-pilus and TolA receptor [30]. For details on the structure and assembly of filamentous bacteriophages we refer the reader to an extensive review on the topic [31]. Typically, phage propagation is uncoupled from expression and display of desired peptide on the phage particle. This is accomplished through hybrid systems where a phagemid is used for library construction and helper phage is added to provide the information needed for assembly of the phage particle [32].

The M13 phage is a highly versatile system as distinct coat proteins can be used for N- or C-terminal display and for monovalent or multivalent display, respectively [33]. Commonly, the pIII protein is used for low valency display (one to five copies per phage) and the pVIII for high valency display, with up to 1,000 copies per phage in evolved hybrid systems [23]. Highly diverse M13 phage libraries (up to  $10^{10}$ ) can be constructed due to the fact that M13 has a circular ssDNA. For a typical library construction, an oligonucleotide library is designed complementary to the ssDNA with flanking regions corresponding to the phagemid vector. The oligonucleotides are then annealed to the vector and the complementary strand is synthesized and ligated together to form a circular, double stranded DNA vector, which is then electroporated into *E. coli* [34].

**2.2. T7 Phage.** T7 phage is an icosahedral virus of the Podoviridae family and has a linear double stranded (ds) DNA genome. In contrast to M13, T7 is not secreted but released from the host cell through lysis (Figure 2(b)). The T7 phage starts to reproduce immediately upon infection, which is continuous until the point of cell lysis. The major capsid protein (gp10) is encoded by gene 10 and makes up about 90% of the icosahedron capsid. This gene yields two products, 10A and 10B, in a nine-to-one ratio. The minor protein 10B results from a frame shift in the end of the gene that makes

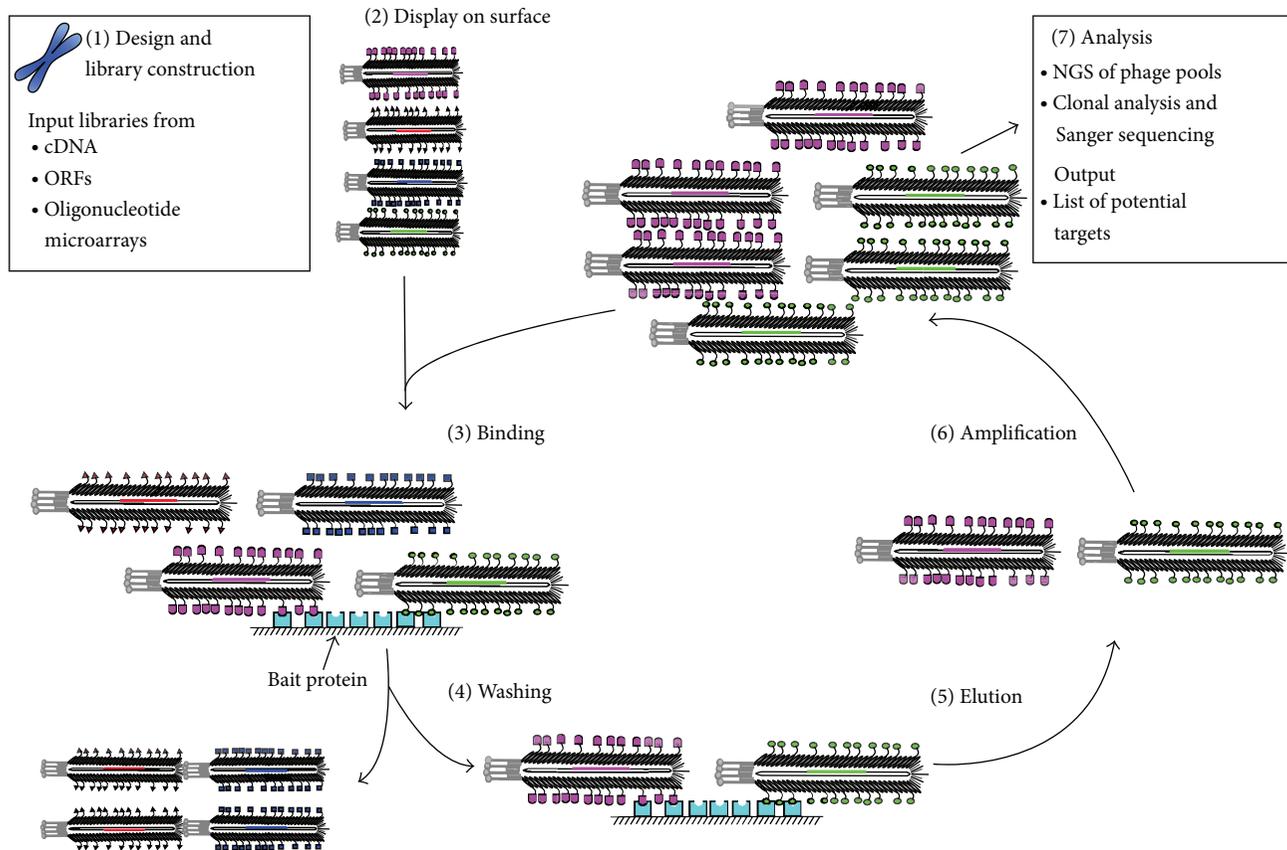


FIGURE 1: Schematic representation of proteomic phage display using the M13 phage. Input phage display libraries are constructed from cDNA, ORFs, or oligonucleotide arrays designed from a proteome of interest (1). Peptides are displayed on pVIII (2). Bait proteins are immobilized on a solid surface and incubated with the naïve input phage library (3). Binding of phage occurs through interactions between displayed peptides and bait proteins, but nonspecific interactions cause noise in the selection (not shown). Unbound phage is washed away (4) and bound phage is eluted through acidic or basic conditions or by the addition of actively growing host bacteria (5). Eluted phage is amplified (6) and used for repeated (typically 3–5) cycles of selection, which is necessary to amplify specifically bound phage over nonspecific binders. Sanger sequencing of confirmed binders and/or NGS of the retained phage pools provides lists of binders from the target proteome (7).

the capsid protein 52 residues longer [35]. Fusion proteins are displayed on protein 10B C-terminally of the 52 extra residues. Depending on the system used, up to 1,200 amino acid inserts can be displayed at low valency (5–15 copies per virion) or shorter inserts (up to 50 amino acids) at higher valency (up to 415 copies) [36]. The linear genome makes it more challenging to construct T7 phage libraries as compared to M13 libraries. Library construction includes two-step ligations and the *in vitro* packing of DNA into the phage, which in the T7 select system (Novagen) is accomplished by the addition of DNA to commercially available packaging extract. The packaging extract is sensitive to work with and rather costly if larger libraries are prepared [33].

**2.3. Lambda Phage.** The temperate  $\lambda$  phage has an icosahedral head. The main structure of the shell is built from the major coat protein gpE (415 copies) and is stabilized by the capsid protein gpD (402–420 copies) [37]. The head is linked to a flexible helical tail constructed by disks of the major tail protein gpV. Its linear dsDNA is packed in the bacteriophage

head. The DNA is injected into the host bacteria and is stably integrated into the host chromosome during the lysogenic state. When triggered correctly, the  $\lambda$  phage starts a lytic cycle [38].

Both the tail protein gpV and the head protein gpD have been used for phage display. Initially, the  $\lambda$ foo vector was constructed for the C-terminal display on gpV, with a low display level that made it suitable for capturing high-affinity interactions [39]. Later, systems were developed for the display of peptides N-terminally or C-terminally to the major coat protein gpD [40–42]. Libraries with diversities in the range of  $10^7$ – $10^8$  are constructed using commercially available *in vitro* packaging systems.

### 3. Proteomic Phage Display

Over the years, different approaches towards proteomic phage display have been taken, from cDNA and ORF display to the display of the expression products of highly defined synthetic oligonucleotide libraries, as detailed in the following section.

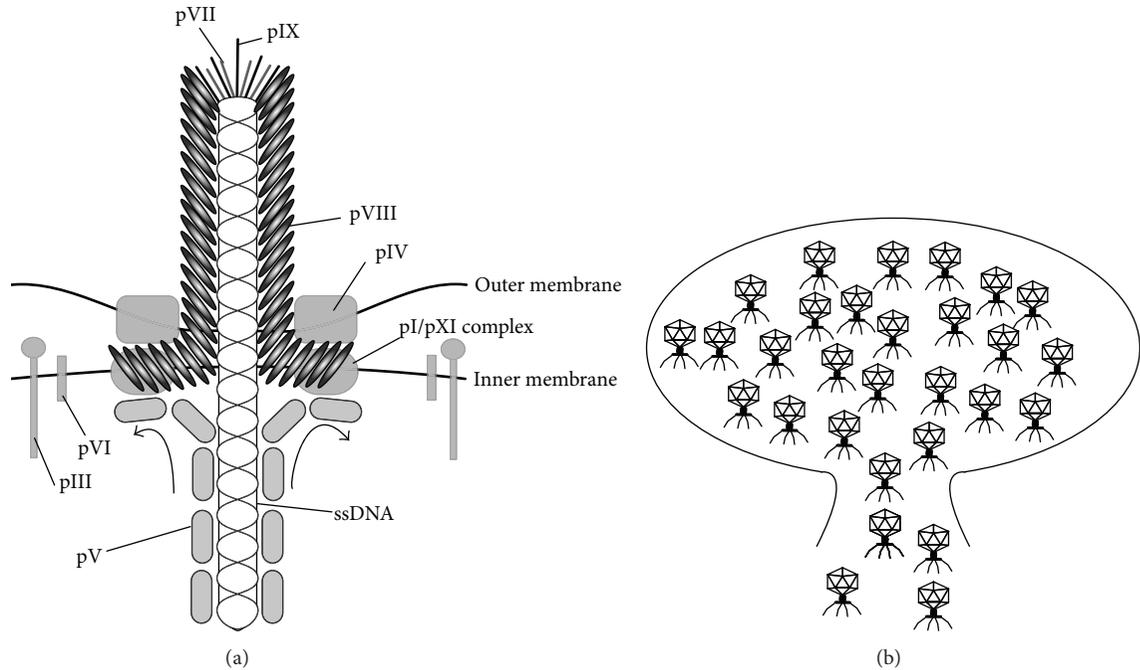


FIGURE 2: Schematic representation of the assembly and excretion of the M13 filamentous phages and the exit through cell lysis of the lytic T7 phage. (a) The filamentous M13 phage is assembled at the cell membrane of the bacterial host. In the host cell, the ssDNA is protected by association with protein pV, which detaches at the membrane upon assembly. At the start of the assembly, membrane-associated pVII/pIX bind a specific DNA packing signal. Membrane-bound protein pVIII binds to the DNA and is transferred across the membranes. The transport is facilitated by a complex of pI and pXI situated in the inner membrane and protein pIV that makes a pore through the outer membrane of the bacteria for the phage to pass through. As a final step, pVI and pIII that span the inner membrane are assembled on the phage. The figure was created based on [31]. (b) The lytic T7 phage, schematically shown with its typical icosahedral head, is assembled in the cytosol of the host cell. It is multiplied to such extent that the host cell finally bursts and the phage is released to the surrounding.

**3.1. cDNA/ORF Display.** In cDNA display, a gene, a cDNA, or a complete genome is displayed on phage particles. Theoretically, this is a straightforward technique. However, it suffers from difficulties in obtaining high-quality libraries [43]. This is a consequence of the transcriptional stop codons at the 3'-end of coding regions, the polyA tail of mRNA, and the often nondirectional cloning. The fraction of clones expressing peptides in frame in a naïve cDNA library may be as low as 6%. Additionally, phage with truncated constructs tends to outgrow clones with correctly displayed sequences [44, 45]. The quality can be improved by using ORF enriched DNA collections for library construction [43, 46]. Library quality can further be improved by fragmentation of the DNA by, for example, treatment with Deoxyribonuclease I or by sonication before cloning [26, 47]. Despite the quality issues, a variety of libraries based on human or pathogen cDNA/ORFs have been derived and used for PPI screening.

**3.1.1. cDNA/ORF Display Using the M13 Filamentous Phage.** A number of studies have employed the M13 filamentous phage system for cDNA display. In most cases, the expression products are displayed on the pIII protein either indirectly through the Jun-Fos system or directly [48]. There are also reports of N-terminal multivalent display on pVIII [49] and monovalent C-terminal display on pVI [50]. When displaying inserts N-terminally, a main limitation is that the inserts have

to be in the same reading frame as the pIII or pVIII proteins and that there can be no in frame stop codons. A way to enrich for ORFs and the correct presentation of encoded sequences and thereby improve the quality is the so-called Hyperphage [51]. In this system, the helper phage has a truncated g3 so that the phagemid pIII fusion is the only source of pIII, as originally described by Kristensen and Winter [52]. This strategy has been used for the successful identification of immunogenic polypeptides of *Mycoplasma hyopneumoniae* [53]. Recently, a novel trypsin-sensitive helper phage was derived for a similar purpose [54].

**Indirect Display on pIII: The Jun-Fos System.** A system for indirect cDNA display on pIII termed pJun-Fos was engineered in 1993. This system takes advantage of the strong association between the Leucine Zipper Jun and Fos [55]. The pIII-fused Jun and Fos-linked cDNA expression product is assembled in the *E. coli* periplasm, which leads to the indirect display of the functional expression products of cDNA on pIII. The complex is stabilized by disulfide bridges between cysteines engineered at the N- and C-termini of Jun and Fos. In the original publication, enzymatically active alkaline phosphatase was displayed on pIII and the authors discussed the potential of the system as a tool in PPI screening. The Jun-Fos system has since then been a popular cDNA display system for the discovery of antibody epitopes [56] as reviewed elsewhere [48].

A Jun-Fos system modified to ensure cloning in all three reading frames was used to identify host-pathogen protein-protein interactions between the ribonucleoproteins of influenza virus and the expression products of a human cDNA library (inserts >750 bp) [57]. In this study, the authors pinpointed a direct interaction between the A domain of human high mobility group box proteins and the viral bait protein.

*Direct Display on pIII.* Expression products of cDNA/ORF have also been displayed directly on pIII. In an early study, the plasminogen-activator inhibitor 1 was fragmented into 50–200 bp and cloned into the M13 gpIII phagemid vector [58]. This library was used for epitope mapping of a monoclonal antibody raised against this protein. For PPI screening, Hertveldt et al. constructed a phage library by fusing genomic *S. cerevisiae* DNA (100–1,500 bp) to gpIII lacking the N1 domain [59]. From panning of the yeast cDNA library against GAL80, fragments of the known binder GAL4 and three other ligands of potential physiological relevance were retrieved, thereby demonstrating that the system can be used for identification of biologically relevant targets. Around the same time, Yano et al. constructed a fragmented genomic *E. coli* pIII library and identified binders to alkaline phosphatase [60].

Two other studies demonstrate that ORF enriched cDNA display on pIII can be used to identify targets of potential biological relevance. In the first study, the interactomes of the high mobility group A proteins HMG1 and HMG2 were elucidated using an ORF enriched murine cDNA M13 pIII library displaying 200–500 base pair fragments [61]. For these nuclear chromatin factors, four targets were identified, namely, TBP associated factor 3b and chromatin assembly factor I, subunit A, and two previously uncharacterized proteins. For the first two proteins, interactions were confirmed between the full-length proteins through GST-pull down assays and coaffinity purification of overexpressed proteins in HEK293T cells [61].

In the second study, an ORF enriched and fragmented cDNA library displayed on pIII was used for interactome mapping of transglutaminase 2 (TG2) [62]. Through next-generation sequencing (NGS) of selected phage pools a list of potential targets was retrieved. The most frequently occurring ligands interactions were validated through protein complementarity assays with 80% success rate, thus demonstrating the power of the combination of ORF enriched cDNA display and NSG in interactome mapping.

*Posttranslational Modifications and ORF Display on pIII.* PPIs are often controlled by posttranslational modifications, with the most common modifications in eukaryotic proteomes being phosphorylation of Ser/Thr/Tyr residues [63]. These modifications can create or abrogate binding sites or modulate function by more indirect means. A few attempts have been made towards investigating PPIs relying on posttranslational modifications through proteomic phage display. In particular, Cochrane and coworkers used the fyn tyrosine kinase to *in vitro* phosphorylate a fragmented leukocyte cDNA library ( $10^8$ ) displayed on pIII. The phosphorylated

library was used in selections against the phosphotyrosine binding tandem of SH2 domains of SHP-2 [64]. Nonspecific binding clones were removed before *in vitro* phosphorylation and selection using SHP-2 Sepharose. Through competitive ELISA experiments using phosphorylated phage and synthetic peptides, double phosphorylated PECAM-1 was identified and confirmed as a SHP-2 ligand. It thus appears possible to identify natural interactions relying on posttranslational modification through cDNA phage display. However, given the lack of follow-up studies it does not seem like a feasible way to go for high-throughput analysis of PPIs depending on posttranslational modification. Other attempts to tackle posttranslational modification involved the system for the production and enrichment of phage displaying N-glycoproteins [65].

*cDNA Display on pVI.* A limited set of studies has employed C-terminal cDNA display on pVI, thus circumventing issues related to the presence of premature stop codons. Using a pVI cDNA library of the hookworm *Ancylostoma caninum* ligands were identified for two serine proteases [50]. A few years later, a rat liver cDNA library fused to pVI was used for the identification of peroxisomal proteins by panning the library against antibodies raised against peroxisomal subfractions [66]. In another study, a pVI cDNA library from the colorectal cancer cell line HT-29 was used to identify a panel of candidate tumor antigens [67]. Other studies have reported the discovery of autoantigens for diseases such as multiple sclerosis [68] and rheumatoid arthritis [69]. However, at this stage there are no studies that have applied pVI cDNA display for the explicit purpose of interactome analysis. The monovalent display on pVI makes it less suited for capturing low-affinity interactions.

*3.1.2. T7 Phage Display.* T7 phage display has become a popular system for cDNA/ORF display, starting from the identification of RNA binding proteins from cDNA displayed on the C-terminus of the capsid protein 10B [70]. It has typically been used for antigen discovery [71]. For example, T7 cDNA display of sea snake venom gland mRNA identified rabbit anti-sea snake venom IgGs as well as new toxins [72]. T7 cDNA display has also been used to explore interactions between parasite proteins and host enterocytes [73, 74].

High-quality ORF T7 display libraries have been used for interactome analysis. In particular, Caberoy and coworkers created a library by combining dual phage display with specific elution of bound phage by protease cleavage [75]. In this system a biotin tag is expressed C-terminally of the inserts and thus is only present when the inserts are in frame. The tag is biotinylated by the *E. coli* BirA enzyme, which enables the selection of ORF clones using immobilized streptavidin. Bound phage is eluted by cleavage with 3C protease. Following this approach, novel tubby binding proteins were identified and then validated through complementary approaches. Of 14 potential target proteins tested, 10 were confirmed as ligands by Y2H and/or pull down assays [75]. The same group used their T7 high-quality ORF library to identify tubby and tubby-like protein 1 as eat-me signals stimulating phagocytosis [76] as well as substrates for the protease calpain 2 [77].

A final example is provided by a study on the suppressor of cytokine signaling 3 (SOC3) [78]. A potential ligand of SOC3, an 11-mer C-terminal peptide of the very long chain acyl-CoA dehydrogenase (VLAD), was found through selections against a human liver cDNA T7 phage library. The interaction was confirmed *in vitro* and in cell-based experiments and was further validated in animal experiments. Based on the results, the authors proposed that SOC3 is an important factor for lipid metabolism.

**3.1.3. Phage  $\lambda$  cDNA/ORF Display.** Phage  $\lambda$  cDNA/ORF display has found use in antigen discovery, as reviewed elsewhere [79]. Already in 1997, the  $\lambda$ foo system was used for epitope mapping of human galectin [80]. In this study, a library was constructed from fragmented cDNA of galectin-3 and screened against immobilized monoclonal antibodies, leading to the identification of two distinct epitopes of nine and eleven amino-acid residues. This method was shown to outperform a random peptide phage library. Other studies report on epitope mapping of monoclonal and polyclonal antibodies with cDNA phage  $\lambda$  libraries from human brain and mouse embryo [41, 81, 82]. However, to our knowledge there are at this stage no papers on interactome analysis using phage  $\lambda$ .

**3.2. Proteomic Peptide Phage Display Libraries from Oligonucleotide Array.** Recently, the advances in oligonucleotide microarray synthesis [83] in combination with bioinformatics and NGS have opened new avenues for the construction of highly defined phage libraries. The pioneering study published in 2011 by Larman and coworkers reported the creation of a T7 library displaying 36-mer peptides representing the complete human proteome, with seven amino acids overlaps [84]. With this library, the authors developed a phage immunoprecipitation sequencing platform for the discovery of autoantibodies. They also demonstrated a more general use for interactome mapping by identifying targets for thereplication protein A2.

In a recent study, a previously engineered pVIII phagemid for multivalent C-terminal display [85] was used to create two distinct proteomic peptide phage display (ProP-PD) libraries. The first library was designed to contain all human C-terminal 7-mer peptides whereas the second library contained all C-termini of known viral proteins. After confirmation of composition and coverage of the libraries through NGS they were used in selection against nine PDZ domains of densin-180, DLG1, erbin, and scribble. Phage pools retained after different selection rounds were analyzed through NGS, which provided detailed information on the progress of the selections. Between two and thirty ligands were obtained for each PDZ domain after the fifth round of selection. Of these, more than 50% of the ligands retained for DLG1, densin-180, and erbin were previously known targets. In contrast, only 13% of the scribble ligands were known since previously. Interactions between full-length scribble and the novel ligands plakophilin-4, mitogen-activated protein kinase 12, and guanylate cyclase soluble subunit alpha-2 were confirmed through colocalizations and coimmunoprecipitations, suggesting that ProP-PD identified biologically relevant targets and that the approach can be used to complement

PPI networks. The ligands retrieved from the selections against the library designed from virus proteins were mostly established biologically relevant ligands, thus demonstrating that the approach can efficiently identify host-pathogen PPIs of biological relevance. Taken together, the proteomic peptide phage display appears to be a highly useful tool for proteome wide screening of domain-motif interactions.

#### 4. Concluding Remarks and Further Perspectives

Various systems for proteomic phage display have been evolved over the last 20 years, with different approaches taken to improve the quality of the displayed sequences. The preferred systems have been the filamentous M13 and the lytic T7 system. The displayed regions range from 7 to 1,500 bps, thus allowing the proteomic identification of peptide ligands as well as interactions involving folded domains. Although most studies have focused on mapping antibody epitopes, it appears as if cDNA/ORF phage display has the potential to successfully identify PPIs of putative biological relevance. This is evident from the validation range of 50–80%, which is considerably better than for techniques such as Y2H [86]. However, cDNA/ORF phage display has had limited use as a method for interactome analysis. The main issue of cDNA/ORF phage display is the lack of control over the displayed sequences, which affects library quality and likely results in the display of a high percentage of unfolded/misfolded proteins and of stretches that are typically inside of folded proteins and not normally available for binding. In most cases, there is a lack of information on the library quality and the coverage of the target genome. Despite the advent of NGS we did not find any publication reporting on the complete sequencing of a cDNA/ORF library, which would provide valuable insights into the quality of the libraries and a better understanding of the interaction space covered during the experiments.

Proteomic peptide phage libraries, created using a combination of bioinformatics and synthetic oligonucleotide libraries, and analyzed through NGS, offer the advantage of full control of displayed regions [87]. At this stage, the cost of highly diverse high-quality oligonucleotide libraries is still rather high. However, given the rapid advances in large-scale *de novo* DNA synthesis [83] we foresee that the cost will go down and that this approach will become increasingly popular for proteomic screening of domain-motif interactions. This will be particularly feasible as phage display can be scaled to hundreds of proteins in parallel [88] and can be paired with NGS of the naïve phage libraries [89] as well as the selected pools, thus providing comprehensive information on the library composition as well as on the retained targets.

By performing proteomic phage display in parallel with other high-throughput methods such as AP-MS or Y2H it is possible to enrich PPI networks with additional interactions and insights on the domain-motif level. Such attempts have previously been made using combinatorial peptide phage display, with an excellent example provided by the Tong et al. study that elucidated the yeast SH3 interactome [90]. More recently, the binding specificities of the worm *Caenorhabditis*

*elegans* SH3 domains were elucidated via high-throughput peptide phage display. The results were combined with the SH3 interactome that was mapped through Y2H experiment [91]. The use of proteomic libraries rather than combinatorial phage libraries for this kind of analysis will obviate the need for predictions as it directly identifies the target protein based on the selected ligands. ProP-PD will be particularly useful in providing unbiased information on domain-motif interactions. This will give novel insights into the function of unexplored motifs in the human proteome. As these experiments can be performed in high-throughput the limiting factor for elucidating domain-motif interactions will be the access to recombinant proteins and the downstream cell biological validations.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgments

This work was supported by grants from the Swedish Research Council and the Åke Wiberg Foundation (Sweden).

### References

- [1] K. Venkatesan, J. F. Rual, A. Vazquez et al., "An empirical framework for binary interactome mapping," *Nature Methods*, vol. 6, no. 1, pp. 83–90, 2009.
- [2] M. Vidal, M. E. Cusick, and A.-L. Barabási, "Interactome networks and human disease," *Cell*, vol. 144, no. 6, pp. 986–998, 2011.
- [3] A.-C. Gavin, K. Maeda, and S. Kühner, "Recent advances in charting protein-protein interaction: Mass spectrometry-based approaches," *Current Opinion in Biotechnology*, vol. 22, no. 1, pp. 42–49, 2011.
- [4] W. H. Dunham, M. Mullin, and A.-C. Gingras, "Affinity-purification coupled to mass spectrometry: basic principles and strategies," *Proteomics*, vol. 12, no. 10, pp. 1576–1590, 2012.
- [5] M. Barrios-Rodiles, K. R. Brown, B. Ozdamar et al., "High-throughput mapping of a dynamic signaling network in mammalian cells," *Science*, vol. 307, no. 5715, pp. 1621–1625, 2005.
- [6] J.-F. Rual, K. Venkatesan, T. Hao et al., "Towards a proteome-scale map of the human protein-protein interaction network," *Nature*, vol. 437, no. 7062, pp. 1173–1178, 2005.
- [7] B. A. Liu, B. W. Engelmann, and P. D. Nash, "High-throughput analysis of peptide-binding modules," *Proteomics*, vol. 12, no. 10, pp. 1527–1546, 2012.
- [8] P. Tompa, N. E. Davey, T. J. Gibson, and M. M. Babu, "A million peptide motifs for the molecular biologist," *Molecular Cell*, vol. 55, no. 2, pp. 161–169, 2014.
- [9] J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton, and D. T. Jones, "Prediction and functional analysis of native disorder in proteins from the three kingdoms of life," *Journal of Molecular Biology*, vol. 337, no. 3, pp. 635–645, 2004.
- [10] H. Dinkel, K. van Roey, and S. Sushama Michael, "The eukaryotic linear motif resource ELM: 10 years and counting," *Nucleic Acids Research*, vol. 42, pp. D259–D266, 2014.
- [11] Y. Ivarsson, "Plasticity of PDZ domains in ligand recognition and signaling," *FEBS Letters*, vol. 586, no. 17, pp. 2638–2647, 2012.
- [12] J. Jin and T. Pawson, "Modular evolution of phosphorylation-based signalling systems," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 367, no. 1602, pp. 2540–2555, 2012.
- [13] N. E. Davey, G. Travé, and T. J. Gibson, "How viruses hijack cell regulation," *Trends in Biochemical Sciences*, vol. 36, no. 3, pp. 159–169, 2011.
- [14] B. Uyar, R. J. Weatheritt, H. Dinkel, N. E. Davey, and T. J. Gibson, "Proteome-wide analysis of human disease mutations in short linear motifs: neglected players in cancer?" *Molecular BioSystems*, 2014.
- [15] Y. Zhang, B. A. Appleton, C. Wiesmann et al., "Inhibition of Wnt signaling by Dishevelled PDZ peptides," *Nature Chemical Biology*, vol. 5, no. 4, pp. 217–219, 2009.
- [16] J. Teyra, S. S. Sidhu, and P. M. Kim, "Elucidation of the binding preferences of peptide recognition modules: SH3 and PDZ domains," *FEBS Letters*, vol. 586, no. 17, pp. 2631–2637, 2012.
- [17] K. Luck and G. Travé, "Phage display can select over-hydrophobic sequences that may impair prediction of natural domain-peptide interactions," *Bioinformatics*, vol. 27, no. 7, pp. 899–902, 2011.
- [18] C. Rhyner, M. Weichel, S. Flückiger, S. Hemmann, T. Kleber-Janke, and R. Cramer, "Cloning allergens via phage display," *Methods*, vol. 32, no. 3, pp. 212–218, 2004.
- [19] A. Fosså, L. Alsøe, R. Cramer, S. Funderud, G. Gaudernack, and E. B. Smeland, "Serological cloning of cancer/testis antigens expressed in prostate cancer using cDNA phage surface display," *Cancer Immunology, Immunotherapy*, vol. 53, no. 5, pp. 431–438, 2004.
- [20] B. van Dorst, J. Mehta, E. Rouah-Martin et al., "cDNA phage display as a novel tool to screen for cellular targets of chemical compounds," *Toxicology in Vitro*, vol. 24, no. 5, pp. 1435–1440, 2010.
- [21] W. Guo, S. Liu, J. Peng et al., "Examining the interactome of huperzine A by magnetic biopanning," *PLoS ONE*, vol. 7, no. 5, Article ID e37098, 2012.
- [22] L. J. Matthews, R. Davis, and G. P. Smith, "Immunogenically fit subunit vaccine components via epitope discovery from natural peptide libraries," *Journal of Immunology*, vol. 169, no. 2, pp. 837–846, 2002.
- [23] G. A. Weiss and S. S. Sidhu, "Design evolution of artificial M13 coat proteins," *Journal of Molecular Biology*, vol. 300, no. 1, pp. 213–219, 2000.
- [24] Y. Weiss-Ottolenghi and J. M. Gershoni, "Profiling the IgOme: meeting the challenge," *FEBS Letters*, vol. 588, no. 2, pp. 318–325, 2014.
- [25] L. M. Mullen, S. P. Nair, J. M. Ward, A. N. Rycroft, and B. Henderson, "Phage display in the study of infectious diseases," *Trends in Microbiology*, vol. 14, no. 3, pp. 141–147, 2006.
- [26] J. Kugler, J. Zantow, T. Meyer et al., "Oligopeptide m13 phage display in pathogen research," *Viruses*, vol. 5, no. 10, pp. 2531–2545, 2013.
- [27] L. R. H. Krumpke, A. J. Atkinson, G. W. Smythers et al., "T7 lytic phage-displayed peptide libraries exhibit less sequence bias than M13 filamentous phage-displayed peptide libraries," *Proteomics*, vol. 6, no. 15, pp. 4210–4222, 2006.
- [28] G. P. Smith, "Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface," *Science*, vol. 228, no. 4705, pp. 1315–1317, 1985.

- [29] P. Malik, T. D. Terry, L. R. Gowda et al., "Role of capsid structure and membrane protein processing in determining the size and copy number of peptides displayed on the major coat protein of filamentous bacteriophage," *Journal of Molecular Biology*, vol. 260, no. 1, pp. 9–21, 1996.
- [30] J. Lubkowski, F. Hennecke, A. Plückthun, and A. Wlodawer, "Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA," *Structure*, vol. 7, no. 6, pp. 711–722, 1999.
- [31] D. A. Marvin, M. F. Symmons, and S. K. Straus, "Structure and assembly of filamentous bacteriophages," *Progress in Biophysics and Molecular Biology*, vol. 114, no. 2, pp. 80–122, 2014.
- [32] J. Vieira and J. Messing, "Production of single-stranded plasmid DNA," *Methods in Enzymology*, vol. 153, pp. 3–11, 1987.
- [33] K. Fukunaga and M. Taki, "Practical tips for construction of custom peptide libraries and affinity selection by using commercially available phage display cloning systems," *Journal of Nucleic Acids*, vol. 2012, Article ID 295719, 9 pages, 2012.
- [34] S. Rajan and S. S. Sidhu, "Simplified synthetic antibody libraries," *Methods in Enzymology*, vol. 502, pp. 3–23, 2012.
- [35] B. G. Condrón, J. F. Atkins, and R. F. Gesteland, "Frameshifting in gene 10 of bacteriophage T7," *Journal of Bacteriology*, vol. 173, no. 21, pp. 6998–7003, 1991.
- [36] S. C. Sharma, A. Memic, C. N. Rupasinghe, A.-C. E. Duc, and M. R. Spaller, "T7 phage display as a method of peptide ligand discovery for PDZ domain proteins," *Biopolymers*, vol. 92, no. 3, pp. 183–193, 2009.
- [37] F. Yang, P. Forrer, Z. Dauter et al., "Novel fold and capsid-binding properties of the  $\lambda$ -phage display platform protein gpD," *Nature Structural Biology*, vol. 7, no. 3, pp. 230–237, 2000.
- [38] J. Nicastro, K. Sheldon, and R. A. Slavcev, "Bacteriophage lambda display systems: developments and applications," *Applied Microbiology and Biotechnology*, vol. 98, no. 7, pp. 2853–2866, 2014.
- [39] I. N. Maruyama, H. I. Maruyama, and S. Brenner, " $\lambda$ foo: a  $\lambda$  phage vector for the expression of foreign proteins," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, no. 17, pp. 8273–8277, 1994.
- [40] N. Sternberg and R. H. Hoess, "Display of peptides and proteins on the surface of bacteriophage  $\lambda$ ," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, no. 5, pp. 1609–1613, 1995.
- [41] E. Santi, S. Capone, C. Mennuni et al., "Bacteriophage lambda display of complex cDNA libraries: a new approach to functional genomics," *Journal of Molecular Biology*, vol. 296, no. 2, pp. 497–508, 2000.
- [42] Y. G. Mikawa, I. N. Maruyama, and S. Brenner, "Surface display of proteins on bacteriophage  $\lambda$  heads," *Journal of Molecular Biology*, vol. 262, no. 1, pp. 21–30, 1996.
- [43] W. Li, "ORF phage display to identify cellular proteins with different functions," *Methods*, vol. 58, no. 1, pp. 2–9, 2012.
- [44] Z. Kalniņa, K. Siliņa, I. Meistere et al., "Evaluation of T7 and lambda phage display systems for survey of autoantibody profiles in cancer patients," *Journal of Immunological Methods*, vol. 334, no. 1–2, pp. 37–50, 2008.
- [45] H.-S. Lin, H. S. Talwar, A. L. Tarca et al., "Autoantibody approach for serum-based detection of head and neck cancer," *Cancer Epidemiology Biomarkers and Prevention*, vol. 16, no. 11, pp. 2396–2405, 2007.
- [46] P. Zacchi, D. Sblattero, F. Florian, R. Marzari, and A. R. M. Bradbury, "Selecting open reading frames from DNA," *Genome Research*, vol. 13, no. 5, pp. 980–990, 2003.
- [47] P. H. Faix, M. A. Burg, M. Gonzales, E. P. Ravey, A. Baird, and D. Larocca, "Phage display of cDNA libraries: enrichment of cDNA expression using open reading frame selection," *BioTechniques*, vol. 36, no. 6, pp. 1018–1029, 2004.
- [48] Y. Georgieva and Z. Konthur, "Design and Screening Of M13 phage display cDNA libraries," *Molecules*, vol. 16, no. 2, pp. 1667–1681, 2011.
- [49] K. Jacobsson and L. Frykberg, "Gene VIII-based, phage-display vectors for selection against complex mixtures of ligands," *BioTechniques*, vol. 24, no. 2, pp. 294–301, 1998.
- [50] L. S. Jaspers, J. H. Messens, A. de Keyser et al., "Surface expression and ligand-based selection of cdnas fused to filamentous phage gene VI," *Nature Biotechnology*, vol. 13, no. 4, pp. 378–382, 1995.
- [51] M. Hust, M. Meysing, T. Schirrmann et al., "Enrichment of open reading frames presented on bacteriophage M13 using hyperphage," *BioTechniques*, vol. 41, no. 3, pp. 335–342, 2006.
- [52] P. Kristensen and G. Winter, "Proteolytic selection for protein folding using filamentous bacteriophages," *Folding and Design*, vol. 3, no. 5, pp. 321–328, 1998.
- [53] J. Kügler, S. Nieswandt, G. F. Gerlach, J. Meens, T. Schirrmann, and M. Hust, "Identification of immunogenic polypeptides from a *Mycoplasma hyopneumoniae* genome library by phage display," *Applied Microbiology and Biotechnology*, vol. 80, no. 3, pp. 447–458, 2008.
- [54] A. Gupta, N. Shrivastava, P. Grover et al., "A novel helper phage enabling construction of genome-scale ORF-enriched phage display libraries," *PLoS ONE*, vol. 8, no. 9, Article ID e75212, 2013.
- [55] R. Cramer and M. Suter, "Display of biologically active proteins on the surface of filamentous phages: a cDNA cloning system for the selection of functional gene products linked to the genetic information responsible for their production [Gene 137 (1993) 69–75]," *Gene*, vol. 160, no. 1, 139 pages, 1995.
- [56] S. Cramer, R. Jaussi, G. Menz, and K. Blaser, "Display of expression products of cDNA libraries on phage surfaces. A versatile screening system for selective isolation of genes by specific gene-product/ligand interaction," *European Journal of Biochemistry*, vol. 226, no. 1, pp. 53–58, 1994.
- [57] D. Moisy, S. V. Avilov, Y. Jacob et al., "HMGB1 protein binds to influenza virus nucleoprotein and promotes viral replication," *Journal of Virology*, vol. 86, no. 17, pp. 9122–9133, 2012.
- [58] A.-J. van Zonneveld, B. M. M. van den Berg, M. van Meijer, and H. Pannekoek, "Identification of functional interaction sites on proteins using bacteriophage-displayed random epitope libraries," *Gene*, vol. 167, no. 1–2, pp. 49–52, 1995.
- [59] K. Hertveldt, M. L. Dechassa, J. Robben, and G. Volckaert, "Identification of Gal80p-interacting proteins by *Saccharomyces cerevisiae* whole genome phage display," *Gene*, vol. 307, no. 1–2, pp. 141–149, 2003.
- [60] K. Yano, T. Yoshino, M. Shionoya, S. Y. Sawata, K. Ikebukuro, and I. Karube, "Preparation of a whole genome phage library using fragmented *Escherichia coli* genome and its characterization of protein binding properties by surface plasmon resonance," *Biosensors and Bioelectronics*, vol. 18, no. 10, pp. 1201–1207, 2003.
- [61] E. Malini, E. Maurizio, S. Bembich, R. Sgarra, P. Edomi, and G. Manfioletti, "HMGA interactome: new insights from phage display technology," *Biochemistry*, vol. 50, no. 17, pp. 3462–3468, 2011.

- [62] R. di Niro, A. M. Sulic, F. Mignone et al., "Rapid interactome profiling by massive sequencing," *Nucleic Acids Research*, vol. 38, no. 9, pp. e110–e110, 2010.
- [63] J. A. Ubersax and J. E. Ferrell Jr., "Mechanisms of specificity in protein phosphorylation," *Nature Reviews Molecular Cell Biology*, vol. 8, no. 7, pp. 530–541, 2007.
- [64] D. Cochrane, C. Webster, G. Masih, and J. McCafferty, "Identification of natural ligands for SH2 domains from a phage display cDNA library," *Journal of Molecular Biology*, vol. 297, no. 1, pp. 89–97, 2000.
- [65] C. Drr, H. Nothaft, C. Lizak, R. Glockshuber, and M. Aebi, "The *Escherichia coli* glyco-phage display system," *Glycobiology*, vol. 20, no. 11, pp. 1366–1372, 2010.
- [66] M. Fransens, P. P. van Veldhoven, and S. Subramani, "Identification of peroxisomal proteins by using M13 phage protein VI phage display: molecular evidence that mammalian peroxisomes contain a 2,4-dienoyl-CoA reductase," *Biochemical Journal*, vol. 340, part 2, pp. 561–568, 1999.
- [67] V. A. Somers, R. J. Brandwijk, B. Joosten et al., "A panel of candidate tumor antigens in colorectal cancer revealed by the serological selection of a phage displayed cDNA expression library," *Journal of Immunology*, vol. 169, no. 5, pp. 2772–2780, 2002.
- [68] C. Govarts, K. Somers, R. Hupperts, P. Stinissen, and V. Somers, "Exploring cDNA phage display for autoantibody profiling in the serum of multiple sclerosis patients: optimization of the selection procedure," *Annals of the New York Academy of Sciences*, vol. 1109, pp. 372–384, 2007.
- [69] K. Somers, P. Geusens, D. Elewaut et al., "Novel autoantibody markers for early and seronegative rheumatoid arthritis," *Journal of Autoimmunity*, vol. 36, no. 1, pp. 33–46, 2011.
- [70] S. Danner and J. G. Belasco, "T7 phage display: a novel genetic selection system for cloning RNA-binding proteins from cDNA libraries," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 23, pp. 12954–12959, 2001.
- [71] Y. Kim, N. B. Caberoy, G. Alvarado, J. L. Davis, W. J. Feuer, and W. Li, "Identification of Hnrph3 as an autoantigen for acute anterior uveitis," *Clinical Immunology*, vol. 138, no. 1, pp. 60–66, 2011.
- [72] T. Tan, X. Xiang, H. Qu, S. Zhu, and Q. Bi, "The study on venom proteins of *Lapemis hardwickii* by cDNA phage display," *Toxicology Letters*, vol. 206, no. 3, pp. 252–257, 2011.
- [73] S. B. Lauterbach, R. Lanzillotti, and T. L. Coetzer, "Construction and use of *Plasmodium falciparum* phage display libraries to identify host parasite interactions," *Malaria Journal*, vol. 2, article 1, 2003.
- [74] H. J. Ren, R. D. Liu, Z. Q. Wang, and J. Cui, "Construction and use of a *Trichinella spiralis* phage display library to identify the interactions between parasite and host enterocytes," *Parasitology Research*, vol. 112, no. 5, pp. 1857–1863, 2013.
- [75] N. B. Caberoy, Y. Zhou, X. Jiang, G. Alvarado, and W. Li, "Efficient identification of tubby-binding proteins by an improved system of T7 phage display," *Journal of Molecular Recognition*, vol. 23, no. 1, pp. 74–83, 2010.
- [76] N. B. Caberoy, D. Maignel, Y. Kim, and W. Li, "Identification of tubby and tubby-like protein 1 as eat-me signals by phage display," *Experimental Cell Research*, vol. 316, no. 2, pp. 245–257, 2010.
- [77] N. B. Caberoy, G. Alvarado, and W. Li, "Identification of calpain substrates by ORF phage display," *Molecules*, vol. 16, no. 2, pp. 1739–1748, 2011.
- [78] B. Luo, T. Zou, N. Lu et al., "Role of suppressor of cytokine signaling 3 in lipid metabolism: analysis based on a phage-display human liver cDNA library," *Biochemical and Biophysical Research Communications*, vol. 416, no. 1-2, pp. 39–44, 2011.
- [79] E. Beghetto and N. Gargano, "Lambda-display: a powerful tool for antigen discovery," *Molecules*, vol. 16, no. 4, pp. 3089–3105, 2011.
- [80] I. Kuwabara, H. Maruyama, Y. G. Mikawa, R. I. Zuberi, F. T. Liu, and I. N. Maruyama, "Efficient epitope mapping by bacteriophage lambda surface display," *Nature Biotechnology*, vol. 15, no. 1, pp. 74–78, 1997.
- [81] E. Beghetto, N. Gargano, S. Ricci et al., "Discovery of novel *Streptococcus pneumoniae* antigens by screening a whole-genome  $\lambda$ -display library," *FEMS Microbiology Letters*, vol. 262, no. 1, pp. 14–21, 2006.
- [82] H. Ansuini, C. Cicchini, A. Nicosia, M. Tripodi, R. Cortese, and A. Luzzago, "Biotin-tagged cDNA expression libraries displayed on lambda phage: a new tool for the selection of natural protein ligands," *Nucleic acids research*, vol. 30, no. 15, p. e78, 2002.
- [83] S. Kosuri and G. M. Church, "Large-scale de novo DNA synthesis: technologies and applications," *Nature Methods*, vol. 11, no. 5, pp. 499–507, 2014.
- [84] H. B. Larman, Z. Zhao, U. Laserson et al., "Autoantigen discovery with a synthetic human peptidome," *Nature Biotechnology*, vol. 29, no. 6, pp. 535–541, 2011.
- [85] G. Fuh, M. T. Pisabarro, Y. Li, C. Quan, L. A. Lasky, and S. S. Sidhu, "Analysis of PDZ domain-ligand interactions using carboxyl-terminal phage display," *The Journal of Biological Chemistry*, vol. 275, no. 28, pp. 21486–21491, 2000.
- [86] R. Mrowka, A. Patzak, and H. Herzel, "Is there a bias in proteome research?" *Genome Research*, vol. 11, no. 12, pp. 1971–1973, 2001.
- [87] Y. Ivarsson, R. Arnold, M. McLaughlin et al., "Large-scale interaction profiling of PDZ domains through proteomic peptide-phage display using human and viral phage peptidomes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 7, pp. 2542–2547, 2014.
- [88] H. Huang and S. S. Sidhu, "Studying binding specificities of peptide recognition modules by high-throughput phage display selections," *Methods in Molecular Biology*, vol. 781, pp. 87–97, 2011.
- [89] W. L. Matochko, K. Chu, B. Jin, S. W. Lee, G. M. Whitesides, and R. Derda, "Deep sequencing analysis of phage libraries using Illumina platform," *Methods*, vol. 58, no. 1, pp. 47–55, 2012.
- [90] A. H. Y. Tong, B. Drees, G. Nardelli et al., "A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules," *Science*, vol. 295, no. 5553, pp. 321–324, 2002.
- [91] X. Xin, D. Gfeller, J. Cheng et al., "SH3 interactome conserves general function over specific form," *Molecular Systems Biology*, vol. 9, article 652, 2013.

## Review Article

# Applications of Flow Cytometry to Characterize Bacterial Physiological Responses

Verónica Ambriz-Aviña, Jorge A. Contreras-Garduño, and Mario Pedraza-Reyes

Departamento de Biología, División de Ciencias Naturales y Exactas, Universidad de Guanajuato, Noria Alta S/N, 36050 Guanajuato, GTO, Mexico

Correspondence should be addressed to Mario Pedraza-Reyes; pedrama@ugto.mx

Received 27 June 2014; Revised 13 August 2014; Accepted 13 August 2014; Published 9 September 2014

Academic Editor: Gerhard Litscher

Copyright © 2014 Verónica Ambriz-Aviña et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Although reports of flow cytometry (FCM) applied to bacterial analysis are increasing, studies of FCM related to human cells still vastly outnumber other reports. However, current advances in FCM combined with a new generation of cellular reporter probes have made this technique suitable for analyzing physiological responses in bacteria. We review how FCM has been applied to characterize distinct physiological conditions in bacteria including responses to antibiotics and other cytotoxic chemicals and physical factors, pathogen-host interactions, cell differentiation during biofilm formation, and the mechanisms governing development pathways such as sporulation. Since FCM is suitable for performing studies at the single-cell level, we describe how this powerful technique has yielded invaluable information about the heterogeneous distribution of differently and even specialized responding cells and how it may help to provide insights about how cell interaction takes place in complex structures, such as those that prevail in bacterial biofilms.

## 1. Introduction

The study of bacterial physiological responses using approaches that assess the overall population response considering it as homogeneous is becoming a thing of the past. Current experimental evidence indicates that an ordinary laboratory culture considered as being composed of isogenic bacteria is actually constituted of heterogeneous subpopulations [1, 2] that respond differentially to changes in their environment [3]. Although an invaluable amount of knowledge has been obtained by studying batch cultures, it should be considered that intracellular heterogeneity and cell-to-cell interaction (e.g., *quorum sensing*) exist in an isogenic bacterial culture. Therefore, analysis focused on subpopulations combined with single-cell analysis techniques will help progress our understanding of the bacterial subpopulation behavior in laboratory cultures.

Current available technologies like flow cytometry (FCM) provide more information regarding the individual events that may rule out the overall population response [4]. As reviewed and discussed by Lidstrom and Konopka [5], a normal distribution in the response across the bacterial cells in

a culture does not exist, but there are *on*, *off*, and *intermediate* states that depend on a threshold response mechanism, giving rise to physiologically distinct populations. The extent or probability that a well-differentiated subpopulation response could impact on the overall population behavior or fate of a bacterial culture will depend on the nature of the response itself. For instance, when the subpopulation produces and secretes an inhibitor, a growth factor, or an autoinducer into the medium, it causes a response from the rest of the cells [6]. The origins of such heterogeneity are, according to the literature [7], due to differences in microenvironments [8], created by a large number of bacterial cells *growing in* and *modifying a* culture, giving rise to what is called extrinsic noise, in combination with intrinsic cellular noise, due to the fact that each cell possess a different spatiotemporal distribution of cellular components, that is, the threshold levels of molecules that switch *on/off* gene expression arose at different times in each cell, since gene expression is stochastic [9].

In this context, if we are interested in assessing a bacterial response to certain stimuli, using a fluorescent tracer, and we find no difference between the treated and non-treated samples, it may be due to an absence of response or perhaps

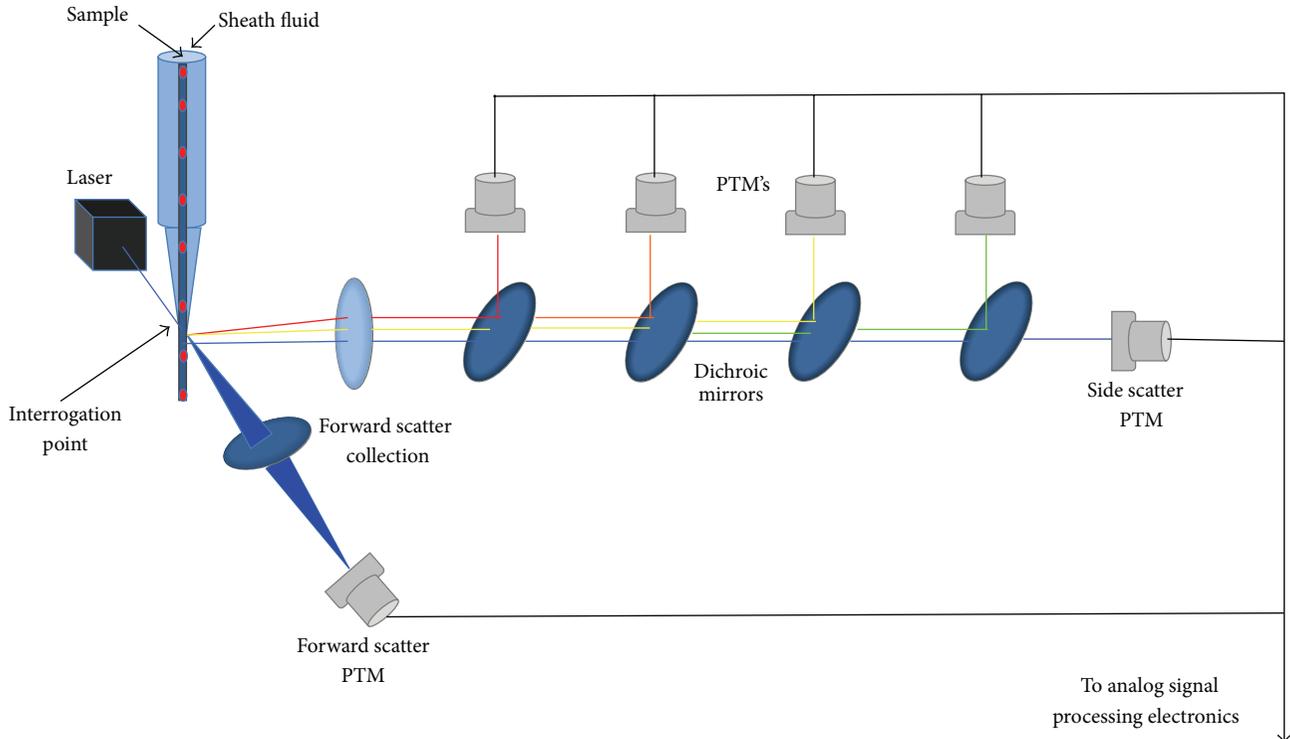


FIGURE 1: Drawing of the main elements that compose a flow cytometer.

to a response in a few percentage of the cells. In a distinct scenario, we could measure the treated culture, if we find in the treated sample half the fluorescence compared to the non-treated sample, we may interpret that all the cells diminished the response by half or perhaps that half the cells do not fluoresce at all or perhaps all the intermediary situations. In order to solve this question, it is necessary to measure the fluorescence emitted by each cell in the culture.

FCM is a useful tool for accomplishing this purpose; this technology was first used in the 1960's [10], and since then a huge increase in its development and applications in different fields can be found in the literature. In 1996, Davey and Kell wrote a comprehensive review about the application of FCM in studying heterogeneous microbial populations [10]; therefore, the present review is primarily focused on the information published after that report, emphasizing the applications of FCM when studying bacterial physiological responses.

## 2. Application of Flow Cytometry

**2.1. Flow Cytometry: An Overview.** A flow cytometer is an apparatus that makes cells or micrometric particles pass through an interrogation point, where a laser beam impacts them and the light that the particle absorbs, scatters or emits due to its intrinsic or extrinsic physical properties are measured. There are a variety of bibliographic resources [11, 12] that explain at length how a flow cytometer works, but for practical purposes, here we only present a brief and simple

description. A flow cytometer is composed of three major parts: fluidics, optics, and electronics systems (Figure 1). The goal of the fluidics system is to make the cell or particle of interest pass through the interrogation point, one by one, in the center of a core stream, within which the diameter of the particle's trajectory is controlled by means of the laminar flow exerted by a surrounding sheath stream so that ideally just one cell passes at a time. The optics system consists of a light source: the most typical one is a 488 nm, (blue) Argon ion laser; this beam is focused by means of lens to impact on the cell in the interrogation point. The particle under study scatters light to all the angles, and the light scattered at acute angles, called forward scatter, is (in general) indicative of a particle's size. The light scattered at wide angles ( $90^\circ$ ), which is called side scatter, is proportional to the particles' roughness and complexity. Forward scattered light is detected in front of the incident laser beam trajectory by a photodiode or by a photomultiplier tube (PMT), and side scattered light and fluorescence are collected, collimated, and directed  $90^\circ$  through a pathway between a series of dichroic mirrors or beam splitters that permit the pass of certain, unwanted wavelength light and reflect the wanted light to a band pass filter (red, orange, yellow, and green light) and finally to a PMT (Figure 1). In the electronics system, the PMT collects photons and expels electrons, amplifying the signal because a PMT may produce a few hundred electrons for each photon. The photo detectors then produce brief current pulses that are amplified, converted to voltage, and finally converted to numbers by the analog signal processing electronics. There are several ways to represent all the hundreds of events

generated per second, and ultimately, although the user knows that he puts in cells in a flow cytometer, after the analysis he obtains a series of numbers and plots to interpret [11]. Here is where a subsequent fluorescence activated cell sorting (FACS) and/or complementary (e.g., microscopy) analysis would be useful.

As noted above, FCM provides useful information regarding the intrinsic characteristics of a cell or particle; thus, the amount of light scattered at low angles or forward scattering (FS) keeps in most cases a direct relationship with the size of the particle, whereas the complexity, granularity, and protein content can be generally estimated on basis of the light scattered at high angles, a property referred to as side scattering (SS) [11]. Although light scattering is a function of intrinsic characteristics of cells it is difficult to distinguish between different bacterial species employing only this property, therefore, to accomplish this task, cells must be marked with fluorescent dyes [13]. However, light scattering has been successfully used to determine distinct biological properties in bacteria, including, poli( $\beta$ -hydroxybutyrate) accumulation [14, 15] as well as cell filamentation and death [16, 17]. Moreover, changes in light scattering has been successfully used for monitoring and sorting the best clones in an *E. coli* culture capable of producing interferon or growth hormone [18]. The capacity of conventional flow cytometers to measure intrinsic and/or extrinsic fluorescence in cells has been used as a property to assess cell viability, protein identity, and enzymatic activity [10, 11, 19]. Discussions regarding the specific properties of fluorescent probes have been published elsewhere [10–12]. Moreover, a complete list of fluorescent dyes can be found in the catalog of Molecular Probes [20]. Readers can find the most appropriate dye for specific FCM needs using these resources, including the target, the type of cell under study, and the excitation/emission wavelengths of the fluorescent compound. One step that may be problematic for FCM analysis during staining is that permeabilization and fixation may affect the viability of the cells; specific examples of how these problems can be overcome have been previously published [16]. Gene expression analysis in bacteria has been assessed by FCM employing fluorescent proteins including the *Aequorea* green fluorescent protein (GFP) [21]. Special forms of GFP have been specifically tailored for its application in FCM [22, 23].

The small size of the bacteria could pose a limitation for FCM analysis specifically for the difficulty of distinguishing between small cells and cellular debris. To overcome this, the literature recommends using both forward scatter and fluorescence as dual trigger signals [10, 11]. Moreover, nucleic acid staining for distinguishing between cells and abiotic particles and the employment of polymer beads for standardization have been found useful in overcoming this limitation [16, 17, 24]. Finally, cell aggrupration, including the formation of bacterial chains or clusters of cells, may be problematic during FCM analysis since this technique measures events and cannot distinguish between a single cell and groups of cells passing through the interrogation point. Therefore, it is necessary to disaggregate and homogenize the sample before analysis; mild sonication has been applied to accomplish this purpose [25–27].

Fluorescence activated cell sorting (FACS) is an invaluable tool for separating the subpopulation(s) of interest that possess certain measurable characteristics. A cell sorter is basically a flow cytometer that has the option of separating cells. When the cell sorter detects a particle with the characteristics chosen by the operator, a charge is given to the droplet where that particle is, and as the droplets are passing through an electrostatic field formed between two charged plates, charged droplets are deflected appropriately, whereas uncharged droplets continue on their original course to the waste tank. With this methodology, around 40 000 cells can be separated per second [28]. As shown in Figure 2, what makes this technique relevant is the possibility of separating subpopulations for subsequent physiological and/or molecular analysis [29, 30]; moreover, cells can be recovered alive for growth and physiological analysis purposes [31]. However, if the cell population of interest exists in a very low abundance, it may be very difficult in principle to set the correct gate for sorting it. Secondly, it may take a long time to obtain a useful quantity of cells for downstream analysis; for example, at least 1000 to 10 000 cells may be needed for obtaining detectable PCR products [30], and if the population of interest constitutes 0.1% of the total sample, it would take us approximately 4 minutes to obtain 10 000 cells at a sorting rate of 40 000 cells per second, and it would take us up to 41 minutes to obtain the sufficient number of cells if the abundance of such population is as low as 0.01%. Furthermore, to perform proteomics, up to  $5 \times 10^6$  to  $10^9$  bacterial cells may be required, and it took 3 days and 3 weeks, respectively, to collect that number of cells of interest [29, 32, 33]. Both the time and gating problems have proven to be diminished by repeated sort cycles, enriching first the population of interest using a wide gate and after that using a narrower gate to purify it [30]. Otherwise, if we are interested in certain subpopulation that possess a transient physiological state for further analysis, the lapse of time required for separate those cells could be too long that the transient phenotype could be missed before reaching the number of cells needed.

Although FCM emerged from 1960's as a tool for analyzing blood cells, advances in flow, optics, electronics, and probing techniques have done this approach suitable for study of a great number of purposes in different biological systems [12]. Quixabeira et al. [34] made an analysis of published literature regarding FCM applied to genetic studies, from 1991 to 2007, and they reported that more than a half of these studies are human related, the majority are related to diseases, and just a small proportion has been devoted to studying bacterial responses (viruses/bacteria, 8.4%). The main factor that has limited the use of FCM in prokaryote field is the small size of bacteria which makes them difficult for detection [4, 34, 35]. However, current improvements have allowed applying this technique to characterize the physiological events underlying the bacterial responses as described below.

**2.2. Flow Cytometry Analysis of Bacterial Response to Antibiotic Agents.** Because an increasing number of resistant strains are emerging [36, 37], huge efforts are dedicated to new

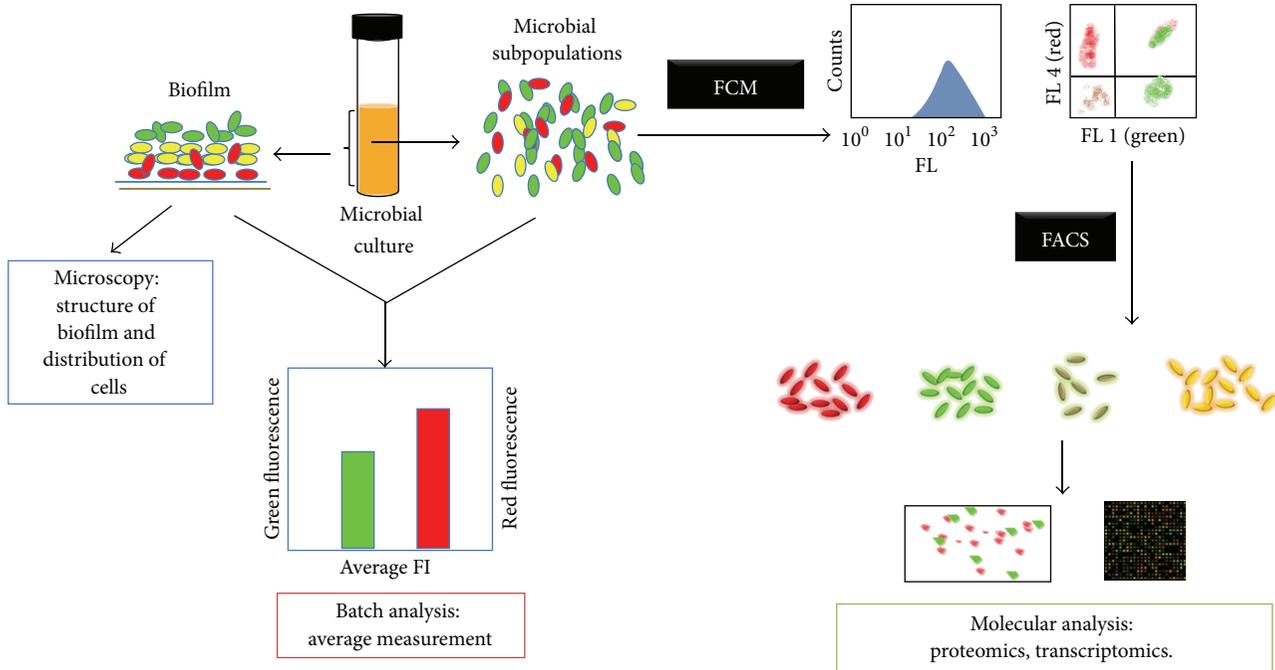


FIGURE 2: *Left.* Structure and cell distribution in biofilms can be studied by microscopy. *Bottom left.* Analysis of a bacterial culture sample allows obtaining average measurements of a physiological response. *Right.* FCM allows analyzing subpopulations of cultured bacterial cells with different physiological states; FACS permits separating those subpopulations that can be independently characterized by high throughput molecular techniques.

antibacterial drug research [38, 39] using automated methods to assess the number of candidates required, besides the necessity to find in terms of hours rather than days the appropriate drug for treating bacterial infections in clinical practice [40, 41].

When studying response to antibiotic agents, susceptibility is traditionally assessed in bulk, by the growth capability of the bacteria. It is possible to study viability as a parameter of antibiotic susceptibility at single-cell level, measuring criteria such as impermeability of membrane to dyes, maintenance of membrane potential, and the presence of metabolic activity measured by the production of a fluorescent metabolite from a nonfluorescent precursor [12].

Walberg et al. [24] employed FCM to assess the susceptibility of growing *E. coli* cells to mecillinam and ampicillin by using DNA staining with a combination of mithramycin and ethidium bromide (first reported by Steen et al., 1994 [42]). Prior to staining, the cells had to be permeabilized with ice-cold ethanol treatment. Fluorescence was recorded to measure DNA content, forward scatter recorded as cell size indicative, and side scatter as proportional to dry weight or protein content. These three parameters were augmenting in cells exposed to antibiotic since 30 minutes (and doubled at 60 min) incubated with minimal inhibitory concentration (MIC) of mecillinam and ampicillin. Thus showing that DNA and protein synthesis continued but cells were unable to complete division, since cell number remains constant. Such results were in agreement with the known mode of action of penicillin, which interferes with cell wall synthesis.

Walberg's results indicate that, under the conditions tested, drug responses can be detected by light scatter alone, without staining the cells, but, as Walberg points out, in clinical samples, fluorescence measuring is important in order to distinguish between cells and other particles like debris. In an ongoing study [16], the same group demonstrated the applicability of flow cytometry measurement of DNA content and light scattering to assesses the susceptibility of antibiotic drugs, with different modes of action, including, ceftazidime a beta lactam that works as a cell wall antagonist, ciprofloxacin, a quinolone that targets DNA gyrase, and gentamicin, an aminoglycoside that irreversibly binds to ribosomes. Their results, supported by microscopic observations, were consistent with the action mode of the tested drugs. They reported for these drugs that the cell size/dry weight ratio and DNA content augment as a function of drug concentration and exposure time. The same was true in fluorescence versus forward scatter plot with the existence of populations of filamenting cells, disintegrating filaments, and debris after treatment with the drugs. In a subsequent study, Walberg et al. [17] reported the applicability of the previously exploited technique in studying a heterogeneous drug response, to a mixed culture of two clinically important bacteria species: *E. coli* and *Klebsiella pneumoniae*. These microorganisms are associated with polymicrobial urinary tract infection and both have different susceptibility to the drug tested, ampicillin [17]. This study was found to be valuable in detecting polymicrobial infections; by measuring cell number, light scattering, and DNA content-associated

fluorescence, the findings show that it was possible to detect susceptible and resistant cells within an hour of incubation with ampicillin in the same sample.

In another study, Dessus-Babus et al. developed a flow cytometric method to calculate the minimal inhibitory concentration (MIC) to distinct antibiotics in clinical isolates of *Chlamydia trachomatis* and a reference strain [43]. They infected McCoy cells and later treated them with increasing doses of doxycycline, ofloxacin, and erythromycin; after incubation with the antibiotic, they immunostained McCoy cells with a fluorescein isothiocyanate- (FITC-) conjugated antibody to detect *Chlamydia* inclusions. Dessus-Babus and colleagues were able to quantify inclusion forming colonies (IFC)/mL by microscopy and by FCM; the latter method was not as sensitive as microscopy, but it did have the advantages of being specific and reproducible. Although infected McCoy cells were quantified instead of bacterial cells, this study demonstrated that flow cytometry is useful because the interpretation of results did not depend entirely on a skilled and experienced observer. Moreover, since FCM is an automated method, it is time-saving and results can be statistically improved.

Another application of flow cytometry to determine MIC of antibiotic drugs was reported by Assunção et al. in 2007. These authors succeeded in determining the MIC of nine different antibiotics in *Mycoplasma hyopneumoniae*, quantifying the total number of cells by means of staining DNA with SYBR Green. The results obtained with this approach were comparable with those obtained by the classical broth dilution method, which is dependent on the observation of a color change or increasing turbidity in the growth medium. However, the FCM approach was faster (12 h) than the broth dilution method (48 h), and furthermore, it was concluded that at 48 h FCM was more sensitive for tylosin and at 72 h for oxytetracycline and streptomycin MIC determination [44].

Recently, Soejima et al. reported for the first time an FCM based methodology to distinguish between live-injured and dead *Listeria monocytogenes* cells after treatment with antibiotics [45]. Furthermore, these authors were able to identify both states of bacterial cells in clinical blood samples, revealing the potential value of this approach to the opportune evaluation of bacteremia and the assessment of drug treatment in real time. This methodology employed photoactivated ethidium monoazide (EMA) that cleaves DNA of injured, not living cells, in combination with SYTO9 that enters live and dead cells and PI that penetrates dead cells that had lost membrane integrity [45].

Besides susceptibility to antibiotic assessment, relevant information can be obtained by FCM analysis in order to establish antibacterial drug action mechanisms [46–51]. Such information cannot be obtained by traditional culture-based techniques [52]. The use of fluorescent probes to detect specific cell changes [12], including, permeabilization and changes, in membrane potential [53–55], DNA content [16, 17, 24], and metabolic activity [53, 55], are useful parameters to assess viability and thus antibiotic susceptibility. Thus, Suller and Lloyd [52] performed a study to evaluate the effects of ceftazidime on *Pseudomonas aeruginosa*, ampicillin on *E. coli*, and vancomycin on *Staphylococcus aureus*,

using the fluorescent probes bis(1,3-dibutylbarbituric acid), trimethine oxonol (DiBAC<sub>4</sub>(3)), and SYTOX Green to measure membrane potential, the redox dye cyano-2,3-ditolyl tetrazolium chloride (CTC) to measure actively respiring bacteria, and the BacLite viability kit (Molecular Probes, Life Technologies, Grand Island, NY) to test viability. Results showed that the use of these fluorophores is effective to assess the antibacterial activities of the drugs tested although different responses between dyes were apparent. Thus, CTC was more efficient in detecting different subpopulations than DiBAC<sub>4</sub>(3) and SYTOX Green when fluorescence values were plotted as a function of forward scatter; moreover, when the dyes (SYTO9 and propidium iodide) of the BacLite viability kit were used, no populations of differently responding cells were detected. Additionally, the use of these dyes has the advantage that no pretreatments of cells were required. Other studies have succeeded in employing fluorescent probes to measure the efficiency of new antibacterial drugs [12, 56, 57]. For instance, Ghosh et al. [57] demonstrated by FCM the antibacterial activity of two bioactive compounds isolated from seeds of *Alpinia nigra* against seven pathogenic bacteria. Ghosh and colleagues found by measuring the extent of propidium iodide DNA staining of treated cells that both compounds caused significant increase in fluorescence intensity compared to the controls, which indicates membrane damage. The interpretation of this finding was supported by field emission scanning electronic microscopy; they observed membrane disintegration and significant damage to the cell wall [57].

Since multiparametric flow cytometric analysis gives precise, reproducible, and accurate information regarding cell function at a single-cell level, the validity of growth based methods to assess cell viability has been questioned [58]. However, it has also been questioned whether cytometric measurements correlate with cell viability in the culture [12]; thus, it is reasonable to emphasize that there is no a universal formula that can be applied to every bacteria and every drug, but it is clear that FCM is a very sensitive, accurate, and time-saving method for assessing bacterial responses to antibiotics [16, 50]. Inconsistencies between cytometric analysis and growth based approaches can be explained by the existence of metabolically active but noncultivable cells (VNBC) in a cell culture [58, 59]. In fact, FCM has been successfully applied to detect VBNC cells from a variety of sanitary and clinically important sources, when the isolation and growth of viable cells are not possible [60, 61]. Cell sorting (by FACS) and further analysis of cells allows us to do further tests in order to corroborate the cytometric results with traditional methods, such as growth in culture and high-resolution single-cell microscopic analysis [58, 62].

As noted above, GFP reporter fusions are very useful for studying gene expression, and Sánchez-Romero and Casadesús [63] quantified GFP fluorescence by flow cytometry, in a liquid isogenic culture of *Salmonella enterica* expressing a *ompC:gfp* reporter gene fusion. In this study, heterogeneity was observed in the expression levels of the outer membrane porin coding gene; moreover, after sorting populations with high and low expression levels and assessing the susceptibility of such cell populations to kanamycin, it

was found that low levels of *ompC* expression correlate with high kanamycin resistance. They proposed that the noisy expression of *ompC* is a mechanism contributing to the adaptive resistance to lethal concentrations of kanamycin [63]. In another study, Cui et al. used a GFP fusion as reporter of *graF* expression; the researchers sorted *S. aureus* cells with different levels of fluorescence by FACS, further susceptibility analysis to glycopeptide antibiotics, and morphology studies of *graF* overexpressing cells were performed. They found that upregulated activity of *graF* promoter is consistent with reduced susceptibility to glycopeptides due to an increased cell wall thickness [64]. Another approach to detect resistant bacteria to antibiotics is using fluorescent drug analogs called reporter enzyme fluorescence [65, 66] or using drug fluorescent analogs with FRET pairs (fluorescence resonance energy transfer), that upon enzyme cleavage, the quenching molecule is eliminated and the fluorophore then emits light [67, 68]. An example of this approach was reported by Shao et al. who designed optical probes for detecting  $\beta$ -lactamase activity and covalent fluorescent labeling of antibiotic-resistant bacteria [69]. Therefore, an increasing number of future applications in research and clinical fields using the above-mentioned techniques may be anticipated, which can be substantially potentiated by employing personalized probes or dyes for monitoring particular targets and cellular functions.

**2.3. Measurement of Bacterial Responses to Other Chemical and Physical Stresses.** Together with the search for chemical agents that are useful to kill bacteria, other strategies have also considered the idea of causing cell stress and stopping bacterial growth. One of such approaches employs photodynamic inactivation (PDI), which consists of the use of a nontoxic photosensitizer that is activated by harmless visible light. This then produces reactive oxygen species that causes fatal damage to target bacteria [70]. This approach has been tested *in vitro* [71] and *in vivo* [72]. Since PDI causes nonspecific injuries to the pathogens, it is very unlikely that bacteria can acquire resistance. One work related to ROS production by photoactivation of hypocrellin B (a component of a traditional Chinese herb *H. bambuase*) was reported by Jiang et al. [73]. In this report, clinically isolated *S. aureus* cells were incubated with hypocrellin B and after photodynamic treatment they used DCFH-DA (dichlorodihydrofluorescein diacetate) to measure by FCM the production of ROS [74]. ROS levels increased substantially in cells treated with PDI. The treatment caused reduction of viability as demonstrated by CFU counts and cellular damage, including membrane damage and cytoplasm leakage that was corroborated by confocal laser scanning microscopy [73].

FCM has also been applied in assessing the effectiveness of the treatments that food products for human consumption receive to eliminate and prevent pathogen dissemination and spoiling bacteria, including *Listeria monocytogenes*. This bacterium possesses a remarkable adaptability to stress conditions during food processing such as acidic pH, high salt concentration, and extreme temperatures [75]. It has been shown that, in order to cope with these stresses, this microorganism

activates the expression of genes from the general stress sigma B ( $\sigma^B$ ) regulon. To clarify the role of  $\sigma^B$  during adaptation to low temperature, Utratna et al. analyzed the expression of a reporter gene fusion between the promoter of the  $\sigma^B$ -dependent *Imo2230* gene and GFP at normal and low temperatures. The FCM analysis revealed a heterogeneous activation of the  $\sigma^B$ -dependent GFP fusion expression occurring since the early exponential phase, although the maximal expression occurred when cells entered into the stationary phase of growth; however, such a result was found to occur at both temperatures tested. Therefore, they concluded that  $\sigma^B$  does not play a pivotal role in adaptation to cold temperatures [76].

Another bacterium receiving a lot of attention in regard to stress response characterization, being a causal agent of food borne diseases, is *Bacillus cereus* [77]. Flow cytometry among other techniques has been useful in studying *B. cereus* responses to several chemical and physical stresses, such as low pH, which are encountered by the bacterium in several situations including the treatment that alimentary products are subjected to [78], in the host's gastrointestinal system [77, 79], and to other sporistatic/sporicidal physicochemical treatments [80, 81].

FCM has also been applied to investigate the way in which bacteria respond and adapt to stressful conditions that they face during industrial and remediation bioprocesses [82]. *Streptococcus macedonicus*, *Bacillus licheniformis*, and *Lactobacillus rhamnosus* are examples of bacteria employed in industrial fermentations. Papadimitriou et al. [55] reported the use of FCM to assess *in situ* the physiological status of *S. macedonicus* in response to acid stress. This microorganism that is a member of the lactic acid bacteria group is widely used in food industry. In this report, Papadimitriou and colleagues employed FCM to analyze characteristics such as membrane potential with DiBAC<sub>4</sub>(3), membrane integrity with propidium iodide, and enzymatic activity as well as membrane integrity with cFDA, carboxyfluorescein diacetate. Finally, they determined cultivability after cell sorting. From these analyses, the coexistence of three distinct subpopulations was observed: intact/culturable, permeabilized/dead, and potentially injured with decreased cultivability [55].

Sunny-Roberts and Knorr monitored changes in membrane integrity using propidium iodide and esterase activity with carboxyfluorescein diacetate of *L. rhamnosus* in response to osmotic stress, which is a condition that this bacterium has to challenge in natural habitats as well as in food formulations and processes where a probiotic bacterium is used. By FCM and conventional culture techniques, it was found that this microorganism is able to tolerate even extreme sucrose concentrations [83], and such a characteristic is exploitable for its use in food processing and formulations.

In another study, FCM was employed to assess physiological responses of the industrially important bacterium *B. licheniformis*, used in bioremediation processes. Using the same technique, researchers have monitored population dynamics of continuous cultivations, responses to starvation conditions, and responses to glucose and lactose pulses [84, 85].

The heterologous expression of proteins in *E. coli* is a routine method for obtaining large amounts of recombinant proteins. As this process may represent a stressful condition to *E. coli* cells, Borth et al. analyzed by FCM the physiological changes in response to the metabolic stress that represents the production of foreign proteins at high yield. They measured total DNA and RNA content (indicators of cell division and protein synthesis activity, resp.), total protein content (cell size), and intracellular SOD content (foreign protein). They found that after induction of foreign protein synthesis, cells increased their biomass but did not divide. Thus, they conclude that it is important to fine-tune the expression system in order to prolong lifetime and therefore production yield [86]. Thus, FCM could help monitoring physiological changes in response to fine-tuning. Trip et al. developed an elegant, FCM-FACS based system to detect and separate *B. subtilis* cells that secrete heterologous proteins in large amounts. This system is proposed to be used in screening for biotechnologically important secreted proteins from genomic libraries of uncharacterized bacteria. Such a system consists of a *phtrA-gfp* reporter system, whose expression is activated by the CssRS secretion stress response; thus, when the reporter *B. subtilis* strain over produces and secretes proteins, the CssRS system is activated and promotes the expression of *gfp* under control of *htrA promoter*, thus producing a green fluorescent *B. subtilis* cell that can be separated by FACS for further propagation and analysis [87].

FCM has been also employed to characterize the physiological response of bacteria to organic compounds, such as phenol and toluene. A study that aimed to investigate these points was carried out by Wiacek et al. [33]. It was observed that a culture of *Cupriavidus necator* exhibited a heterogeneous response to harmful phenol concentration treatment determined by measuring chromosomal DNA and PHB content (with DAPI and Nile red, resp.), and after sorting different responding subpopulations they performed functional proteomics analysis [33]. The same group also investigated later the mechanism by which some bacteria uptake toluene. By means of a fluorescent labeled toluene analogue dye (NBDT), they measured toluene uptake using FCM, and after cell sorting, proteomics analysis revealed the presence of toluene transporting porins in *P. putida mt-2* when grown on toluene but not when grown in glucose [88].

The latest works exploited the advantage that cell sorting offers, which includes further analysis of subpopulations, either for obtaining more accurate results than those obtained by analyzing batch cultures or for comparing and characterizing the physiological state of different subpopulations.

**2.4. Host-Pathogen Interaction.** FCM is an automated technique suitable for studying bacterial interactions with host cells, useful for evaluating exclusively the interaction process or even for characterizing such interaction.

Analysis of expression of GFP reporter fusions during analysis of host-pathogen interactions is one of the most common methods employed; following this approach neither the physiology nor the pathogenesis course is altered [89–93].

Stapleton et al. established and standardized the use of FCM to analyze the relative adherence of uropathogenic *E. coli* to clinical samples of exfoliated vaginal epithelial cells (VEC). The researchers employed three GFP-expressing *E. coli* strains isolated from women with urinary tract infections, expressing different types of fimbria and with different *papG* classes. They counted PI stained VEC's, measured adhered bacterial associated GFP fluorescence, and determined the percentage of positive VEC's and the number of adhered bacterial cells. Their results showed different adherence patterns among *E. coli* strains expressing different adhesins [94].

Hara-Kaonga and Pistole's study is another example of the application of FCM to detect bacteria-host cell interactions. They were interested in clarifying whether *ompC* and *ompD* from *Salmonella enterica* were involved in the recognition by macrophages and epithelial cells. Specifically, the researchers analyzed whether these cells differentially recognize single and double *ompC ompD* null mutants. To this end, bacterial cells were stained with FITC and were later incubated with macrophages or epithelial cells. Then, the fluorescent human cells were detected by FCM and the number of adhered bacterial cells was determined. Their results showed that *OmpD*, but not *OmpC* porin, is involved in the recognition of *S. enterica* serovar *typhimurium* by human macrophages and epithelial cells [95]. The same authors later reported a method for assessing the adherence of *S. enterica* to host cells by measuring FITC associated fluorescence of bacteria and lipophilic dye PKH-26 associated fluorescence of eukaryotic cells, therefore eliminating the inaccuracies due to bacterial cell aggregates that mimic forward scattering of eukaryotic cells [96].

Another application of FCM in the study of pathogen-host interaction is analyzing genes that are specifically expressed when this interaction takes place. An example was reported by Bent et al. who investigated the expression of *yspP* and *orf6* genes, codifying, respectively, an effector protein secreted by and a structural component of the T3SS (type 3 secretion system) apparatus. The *yspP* and *orf6* expressions were determined by GFP reporter fusions to those gene's promoters. The reported *Yersinia enterocolitica* cells were used to infect mice. Infected murine cells were subjected to flow cytometric analysis in order to determine GFP fluorescence associated to *yspP* and *orf6* genes expression; this approach combined with qRT-PCR demonstrated that *Ysa T3SS* is expressed in infected mice [97].

With the purpose of separating cells of interest for further analysis using FACS technology, reports have been published that analyze gene expression during host-pathogen interaction. One common approach was to make a fusion library of chromosome fragments and to clone the fragments in a vector carrying a promoterless *gfp* gene to capture promoters. The approach is the following: bacterial cells transformed with plasmids that express GFP when interacting with host cells are separated by FACS and are later recovered, and their promoter is then at least partially sequenced to identify genes expressed during host-pathogen interaction. With this experimental approach, Barker et al. investigated the genes of *Mycobacterium marinum* that are differentially expressed

during macrophage phagocytosis. They sorted fluorescent phagosomes, and after 2 to 3 days after infection, they separated vesicles containing single cells; following this approach, they identified 12 clones containing GFP fusions differentially expressed including membrane proteins and biosynthetic enzymes [98]. Following a similar approach, Wilson et al. analyzed the *in vivo* transcriptional response when *Listeria monocytogenes* infected a murine macrophage model [99].

Thus, flow cytometry offers a rapid, accurate, and reproducible method to quantify the number of wild type or mutant bacterial cells adhered or phagocyted by host cells, with technical complications that each bacterium and host cell type represents. On the other hand, the expression of selected genes during host-pathogen interaction can be assessed, or even genes participating in infection processes can be identified using FCM coupled to cell sorting for further analysis. Although fluorescence microscopy could be an alternative for these purposes, FCM offers the advantage of analyzing and counting a high number of cells, thereby improving the statistics. However, it must be pointed that FCM does not provide information regarding the distribution or location of bacteria within the host cell, and hence a microscopic analysis could be an excellent complement to this technique.

The pathogenic bacteria *Shigella flexneri* are capable of causing dysentery by invading the epithelial cells of the colon. To identify the factors that allow this bacterium to infect the epithelial layer, an FACS based approach was employed to sort mutant clones incapable of spreading within and between epithelial cells. At least three different classes of mutants were identified with this approach, namely, those that presented an altered lipopolysaccharide structure, clones that were affected in intracellular motility, or those that exhibited defect in its capacity to invade the cell hosts [100]. With the purpose of understanding the molecular basis of bacterial virulence, genetic approaches combined with FCM have been applied to identify and isolate nonvirulent bacteria. In a recent study, a liquid culture of *Vibrio cholerae* was subjected to random transposon mutagenesis with the purpose of generating mutants deficient in virulence activation. The entire population was subjected to FACS to successfully separate the cells that were deficient for virulence activation [101]. In another report, FACS was applied to separate recombinant *Mycobacterium* capable of expressing high levels of a foreign ovalbumin epitope; notably, the sorted clones were more efficient in inducing an immunogenic response. These studies demonstrated that recombinant *Mycobacterium* has the potential to be employed as vehicles to delivering pathogen antigen peptides and inducing systemic and mucosal immune responses [102].

### 2.5. Analysis of Bacterial Physiology at a Single-Cell Level.

FCM has also been applied to understand the physiological changes occurring in bacterial cells in response to environment changes. One of the most regulated processes in bacteria is cell division. In Gram-positive bacteria, the set of proteins required for cell division is known; these cell division proteins (CDPs) must act in a specific yet known order.

Trip et al. were interested in investigating if transcription of those CDPs is the underlying process that regulates speed of cell division. To investigate this aspect, transcriptional gene fusions between CPD promoters and the GFP encoding gene were recombined into a neutral locus of the chromosome of *B. subtilis*. The GFP-fusion containing strains were employed to measure by FCM and qRT-PCR if CDPs expression varies or not under different culture conditions that modify growth rate. Constant expression levels of the fusions were found when tested independently of the growth rate and cell cycle; therefore, it was concluded that cell division in *B. subtilis* is regulated mainly at a posttranslational level and is influenced by other factors that do fluctuate, such as metabolic state and substrate availability, but not by transcription [103].

DNA staining and subsequent FCM quantitative analysis has been applied to gain insights into mechanisms that underlie cell cycle regulation and chromosome biology in bacterial cells. Such knowledge impacts on the understanding of the life cycle of the bacterial cell and can be used to determine cell states quickly and thus predicts the metabolic and survival behavior of microorganisms [4, 104]. DNA quantification patterns are indicative of the number of chromosomes, and by FCM analysis information can be obtained about the number of individual cells that possess an  $n$  number of chromosomes in an asynchronous growing population. These patterns are characteristic of distinct bacterial species and change under different environment conditions (reviewed by [105]). When studying protein production, function, and localization in batch cultures, it is important to synchronize the cells. FCM analysis of DNA stained with chromomycin has been used to ensure the quality of synchronous population of *Caulobacter crescentus* for studying morphogenesis and cell cycle regulatory proteins that function along the cell cycle [106]. DNA staining histograms obtained by FCM, after treating the cells with drugs that inhibit replication initiation and cell division, are an approach employed to find the number of chromosomes per cell. Thus, variations in these distribution patterns may be analyzed as a function of different conditions and genetic backgrounds to gain insights into cell cycle regulation [107, 108].

There is also a wide interest in studying bacterial differentiation due to the impact of resistant spores and antibiotics resistance related to competence development over clinical, sanitary, and biotechnological issues. *B. subtilis* is a suitable model for studying the molecular mechanisms underlying natural competence development and sporulation [109]. Both developmental pathways are interconnected and appear to be mutually exclusive. Chung et al. (1994) published one of the pioneering works that applied FCM to investigate the heterogeneous development of spore differentiation pathways in *B. subtilis*. In this work, using *lacZ* reporter fusions with early spore development genes *spoVG*, *spoIIG*, and *spoIID*, they measured  $\beta$ -galactosidase activity at the single-cell level, revealed by the hydrolysis of the fluorogenic compound C8-FDG (5-octanoylamino fluorescein-di- $\beta$ -D-galactopyranoside) that releases C8-fluorescein which emits green light when excited at 488 nm. They found with this approach that, within a culture of sporulating *B. subtilis* cells, there are two distinct subpopulations and only one of

them initiates and concludes with the spore developmental program [110]. FCM and fluorescent reporter fusions to promoters have been employed to gain insights into the pathways that control sporulation and competence development and in the molecular mechanisms by which these processes occur in heterogeneous populations within an isogenic culture; this is important because only a small fraction of cells may experience those developmental processes. Smits et al., employing a (ComK activated)  $P_{\text{comK}}-gfp$  to analyze competence development at single-cell level by FCM, showed the importance of the transcriptional autostimulation of *comK* in the development of competence and also demonstrated that this state occurs in a population of cells that reach a threshold level of ComK [111]. Later, following a single-cell analysis approach, that analyzed the expression of a  $P_{\text{spoIIA}}-gfp$  fusion activated by Spo0A, allowed them to conclude that the autostimulatory activation of Spo0A is responsible for the bistable expression pattern in sporulating cultures [112]. Using the same reporter fusion, this research group performed single-cell analyses on the expression patterns of both competence development and sporulation of a *B. subtilis* culture of a 1:1 mix spore induction: competence induction in chemically defined liquid media and also in biofilms. They found that, in the 1:1 mix, both processes are sequentially initiated; first competence develops and sporulation activates later. In another experimental approach, it was found that a small fraction of spores were also capable of acquiring competence. They also found that sporulation is more effectively initiated in biofilms than in planktonic cells. Moreover, this group reported that, under conditions that do not usually trigger sporulation or competence, there are few cells that form spores or that become competent respectively; that is, both differentiation pathways are noisy. Of note, these analyses were only possible by employing FCM *noise measurement*, since at least 100 000 cells must be analyzed under rigorous gating to detect positive cells [109]. Later, this group identified RapH as a novel factor involved in the temporal separation of competence and sporulation. They observed, by using approaches such as single-cell analysis, that the overproduction of RapH provokes a drastic reduction of competence and sporulation gene expression and that *rapH* genetic disruption causes a significant increase in the frequency of cells simultaneously expressing both sporulation and competence fusion reporters; therefore, in a *rapH* mutant sporulation initiation and competence development are not strictly separated [113]. The mechanisms governing the production of exoproteases in nonsporulating cells of *B. subtilis*, within heterogeneous populations, were analyzed by FCM using strains harboring GFP reporter fusions. It was found that only a fraction of the vegetative cells turned on the expression of the bacillopeptidase (*bpr*) and subtilisin (*aprE*) encoding genes under control of the regulator DegU [114].

In a different bacterium, Stecchini et al. [8] investigated the effects of changes in humidity and viscosity in growth cultures of *B. cereus*, a human pathogen [77] over cell motility, spore dimension, and thermal resistance. Such parameters were measured by forward and side scattering (assess dimension) and propidium iodide (thermal damage) [8].

Cronin and Wilkinson, using differential staining, microscopy, and FCM analyses, reported the establishment of a new methodology to identify, quantify, and assess changes in permeability and metabolism of germinating *B. cereus* endospores. This methodology consisted in FCM single-cell level measurement of CFDA/Hoechst 33342 to estimate overall germination rate; and by measuring side-scatter and SYTO9 staining, they quantified ungerminated, germinating, and outgrowing endospores [115].

The list of discoveries enriched and supported by cytometric analysis related to physiological bacterial functions is still growing. Another interesting issue is *quorum sensing* (QS), a way by which bacteria communicate by secreting and responding to signal molecules or autoinducers. Such processes are relevant because in pathogenic bacteria certain virulence factors encoding genes are regulated by *quorum sensing* [116]. Using FCM analysis of reporter strains expressing promoter-*gfp* fusions, Anetzberger et al. monitored the induction/repression of autoinducer-regulated genes in the shrimp pathogen *Vibrio harveyi*. It was found that, besides luminescence, exoprotease gene expression is also regulated by cell density and evidenced by single-cell analysis of simultaneously assessed luminescence and exoprotease expression which they observed functional heterogeneity within the population [117].

In nature, most bacteria exist as aggregates surrounded by an extracellular matrix that is in a biofilm [118]. Biofilms are conformed by specialized subpopulations, with different physiological characteristics compared to those exhibited as planktonic cells [25, 26, 119]. In several contexts, including clinical, industrial, and environmental ones, there is a great interest in understanding how a biofilm is formed and how the cells composing it behave [119].

Given the nature of the biofilm, microscopy has been used as the choice method to analyze the structure, localization, or distribution of the biofilm components; however, FCM has also been successfully utilized for analyzing the physiological conditions of the cells in a biofilm [25]. A requirement for FCM analysis is that cells must be free, in a suspension; to accomplish this, the biofilm has to be disrupted, for instance by passing it by trough a pipette or a needle or by mild sonication [25, 120].

The importance of heterogeneous subpopulations in biofilm formation, using *B. subtilis* as model, has been investigated by FCM. Using this technique, it was discovered that the specialized matrix which is essential for biofilm formation is composed by TasA amyloid-like fibers, exopolysaccharide, and hydrophobin BslA and is produced by a specific subpopulation [27, 121–125]. Garcia-Betancur et al. reported the use of fluorescence microscopy combined with FCM to visualize and quantify the subpopulations of matrix producers and surfactin secretors (signaling molecule that triggers differentiation of matrix producers) within biofilms of *B. subtilis*, with the use of fluorescent reporter fusions with promoters of genes required for matrix ( $P_{\text{tapA}}$ -CFP) and surfactin ( $P_{\text{srfAA}}$ -YFP) production. They found that the reporter fusions are expressed only in certain subpopulations, under biofilm formation induction, and in the case of the double-labeled strain showed a single population

of fluorescent cells expressing both CFP and YFP, which indicates that in the same population both differentiation pathways are coordinately activated [25].

Using the same microorganism, Marlow et al. reported another type of specialized cells involved in biofilm formation. These authors applied FCM and microscopy to detect exoprotease-producing cells within the biofilm. To this end, fluorescent proteins fusions to  $P_{bpr}$  as reporter of expression of exoprotease-encoding gene  $bpr$  and other fluorescent reporter fusion  $P_{tapA}$ - $mKate2$  to monitor matrix producing cells were employed. It was found that the number of exoprotease producing cells increased as the biofilm matures and that exoprotease production is dependent on the levels of the response regulator phosphorylated DegU (that controls swarming motility, biofilm formation, and exoprotease production); moreover, it was found that exoprotease producing cells arise from both matrix-producing and nonproducing cells. Using microscopic analysis of a cross-section of a mature biofilm, they found that the subpopulation that produced exoprotease is enriched in the air interface more than that in the agar interface [126].

Although the above-mentioned works were performed *in vitro*, there are also *in vivo* studies; for example, Beauregard et al. demonstrated that *B. subtilis* colonizes *Arabidopsis thaliana* roots forming biofilms. This colonization has beneficial consequences for plant development. Although the usefulness of this bacterium as a biofertilizer is well known, Beauregard's results established the importance of biofilm formation in plant colonizing. They determined by FCM that plant polysaccharides, like arabinogalactan, pectin, and xylan, play signaling roles during biofilm formation and as a source of sugars for the synthesis of extracellular matrix. This was made by measuring the levels of fluorescence of biofilm forming cells expressing the  $P_{tapA}$ - $yfp$  fusion [127]. These results were consistent with *in vitro* pellicles formation assays and *in vivo* biofilm formation as was observed by fluorescence microscopy.

*B. licheniformis* is a useful bacterium for treating waste residues in water generated from alimentary industrial processes and is therefore important to optimize culture conditions for improved cell aggregation and biomass separation. da Silva et al., using FCM combined with biomass quantitation and confocal microscopy, established such optimal conditions, by assessing changes in the limiting nutrient, dilution rate, and agitation intensity looking for those that gave the cell aggregates where the majority of cells were metabolically active [128].

The employment of FCM in studying biofilms is very useful but not sufficient. Although this technique contributes to the statistics and accuracy of determinations, at single-cell level, it does not provide information regarding the structure in the biofilm and distribution of the distinct subpopulations.

### 3. Concluding Remarks

FCM is a tool that is currently applied to bacterial analysis from detecting and counting bacteria, to determining

changes in cellular functions and metabolic activity, and even in identifying genes that are expressed specifically under certain conditions. Although these analyses can be applied to a sample of cultured bacteria following the classic dilution and plating based methodology, FCM offers real measurements for each cell assessed, and in less time. With an appropriate combination of dyes, the damage caused by antibiotics and other threatening agents may be deduced and the proportion of affected cells determined; moreover, using fluorescent postenzymatic cleavage substrate analogs, metabolic activity can also be determined. FCM is an automated technique: it is time-saving, accurate, sensitive, and as thousands of cells can be processed per second, statistics are improved; this point is relevant and it is what makes FCM indispensable for certain purposes, such as identifying and quantifying cells present in very low abundance in a population. However, perhaps the most remarkable potential of FCM when coupled to cell sorting is that offers a unique opportunity to separate specific subpopulations of bacteria that exhibit differential physiological states, from entire cultures. Applying proteomics and/or transcriptomic analysis to characterize such subpopulations makes this approach a powerful tool to understand the cellular and molecular mechanisms that govern cell differentiation in bacteria. Finally, it must be pointed out that FCM and FACS require expensive equipment and skilled personal to operate it and interpret the results. With the appearance of low cost equipment in the market, it is expected that, in the near future, more studies can be carried out to study physiological responses in bacteria using these powerful approaches.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgments

This work was supported by the University of Guanajuato (Grant no. DAIP-324-2013) and by the Consejo Nacional de Ciencia y Tecnología (CONACYT) (Grants nos. 205744 and 221231) of México. Verónica Ambriz-Aviña was supported by scholarships from CONACYT and the University of Guanajuato. The authors wish to acknowledge the support of DAIP-UGTO with professional editing assistance.

### References

- [1] P. B. Rainey, H. J. E. Beaumont, G. C. Ferguson et al., "The evolutionary emergence of stochastic phenotype switching in bacteria," *Microbial Cell Factories*, vol. 10, article S14, no. 1, 2011.
- [2] D. Fraser and M. Kærn, "A chance at survival: gene expression noise and phenotypic diversification strategies," *Molecular Microbiology*, vol. 71, no. 6, pp. 1333–1340, 2009.
- [3] I. G. de Jong, J. Veening, and O. P. Kuipers, "Single cell analysis of gene expression patterns during carbon starvation in *Bacillus subtilis* reveals large phenotypic variation," *Environmental Microbiology*, vol. 14, no. 12, pp. 3110–3121, 2012.

- [4] S. Müller and H. Davey, "Recent advances in the analysis of individual microbial cells," *Cytometry A*, vol. 75, no. 2, pp. 83–85, 2009.
- [5] M. E. Lidstrom and M. C. Konopka, "The role of physiological heterogeneity in microbial population behavior," *Nature Chemical Biology*, vol. 6, no. 10, pp. 705–712, 2010.
- [6] D. Garmyn, L. Gal, R. Briandet et al., "Evidence of autoinduction heterogeneity via expression of the *Agr* system of *Listeria monocytogenes* at the single-cell level," *Applied and Environmental Microbiology*, vol. 77, no. 17, pp. 6286–6289, 2011.
- [7] B. Snijder and L. Pelkmans, "Origins of regulated cell-to-cell variability," *Nature Reviews Molecular Cell Biology*, vol. 12, no. 2, pp. 119–125, 2011.
- [8] M. L. Stecchini, M. Spaziani, M. D. Torre, and S. Pacor, "*Bacillus cereus* cell and spore properties as influenced by the microstructure of the medium," *Journal of Applied Microbiology*, vol. 106, no. 6, pp. 1838–1848, 2009.
- [9] C.-C. Shu, A. Chatterjee, G. Dunny, W.-S. Hu, and D. Ramkrishna, "Bistability versus bimodal distributions in gene regulatory processes from population balance," *PLoS Computational Biology*, vol. 7, no. 8, Article ID e1002140, 2011.
- [10] H. M. Davey and D. B. Kell, "Flow cytometry and cell sorting of heterogeneous microbial populations: the importance of single-cell analyses," *Microbiological Reviews*, vol. 60, no. 4, pp. 641–696, 1996.
- [11] R. C. Leif, *Practical Flow Cytometry*, by M. Howard and M.D. Shapiro, Wiley-Liss, Inc, New York, NY, USA, 3rd edition, 1995.
- [12] H. M. Shapiro, "Microbial analysis at the single-cell level: tasks and techniques," *Journal of Microbiological Methods*, vol. 42, no. 1, pp. 3–16, 2000.
- [13] D. A. Veal, D. Deere, B. Ferrari, J. Piper, and P. V. Attfeld, "Fluorescence staining and flow cytometry for monitoring microbial cells," *Journal of Immunological Methods*, vol. 243, no. 1–2, pp. 191–210, 2000.
- [14] P. Fouchet, S. Jan, J. Courtois, B. Courtois, G. Frelat, and J. N. Barbotin, "Quantitative single-cell detection of poly( $\beta$ -hydroxybutyrate) accumulation in *Rhizobium meliloti* by flow cytometry," *FEMS Microbiology Letters*, vol. 126, no. 1, pp. 31–35, 1995.
- [15] F. Sreenc, B. Arnold, and J. E. Bailey, "Characterization of intracellular accumulation of poly- $\beta$ -hydroxybutyrate (PHB) in individual cells of *Alcaligenes eutrophus* H16 by flow cytometry," *Biotechnology and Bioengineering*, vol. 26, no. 8, pp. 982–987, 1984.
- [16] M. Walberg, P. Gaustad, and H. B. Steen, "Rapid assessment of ceftazidime, ciprofloxacin, and gentamicin susceptibility in exponentially-growing *E. coli* cells by means of flow cytometry," *Cytometry*, vol. 27, no. 2, pp. 169–178, 1997.
- [17] M. Walberg, P. Gaustad, and H. B. Steen, "Rapid discrimination of bacterial species with different ampicillin susceptibility levels by means of flow cytometry," *Cytometry*, vol. 29, no. 3, pp. 267–272, 1997.
- [18] K. D. Wittrup, M. B. Mann, D. M. Fenton, L. B. Tsai, and J. E. Bailey, "Single-cell light scatter as a probe of refractile body formation in recombinant *Escherichia coli*," *Nature Biotechnology*, vol. 6, no. 4, pp. 423–426, 1988.
- [19] H. B. Steen, "Light scattering measurement in an arc lamp-based flow cytometer," *Cytometry*, vol. 11, no. 2, pp. 223–230, 1990.
- [20] R. P. Haugland, *Handbook of Fluorescent Probes and Research Products*, vol. 1, Molecular Probes, Eugene, Ore, USA, 9th edition, 2002.
- [21] M. Chalfie, Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher, "Green fluorescent protein as a marker for gene expression," *Science*, vol. 263, no. 5148, pp. 802–805, 1994.
- [22] A. B. Cubitt, R. Heim, S. R. Adams, A. E. Boyd, L. A. Gross, and R. Y. Tsien, "Understanding, improving and using green fluorescent proteins," *Trends in Biochemical Sciences*, vol. 20, no. 11, pp. 448–455, 1995.
- [23] M. Zimmer, "Green fluorescent protein (GFP): applications, structure, and related photophysical behavior," *Chemical Reviews*, vol. 102, no. 3, pp. 759–781, 2002.
- [24] M. Walberg, P. Gaustad, and H. B. Steen, "Rapid flow cytometric assessment of mecillinam and ampicillin bacterial susceptibility," *Journal of Antimicrobial Chemotherapy*, vol. 37, no. 6, pp. 1063–1075, 1996.
- [25] J. C. Garcia-Betancur, A. Yepes, J. Schneider, and D. Lopez, "Single-cell analysis of *Bacillus subtilis* biofilms using fluorescence microscopy and flow cytometry," *Journal of Visualized Experiments*, no. 60, Article ID e3796, 2012.
- [26] H. Vlamakis, C. Aguilar, R. Losick, and R. Kolter, "Control of cell fate by the formation of an architecturally complex bacterial community," *Genes & Development*, vol. 22, no. 7, pp. 945–953, 2008.
- [27] S. S. Branda, F. Chu, D. B. Kearns, R. Losick, and R. Kolter, "A major protein component of the *Bacillus subtilis* biofilm matrix," *Molecular Microbiology*, vol. 59, no. 4, pp. 1229–1238, 2006.
- [28] L. A. Herzenberg and R. G. Sweet, "Fluorescence-activated cell sorting," *Scientific American*, vol. 234, no. 3, pp. 108–117, 1976.
- [29] M. Jahn, J. Seifert, M. von Bergen, A. Schmid, B. Bühler, and S. Müller, "Subpopulation-proteomics in prokaryotic populations," *Current Opinion in Biotechnology*, vol. 24, no. 1, pp. 79–87, 2013.
- [30] G. Wallner, B. Fuchs, S. Spring, W. Beisker, and R. Amann, "Flow sorting of microorganisms for molecular analysis," *Applied and Environmental Microbiology*, vol. 63, no. 11, pp. 4223–4231, 1997.
- [31] T. U. Sørensen, G. J. Gram, S. D. Nielsen, and J. S. Hansen, "Safe sorting of GFP-transduced live cells for subsequent culture using a modified FACS vantage," *Cytometry*, vol. 37, no. 4, pp. 284–290, 1999.
- [32] N. Jehmlich, T. Hübschmann, M. Gesell Salazar et al., "Advanced tool for characterization of microbial cultures by combining cytomics and proteomics," *Applied Microbiology and Biotechnology*, vol. 88, no. 2, pp. 575–584, 2010.
- [33] C. Wiacek, S. Müller, and D. Benndorf, "A cytomic approach reveals population heterogeneity of *Cupriavidus necator* in response to harmful phenol concentrations," *Proteomics*, vol. 6, no. 22, pp. 5983–5994, 2006.
- [34] V. B. L. Quixabeira, J. C. Nabout, and F. M. Rodrigues, "Trends in genetic literature with the use of flow cytometry," *Cytometry A*, vol. 77, no. 3, pp. 207–210, 2010.
- [35] G. Nebe-von-Caron, "Standardization in microbial cytometry," *Cytometry Part A*, vol. 75, no. 2, pp. 86–89, 2009.
- [36] J. A. Heinemann, R. G. Ankenbauer, and C. F. Amabile-Cuevas, "Do antibiotics maintain antibiotic resistance?" *Drug Discovery Today*, vol. 5, no. 5, pp. 195–204, 2000.
- [37] B. Spellberg, R. Guidos, D. Gilbert et al., "The epidemic of antibiotic-resistant infections: a call to action for the medical community from the infectious diseases society of America," *Clinical Infectious Diseases*, vol. 46, no. 2, pp. 155–164, 2008.
- [38] P. S. Negi, G. K. Jayaprakasha, and B. S. Jena, "Antibacterial activity of the extracts from the fruit rinds of *Garcinia cowa*

- and *Garcinia pedunculata* against food borne pathogens and spoilage bacteria," *LWT: Food Science and Technology*, vol. 41, no. 10, pp. 1857–1861, 2008.
- [39] C. Gaspar-Marques, P. Rijo, M. F. Simões, M. A. Duarte, and B. Rodriguez, "Abietanes from *Plectranthus grandidentatus* and *P. hereroensis* against methicillin- and vancomycin-resistant bacteria," *Phytomedicine*, vol. 13, no. 4, pp. 267–271, 2006.
- [40] J. Barenfanger, C. Drake, and G. Kacich, "Clinical and financial benefits of rapid bacterial identification and antimicrobial susceptibility testing," *Journal of Clinical Microbiology*, vol. 37, no. 5, pp. 1415–1418, 1999.
- [41] H. M. Shapiro, "Multiparameter flow cytometry of bacteria: implications for diagnostics and therapeutics," *Cytometry*, vol. 43, no. 3, pp. 223–226, 2001.
- [42] H. B. Steen, M. W. Jernaes, K. Skarstad, and E. Boye, "Chapter 28 staining and measurement of DNA in bacteria," *Methods in Cell Biology*, vol. 42, pp. 477–487, 1994.
- [43] S. Dessus-Babus, F. Belloc, C. M. Bébér et al., "Antibiotic susceptibility testing for *Chlamydia trachomatis* using flow cytometry," *Cytometry*, vol. 31, no. 1, pp. 37–44, 1998.
- [44] P. Assunção, N. T. Antunes, R. S. Rosales et al., "Application of flow cytometry for the determination of minimal inhibitory concentration of several antibacterial agents on *Mycoplasma hyopneumoniae*," *Journal of Applied Microbiology*, vol. 102, no. 4, pp. 1132–1137, 2007.
- [45] T. Soejima, J. Minami, and K. Iwatsuki, "The exclusive use of flow cytometry to evaluate the antibiotic-susceptibility," *Biochimica et Biophysica Acta—General Subjects*, vol. 1820, no. 12, pp. 1980–1986, 2012.
- [46] D. J. Mason, E. G. M. Power, H. Talsania, I. Phillips, and V. A. Gant, "Antibacterial action of ciprofloxacin," *Antimicrobial Agents and Chemotherapy*, vol. 39, no. 12, pp. 2752–2758, 1995.
- [47] B. L. Roth, M. Poot, S. T. Yue, and P. J. Millard, "Bacterial viability and antibiotic susceptibility testing with SYTOX green nucleic acid stain," *Applied and Environmental Microbiology*, vol. 63, no. 6, pp. 2421–2431, 1997.
- [48] M. T. E. Suller, J. M. Stark, and D. Lloyd, "A flow cytometric study of antibiotic-induced damage and evaluation as a rapid antibiotic susceptibility test for methicillin-resistant *Staphylococcus aureus*," *Journal of Antimicrobial Chemotherapy*, vol. 40, no. 1, pp. 77–83, 1997.
- [49] F. C. Mortimer, D. J. Mason, and V. A. Gant, "Flow cytometric monitoring of antibiotic-induced injury in *Escherichia coli* using cell-impermeant fluorescent probes," *Antimicrobial Agents and Chemotherapy*, vol. 44, no. 3, pp. 676–681, 2000.
- [50] D. J. Novo, N. G. Perlmutter, R. H. Hunt, and H. M. Shapiro, "Multiparameter flow cytometric analysis of antibiotic effects on membrane potential, membrane permeability, and bacterial counts of *Staphylococcus aureus* and *Micrococcus luteus*," *Antimicrobial Agents and Chemotherapy*, vol. 44, no. 4, pp. 827–834, 2000.
- [51] H. J. Wickens, R. J. Pinney, D. J. Mason, and V. A. Gant, "Flow cytometric investigation of filamentation, membrane patency, and membrane potential in *Escherichia coli* following ciprofloxacin exposure," *Antimicrobial Agents and Chemotherapy*, vol. 44, no. 3, pp. 682–687, 2000.
- [52] M. T. Suller and D. Lloyd, "Fluorescence monitoring of antibiotic-induced bacterial damage using flow cytometry," *Cytometry*, vol. 35, no. 3, pp. 235–241, 1999.
- [53] C. Laflamme, S. Lavigne, J. Ho, and C. Duchaine, "Assessment of bacterial endospore viability with fluorescent dyes," *Journal of Applied Microbiology*, vol. 96, no. 4, pp. 684–692, 2004.
- [54] J. Lehtinen, J. Nuutila, and E. Lilius, "Green fluorescent protein-propidium iodide (GFP-PI) based assay for flow cytometric measurement of bacterial viability," *Cytometry A*, vol. 60, no. 2, pp. 165–172, 2004.
- [55] K. Papadimitriou, H. Pratsinis, G. Nebe-von-Caron, D. Kletsas, and E. Tsakalidou, "Rapid assessment of the physiological status of *Streptococcus macedonicus* by flow cytometry and fluorescence probes," *International Journal of Food Microbiology*, vol. 111, no. 3, pp. 197–205, 2006.
- [56] I. Faria-Ramos, M. J. Espinar, R. Rocha et al., "A novel flow cytometric assay for rapid detection of extended-spectrum beta-lactamases," *Clinical Microbiology and Infection*, vol. 19, no. 1, pp. E8–E15, 2013.
- [57] S. Ghosh, K. Indukuri, S. Bondalapati, A. K. Saikia, and L. Rangan, "Unveiling the mode of action of antibacterial labdane diterpenes from *Alpinia nigra* (Gaertn.) B. L. Burtt seeds," *European Journal of Medicinal Chemistry*, vol. 66, pp. 101–105, 2013.
- [58] G. Nebe-Von-Caron, P. J. Stephens, C. J. Hewitt, J. R. Powell, and R. A. Badley, "Analysis of bacterial function by multi-colour fluorescence flow cytometry and single cell sorting," *Journal of Microbiological Methods*, vol. 42, no. 1, pp. 97–114, 2000.
- [59] J. D. Oliver, "The viable but nonculturable state in bacteria," *Journal of Microbiology*, vol. 43, pp. 93–100, 2005.
- [60] B. H. Pyle, S. C. Broadaway, and G. A. McFeters, "Sensitive detection of *Escherichia coli* O157:H7 in food and water by immunomagnetic separation and solid-phase laser cytometry," *Applied and Environmental Microbiology*, vol. 65, no. 5, pp. 1966–1972, 1999.
- [61] M. M. Taimur Khan, B. H. Pyle, and A. K. Camper, "Specific and rapid enumeration of viable but nonculturable and viable-culturable gram-negative bacteria by using flow cytometry," *Applied and Environmental Microbiology*, vol. 76, no. 15, pp. 5088–5096, 2010.
- [62] R. Santiso, M. Tamayo, J. Gosálvez, G. Bou, M. C. Fernández, and J. L. Fernández, "A rapid in situ procedure for determination of bacterial susceptibility or resistance to antibiotics that inhibit peptidoglycan biosynthesis," *BMC Microbiology*, vol. 11, article 191, 2011.
- [63] M. A. Sánchez-Romero and J. Casadesús, "Contribution of phenotypic heterogeneity to adaptive antibiotic resistance," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 1, pp. 355–360, 2014.
- [64] L. Cui, J. Lian, H. Neoh, E. Reyes, and K. Hiramatsu, "DNA microarray-based identification of genes associated with glycopeptide resistance in *Staphylococcus aureus*," *Antimicrobial Agents and Chemotherapy*, vol. 49, no. 8, pp. 3404–3413, 2005.
- [65] H. Xie, J. Mire, Y. Kong et al., "Rapid point-of-care detection of the tuberculosis pathogen using a *BlaC*-specific fluorogenic probe," *Nature Chemistry*, vol. 4, no. 10, pp. 802–809, 2012.
- [66] J. Zhang, Y. Shen, S. L. May, D. C. Nelson, and S. Li, "Ratiometric fluorescence detection of pathogenic bacteria resistant to broad-spectrum  $\beta$ -lactam antibiotics," *Angewandte Chemie International Edition*, vol. 51, no. 8, pp. 1865–1868, 2012.
- [67] S. Mizukami, S. Watanabe, Y. Akimoto, and K. Kikuchi, "No-wash protein labeling with designed fluorogenic probes and application to real-time pulse-chase analysis," *Journal of the American Chemical Society*, vol. 134, no. 3, pp. 1623–1629, 2012.
- [68] Q. Shao and B. Xing, "Enzyme responsive luminescent ruthenium(ii) cephalosporin probe for intracellular imaging and photoinactivation of antibiotics resistant bacteria," *Chemical Communications*, vol. 48, no. 12, pp. 1739–1741, 2012.

- [69] Q. Shao, Y. Zheng, X. Dong, K. Tang, X. Yan, and B. Xing, "A covalent reporter of  $\beta$ -lactamase activity for fluorescent imaging and rapid screening of antibiotic-resistant bacteria," *Chemistry*, vol. 19, no. 33, pp. 10903–10910, 2013.
- [70] X.-J. Fu, Y. Fang, and M. Yao, "Antimicrobial photodynamic therapy for methicillin-resistant *Staphylococcus aureus* infection," *BioMed Research International*, vol. 2013, Article ID 159157, 9 pages, 2013.
- [71] H. M. Tang, M. R. Hamblin, and C. M. N. Yow, "A comparative in vitro photoinactivation study of clinical isolates of multidrug-resistant pathogens," *Journal of Infection and Chemotherapy*, vol. 13, no. 2, pp. 87–91, 2007.
- [72] P. S. Zolfaghari, S. Packer, M. Singer et al., "In vivo killing of *Staphylococcus aureus* using a light-activated antimicrobial agent," *BMC Microbiology*, vol. 9, article 27, 2009.
- [73] Y. Jiang, A. W. Leung, X. Wang, H. Zhang, and C. Xu, "Inactivation of *Staphylococcus aureus* by photodynamic action of hypocrellin B," *Photodiagnosis and Photodynamic Therapy*, vol. 10, no. 4, pp. 600–606, 2013.
- [74] A. Aranda, L. Sequedo, L. Tolosa et al., "Dichloro-dihydro-fluorescein diacetate (DCFH-DA) assay: a quantitative method for oxidative stress assessment of nanoparticle-treated cells," *Toxicology in Vitro*, vol. 27, no. 2, pp. 954–963, 2013.
- [75] M. Gandhi and M. L. Chikindas, "Listeria: a foodborne pathogen that knows how to survive," *International Journal of Food Microbiology*, vol. 113, no. 1, pp. 1–15, 2007.
- [76] M. Utratna, E. Cosgrave, C. Baustian, R. H. Ceredig, and C. P. O'Byrne, "Effects of growth phase and temperature on  $\sigma^B$  activity within a *Listeria monocytogenes* population: evidence for RsbV-independent activation of  $\sigma^B$  at refrigeration temperatures," *BioMed Research International*, vol. 2014, Article ID 641647, 11 pages, 2014.
- [77] M. Mols and T. Abee, "*Bacillus cereus* responses to acid stress," *Environmental Microbiology*, vol. 13, no. 11, pp. 2835–2843, 2011.
- [78] U. P. Cronin and M. G. Wilkinson, "Physiological response of *Bacillus cereus* vegetative cells to simulated food processing treatments," *Journal of Food Protection*, vol. 71, no. 11, pp. 2168–2176, 2008.
- [79] M. Mols, R. van Kranenburg, C. C. J. van Melis, R. Moezelaar, and T. Abee, "Analysis of acid-stressed *Bacillus cereus* reveals a major oxidative response and inactivation-associated radical formation," *Environmental Microbiology*, vol. 12, no. 4, pp. 873–885, 2010.
- [80] U. P. Cronin and M. G. Wilkinson, "Monitoring changes in germination and permeability of *Bacillus cereus* endospores following chemical, heat and enzymatic treatments using flow cytometry," *Journal of Rapid Methods and Automation in Microbiology*, vol. 16, no. 2, pp. 164–184, 2008.
- [81] U. P. Cronin and M. G. Wilkinson, "*Bacillus cereus* endospores exhibit a heterogeneous response to heat treatment and low-temperature storage," *Food Microbiology*, vol. 25, no. 2, pp. 235–243, 2008.
- [82] A. Neumeyer, T. Hübschmann, S. Müller, and J. Frunzke, "Monitoring of population dynamics of *Corynebacterium glutamicum* by multiparameter flow cytometry," *Microbial Biotechnology*, vol. 6, no. 2, pp. 157–167, 2013.
- [83] E. O. Sunny-Roberts and D. Knorr, "Evaluation of the response of *Lactobacillus rhamnosus* VTT E-97800 to sucrose-induced osmotic stress," *Food Microbiology*, vol. 25, no. 1, pp. 183–189, 2008.
- [84] A. Reis, T. L. da Silva, C. A. Kent, M. Kosseva, J. C. Roseiro, and C. J. Hewitt, "Monitoring population dynamics of the thermophilic *Bacillus licheniformis* CCMI 1034 in batch and continuous cultures using multi-parameter flow cytometry," *Journal of Biotechnology*, vol. 115, no. 2, pp. 199–210, 2005.
- [85] T. da Silva, A. Reis, C. Kent, M. Kosseva, J. C. Roseiro, and C. Hewitt, "Stress-induced physiological responses to starvation periods as well as glucose and lactose pulses in *Bacillus licheniformis* CCMI 1034 continuous aerobic fermentation processes as measured by multi-parameter flow cytometry," *Biochemical Engineering Journal*, vol. 24, no. 1, pp. 11–15, 2005.
- [86] N. Borth, R. Mitterbauer, D. Mattanovich, W. Kramer, K. Bayer, and H. Katinger, "Flow cytometric analysis of bacterial physiology during induction of foreign protein synthesis in recombinant *Escherichia coli* cells," *Cytometry*, vol. 31, no. 2, pp. 125–129, 1998.
- [87] H. Trip, P. J. van der Veek, T. C. Renniers et al., "A novel screening system for secretion of heterologous proteins in *Bacillus subtilis*," *Microbial Biotechnology*, vol. 4, no. 5, pp. 673–682, 2011.
- [88] H. Sträuber, T. Hübschmann, N. Jehmlich et al., "NBDT (3-(N-(7-nitrobenz-2-oxa-1,3-diazol-4-yl)amino)3-toluene)—a novel fluorescent dye for studying mechanisms of toluene uptake into vital bacteria," *Cytometry A*, vol. 77, no. 2, pp. 113–120, 2010.
- [89] R. H. Valdivia, A. E. Hromockyj, D. Monack, L. Ramakrishnan, and S. Falkow, "Applications for green fluorescent protein (GFP) in the study of host-pathogen interactions," *Gene*, vol. 173, no. 1, pp. 47–52, 1996.
- [90] S. Dhandayuthapani, L. E. Via, C. A. Thomas, P. M. Horowitz, D. Deretic, and V. Deretic, "Green fluorescent protein as a marker for gene expression and cell biology of mycobacterial interactions with macrophages," *Molecular Microbiology*, vol. 17, no. 5, pp. 901–912, 1995.
- [91] L. Kremer, A. Baulard, J. Estaquier, O. Poulain-Godefroy, and C. Locht, "Green fluorescent protein as a new expression marker in mycobacteria," *Molecular Microbiology*, vol. 17, no. 5, pp. 913–922, 1995.
- [92] R. H. Valdivia and S. Falkow, "Bacterial genetics by flow cytometry: rapid isolation of *Salmonella typhimurium* acid-inducible promoters by differential fluorescence induction," *Molecular Microbiology*, vol. 22, no. 2, pp. 367–378, 1996.
- [93] S. Aymanns, S. Mauerer, G. van Zandbergen, C. Wolz, and B. Spellerberg, "High-level fluorescence labeling of Gram-positive pathogens," *PLoS ONE*, vol. 6, no. 6, Article ID e19822, 2011.
- [94] A. E. Stapleton, C. L. Fennell, D. M. Coder, C. L. Wobbe, P. L. Roberts, and W. E. Stamm, "Precise and rapid assessment of *Escherichia coli* adherence to vaginal epithelial cells by flow cytometry," *Clinical Cytometry*, vol. 50, no. 1, pp. 31–37, 2002.
- [95] B. Hara-Kaonga and T. G. Pistole, "OmpD but not OmpC is involved in adherence of *Salmonella enterica* serovar Typhimurium to human cells," *Canadian Journal of Microbiology*, vol. 50, no. 9, pp. 719–727, 2004.
- [96] B. Hara-Kaonga and T. G. Pistole, "A dual fluorescence flow cytometric analysis of bacterial adherence to mammalian host cells," *Journal of Microbiological Methods*, vol. 69, no. 1, pp. 37–43, 2007.
- [97] Z. W. Bent, S. S. Branda, and G. M. Young, "The *Yersinia enterocolitica* Ysa type III secretion system is expressed during infections both in vitro and in vivo," *Microbiologyopen*, vol. 2, no. 6, pp. 962–975, 2013.
- [98] L. P. Barker, D. M. Brooks, and P. L. C. Small, "The identification of *Mycobacterium marinum* genes differentially expressed in macrophage phagosomes using promoter fusions to green

- fluorescent protein," *Molecular Microbiology*, vol. 29, no. 5, pp. 1167–1177, 1998.
- [99] R. L. Wilson, A. R. Tvinnereim, B. D. Jones, and J. T. Harty, "Identification of *Listeria monocytogenes* in vivo-induced genes by fluorescence-activated cell sorting," *Infection and Immunity*, vol. 69, no. 8, pp. 5016–5024, 2001.
- [100] M. Rathman, N. Jourhi, A. Allaoui, P. Sansonetti, C. Parsot, and G. T. Van Nhieu, "The development of a FACS-based strategy for the isolation of *Shigella flexneri* mutants that are deficient in intercellular spread," *Molecular Microbiology*, vol. 35, no. 5, pp. 974–990, 2000.
- [101] B. H. Abuaita and J. H. Withey, "Genetic screening for bacterial mutants in liquid growth media by fluorescence-activated cell sorting," *Journal of Microbiological Methods*, vol. 84, no. 1, pp. 109–113, 2011.
- [102] J. S. Yu, J. Whitesides, S. H. Lee et al., "Flow cytometry sorting of recombinant mycobacterial species yields bacterial clones with enhanced insert expression," *Clinical and Vaccine Immunology*, vol. 18, no. 1, pp. 43–49, 2011.
- [103] E. N. Trip, J.-W. Veening, E. J. Stewart, J. Errington, and D.-J. Scheffers, "Balanced transcription of cell division genes in *Bacillus subtilis* as revealed by single cell analysis," *Environmental Microbiology*, vol. 15, no. 12, pp. 3196–3209, 2013.
- [104] S. Müller and G. Nebe-Von-Caron, "Functional single-cell analyses: Flow cytometry and cell sorting of microbial populations and communities," *FEMS Microbiology Reviews*, vol. 34, no. 4, pp. 554–587, 2010.
- [105] S. Müller, "Modes of cytometric bacterial DNA pattern: a tool for pursuing growth," *Cell Proliferation*, vol. 40, no. 5, pp. 621–639, 2007.
- [106] C. Jacobs, D. Hung, and L. Shapiro, "Dynamic localization of a cytoplasmic signal transduction response regulator controls morphogenesis during the *Caulobacter* cell cycle," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 7, pp. 4095–4100, 2001.
- [107] I. O. Odsbu and K. Skarstad, "Growth rate dependent numbers of SeqA structures organize the multiple replication forks in rapidly growing *Escherichia coli*," *Genes to Cells*, vol. 14, no. 5, pp. 643–657, 2009.
- [108] G. Demarre and D. K. Chattoraj, "DNA adenine methylation is required to replicate both *Vibrio cholerae* chromosomes once per cell cycle," *PLoS genetics*, vol. 6, Article ID e1000939, 2010.
- [109] J. W. Veening, W. K. Smits, L. W. Hamoen, and O. P. Kuipers, "Single cell analysis of gene expression patterns of competence development and initiation of sporulation in *Bacillus subtilis* grown on chemically defined media," *Journal of Applied Microbiology*, vol. 101, no. 3, pp. 531–541, 2006.
- [110] J. D. Chung, G. Stephanopoulos, K. Ireton, and A. D. Grossman, "Gene expression in single cells of *Bacillus subtilis*: evidence that a threshold mechanism controls the initiation of sporulation," *Journal of Bacteriology*, vol. 176, no. 7, pp. 1977–1984, 1994.
- [111] W. K. Smits, C. C. Eschevins, K. A. Susanna, S. Bron, O. P. Kuipers, and L. W. Hamoen, "Stripping *Bacillus*: comK auto-stimulation is responsible for the bistable response in competence development," *Molecular Microbiology*, vol. 56, no. 3, pp. 604–614, 2005.
- [112] J. Veening, L. W. Hamoen, and O. P. Kuipers, "Phosphatases modulate the bistable sporulation gene expression pattern in *Bacillus subtilis*," *Molecular Microbiology*, vol. 56, no. 6, pp. 1481–1494, 2005.
- [113] W. K. Smits, C. Bongiorno, J. Veening, L. W. Hamoen, O. P. Kuipers, and M. Perego, "Temporal separation of distinct differentiation pathways by a dual specificity Rap-Phr system in *Bacillus subtilis*," *Molecular Microbiology*, vol. 65, no. 1, pp. 103–120, 2007.
- [114] J. W. Veening, O. A. Igoshin, R. T. Eijlander, R. Nijland, L. W. Hamoen, and O. P. Kuipers, "Transient heterogeneity in extracellular protease production by *Bacillus subtilis*," *Molecular Systems Biology*, vol. 4, article 184, 2008.
- [115] U. P. Cronin and M. G. Wilkinson, "The use of flow cytometry to study the germination of *Bacillus cereus* endospores," *Cytometry A*, vol. 71, no. 3, pp. 143–153, 2007.
- [116] J. M. Henke and B. L. Bassler, "Quorum sensing regulates type III secretion in *Vibrio harveyi* and *Vibrio parahaemolyticus*," *Journal of Bacteriology*, vol. 186, no. 12, pp. 3794–3805, 2004.
- [117] C. Anetzberger, U. Schell, and K. Jung, "Single cell analysis of *Vibrio harveyi* uncovers functional heterogeneity in response to quorum sensing signals," *BMC Microbiology*, vol. 12, article 209, 2012.
- [118] G. O'Toole, H. B. Kaplan, and R. Kolter, "Biofilm formation as microbial development," *Annual Review of Microbiology*, vol. 54, pp. 49–79, 2000.
- [119] S. J. Pamp, C. Sternberg, and T. Tolker-Nielsen, "Insight into the microbial multicellular lifestyle via flow-cell technology and confocal microscopy," *Cytometry A*, vol. 75, no. 2, pp. 90–103, 2009.
- [120] G. Nebe-Von Caron and R. A. Badley, "Viability assessment of bacteria in mixed populations using flow cytometry," *Journal of Microscopy*, vol. 179, no. 1, pp. 55–66, 1995.
- [121] S. S. Branda, Å. Vik, L. Friedman, and R. Kolter, "Biofilms: the matrix revisited," *Trends in Microbiology*, vol. 13, no. 1, pp. 20–26, 2005.
- [122] S. S. Branda, J. E. González-Pastor, S. Ben-Yehuda, R. Losick, and R. Kolter, "Fruiting body formation by *Bacillus subtilis*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 20, pp. 11621–11626, 2001.
- [123] U. Bai, I. Mandic-Mulec, and I. Smith, "SinI modulates the activity of SinR, a developmental switch protein of *Bacillus subtilis*, by protein-protein interaction," *Genes and Development*, vol. 7, no. 1, pp. 139–148, 1993.
- [124] D. B. Kearns, F. Chu, S. S. Branda, R. Kolter, and R. Losick, "A master regulator for biofilm formation by *Bacillus subtilis*," *Molecular Microbiology*, vol. 55, no. 3, pp. 739–749, 2005.
- [125] Y. Chai, F. Chu, R. Kolter, and R. Losick, "Bistability and Biofilm formation in *Bacillus subtilis*," *Molecular Microbiology*, vol. 67, no. 2, pp. 254–263, 2008.
- [126] V. L. Marlow, F. R. Cianfanelli, M. Porter, L. S. Cairns, and J. K. Dale, "The prevalence and origin of exoprotease-producing cells in the *Bacillus subtilis* biofilm," *Microbiology*, vol. 160, pp. 56–66, 2014.
- [127] P. B. Beauregard, Y. Chai, H. Vlamakis, R. Losick, and R. Kolter, "*Bacillus subtilis* biofilm induction by plant polysaccharides," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 17, pp. E1621–E1630, 2013.
- [128] T. L. da Silva, A. Reis, C. A. Kent, J. C. Roseiro, and C. J. Hewitt, "The use of multi-parameter flow cytometry to study the impact of limiting substrate, agitation intensity, and dilution rate on cell aggregation during *Bacillus licheniformis* CCM1 1034 aerobic continuous culture fermentations," *Biotechnology and Bioengineering*, vol. 92, no. 5, pp. 568–578, 2005.

## Research Article

# Female Aging Alters Expression of Human Cumulus Cells Genes that Are Essential for Oocyte Quality

**Tamadir Al-Edani,<sup>1,2</sup> Said Assou,<sup>1,2</sup> Alice Ferrières,<sup>1,3</sup> Sophie Bringer Deutsch,<sup>1,3</sup>  
Anna Gala,<sup>1,3</sup> Charles-Henri Lecellier,<sup>4</sup> Ounissa Aït-Ahmed,<sup>1,2</sup> and Samir Hamamah<sup>1,2,3</sup>**

<sup>1</sup> UFR de Médecine, Université Montpellier 1, 34295 Montpellier, France

<sup>2</sup> CHU Montpellier, Institut pour la Médecine Régénérative et Biothérapies, Hôpital Saint-Eloi, INSERM U1040, 34295 Montpellier, France

<sup>3</sup> ART-PGD Department, CHU Montpellier, Hôpital Arnaud de Villeneuve, 34295 Montpellier, France

<sup>4</sup> Institute of Molecular Genetics of Montpellier, 34293 Montpellier, France

Correspondence should be addressed to Samir Hamamah; [s-hamamah@chu-montpellier.fr](mailto:s-hamamah@chu-montpellier.fr)

Received 2 July 2014; Revised 15 July 2014; Accepted 17 July 2014; Published 3 September 2014

Academic Editor: Calvin Yu-Chian Chen

Copyright © 2014 Tamadir Al-Edani et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Impact of female aging is an important issue in human reproduction. There was a need for an extensive analysis of age impact on transcriptome profile of cumulus cells (CCs) to link oocyte quality and developmental potential with patient's age. CCs from patients of three age groups were analyzed individually using microarrays. RT-qPCR validation was performed on independent CC cohorts. We focused here on pathways affected by aging in CCs that may explain the decline of oocyte quality with age. In CCs collected from patients >37 years, angiogenic genes including *ANGPTL4*, *LEPR*, *TGFBR3*, and *FGF2* were significantly overexpressed compared to patients of the two younger groups. In contrast genes implicated in TGF- $\beta$  signaling pathway such as *AMH*, *TGFBI*, inhibin, and activin receptor were underexpressed. CCs from patients whose ages are between 31 and 36 years showed an overexpression of genes related to insulin signaling pathway such as *IGFBP3*, *PIK3RI*, and *IGFBP5*. A bioinformatic analysis was performed to identify the microRNAs that are potential regulators of the differentially expressed genes of the study. It revealed that the pathways impacted by age were potential targets of specific miRNAs previously identified in our CCs small RNAs sequencing.

## 1. Introduction

In developing countries, the first baby is conceived with a delay that keeps increasing. With aging there is natural decline in female fertility, which raises crucial issues for the society. The fertility decline is slow and steady in 30 to 35 years old women. However, this decline accelerates past 35 years due to the decrease in oocyte quality and ovarian reserve [1, 2]. Therefore female age is crucial and oocyte aging is a common cause of assisted reproduction technology failures [3]. MII oocyte stores large quantities of mRNA and proteins and contains a high number of mitochondria [4, 5]. Oocytes from women with an advanced reproductive age may have an increase of oxidative stress with consequences on mitochondrial DNA (mtDNA) integrity, resulting in

mitochondrial dysfunction [6, 7]. Interestingly transcriptome profiles showed a substantial difference between younger and older human oocytes [8]. Moreover the increase of aneuploidy due to aging is well documented. Indeed, the link between female age and oocyte aneuploidy prevalence was extensively studied [9]. However both intrinsic (oocyte) and/or extrinsic (follicular) factors may be involved in the oocyte quality decline. The ovarian follicular microenvironment, mediated through cumulus cells (CCs), is crucial for the development of competent oocytes [10]. The CCs are in physical contact with the oocyte; together they form the cumulus-oocyte complex (COC) and undergo a cross-talk [11]. The oocyte controls the differentiation and expansion of CCs, which in turn are responsible for the metabolism of the glucose and pyruvate used for energy production in

the oocyte [12]. An aged follicular microenvironment could impact oocytes and leave a characteristic transcriptional footprint in the surrounding CCs. Indeed, the use of human CC gene expression has proved powerful as a noninvasive approach to predict oocyte quality and developmental potential [13–16]. The analysis of gene expression in human CCs in relation to female age is based on the same rationale [17–19]. However, with the exception of one proteomic analysis [17], no high throughput study based on gene expression profile in relation to female age was performed on cumulus cells. Our hypothesis here is based on the assumption that female age may have a wide impact on gene expression and may specifically affect pathways that are critical for oocyte quality and development. The purposes of this study were (i) to thoroughly evaluate impact of maternal age on gene expression profiles using individual CCs isolated from the periovulatory follicles of three age categories of patients, (ii) to characterize the pathways that were significantly affected by female aging, and (iii) to identify their miRNAs regulators.

## 2. Materials and Methods

**2.1. Sample Characterization and Collection.** The Review Board of the Institute of Research in Biotherapy approved this project. All patients provided their written informed consent for the use of CC samples for research.

CC samples were collected from patients who participated to the multicentric trial previously described [20] and from Montpellier ART centre. Patients were stimulated with a combination of GnRH antagonist protocol with recombinant FSH or with HP-hMG before undergoing intracytoplasmic sperm injection (ICSI) procedure for male infertility. Cumulus oocyte complexes (COCs) were recovered under ultrasound echo-guidance 36 h after human Chorionic Gonadotrophin (5,000 UI, hCG) administration. CCs were separated mechanically from the corresponding oocyte as previously described [14]. For microarray 28 individual CC samples obtained from 16 patients were classified into three age groups: <30 years ( $CC_{\text{younger}}$ ), 31–34 years ( $CC_{\text{median}}$ ), and >37 years ( $CC_{\text{older}}$ ). The qRT-PCR analyses were performed on 15 independent CCs from the above groups and 4 CCs from a 35–36 additional group.

**2.2. RNA Extraction and Microarray Processing.** CCs were frozen at  $-80^{\circ}\text{C}$  in RLT buffer before RNA extraction. Then the RNeasy Micro kit (ref: 74004; Qiagen) was used to extract total RNA from each CC sample, according to the manufacturers' recommended protocols. The quantity and purity of the total RNAs were determined by using a NanoDrop ND-1000 spectrophotometer (NanoDrop ND-Thermo Fisher Scientific, Wilmington, DE, USA) and their integrity determined by using the Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, <http://www.agilent.com>). All RNA samples were stored at  $-80^{\circ}\text{C}$ . Microarray experiments were performed on the microarray platform of Institute of Research in Biotherapy at the Montpellier University Hospital. The Affymetrix 3' IVT express protocol (reference 901229) was used as previously described [20].

**2.3. Microarray Data Analysis.** After image processing with the Affymetrix GeneChip Operating 1.4 software, the CEL files were analyzed using the Affymetrix Expression Console Software v1.3.1 and normalized with the MAS5.0 algorithm by scaling each array to a target value of 100 using the global scaling method. This algorithm also determines whether a gene is expressed with a defined "detection call." This "call" can either be "present" (when the perfect match probes are significantly more hybridized than the mismatch probes,  $P < 0.04$ ), "marginal" ( $0.04 < P < 0.06$ ), or "absent" ( $P > 0.06$ ). Gene annotation was performed using NetAffx (<http://www.affymetrix.com>; March 2009). A first selection using the detection call (present in at least seven samples) and variation coefficient ( $\geq 40\%$ ) of CC samples identified 9,802 transcripts. Then, to compare the three groups of CCs according to maternal age, a Significance Analysis of Microarrays-Multiclass (SAM-M) (<http://statweb.stanford.edu/~tibs/SAM/>) was used. SAM-M handed the significantly expressed genes with a  $q$ -value  $< 5\%$  in the three age categories. CLUSTER and TREEVIEW software packages were used for the hierarchical clustering analysis. SPSS 12.0 (SPSS, Chicago, IL) software was used for box-and-whisker plots representation of expression levels of specific genes. The miRNA target predictions were performed with GeneGo MetaCore analysis software (St. Joseph, MI). Ingenuity Pathway Analysis software and DAVID (<http://david.abcc.ncifcrf.gov/>) were used for functional annotation.

**2.4. Quantitative RT-PCR.** Reverse transcription (RT) was performed as recommended by the manufacturer (Invitrogen) with 150 ng of RNA in a  $20\ \mu\text{L}$  reaction volume that included Superscript II (ref. 18064-014, Invitrogen), oligo-dT primer, dNTP mixture,  $\text{MgCl}_2$ , and RNase inhibitor. Quantitative PCR was performed using the SYBR Green I Master kit (Roche Diagnostics, Mannheim, Germany) with  $2\ \mu\text{L}$  of 1/20 dilution of the RT reaction product and 0.5 mM primer (SIGMA Genosys) in a total volume of  $10\ \mu\text{L}$ . The amplification was run in a LightCycler 480 apparatus as follows: after the denaturation step for 10 min at  $95^{\circ}\text{C}$ , cycling conditions were 10 s at  $95^{\circ}\text{C}$ , 30 s at  $65^{\circ}\text{C}$  and 1 s at  $72^{\circ}\text{C}$  for 45 cycles. Gene expression levels were normalized to the housekeeping gene Glyceraldehyde 3-Phosphate Dehydrogenase (*GAPDH*) using the following formula  $100/2^{\Delta\Delta\text{Ct}}$  where  $\Delta\Delta\text{Ct} = \Delta\text{Ct}_{\text{unknown}} - \Delta\text{Ct}_{\text{positive control}}$ . The primer sequences are shown in (see Table SI in Supplementary Material available online at <http://dx.doi.org/10.1155/2014/964614>).

**2.5. Statistical Analysis.** Statistical analysis was performed with the GraphPad InStat 3 software. For qRT-PCR, the Kruskal-Wallis nonparametric test was used. The differences among the groups were considered significant when the  $P$  value is  $< 0.05$ .

## 3. Results

**3.1. Gene Expression Profiles of CCs according to Female Age.** In order to gain insight into the molecular basis of age

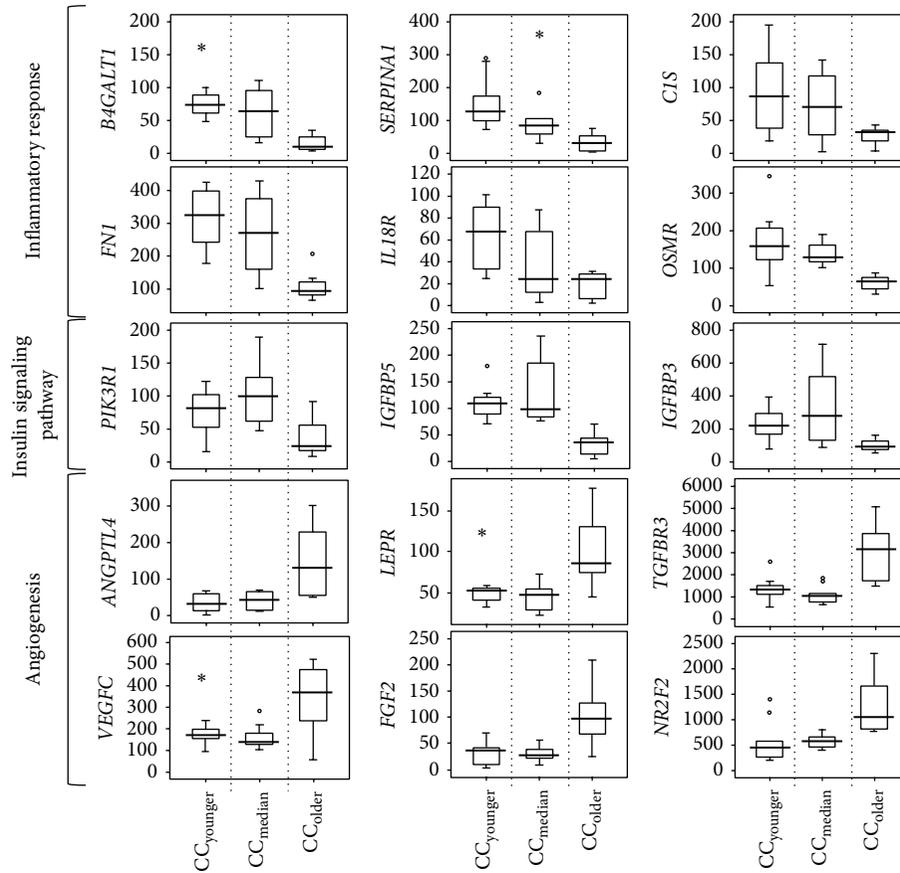
impact on COCs, we analyzed the transcriptomes of CCs from women with different age categories. A first selection based on the detection call and variation coefficient of all the CC samples from aged and young patients delineated 9,802 transcripts. Then, using SAM-M and after having discarded 35 genes that we previously showed to be affected by the COS protocols [20], we identified a total of 2,186 transcripts (corresponding to 1,874 genes) with a  $q$ -value <5% that significantly distinguished the three CC groups according to female age (Supplementary Table SII). The analysis of the transcriptome data revealed a characteristic molecular signature for each one of the three age categories (Figure 1). The expression patterns of the genes that best represent these categories are illustrated in the box-plots (Figure 1(a)). In  $CC_{\text{younger}}$  group, overexpression was observed for inflammatory response genes such as *B4GALTI*, *SERPINA1*, *CIS*, *IL18R1*, *FNI*, and *OSMR*. The  $CC_{\text{median}}$  group revealed overexpression of genes involved in insulin signaling pathway, the most representative being *IGFBP3*, *IGFBP5* and *PIK3R1*. Finally the  $CC_{\text{older}}$  group was significantly enriched with genes that are important for angiogenesis such as *ANGPTL4*, *LEPR*, *TGFBR3*, *VEGFC*, *FGF2* and *NR2F2*. In addition, a list of 20 genes with the highest contrast and lowest  $q$ -value according to SAM-M, were chosen for each category to perform the hierarchical clustering (Supplementary Table SIII). Interestingly,  $CC_{\text{older}}$  samples distantly located from the  $CC_{\text{younger}}$  and  $CC_{\text{median}}$  samples (Figure 1(b)).

**3.2. Validation of Gene Expression by Quantitative RT-PCR.** Nine differentially expressed genes were selected for validation on the basis of relevant functional annotations. Hence, three genes involved in the inflammatory process (*B4GALTI*, *SERPINA1*, and *CIS*), three genes of the insulin signaling (*IGFBP3*, *IGFBP5*, and *PIK3R1*) and three genes of the angiogenesis process (*ANGPTL4*, *LEPR*, and *TGFBR3*) were chosen for qRT-PCR validation. Analysis of the qRT-PCR data on independent cohorts of CCs indicated that all the selected genes were differentially expressed in the three age categories and in agreement with the microarray findings (Figure 2). Using qRT-PCR we aimed to test the expression level of the above genes in individual CCs from 35 and 36 old patients. These CCs clearly displayed an expression pattern similar to the  $CC_{\text{median}}$  age category (Figure SI) suggesting that the switch for these genes occurs after the age of 36.

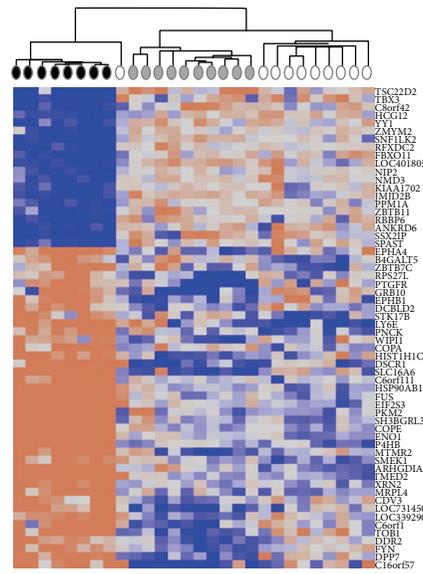
**3.3. Deregulation in  $CC_{\text{older}}$  Genes that Are Essential for the Oocyte Quality and Competence.** Many biological pathways were reported to be crucial for their impact on the oocyte development. They include transforming growth factor  $\beta$  (TGF- $\beta$ ) signaling, steroidogenesis and metabolic pathways. Interestingly the key members of these pathways displayed significant changes in their gene expression (Table SII). As shown in Figure 3(a), many genes of the TGF- $\beta$  signaling pathway were underexpressed in  $CC_{\text{older}}$  compared with  $CC_{\text{younger}}$  and  $CC_{\text{median}}$ , including *AMH* (Anti-Mullerian Hormone), *TGFBI*, inhibin (*INHA*) and activin receptor (*ACVR2B*). In contrast overexpression was observed in  $CC_{\text{older}}$  for several genes that are

involved in steroidogenesis and fatty acid metabolism (*HSD17B1*, *HSD17B6*, *NSDHL*, *SRA1*, *CYP19A1*, *PPARA*), glucose metabolism (*ALG13*, *GLT8D3*) and glucose transporters (*SLC2A3*, *SLC2A1*, *SLC2A13*, *SLC2A8*). It is noteworthy that several genes that play an essential role in the cumulus-oocyte dialog (*INHA*, *CD200* and *IL6ST*) were downregulated in  $CC_{\text{older}}$  (Figure 3(b)). Moreover,  $CC_{\text{older}}$  may be distinguished from the two younger age categories by a downregulation of genes that are essential for genome integrity, in particular *MSRB3*, *UCHL5IP*, *POLH*, *OBFC2B*, and *CHAFIA* that are essential for antioxidative and DNA repair functions.

**3.4. Potential miRNA Regulators of the Differentially Expressed Genes of the Study.** Using the GenGo Metacore software, we first aimed to identify which miRNAs regulate the genes that were overexpressed in each of the three age categories,  $CC_{\text{younger}}$ ,  $CC_{\text{median}}$  and  $CC_{\text{older}}$  (Figure 4(a)). We identified altogether 286 miRNAs that are putative regulators of the differentially expressed genes identified in this study, among which 176 are common putative regulators of the genes overexpressed in the three age categories, 71 for the genes whose expression was higher in  $CC_{\text{younger}}$  and  $CC_{\text{median}}$ . Only one miRNA was shared by  $CC_{\text{median}}$  and  $CC_{\text{older}}$  categories specifically; similarly genes overexpressed in  $CC_{\text{younger}}$  and  $CC_{\text{older}}$  had one specific miRNA in common. Interestingly this analysis also discriminates the  $CC_{\text{older}}$  from  $CC_{\text{younger}}$  and  $CC_{\text{median}}$ , which may be considered as a super-group with common features. Some miRNAs were specific for one of the three age categories. Thirty-three miRNAs were identified as putative regulators of the genes overexpressed in  $CC_{\text{older}}$ , one for the  $CC_{\text{younger}}$  and 3 for the  $CC_{\text{median}}$  categories (for the comprehensive lists, see Supplementary Table SIV). Among all the miRNAs retrieved by GenGo, 87% were identified by sequencing in CCs [21]. The fact that only the differentially expressed genes were submitted to GenGO may account for the missing 13%. There is another discrepancy between the list of the potential regulators and the miRNAs actually present in the CCs as identified in our previous work [21]. It is illustrated in Figure 4(a) for the two categories that stand out in the present study, namely the  $CC_{\text{younger}}-CC_{\text{median}}$  super-group (71) on the one hand and the  $CC_{\text{older}}$  (33) on the other hand. Among these potential miRNA regulators, only 6 are actually expressed in CCs: *MIR425*, *MIR744*, *MIR146b*, *Let-7d* for the  $CC_{\text{younger}}-CC_{\text{median}}$  super group and *MIR202*, *Let-7e* for the  $CC_{\text{older}}$ . This discrepancy might reflect a tissue specific expression of miRNAs. Interestingly *MIR202* is a potential regulator of the hyaluronan synthase-encoding gene *HAS2* that is related to aging and angiogenesis [22] and *MIR744* is a *TGFBI* validated regulator [23]. The largest set of miRNAs retrieved by GenGo was common to the three age categories (176). This set was crossed with those effectively expressed in CCs [21], resulting in a list of 22 miRNAs. We were interested in those that regulate significant gene members of the pathways and processes impacted by female age and that were also experimentally validated. The results of this analysis are shown in Figure 4(b). None fulfills these criteria for the validated genes of the inflammatory process overexpressed in the  $CC_{\text{younger}}$ . In  $CC_{\text{median}}$ , *IGFBP3*,



(a)



● CC<sub>older</sub> ○ CC<sub>younger</sub>  
● CC<sub>median</sub>

(b)

FIGURE 1: (a) Expression of cumulus cells genes according to female age. Box-and-whisker plots that represent expression of genes implicated in different biological processes and signaling pathways in the three female age categories, CC<sub>younger</sub>, CC<sub>median</sub>, and CC<sub>older</sub>. The signal intensity of each gene is shown on the y axis as arbitrary units determined by the Affymetrix GCOS software. (b) Heat map and cluster dendrograms of differentially expressed genes. Hierarchical clustering is shown for 20 genes with the highest expression level in each of the 3 age categories of individual CCs. Overexpressed and underexpressed genes were marked in blue and pink, respectively. The three age categories are shown in white for CC<sub>younger</sub>, grey for CC<sub>median</sub>, and black for CC<sub>older</sub>.

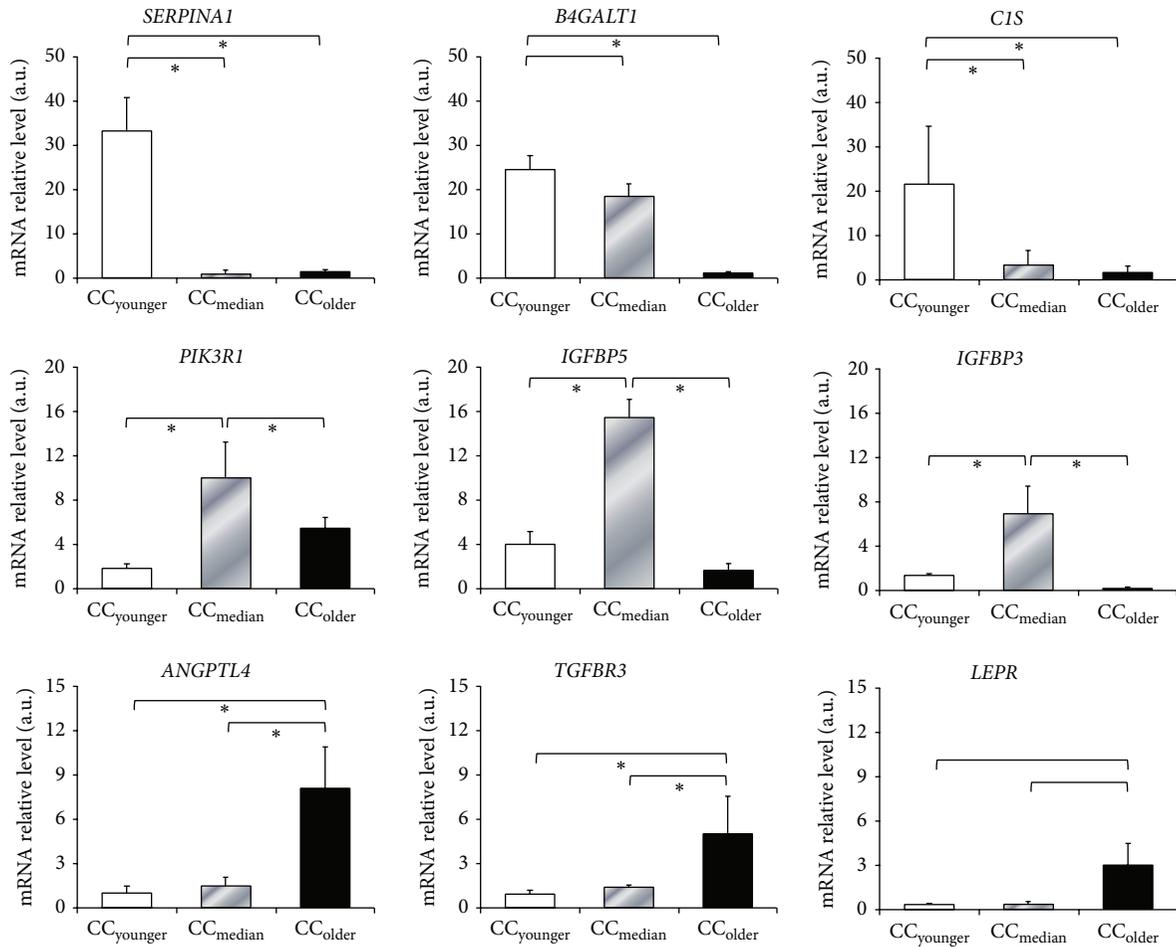


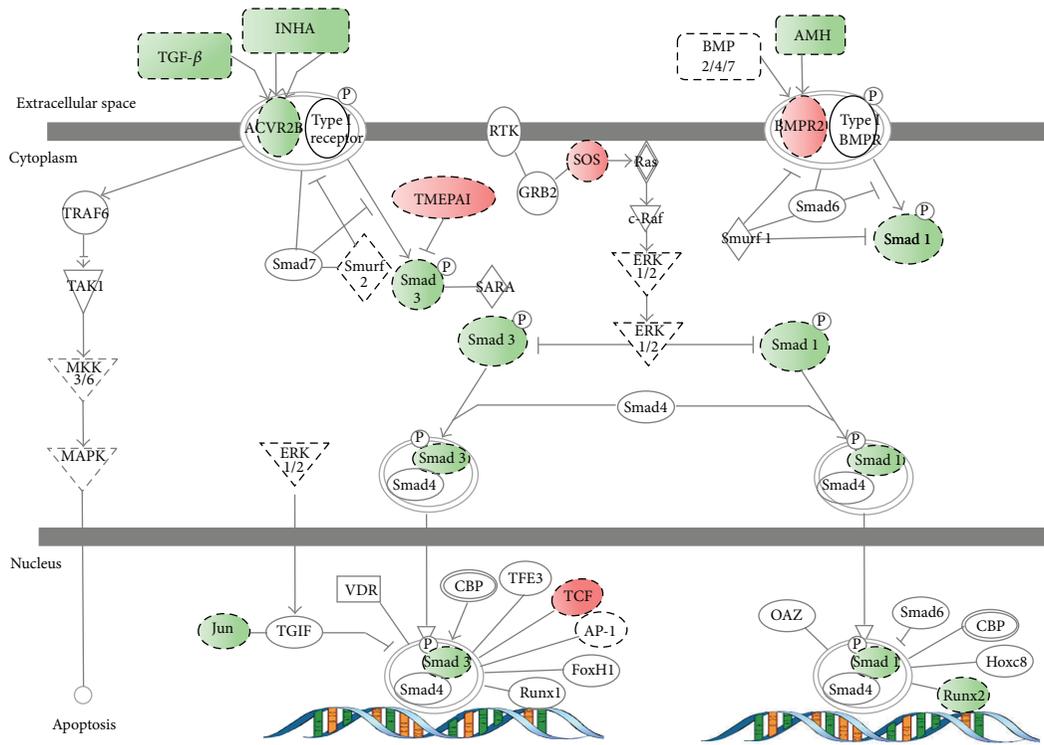
FIGURE 2: Validation by qRT-PCR of some gene members of key pathways that are differentially expressed in the three age categories. This figure shows the mRNA relative abundance of three genes implicated in inflammatory response (*SERPINA1*, *B4GALT1*, and *CIS*), three genes in insulin signaling (*PIK3R*, *IGFBP3*, and *IGFBP5*), and three genes in angiogenesis process (*ANGPTL4*, *TGFBR3*, and *LEPR*). The signal intensity for each gene is shown on the y-axis in arbitrary units determined by RT-qPCR analysis. \* indicates a significant difference of gene expression between CCs categories (\*  $P < 0.05$ ). Results were presented as the mean  $\pm$  SEM. CC<sub>younger</sub> (white, age: <30 years), CC<sub>median</sub> (grey, age: 31–34 years), and CC<sub>older</sub> (Black, age >37 years).

and *IGFBP5* of the insulin-signaling pathway are targets of *MIR210* and *MIR140*, respectively. Finally in CC<sub>older</sub>, genes implicated in angiogenesis *LEPR* and *TGFBR3* are *MIR21* targets whereas *FGF2* is targeted by *MIR424*. For more details see Table SV.

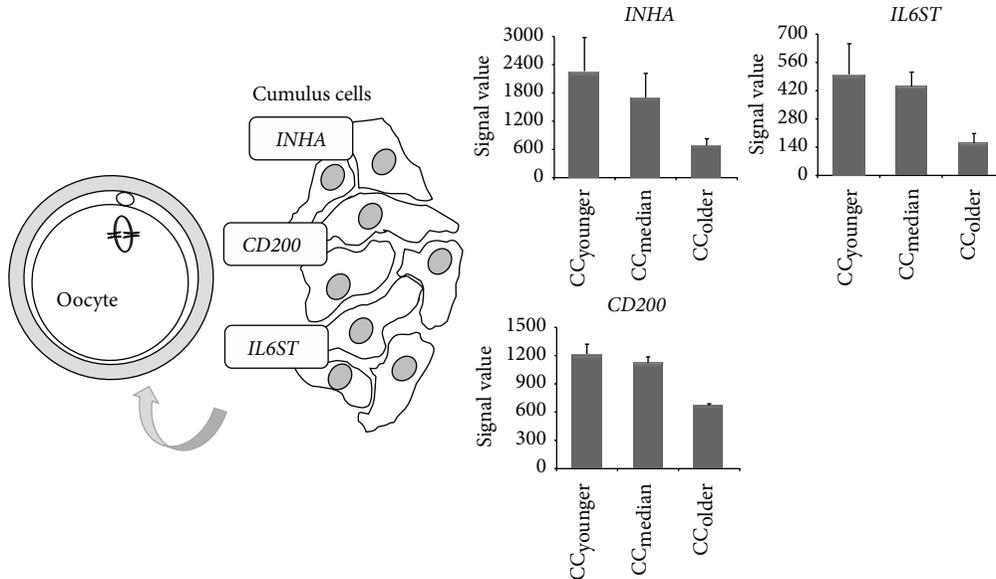
#### 4. Discussion

Acquisition of oocyte competence is a gradual and complex process, which depends on the follicular microenvironment. Within this microenvironment, the bidirectional communication between the CCs and the oocyte plays a crucial role. Therefore gene expression in CCs mirrors the oocyte physiology. In order to gain insight into the mechanisms that underlie oocyte quality decline with age, we first investigated the transcriptome profiles in CCs from women of three age categories. Our objective was to identify molecular signatures characteristic of each age category and investigate

their biological relevance to oocyte quality. DNA microarray analysis revealed a significantly distinct molecular signature of 1,874 genes among the three age groups, suggesting a wide impact of female age on the CC gene-expression profile. It is noteworthy that the inflammatory genes emerged in the CC<sub>younger</sub> group such as *IL18R1*, *IL1R1*, *IL1R2*, *SERPINA1*, and *B4GALT1*. Inflammatory reaction is known to induce ovulation through infiltration of leukocytes into the area surrounding the follicle [24]. Cytokines are important in the regulation of ovarian function and oocyte quality [25]. On the other hand interleukins *IL18* and *IL1 $\beta$*  were reported to be present in floating granulosa cells of human preovulatory follicles [26]. CC<sub>median</sub> group may be characterized by an overexpression of gene members of the “insulin-signaling pathway”, such as *IGFBP3* and *IGFBP5* whereas *INSR* was overexpressed in both the CC<sub>median</sub> and CC<sub>older</sub> groups. Several studies have shown that insulin and *IGF* system play an important role in folliculogenesis [27–29] and in oocyte maturation [30]. IGF-binding proteins (*IGFBPs*)



(a)



(b)

FIGURE 3: (a) TGF- $\beta$  signaling pathway was deregulated in older CCs. The Ingenuity Pathway Analysis software was used to analyze impact of maternal age on TGF- $\beta$  signaling. Downregulated genes in older CCs are shown in green and upregulated ones in red. Uncolored genes were not differentially expressed by our analysis but were integrated into the computationally generated networks on the basis of the evidence stored in the IPA knowledge memory indicating a relevance to this network. A plain line indicates direct interactions, a dashed line indicates indirect interactions, a line without arrowhead indicates binding only, a line finishing with a vertical line indicates inhibition, and a line with an arrowhead indicates “acts on.” (b) Schematic representation of genes upregulated in CC<sub>younger</sub> and CC<sub>median</sub> and downregulated in CC<sub>older</sub> that are involved in cumulus-oocyte complex and oocyte development. Histograms show signal values of genes (*INHA*, *CD200*, and *IL6ST*) that are differentially expressed between age categories. Gene expression is measured by pan-genomic HG-U133 Plus 2.0 Affymetrix oligonucleotides microarrays, and the signal intensity for each gene is shown on the y-axis as arbitrary units determined by the GCOS 1.2 software (Affymetrix).

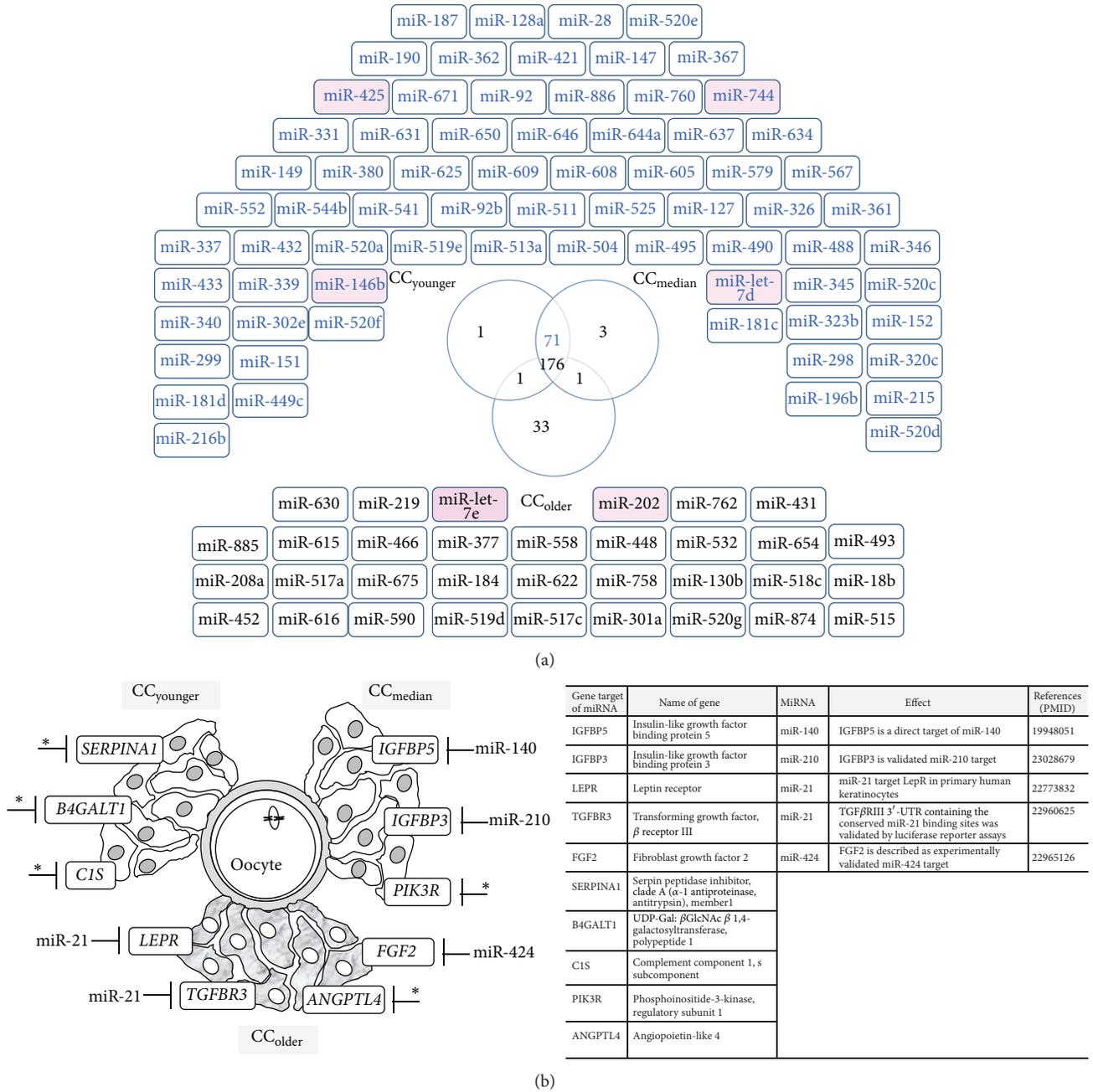


FIGURE 4: (a) Venn diagram representing the number of miRNAs retrieved from the GenGo analysis. The genes overexpressed in each age category were submitted to GenGo to identify their potential miRNA regulators. 249 miRNAs were retrieved for the CC<sub>younger</sub> group, 251 for the CC<sub>median</sub>, and 211 for the CC<sub>older</sub>. The Venn diagram drawn after these lists shows that the majority is common to the three age categories. The miRNAs that are detected in the cumulus cells by using deep-sequencing approach [21] are shown in pink. (b) Schematic representation of some of the validated genes of the three pathways and processes discussed in this work and their miRNA regulators. Only the miRNAs that were found in the CCs small RNA sequencing and reported in the literature to be experimentally validated were represented. \* indicates the validated genes with no miRNA regulator that meets these criteria.

that modulate interactions of IGFs with IGF and insulin receptors [31] have also an antiangiogenic activity [32–34]. Therefore, overexpression of IGFFBPs in CC<sub>median</sub> may be to modulate angiogenesis and maintain a balance. Last, the CC<sub>older</sub> group is precisely characterized by an upregulation of genes associated with angiogenesis (ANGPTL4, LEPR,

TGFBR3, VEGFC, FGF2 and NR2F2). Angiogenesis plays a critical role in the late stages of folliculogenesis by providing nutrients and oxygen to the growing follicles. However, it may be associated with pathology and induced by microenvironmental factors like hypoxia. In this context, the follicular cells synthesize several angiogenic factors [26, 35, 36], among

which the vascular endothelial growth factor C (*VEGFC*) and angiopoietin-like 4 (*ANGPTL4*), which are induced in response to hypoxic stimuli [37–39]. So, the overexpression of angiogenic factors and hypoxia-inducible protein 2 (*HIG2*) in the  $CC_{\text{older}}$  group could be caused by insufficiency of oxygen. Similarly *VEGF* that is shown to increase in follicular fluid with age could be enhanced by hypoxia in old follicles [40, 41]. Most interestingly oocytes from hypoxic follicles have disorganized meiotic spindles [42]. These observations added to the reported increase of aneuploidy with female aging [43] may be revisited in light of our results. Hypoxia might be one of the consequences of aging, which in turn would affect chromosome segregation. Adaptive changes to oxygen availability are critical for cell survival and tissue homeostasis. Therefore, augmentation of angiogenesis in the  $CC_{\text{older}}$  group may be a compensatory process to modulate the deleterious impact of hypoxia. Similarly the upregulation of genes that encode metabolic enzymes (*HSD17B*, *CYP19A1*, *ALG13*, and *GLT8D3*) and glucose transporters (*SLC2A3*, *SLC2A1*, *SLC2A13*, *SLC2A8*) in the  $CC_{\text{older}}$  group could reflect a compensatory mechanism to increase energy production. These results are consistent with the observations reported recently [17, 19]. Indeed, the energy supplied by the CCs is known to be required for oocyte quality [44, 45]. Some members of the *TGF- $\beta$*  superfamily, which are crucial to processes that govern follicle development and oocyte maturation [46], were underexpressed in the  $CC_{\text{older}}$  group such as *AMH* (Anti-Mullerian Hormone), *TGFBI*, inhibin (*INHA*), and activin receptor (*ACVR2B*). Interestingly *AMH* is produced by early primary follicles and its mRNA level is known to decrease with age. Therefore, it represents an early marker of ovarian follicle growth and a reliable marker of ovarian reserve and oocyte quality [47–49].

Another important question we addressed concerns the regulation of the genes that stand out in our study. We focused on the bioinformatic analysis of miRNAs. MiRNAs are noncoding small RNAs (18–25 nucleotides), which regulate cellular genes through RNA degradation or translational inhibition [50, 51]. Not only miRNAs have been shown to regulate the aging process in different tissues and cells [52], but their importance is also well recognized in the control of human cumulus-oocyte crosstalk and ovarian function and aging [21, 53, 54]. Interestingly, *TGF- $\beta$*  signaling is one of the most significant pathways targeted by miRNAs contained in the follicular fluid [55]. Moreover gene members of this pathway are direct targets of *MIR21* that is the most abundant miRNA in CCs [21]. The role of *MIR21* is essential in ovarian function to prevent apoptosis in mouse periovulatory granulosa cells both *in vivo* and *in vitro* [56]. Moreover, it promotes the follicular cell survival during ovulation and is upregulated during luteinization [57]. Interestingly, a recent work reports a correlation between *MIR21* abundance and women age; a significant decrease was observed in follicular fluid of older women [53]. In the current study, two angiogenic genes (*LEPR* and *TGFBR3*) were upregulated in  $CC_{\text{older}}$  where *MIR21* is the least abundant [53]. Finally, the process that is central to this study is angiogenesis that may be induced in response to hypoxia, a major issue in aging follicles. Interestingly, miRNAs play a critical role in

the cellular response to hypoxia [58]. *MIR210* whose overexpression in hypoxic conditions induces angiogenesis [59, 60] directly targets *IGFBP3*, an inhibitor of angiogenesis [32, 61]. Furthermore, *MIR424* that is downregulated in response to hypoxia in primary human trophoblasts [62] targets *FGF2*, an angiogenesis inducer [63, 64]. Taken together these data suggest that in aging follicles angiogenesis may be induced in response to hypoxia by the underexpression of *IGFBP3* and overexpression of *FGF2*.

## 5. Conclusion

The present study reports for the first time an extensive analysis of gene expression in cumulus cells in relation to female age. Specific molecular signatures were characterized for the three age categories. Our findings point to aging as a major player in processes and pathways that are of key biological importance for oocyte growth and genome integrity. Moreover the upregulation of angiogenic genes in  $CC_{\text{older}}$  is very informative on the way the follicle attempts to buffer the deleterious impact of aging associated hypoxia. In addition to the transcriptomes, the comprehensive characterization of the miRNA regulators of the genes impacted by female age represents a valuable resource for future investigations on the biology of aging oocyte.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Authors' Contribution

Tamadir Al-Edani and Said Assou contributed equally to this work.

## Acknowledgments

This work was supported by Ferring Pharmaceuticals A/S. The authors thank the direction of the Montpellier 1 University and University Hospital of Montpellier for their support. They thank all members of ART team for their assistance during this study.

## References

- [1] M. J. Faddy, R. G. Gosden, A. Gougeon, S. J. Richardson, and J. F. Nelson, "Accelerated disappearance of ovarian follicles in mid-life: implications for forecasting menopause," *Human Reproduction*, vol. 7, no. 10, pp. 1342–1346, 1992.
- [2] C. Alviggi, P. Humaidan, C. M. Howles, D. Tredway, and S. G. Hillier, "Biological versus chronological ovarian age: implications for assisted reproductive technology," *Reproductive Biology and Endocrinology*, vol. 7, article 101, 2009.
- [3] A. J. Wilcox, C. R. Weinberg, and D. D. Baird, "Post-ovulatory ageing of the human oocyte and embryo failure," *Human Reproduction*, vol. 13, no. 2, pp. 394–397, 1998.

- [4] R. P. S. Jansen, "Germline passage of mitochondria: quantitative considerations and possible embryological sequelae," *Human Reproduction*, vol. 15, supplement 2, pp. 112–128, 2000.
- [5] T. A. L. Brevini Gandolfi and F. Gandolfi, "The maternal legacy to the embryo: cytoplasmic components and their effects on early development," *Theriogenology*, vol. 55, no. 6, pp. 1255–1276, 2001.
- [6] A. Agarwal, S. Gupta, and R. K. Sharma, "Role of oxidative stress in female reproduction," *Reproductive Biology and Endocrinology*, vol. 3, article 28, 2005.
- [7] U. Eichenlaub-Ritter, M. Wiczorek, S. Lücke, and T. Seidel, "Age related changes in mitochondrial function and new approaches to study redox regulation in mammalian oocytes in response to age or maturation conditions," *Mitochondrion*, vol. 11, no. 5, pp. 783–796, 2011.
- [8] M. L. Grøndahl, C. Y. Andersen, J. Bogstad, F. C. Nielsen, H. Meinertz, and R. Borup, "Gene expression profiles of single human mature oocytes in relation to age," *Human Reproduction*, vol. 25, no. 4, pp. 957–968, 2010.
- [9] E. Fragouli, D. Wells, and J. D. A. Delhanty, "Chromosome abnormalities in the human oocyte," *Cytogenetic and Genome Research*, vol. 133, no. 2–4, pp. 107–118, 2011.
- [10] R. B. Gilchrist, M. Lane, and J. G. Thompson, "Oocyte-secreted factors: regulators of cumulus cell function and oocyte quality," *Human Reproduction Update*, vol. 14, no. 2, pp. 159–177, 2008.
- [11] D. F. Albertini, C. M. H. Combelles, E. Benecchi, and M. J. Carabatsos, "Cellular basis for paracrine regulation of ovarian follicle development," *Reproduction*, vol. 121, no. 5, pp. 647–653, 2001.
- [12] D. L. Russell and R. L. Robker, "Molecular mechanisms of ovulation: co-ordination through the cumulus complex," *Human Reproduction Update*, vol. 13, no. 3, pp. 289–312, 2007.
- [13] S. Assou, D. Haouzi, J. de Vos, and S. Hamamah, "Human cumulus cells as biomarkers for embryo and pregnancy outcomes," *Molecular Human Reproduction*, vol. 16, no. 8, Article ID gaq032, pp. 531–538, 2010.
- [14] S. Assou, D. Haouzi, K. Mahmoud et al., "A non-invasive test for assessing embryo potential by gene expression profiles of human cumulus cells: a proof of concept study," *Molecular Human Reproduction*, vol. 14, no. 12, pp. 711–719, 2008.
- [15] M. Hamel, I. Dufort, C. Robert et al., "Identification of differentially expressed markers in human follicular cells associated with competent oocytes," *Human Reproduction*, vol. 23, no. 5, pp. 1118–1127, 2008.
- [16] S. Assou, I. Boumela, D. Haouzi et al., "Dynamic changes in gene expression during human early embryo development: from fundamental aspects to clinical applications," *Human Reproduction Update*, vol. 17, no. 2, Article ID dmq036, pp. 272–290, 2011.
- [17] S. McReynolds, M. Dzieciatkowska, B. R. McCallie et al., "Impact of maternal aging on the molecular signature of human cumulus cells," *Fertility and Sterility*, vol. 98, no. 6, pp. 1574.e5–1580.e5, 2012.
- [18] M.-S. Lee, C.-H. Liu, T.-H. Lee et al., "Association of creatin kinase B and peroxiredoxin 2 expression with age and embryo quality in cumulus cells," *Journal of Assisted Reproduction and Genetics*, vol. 27, no. 11, pp. 629–639, 2010.
- [19] L. Pacella, D. L. Zander-Fox, D. T. Armstrong, and M. Lane, "Women with reduced ovarian reserve or advanced maternal age have an altered follicular environment," *Fertility and Sterility*, vol. 98, no. 4, pp. 986.e2–994.e2, 2012.
- [20] S. Assou, D. Haouzi, H. Dechaud, A. Gala, A. Ferrieres, and S. Hamamah, "Comparative gene expression profiling in human cumulus cells according to ovarian gonadotropin treatments," *BioMed Research International*, vol. 2013, Article ID 354582, 13 pages, 2013.
- [21] S. Assou, T. Al-Edani, D. Haouzi et al., "MicroRNAs: new candidates for the regulation of the human cumulus-oocyte complex," *Human Reproduction*, vol. 28, no. 11, pp. 3038–3049, 2013.
- [22] C. C. Sprenger, S. R. Plymate, and M. J. Reed, "Aging-related alterations in the extracellular matrix modulate the microenvironment and influence tumor progression," *International Journal of Cancer*, vol. 127, no. 12, pp. 2739–2748, 2010.
- [23] J. Martin, R. H. Jenkins, R. Bennagi et al., "Post-transcriptional regulation of transforming growth factor beta-1 by microRNA-744," *PLoS ONE*, vol. 6, no. 10, Article ID e25044, 2011.
- [24] J. Goto, N. Suganuma, K. Takata et al., "Morphological analyses of interleukin-8 effects on rat ovarian follicles at ovulation and luteinization in vivo," *Cytokine*, vol. 20, no. 4, pp. 168–173, 2002.
- [25] S. Vujišić, S. Ž. Lepej, I. Emedi, R. Bauman, A. Remenar, and M. K. Tiljak, "Ovarian follicular concentration of IL-12, IL-15, IL-18 and p40 subunit of IL-12 and IL-23," *Human Reproduction*, vol. 21, no. 10, pp. 2650–2655, 2006.
- [26] S. Köks, A. Velthut, A. Sarapik et al., "The differential transcriptome and ontology profiles of floating and cumulus granulosa cells in stimulated human antral follicles," *Molecular Human Reproduction*, vol. 16, no. 4, Article ID gap103, pp. 229–240, 2009.
- [27] H. Louhio, O. Hovatta, J. Sjöberg, and T. Tuuri, "The effects of insulin, and insulin-like growth factors I and II on human ovarian follicles in long-term culture," *Molecular Human Reproduction*, vol. 6, no. 8, pp. 694–698, 2000.
- [28] J. Zhao, M. A. Taverne, G. C. van der Weijden, M. M. Bevers, and R. van den Hurk, "Insulin-like growth factor-I (IGF-I) stimulates the development of cultured rat pre-antral follicles," *Molecular Reproduction and Development*, vol. 58, no. 3, pp. 287–296, 2001.
- [29] J. Kwintkiewicz and L. C. Giudice, "The interplay of insulin-like growth factors, gonadotropins, and endocrine disruptors in ovarian follicular development and function," *Seminars in Reproductive Medicine*, vol. 27, no. 1, pp. 43–51, 2009.
- [30] P. L. Lorenzo, M. J. Illera, J. C. Illera, and M. Illera, "Enhancement of cumulus expansion and nuclear maturation during bovine oocyte maturation in vitro by the addition of epidermal growth factor and insulin-like growth factor I," *Journal of Reproduction and Fertility*, vol. 101, no. 3, pp. 697–701, 1994.
- [31] S. M. Firth and R. C. Baxter, "Cellular actions of the insulin-like growth factor binding proteins," *Endocrine Reviews*, vol. 23, no. 6, pp. 824–854, 2002.
- [32] S.-H. Oh, W.-Y. Kim, O.-H. Lee et al., "Insulin-like growth factor binding protein-3 suppresses vascular endothelial growth factor expression and tumor angiogenesis in head and neck squamous cell carcinoma," *Cancer Science*, vol. 103, no. 7, pp. 1259–1266, 2012.
- [33] M. J. Moreno, M. Ball, M. Rukhlova et al., "IGFBP-4 anti-angiogenic and anti-tumorigenic effects are associated with anti-cathepsin B activity," *Neoplasia*, vol. 15, no. 5, pp. 554–567, 2013.
- [34] C. Zhang, L. Lu, Y. Li et al., "IGF binding protein-6 expression in vascular endothelial cells is induced by hypoxia and plays a negative role in tumor angiogenesis," *International Journal of Cancer*, vol. 130, no. 9, pp. 2003–2012, 2012.

- [35] H. M. Fraser, "Regulation of the ovarian follicular vasculature," *Reproductive Biology and Endocrinology*, vol. 4, article 18, 2006.
- [36] C. W. Pugh and P. J. Ratcliffe, "Regulation of angiogenesis by hypoxia: role of the HIF system," *Nature Medicine*, vol. 9, no. 6, pp. 677–684, 2003.
- [37] M. Neeman, R. Abramovitch, Y. S. Schiffenbauer, and C. Tempel, "Regulation of angiogenesis by hypoxic stress: from solid tumours to the ovarian follicle," *International Journal of Experimental Pathology*, vol. 78, no. 2, pp. 57–70, 1997.
- [38] P. González-Muniesa, C. de Oliveira, F. P. de Heredia, M. P. Thompson, and P. Trayhurn, "Fatty acids and hypoxia stimulate the expression and secretion of the adipokine ANGPTL4 (angiopoietin-like protein 4/ fasting-induced adipose factor) by human adipocytes," *Journal of Nutrigenetics and Nutrigenomics*, vol. 4, no. 3, pp. 146–153, 2011.
- [39] S.-H. Kim, Y.-Y. Park, S.-W. Kim, J.-S. Lee, D. Wang, and R. N. DuBois, "ANGPTL4 induction by prostaglandin E 2 under hypoxic conditions promotes colorectal cancer progression," *Cancer Research*, vol. 71, no. 22, pp. 7010–7020, 2011.
- [40] C. I. Friedman, D. R. Danforth, C. Herbosa-Encarnacion, L. Arbogast, B. M. Alak, and D. B. Seifer, "Follicular fluid vascular endothelial growth factor concentrations are elevated in women of advanced reproductive age undergoing ovulation induction," *Fertility and Sterility*, vol. 68, no. 4, pp. 607–612, 1997.
- [41] E. Y. Fujii and M. Nakayama, "The measurements of RAGE, VEGF, and AGEs in the plasma and follicular fluid of reproductive women: the influence of aging," *Fertility and Sterility*, vol. 94, no. 2, pp. 694–700, 2010.
- [42] J. van Blerkom, M. Antczak, and R. Schrader, "The developmental potential of the human oocyte is related to the dissolved oxygen content of follicular fluid: association with vascular endothelial growth factor levels and perifollicular blood flow characteristics," *Human Reproduction*, vol. 12, no. 5, pp. 1047–1055, 1997.
- [43] E. Fragouli, V. Bianchi, P. Patrizio et al., "Transcriptomic profiling of human oocytes: association of meiotic aneuploidy and altered oocyte gene expression," *Molecular Human Reproduction*, vol. 16, no. 8, pp. 570–582, 2010.
- [44] Q. Li, D.-Q. Miao, P. Zhou et al., "Glucose metabolism in mouse cumulus cells prevents oocyte aging by maintaining both energy supply and the intracellular redox potential," *Biology of Reproduction*, vol. 84, no. 6, pp. 1111–1118, 2011.
- [45] D. Brisard, A. Desmarchais, J. L. Touze et al., "Alteration of energy metabolism gene expression in cumulus cells affects oocyte maturation via MOS-mitogen-activated protein kinase pathway in dairy cows with an unfavorable "Fertil-" haplotype of one female fertility quantitative trait locus," *Theriogenology*, vol. 81, no. 4, pp. 599–612, 2014.
- [46] P. G. Knight and C. Glister, "Local roles of TGF- $\beta$  superfamily members in the control of ovarian follicle development," *Animal Reproduction Science*, vol. 78, no. 3–4, pp. 165–183, 2003.
- [47] Z. Merhi, E. Buyuk, D. S. Berger et al., "Leptin suppresses anti-Müllerian hormone gene expression through the JAK2/STAT3 pathway in luteinized granulosa cells of women undergoing IVF," *Human Reproduction*, vol. 28, no. 6, pp. 1661–1669, 2013.
- [48] A. L. L. Durlinger, M. J. G. Gruijters, P. Kramer et al., "Anti-Müllerian hormone inhibits initiation of primordial follicle growth in the mouse ovary," *Endocrinology*, vol. 143, no. 3, pp. 1076–1084, 2002.
- [49] P. Lehmann, M. P. Velez, J. Saumet et al., "Anti-müllerian hormone (AMH): a reliable biomarker of oocyte quality in IVF," *Journal of Assisted Reproduction and Genetics*, vol. 31, no. 4, pp. 493–498, 2014.
- [50] D. P. Bartel, "MicroRNAs: target recognition and regulatory functions," *Cell*, vol. 136, no. 2, pp. 215–233, 2009.
- [51] E. Huntzinger and E. Izaurralde, "Gene silencing by microRNAs: contributions of translational repression and mRNA decay," *Nature Reviews Genetics*, vol. 12, no. 2, pp. 99–110, 2011.
- [52] D. Xu and H. Tahara, "The role of exosomes and microRNAs in senescence and aging," *Advanced Drug Delivery Reviews*, vol. 65, no. 3, pp. 368–375, 2013.
- [53] A. Diez-Fraile, T. Lammens, K. Tilleman et al., "Age-associated differential microRNA levels in human follicular fluid reveal pathways potentially determining fertility and success of in vitro fertilization," *Human Fertility*, vol. 17, no. 2, pp. 90–98, 2014.
- [54] X. Yang, Y. Zhou, S. Peng et al., "Differentially expressed plasma microRNAs in premature ovarian failure patients and the potential regulatory function of mir-23a in granulosa cell apoptosis," *Reproduction*, vol. 144, no. 2, pp. 235–244, 2012.
- [55] J. C. da Silveira, D. N. R. Veeramachaneni, Q. A. Winger, E. M. Carnevale, and G. J. Bouma, "Cell-secreted vesicles in equine ovarian follicular fluid contain mirnas and proteins: a possible new form of cell communication within the ovarian follicle," *Biology of Reproduction*, vol. 86, no. 3, article 71, 2012.
- [56] M. Z. Carletti, S. D. Fiedler, and L. K. Christenson, "MicroRNA 21 blocks apoptosis in mouse periovulatory granulosa cells," *Biology of Reproduction*, vol. 83, no. 2, pp. 286–295, 2010.
- [57] F. X. Donadeu, S. N. Schauer, and S. D. Sontakke, "Involvement of miRNAs in ovarian follicular and luteal development," *Journal of Endocrinology*, vol. 215, no. 3, pp. 323–334, 2012.
- [58] Y. C. Chan, J. Banerjee, S. Y. Choi, and C. K. Sen, "miR-210: the master hypoxamir," *Microcirculation*, vol. 19, no. 3, pp. 215–223, 2012.
- [59] L. Zeng, X. He, Y. Wang et al., "MicroRNA-210 overexpression induces angiogenesis and neurogenesis in the normal adult mouse brain," *Gene Therapy*, vol. 21, no. 1, pp. 37–43, 2014.
- [60] J. Y. Li, T. Y. Yong, M. Z. Michael, and J. M. Gleadle, "MicroRNAs: are they the missing link between hypoxia and pre-eclampsia?" *Hypertens Pregnancy*, vol. 33, no. 1, pp. 102–114, 2014.
- [61] T. Bertero, S. Grosso, K. Robbe-Sermesant et al., "Seed-milarity" confers to hsa-miR-210 and hsa-miR-147b similar functional activity," *PLoS ONE*, vol. 7, no. 9, Article ID e44919, 2012.
- [62] J. F. Mouillet, R. B. Donker, T. Mishima, T. Cronqvist, T. Chu, and Y. Sadovsky, "The unique expression and function of miR-424 in human placental trophoblasts," *Biology of Reproduction*, vol. 89, no. 2, p. 25, 2013.
- [63] T. Ren, Y. Qing, N. Dai et al., "Apyrimidinic endonuclease 1 induced upregulation of fibroblast growth factor 2 and its receptor 3 induces angiogenesis in human osteosarcoma cells," *Cancer Science*, vol. 105, no. 2, pp. 186–194, 2014.
- [64] J. Kim, Y. Kang, Y. Kojima et al., "An endothelial apelin-FGF link mediated by miR-424 and miR-503 is disrupted in pulmonary arterial hypertension," *Nature Medicine*, vol. 19, no. 1, pp. 74–82, 2013.

## Research Article

# Quantitative Analysis of Pork and Chicken Products by Droplet Digital PCR

Yicun Cai, Xiang Li, Rong Lv, Jieli Yang, Jian Li, Yuping He, and Liangwen Pan

*Technical Center for Animal, Plant and Food Inspection and Quarantine, Shanghai Entry-Exit Inspection and Quarantine Bureau of China, No. 1208, Minsheng Road, Pudong New Area, Shanghai 200135, China*

Correspondence should be addressed to Liangwen Pan; [panlw@shciq.gov.cn](mailto:panlw@shciq.gov.cn)

Received 25 June 2014; Revised 15 August 2014; Accepted 15 August 2014; Published 27 August 2014

Academic Editor: Gerhard Litscher

Copyright © 2014 Yicun Cai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this project, a highly precise quantitative method based on the digital polymerase chain reaction (dPCR) technique was developed to determine the weight of pork and chicken in meat products. Real-time quantitative polymerase chain reaction (qPCR) is currently used for quantitative molecular analysis of the presence of species-specific DNAs in meat products. However, it is limited in amplification efficiency and relies on standard curves based Ct values, detecting and quantifying low copy number target DNA, as in some complex mixture meat products. By using the dPCR method, we find the relationships between the raw meat weight and DNA weight and between the DNA weight and DNA copy number were both close to linear. This enabled us to establish formulae to calculate the raw meat weight based on the DNA copy number. The accuracy and applicability of this method were tested and verified using samples of pork and chicken powder mixed in known proportions. Quantitative analysis indicated that dPCR is highly precise in quantifying pork and chicken in meat products and therefore has the potential to be used in routine analysis by government regulators and quality control departments of commercial food and feed enterprises.

## 1. Introduction

In 2013, the horse meat adulteration scandal [1, 2] swept across Europe, a region which is considered to hold the highest food safety standards worldwide. In this scandal, meat foods advertised as containing beef were found to contain undeclared meat, such as horse meat. In some cases, as much as 100% of the meat content was other undeclared meats. Due to the similar textures and processing technologies of different meat products, it is difficult for consumers to identify their authenticity, not to mention the precise proportion of meats within a given product. The temptation for huge profit and quick gain has made meat food adulteration a persistent problem where high quality meat products are replaced with cheaper and inferior counterparts; therefore, state or government quality inspection programs are responsible for ensuring the safety of meat products. Yet, reliable qualitative and quantitative detection methods need to be researched and developed to make this task easier.

Compared to conventional polymerase chain reaction (PCR) [3, 4] which has been widely used as a qualitative

method to detect whether species-specific target DNAs exist in food and feed samples, real-time polymerase chain reaction (qPCR) [5, 6] has the advantage not only to detect but also to quantify the relationship between the cycle threshold (Ct) values and the initial DNA template concentration. Yet, qPCR technology still has several issues [7, 8] that directly affect the accuracy of quantitative analysis, including the PCR amplification efficiency, the use of standard curves based on Ct values, the high background produced by nontarget DNA samples, and problems with the selection of a suitable reference material.

Digital PCR (dPCR) is a novel method for precise quantification of nucleic acids [9, 10], which utilizes a limiting dilution analysis and Poisson distribution analysis to enable the absolute quantification of target DNA copy number [11]. Digital PCR uses a similar amplification reaction system as a standard qPCR system. A droplet generator is used to partition each dPCR reaction mix into 20,000 nanoliter-sized droplets. Each droplet contains zero (negative), one, or more copies of the target DNA (positive). After a conventional PCR procedure, the total number of template-positive or

TABLE 1: Primer and probe sequences for quantitative dPCR assays.

Primer/probe	Sequence/labeling	GenBank accession number
Sus-ACTB-97bp-F	CGTAGGTGCACAGTAGGTCTGAC	Beta-Actin gene
Sus-ACTB-97bp-R	GGCCAGACTGGGGACATG	DQ452569
Sus-ACTB-97bp-P	VIC-CCAGGTCGGGGAGTC-MGB	
Gallus-TGFB3-129bp-F	GGCTGCAAGTCACCGTGGTA	TGFB3 gene
Gallus-TGFB3-129bp-R	CCGCTAGCCAGAAGCTCAGC	AY685072
Gallus-TGFB3-129bp-P	FAM-CAGGAGCCACGTGAGCAGCACAG-BHQ [18]	

-negative individual droplets is counted and recorded by a droplet reader. Finally, the original absolute target DNA copy number (copies/ $\mu\text{L}$ ) of the samples can be directly calculated by following the Poisson distribution law.

As a refinement of conventional qPCR, dPCR has the potential to allow for more accurate and sensitive measurement of the target DNA copy number, especially for low concentration samples, high background samples, and composite samples. As a result, dPCR has been applied in a wide range of areas including quantitative gene expression analysis [12], single nucleotide polymorphism analysis [13], genotyping [14], rare variant and copy number variation detection, pathogen detection [15], drug resistance research [16], and noncoding RNA research [17].

Here, we present a meat product weight measurement system based on dPCR technology for the accurate detection and quantification of specific nucleic acids in chicken or pork. Using this simple amplification and calculation procedure, the weight of meat in the sample can be accurately quantified. This is the first time that the dPCR technique has been utilized to quantify meat products.

## 2. Materials and Methods

**2.1. Meat Sample Preparation.** Fresh lean meat and commercially available products were obtained from the local supermarket. Fresh pork (*Sus scrofa*), chicken (*Gallus gallus*), and commercial products were separately minced, dried in a baking oven (UFE500AO; Memeert, Germany) at  $80^\circ\text{C}$  for 72 h, and then minced to a superfine powder in liquid nitrogen using a Freezer Mixer (6850 freezer/mill; SPEX SamplePrep, USA). The mixed samples of pork and chicken powers with known composition (from 90% to 10%) were used to verify the validity and sensitivity of the method. In order to guarantee that the extracted DNA accurately represents the proportion of different meats, the Freezer Mixer was used to grind the mixtures evenly to ensure complete mixing. The commercially available products were used to test the applicability of this method.

**2.2. DNA Extraction.** For all samples, genomic DNA was extracted from 100 mg powder using the phenol/chloroform method [19]. Briefly speaking, 800  $\mu\text{L}$  histiocyte lysis buffer (Tiangen, China) with 100  $\mu\text{g}$  proteinase K (Tiangen, China) was added to each sample, vortexed, and incubated at  $65^\circ\text{C}$  for 60 min with occasional vigorous shaking; an equal volume of phenol/chloroform was added, mixed, and centrifuged at

12000 rpm for 10 min. The aqueous (upper) layer was transferred to a clean tube; an equal volume of chloroform was added, mixed, and then centrifuged for 5 min at 12000 rpm. The aqueous (upper) layer was transferred to a clean tube; a one-tenth volume of 3 M Na acetate (pH 5.2) and two volumes of ice-cold EtOH (100%) were added, mixed, and incubated at  $-20^\circ\text{C}$  for 30 min and then centrifuged at 12000 rpm for 30 min at  $4^\circ\text{C}$ . The supernatant was removed and the DNA pellet was washed twice with 75% EtOH and centrifuged at 12000 rpm for 2 min at  $4^\circ\text{C}$ ; the supernatant was removed and the pellet was air-dried for 30 min at room temperature, resuspended in 100  $\mu\text{L}$  ddH<sub>2</sub>O, and stored at  $-20^\circ\text{C}$ .

**2.3. Primers and Probes.** To detect pork and chicken, the *Sus Scrofa* beta-actin (*ACTB*) gene (GenBank accession number: DQ452569) [20] and *Gallus gallus* transforming growth factor beta-3 (*TGFB3*) gene (GenBank accession number: AY685072) [18] were selected as the target detection sequences, respectively, as previously described [21–23]. Online tools supported by NCBI were used for sequence search and alignment. Primers and probes were designed using Primer Express Software version 3.0 supported by Applied Biosystems (ABI, Foster City, CA, USA). All of the selected primers and probes passed a specificity and homology evaluation by BLAST searches against the entire GenBank database. The nucleotide sequences of the primers and probes used in this study were designed to meet optimal conditions for dPCR [24]. A probe labeled with the fluorophore VIC (ABI, Foster City, CA, USA) and minor groove binder (MGB) quencher was used to detect pork. The FAM fluorophore and Black Hole Quencher (BHQ) were used to detect chicken (Table 1).

**2.4. Specificity.** In order to verify the specificity of the dPCR system (including the primers and probes), DNA from a wide range of animal samples was isolated and tested by dPCR system.

**2.5. Digital PCR Procedure.** Each 20  $\mu\text{L}$  reaction mixture was prepared as follows: 1.8  $\mu\text{L}$  of each primer (final concentration, 900 nM), 0.5  $\mu\text{L}$  probe (final concentration, 250 nM), and 10  $\mu\text{L}$  ddPCR Master Mix (Bio-Rad, Hercules, CA, USA) were mixed, and then 4  $\mu\text{L}$  (40-fold diluted from the original DNA extraction sample) of template DNA and 1.9  $\mu\text{L}$  of nuclease- and protease-free water (ThermoScientific, Salt Lake City, UT, USA) were added. A Bio-Rad QX100 ddPCR droplet generator (Bio-Rad) was used to divide the 20  $\mu\text{L}$

mixture into approximately 20000 droplets, with the target DNA segments and PCR reagents being randomly distributed among the droplets. Conventional PCR was performed using a T100 Thermal Cycler (Bio-Rad) according to the following cycling protocol: enzyme activation for 10 min at 95°C, followed by 40 cycles of 30 sec denaturation at 94°C; 1 min annealing and extension at 60°C, followed by enzyme inactivation at 98°C for 10 min and hold at 4°C (according to the manufacturer's instructions). After PCR amplification, the droplet reader determines which droplets contain the target DNA amplicon and which do not. The software then calculates the concentration of the target DNA in copies per microliter from the fraction of positive reactions using Poisson distribution analysis.

**2.6. Standard Curve Generation.** A series of meat powders were accurately weighed using a precision electronic balance (BSA224s; Sartorius, Germany). Genomic DNA was extracted and the DNA concentration of each sample was measured using a NanoVue spectrophotometer (GE Healthcare, Little Chalfont, Buckinghamshire, UK). Ten different samples of pork and chicken powder (equally distributed in weight from 10 mg to 100 mg, three replicates per weight) and a nontemplate control (NTC) were analyzed by dPCR. The correlation coefficient of ( $R^2$ ) for the weight of the meat powder and the DNA concentration was calculated using Excel (Microsoft Office 2007; Redmond, WA, USA).

### 3. Results and Discussion

**3.1. Specificity.** We chose single copy nuclear genes that are expressed at relatively stable levels in different cell types as detection targets [25]. First, BlastN searches (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) of the entire NCBI genome database were used to validate the specificity of the primers, probes, and PCR amplicons; all of the pork- and chicken-specific PCR primers, probes, and PCR amplicons bore a high level of species specificity. In practice, a broad range of DNA samples from different animals were isolated and tested as templates for the PCR procedure: cattle (*Bos taurus*), donkey (*Equus asinus*), sheep (*Ovis aries*), goat (*Capra hircus*), horse (*Equus caballus*), elk (*Cervus canadensis*), buffalo (*Bubalis bubalus*), rabbit (*Oryctolagus cuniculus*), duck (*Anas platyrhynchos*), goose (*Anser domesticus*), turkey (*Meleagris gallopavo*), ostrich (*Struthio camelus*), pigeon (*Columba livia*), quail (*Coturnix coturnix*), pheasant (*Phasianus colchicus*), rhesus monkey (*Macaca mulatta*), mice (*Mus musculus*), rat (*Rattus norvegicus*), goldfish (*Carassius auratus*), carp (*Cyprinus carpio*), and trout (*Oncorhynchus mykiss*). Cross amplification from other species was not observed for any primer/probe combination. Therefore, the possibility of cross amplification was excluded from a theoretical and practical perspective.

**3.2. DNA Extraction Efficiency.** Due to the varying and complex definition of meat (fat, skin, internal organs, and so on) in food and feed products, we used fresh lean meat (chicken breast and pork loin) as the standard specimens to extract nucleic acids in this experiment to help minimize

the effect of variation in the quality of the raw meat. In order to establish the relationship between the weight of meat powder (mg) and the corresponding amount of nucleic acid (ng), DNA was extracted from each meat powder sample by proteinase K digestion, phenol/chloroform extraction, and EtOH (100%) precipitation. Ten different weights of pork and chicken reference samples (equally distributed over the range from 10 mg to 100 mg, three replicates per weight) and a NTC sample were extracted. The concentration of each DNA sample was measured using a NanoVue spectrophotometer. In three independent experiments, a linear relationship was observed between the raw meat weight (mg) and the corresponding amount of extracted nucleic acid (ng). The correlation coefficient ( $R^2$ ) was 0.999 for chicken (Figure 1(a)) and 0.998 for pork (Figure 1(b)). These findings indicate that, within the range between 10 mg to 100 mg raw meat powder, the amount of genomic DNA extracted had an approximately linear relationship with the weight of both types of raw meat.

**3.3. Specific Target DNA Detection by dPCR.** In order to explore whether a linear relationship exists between the weight of nucleic acid and the species-specific target DNA copy number, the serially diluted meat DNA samples and a NTC sample were analyzed by dPCR. The dPCR assays were performed on chicken samples containing 40 ng to 320 ng DNA and pork samples containing 80 ng to 800 ng DNA. The maximum concentrations were determined by the detection limit of the dPCR instrument. Each data point was collected based on three replicates per sample in three independent experiments. During the dPCR process, at least 15000 droplets were obtained for each reaction, in compliance with the requirements for absolute quantification. The correlation coefficients ( $R^2$ ) for the nucleic acid weight (ng) and the chicken- or pork-specific DNA copy number were 0.997 and 0.995, respectively (Figures 2(a) and 2(b)). Results from these experiments indicate that, within the range of 40 ng to 320 ng for chicken (Figure 2(a)) and 80 ng to 800 ng for pork (Figure 2(b)), relationships between the nucleic acid weight and specific target DNA copy number were approximately linear. In this step, we found that the width of the linear dynamic range is not more than five orders of magnitude. Compared with the qPCR, dPCR offers a narrow dynamic range as described in the previous articles [24, 26]. But after appropriate dilution, the dynamic range will cover the whole range of quantification needed.

**3.4. Establishment of Quantitative Formulae.** Results from the experiments confirmed two linear relationships: one between the raw meat weight and nucleic acid weight and the other between the nucleic acid weight and specific target DNA copy number. These correlations were essential for establishing the formulae to calculate the raw meat weight. We utilized the nucleic acid weight as an intermediate value to establish the following formulae for calculating the original raw meat weight from the specific DNA copy number: chicken,  $M_{\text{chicken}} = 0.04C - 4$ , and pork,  $M_{\text{pork}} = 0.2C + 2.5$ , where  $C$  is the copy number (copies/ $\mu\text{L}$ ) and  $M$  the raw meat weight (mg).

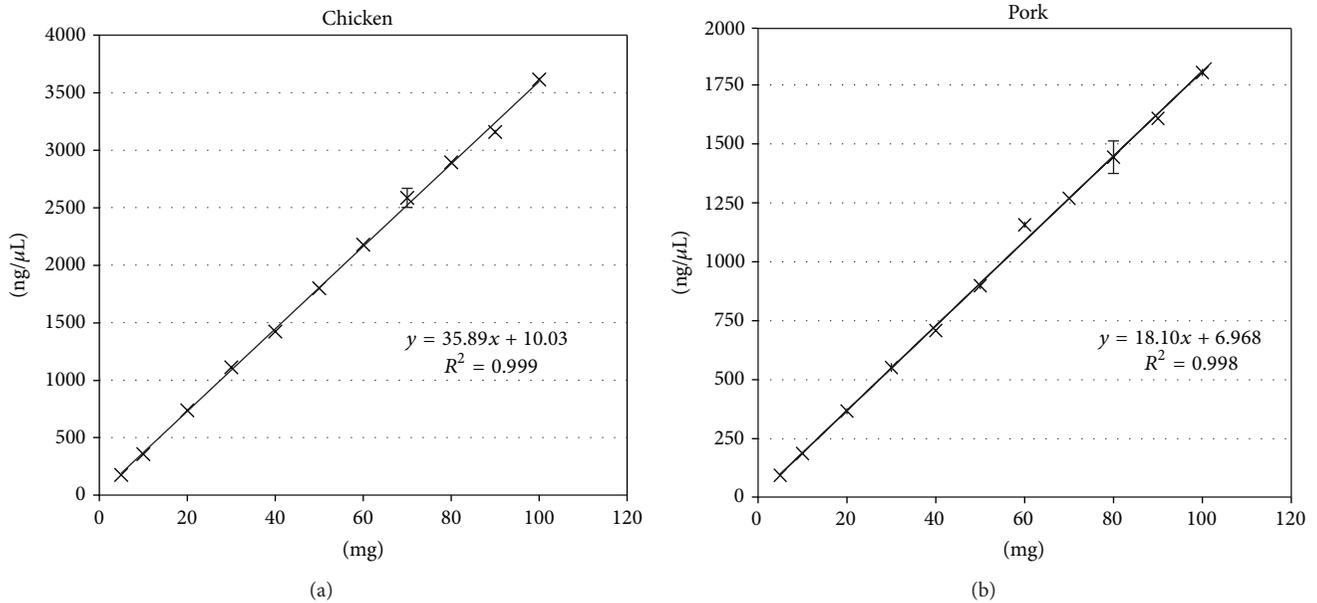


FIGURE 1: Linear relationship between meat quantity (mg) and nucleic acid (ng) content. The efficiency of extracting genomic DNA from chicken and pork was confirmed within the dynamic range. After accurate weighing and DNA extraction, the nucleic acid (ng) content of three replicates for each sample was measured using a NanoVue spectrophotometer. The correlation coefficient ( $R^2$ ) for the initial sample weight (mg) and nucleic acid (ng) content was 0.999 for chicken and 0.998 for pork.

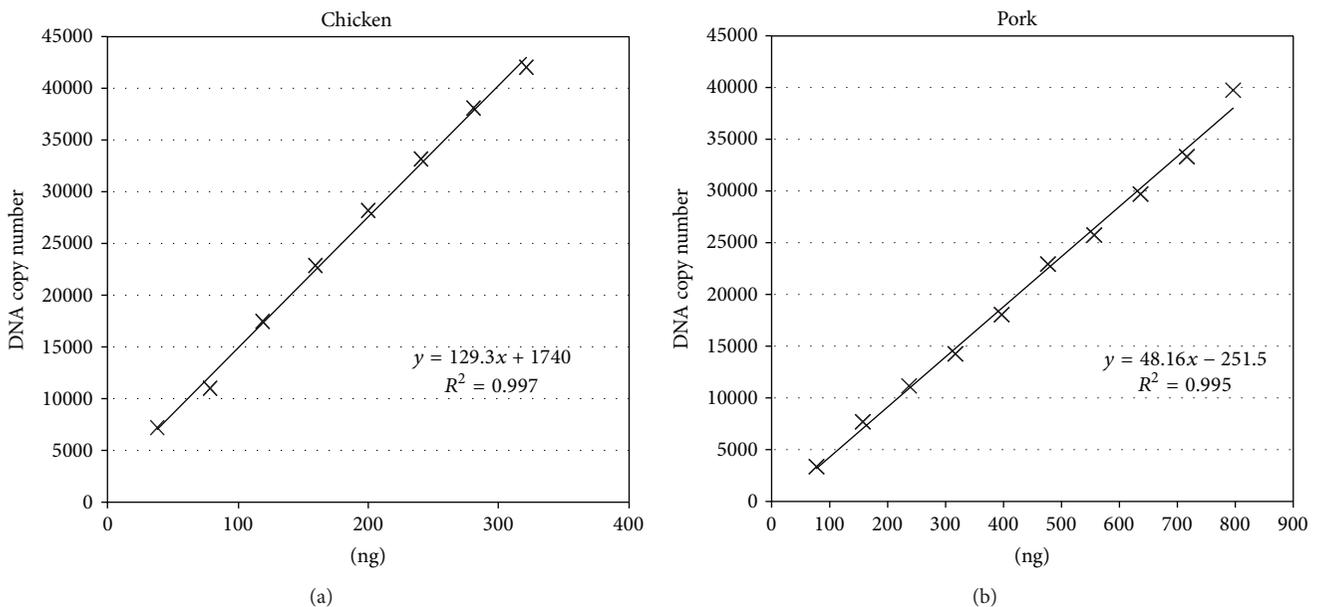


FIGURE 2: Linear relationship between nucleic acid content (ng) and the target DNA copy number. A linear relationship between the quantity of nucleic acid and the target DNA copy number was confirmed by dPCR. DNA samples of known concentration were analyzed by dPCR. Each detection point is the average of triplicate samples from three independent experiments. All experimental data met the quality requirements for dPCR. The correlation coefficient ( $R^2$ ) for the DNA quantity and the DNA copy number was 0.997 for chicken and 0.995 for pork.

3.5. *Analysis of Samples of Known Concentration.* Mixed meat products often appear in food products and various industrial applications. However, the DNA extraction process could be affected by numerous factors, such as the tissue composition, sample treatment, DNA degradation, and even pipetting errors. Therefore, the species-specific DNA may not

truly represent the actual weight proportion of meat(s) in the product. In order to further demonstrate the overall accuracy and applicability of our method, nine mixed meat samples of known composition were used to verify the quantification method. DNA was extracted in duplicate from each mixed meat sample using the same method. Each DNA sample

TABLE 2: The results of quantification of the samples with known concentrations.

	Pork true (mg)	Pork measure (mg)	Pork deviation	Chicken true (mg)	Chicken measure (mg)	Chicken deviation
Sample 1	90	97.50	8.33%	10	8.72	-12.80%
Sample 2	85	84.43	-0.67%	15	16.85	12.36%
Sample 3	70	73.63	5.19%	30	28.59	-4.71%
Sample 4	65	63.50	-2.31%	35	32.33	-7.62%
Sample 5	50	52.37	4.73%	50	49.00	-2.00%
Sample 6	45	46.90	4.22%	55	57.53	4.61%
Sample 7	30	28.83	-3.89%	70	64.25	-8.21%
Sample 8	25	22.30	-10.80%	75	76.81	2.42%
Sample 9	10	11.70	17.00%	90	94.11	4.56%

was diluted 40-fold and 4  $\mu$ L of each sample was analyzed in triplicate in the same dPCR experiment; the data was expressed as the average value. The original meat weight was calculated using the two formulae above. Examined through three different independent experiments, the dPCR technology had a high consistency and reproducibility. More importantly, the final quantitative results for the mixed pork and chicken samples were similar to the true raw meat weights (Table 2). Compared to the qPCR quantification method [18], most of the measured meat weights in this study had a low level of deviation, indicating that the dPCR assay is highly accurate. The variations in the values obtained using dPCR may be due to the heterogeneity of the raw meat samples or artificial operator errors. Therefore, the sample processing steps of dry powder generation, DNA extraction, and sampling uniformity are critically important to the test procedure. Complete drying and full grinding of the test samples are the most basic requirements for accurate quantification.

**3.6. Analysis of Commercial Samples.** A total of 11 commercially available products (Table 3) were collected and analyzed by dPCR quantification system to determine the proportion of the pork or chicken. The result of the experiment validates that the system has good practicability.

## 4. Conclusions

Through a well-designed experiment, this study has demonstrated that the dPCR technique can be used to accurately quantify the weight of specific meats in meat products. A total of 11 commercially available meat products were used to prove the applicability of this system. Some influential factors during the quantification procedure were taken into account. We chose fresh lean meat to guarantee the consistency of DNA content and chose stable (relatively) expression DNA sequence as the detection target, so the quantification accuracy will be guaranteed. Satisfyingly, the relationships between the raw meat weight and DNA weight and between DNA weight and DNA copy number were both close to linear for both pork and chicken. This enabled us to establish formulae to calculate the raw meat weight based on the DNA copy number. This technique has the potential to provide a

TABLE 3: Samples from the local supermarket were analyzed.

Sample	Chicken (%)	Pork (%)
Chicken ham sausage	3.8	0.0
Pork ham sausage	0.0	2.6
Beef ham sausage	0.0	0.0
Minced chicken	48.6	0.0
Minced pork	0.0	35.1
Chicken vegetable dog food	25.1	0.0
Pork spam	0.0	3.2
Canned stewed pork	0.0	41.6
Beef vegetable dog food	0.0	0.0
Fish cat food	0.0	0.0
Chicken flavor	0.0	0.0

convenient way to quantify the meat content of foods or feed. No optimization steps were required during the course of the experiment. The primers and probes from routine qPCR systems can be used directly in a dPCR quantification system. We investigated the accuracy of this technique and found that dPCR could achieve a linear dynamic range for absolute quantification of DNA. In each dPCR reaction, nearly 15000 effective droplets were generated, detected, and analyzed, which ensured the accuracy of the quantification method. Additionally, calculations involved in dPCR are based on absolute data instead of relative data (such as Ct values) and do not require standard curves or reference materials, which improve the accuracy of quantification. This work is the first to demonstrate how to apply the dPCR technology to quantify pork and chicken in meat products. Suitable for routine analysis, this method has the potential to be adapted to quantifying meat of various species. However, a number of technical flaws, such as narrow dynamic range and time consuming nature of the assay, remain to be solved before dPCR can be widely adopted for the routine quantification of the meat content in food and feed products.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## References

- [1] P. J. O'Mahony, "Finding horse meat in beef products—a global problem," *QJM*, vol. 106, no. 6, pp. 595–597, 2013.
- [2] J. Y. Nau, "Horse meat: first lessons of a scandal," *Revue Medicale Suisse*, vol. 9, no. 376, pp. 532–533, 2013.
- [3] N. S. Karabasanavar, S. P. Singh, D. Kumar, and S. N. Shebanavar, "Detection of pork adulteration by highly-specific PCR assay of mitochondrial D-loop," *Food Chemistry*, vol. 145, pp. 530–534, 2014.
- [4] J. Demarquoy, "Meat adulteration: the use of PCR," *Meat Science*, vol. 94, no. 2, 164 pages, 2013.
- [5] M. E. Ali, U. Hashim, S. Mustafa et al., "Analysis of pork adulteration in commercial meatballs targeting porcine-specific mitochondrial cytochrome b gene by TaqMan probe real-time polymerase chain reaction," *Meat Science*, vol. 91, no. 4, pp. 454–459, 2012.
- [6] J. H. Calvo, R. Osta, and P. Zaragoza, "Quantitative PCR detection of pork in raw and heated ground beef and pâté," *Journal of Agricultural and Food Chemistry*, vol. 50, no. 19, pp. 5265–5267, 2002.
- [7] P. Jothikumar, V. Hill, and J. Narayanan, "Design of FRET-TaqMan probes for multiplex real-time PCR using an internal positive control," *BioTechniques*, vol. 46, no. 7, pp. 519–524, 2009.
- [8] M. Yoshimura, S. Nakamura, and H. Ogawa, "TaqMan real-time PCR quantification: conventional and modified methods," *Methods in molecular medicine*, vol. 108, pp. 189–198, 2005.
- [9] C. M. Hindson, J. R. Chevillet, H. A. Briggs et al., "Absolute quantification by droplet digital PCR versus analog real-time PCR," *Nature Methods*, vol. 10, no. 10, pp. 1003–1005, 2013.
- [10] R. Sanders, J. F. Huggett, C. A. Bushell, S. Cowen, D. J. Scott, and C. A. Foy, "Evaluation of digital PCR for absolute DNA quantification," *Analytical Chemistry*, vol. 83, no. 17, pp. 6474–6484, 2011.
- [11] L. B. Pinheiro, V. A. Coleman, C. M. Hindson et al., "Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification," *Analytical Chemistry*, vol. 84, no. 2, pp. 1003–1011, 2012.
- [12] M. Massanella, A. Singhanian, N. Beliakova-Bethell et al., "Differential gene expression in HIV-infected individuals following ART," *Antiviral Research*, vol. 100, no. 2, pp. 420–428, 2013.
- [13] S. Bhat, A. M. Polanowski, M. C. Double, S. N. Jarman, and K. R. Emslie, "The effect of input DNA copy number on genotype call and characterising SNP markers in the humpback whale genome using a nanofluidic array," *PLoS ONE*, vol. 7, no. 6, Article ID e39181, 2012.
- [14] N. B. Tsui, C. A. Hyland, G. J. Gardener et al., "Noninvasive fetal RHD genotyping by microfluidics digital PCR using maternal plasma from two alloimmunized women with the variant RHD(IVS3+IG>A) allele," *Prenatal Diagnosis*, vol. 33, no. 12, pp. 1214–1216, 2013.
- [15] G. S. Brunetto, R. Massoud, E. C. Leibovitch et al., "Digital droplet PCR (ddPCR) for the precise quantification of human T-lymphotropic virus 1 proviral loads in peripheral blood and cerebrospinal fluid of HAM/TSP patients and identification of viral mutations," *Journal of Neurovirology*, vol. 20, no. 4, pp. 341–351, 2014.
- [16] S. Pholwat, S. Stroup, S. Foongladda, and E. Houpt, "Digital PCR to detect and quantify heteroresistance in drug resistant *Mycobacterium tuberculosis*," *PLoS ONE*, vol. 8, no. 2, Article ID e57238, 2013.
- [17] D. W. Dodd, K. T. Gagnon, and D. R. Corey, "Digital quantitation of potential therapeutic target RNAs," *Nucleic Acid Therapeutics*, vol. 23, no. 3, pp. 188–194, 2013.
- [18] R. Köppel, M. Daniels, N. Felderer, and C. Brünen-Nieweler, "Multiplex real-time PCR for the detection and quantification of DNA from duck, goose, chicken, turkey and pork," *European Food Research and Technology*, vol. 236, no. 6, pp. 1093–1098, 2013.
- [19] R. M. Santella, "Approaches to DNA/RNA extraction and whole genome amplification," *Cancer Epidemiology Biomarkers and Prevention*, vol. 15, no. 9, pp. 1585–1587, 2006.
- [20] C. L. Lin, D. G. J. Jennen, S. Ponsuksili et al., "Haplotype analysis of  $\beta$ -actin gene for its association with sperm quality and boar fertility," *Journal of Animal Breeding and Genetics*, vol. 123, no. 6, pp. 384–388, 2006.
- [21] B. Rkjrfa, "Multiplex real-time PCR for the detection and quantification of DNA from beef, pork, chicken and turkey," *European Food Research and Technology*, vol. 227, pp. 1199–1203, 2008.
- [22] M. A. Rodríguez, T. García, I. González et al., "Identification of goose, mule duck, chicken, turkey, and swine in foie gras by species-specific polymerase chain reaction," *Journal of Agricultural and Food Chemistry*, vol. 51, no. 6, pp. 1524–1529, 2003.
- [23] M. A. Rodríguez, T. García, I. González, L. Asensio, P. E. Hernández, and R. Martín, "Quantitation of mule duck in goose foie gras using TaqMan real-time Polymerase Chain Reaction," *Journal of Agricultural and Food Chemistry*, vol. 52, no. 6, pp. 1478–1483, 2004.
- [24] D. Morisset, D. Štebih, M. Milavec, K. Gruden, and J. Žel, "Quantitative analysis of food and feed samples with droplet digital PCR," *PLoS ONE*, vol. 8, no. 5, Article ID e62583, 2013.
- [25] C. A. Heid, J. Stevens, K. J. Livak, and P. M. Williams, "Real time quantitative PCR," *Genome Research*, vol. 6, no. 10, pp. 986–994, 1996.
- [26] M. Baker, "Digital PCR hits its stride," *Nature Methods*, vol. 9, no. 6, pp. 541–544, 2012.

## Research Article

# Methylation-Associated Gene Silencing of *RARB* in Areca Carcinogens Induced Mouse Oral Squamous Cell Carcinoma

Zi-Lun Lai,<sup>1</sup> Yung-An Tsou,<sup>1,2,3</sup> Shin-Ru Fan,<sup>4</sup> Ming-Hsui Tsai,<sup>2,3</sup> Hsiao-Ling Chen,<sup>5</sup> Nai-Wen Chang,<sup>6</sup> Ju-Chien Cheng,<sup>4</sup> and Chuan-Mu Chen<sup>1,7</sup>

<sup>1</sup> Department of Life Sciences, Agricultural Biotechnology Center, National Chung Hsing University, No. 250 Kao-Kuang Road, Taichung 402, Taiwan

<sup>2</sup> Department of Otolaryngology Head and Neck Surgery, China Medical University, Taichung 404, Taiwan

<sup>3</sup> School of Medicine, College of Medicine, China Medical University, Taichung 404, Taiwan

<sup>4</sup> Department of Medical Laboratory Science and Biotechnology, China Medical University, No. 91 Hsueh-Shih Road, Taichung 404, Taiwan

<sup>5</sup> Department of Bioresources, Da-Yeh University, Changhwa 515, Taiwan

<sup>6</sup> Department of Biochemistry, College of Medicine, China Medical University, Taichung 404, Taiwan

<sup>7</sup> Rong Hsing Research Center for Translational Medicine, iEGG Center, National Chung Hsing University, Taichung 402, Taiwan

Correspondence should be addressed to Ju-Chien Cheng; [jccheng@mail.cmu.edu.tw](mailto:jccheng@mail.cmu.edu.tw) and Chuan-Mu Chen; [cmchen7010978@gmail.com](mailto:cmchen7010978@gmail.com)

Received 7 May 2014; Revised 8 June 2014; Accepted 10 June 2014; Published 17 August 2014

Academic Editor: Calvin Yu-Chian Chen

Copyright © 2014 Zi-Lun Lai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Regarding oral squamous cell carcinoma (OSCC) development, chewing areca is known to be a strong risk factor in many Asian cultures. Therefore, we established an OSCC induced mouse model by 4-nitroquinoline-1-oxide (4-NQO), or arecoline, or both treatments, respectively. These are the main two components of the areca nut that could increase the occurrence of OSCC. We examined the effects with the noncommercial MCGI (mouse CpG islands) microarray for genome-wide screening the DNA methylation aberrant in induced OSCC mice. The microarray results showed 34 hypermethylated genes in 4-NQO plus arecoline induced OSCC mice tongue tissues. The examinations also used methylation-specific polymerase chain reaction (MS-PCR) and bisulfite sequencing to realize the methylation pattern in collected mouse tongue tissues and human OSCC cell lines of different grades, respectively. These results showed that retinoic acid receptor  $\beta$  (*RARB*) was indicated in hypermethylation at the promoter region and the loss of expression during cancer development. According to the results of real-time PCR, it was shown that *de novo* DNA methyltransferases were involved in gene epigenetic alternations of OSCC. Collectively, our results showed that *RARB* hypermethylation was involved in the areca-associated oral carcinogenesis.

## 1. Introduction

Throughout the world, oral squamous cell carcinoma (OSCC) is one of the most common types of cancers. It has a high cure rate for small primary tumors and involves the development of second primary tumors and the long-term survival rate is <60% [1]. Furthermore, in Taiwan, according to statistics from the Department of Health, Executive Yuan, Taiwan, OSCC ranks as fourth among the ten leading causes of cancer among males and is the fourth leading cancer in the male population and the number of deaths increases every year [2].

The main risk factor for developing OSCC is chewing areca, especially in many Asian cultures.

In a clinical study, the incidence of oral cancer was elevated 28 times for betel quid users as compared to nonbetel quid users [3]. Cigarette smoking has synergistic effect with areca chewing, and such users have an 89 times higher incidence rate than nonusers. If one has the habit of drinking, smoking, and betel quid chewing combined, there will be a 123 times higher incidence rate of having oral cancer than those average individuals in the general population that are nonusers. There is the longitudinal cohort study

on the alcohol, betel quid, and smoking, to the oral cancer risk. The betel quid partook the significant higher hazard risk to the oral carcinogenesis [4]. The most tumorigenic part of betel quid is the *Piper longum* L. and the calcium hydroxide (slaked lime) which will cause the oral cavity to develop into an alkaline condition which will promote the tumorigenic effect of the Safrole in the *Piper longum* L. The fibers of areca also cause oral mucosa damage and increased mucosa to be exposed to the tumorigenic material in the betel quid. In a clinical survey, oral cancer with areca chewing had far more incidence of oral submucosa fibrosis and erythroplakia; furthermore, the pathologic findings also show severe hyperkeratosis, caries, and gingivitis as compared to nonareca users. The oral cancer patients who had the habit of betel quid chewing were also found to have a higher percentage of dysplastic change surrounding the tumor margin, and skip cancer lesion was frequently noted in the upper aerodigestive tract (tongue, hypopharynx, and esophagus). The condemned mucosa even reached the esophagus. In an average clinical survey, 18% of cases were found to have esophagus cancer diagnosed at the same time when oral cancer presented [5]. Synchronous double cancer (second primary cancer) in the upper aerodigestive tract is frequently noted [6].

There was a stimulating effect when areca nut is chewed along with betel leaf [7]. Furthermore, in Chiang et al. [8], they used areca nut extract (ANE) and saliva-reacted ANE (sANE) to treat three oral carcinoma cell lines, KB (epidermoid carcinoma), SAS (tongue carcinoma), and Ca9-22 (gingival carcinoma). The higher cytotoxic effects involving cell morphologic changes and upregulation of inflammatory signaling in mRNA expression levels were observed in these treatments. In addition, arecoline is the major alkaloid in areca nut extracts and betel quid. It is the primary active ingredient responsible for the central nervous system stimulation that is roughly comparable to that of nicotine, which has a similar chemical structure [9–12]. There is also another carcinogen, 4-nitroquinoline 1-oxide (4-NQO), which effectively induces oral and esophageal cancers that closely resemble early human lesions in mice and rats [13, 14]. In this study, we followed Chang et al. [15] who established an effective mouse model of oral cancer and used this model to identify potential markers of oral tumor progression by utilizing a noncommercial methylation microarray.

The promoter hypermethylation now has a key role for research in the area of human multistage carcinogenesis. Silencing of certain tumor suppressor genes may occur in the absence of genetic change, via aberrant methylation of CpG islands [16–18]. OSCC is believed to arise through the accumulation of numerous genetic and epigenetic alterations [19–21]. There are several methods to determine whether promoter methylation has been developed including combination of bisulfite restriction assay (COBRA) [22], genomic bisulfate sequencing [23], methylation-specific PCR (MS-PCR) [24], and microarray-based methylation analysis [25]. Methylation microarray is a high throughput tool for genome-wide methylation analysis [26–31]. To identify and characterize potential targets for treating oral cancer, a genome-wide approach was taken to quantitatively measure

genomic alterations in OSCC [32, 33]. Consequently, we used home-made mouse CpG island microarray to understand aberrant methylation profile during OSCC tumorigenesis in this study and validated the methylation status by MS-PCR, bisulfite sequencing, and real-time PCR.

In many previous studies, retinoid acid suppresses carcinogenesis and inhibits the growth of human head and neck squamous cell carcinoma (HNSCC) [34, 35]. Loss of retinoids and their receptors has been associated with malignant progression in HNSCC [36]. Their receptors (RAR) are central regulators to the normal growth and differentiation of a variety of epithelial cells. RAR changes have been associated with cell immortalization, and re-expression of *RAR-beta* (*RARB*) leads to growth inhibition in some circumstances [19]. Loss of *RARB* expression is associated with a change in proliferative life span potential from mortality to immortality in HNSCC [37–39]. The promoter hypermethylation of *RARB* could inhibit the gene expression when added to the methylation inhibitor and deacetylation inhibitor like 5'-aza-2'-deoxycytidine (5-aza-dC) and trichostatin A (TSA) which could recover the gene expression and inhibit tumor cells growth [36, 39].

In the present study, we investigated the role of *RARB* hypermethylation of CpG islands in OSCC mouse model and its association with *RARB* expression in human oral cancer cell lines. In addition, we examined whether the repression of *RARB* transcription could be reversed by 5-aza-dC in human oral cancer cell lines, and finally, we evaluated the three main DNA methyltransferases that were involved in *RARB* hypermethylation.

## 2. Materials and Methods

**2.1. Mouse Model for Oral Cancer.** The mouse model development was modified as highlighted by Chang et al. [15, 40]. Briefly, the OSCC model was established by treating arecoline (Sigma, St. Louis, MO), as well as in combination with 4-NQO (Fluka, St. Louis, MO) in 4-5 week age old of C57BL/6JNarl male mice. The conditions for OSCC formation are 500  $\mu\text{g}/\text{mL}$  arecoline (A), 200  $\mu\text{g}/\text{mL}$  4-NQO (N), and 4-NQO (200  $\mu\text{g}/\text{mL}$ ) combined with arecoline (500  $\mu\text{g}/\text{mL}$ ) (NA) in the drinking water for 8 weeks. The drinking water was changed every day, and mice were allowed access to the drinking water at all times while receiving treatment. After the treatment, the drinking water was changed to ddH<sub>2</sub>O and mice were sacrificed at week 8, 12, 14, 18, 20, 26, and 28, respectively. The tongues were collected and classified into tumor parts (T) and nontumor parts (NT) for mouse CpG island microarray analysis.

**2.2. Cell Culture and 5-aza-dC Treatment.** Normal human oral keratinocytes (NHOK) were cultured in Keratinocyte Growth Medium (KGM, GIBCO, CA, USA). Oral cancer cell lines, DOK, OC2, and Ca9-22, were cultured in DMEM (GIBCO, CA, USA). HSC3 and TW2.6 were cultured in DMEM-F12 (GIBCO, CA, USA). All cells supplemented with 10% fetal calf serum and 1% penicillin-streptomycin and cultured at 37°C with 5% CO<sub>2</sub>. Oral cancer cells were treated

TABLE 1: Primer sets used for RT-PCR, MS-PCR, and bisulfite sequencing.

Primer sets	Sense primer (5' → 3')	Antisense primer (5' → 3')	Tm (°C)	PCR size (bp)
RARB-Hu-RT	AGGAGACTTCGAAGCAAG	GTCAAGGGTTCATGTCCTTC	60	771
<i>Dnmt1</i> -Hu-RT	TACCTGGACGACCCTGACCTC	CGTTGGCATCAAAGATGGACA	60	102
<i>Dnmt3a</i> -Hu-RT	TATGTATGAGCGCACAAAGAGAGC	GGGTGTTCCACCCTAACATTGAG	64	110
<i>Dnmt3b</i> -Hu-RT	GGCAAGTTCTCCGAGGTCTCTG	TGGTACATGGCTTTTCGATAGGA	62	112
RARB-Mo-M	GGATTAGAGTTTTTCGTGCGTCCG	TACCCCGCCGATACCCAAACG	65	90
RARB-Mo-U	GGATTAGAGTTTTTGTGTGTTG	TACCCCAACCAATACCCAAACA	62	90
RARB-Mo-BS	CCACCCAACCTCCATCAAACCTC	CCATACAATCAAACATAATCTC	58	476
RARB-Hu-M	ATGTCGAGAACGCGAGCGATTC	CTCGACCAATCCAACCGAAACG	64	151
RARB-Hu-U	GGATGTTGAGAATGTGAGTGATTT	TACTCAACCAATCCAACCAAAACA	62	155
RARB-Hu-BS	GTGTGATAGAAGTAGTAGGAAG	GTGATAGAAGTGGTAGGAAG	55	401

\* Hu: human, Mo: mouse, RT: real-time RT-PCR, M: methylated set, U: unmethylated set, BS: bisulfite sequencing.

by 5-aza-dC at 2  $\mu$ M to reverse the methylation status as described in [36, 41].

**2.3. DNA and RNA Extraction.** The genomic DNA extraction was conducted as noted in our previous report [42]. We used collected nontumor parts (NT) and tumor parts (T) of tongues for DNA and RNA extractions.

**2.4. RT-PCR and Real-Time PCR.** Total RNA was prepared using the TRI REAGENT (Invitrogen, CA, USA). One microgram of total RNA was treated with 10 units of RQ1 RNase-Free DNase (Promega, WI, USA) and extracted with phenol-chloroform. DNase that treated total RNA (1  $\mu$ g) was reverse transcribed with the ImProm II Reverse Transcription System (Promega, WI, USA). For RT-PCR amplification was performed with 2720 thermal cycler (Applied Biosystems Inc., CA, USA) and Real-Time PCR amplification was performed with Rotor-Gene 6000 (Corbett, CA, USA). The amplification of RT-PCR was repeated for 28 cycles as follows: 95°C, 30 sec for denature of the annealing temperature depending on the pair of gene specific primer sets (Table 1) for 30 sec and 72°C, 30 sec for extension. PCR reactions were performed in triplicate and the transcription level was normalized with the *GAPDH*. For Real-Time PCR, the calculated gene expression fold from CT value was performed according to the previously mentioned study, with a *P* value of less than 0.05 exhibiting an obviously significant difference.

**2.5. Preparation of Mouse CpG Island Microarray and Amplification Generation.** The mouse CpG island microarray was based on previously described human CpG island microarrays [43–46]. A total 2,304 mouse CpG islands library (mCGI) clones were spotted on UltraGAPS Coated Slides (Corning, MA, USA) by the BioDot AD1500 (BIODOT, CA, USA). The amplicons for methylation analysis were prepared as previously described [47, 48].

**2.6. Microarray Hybridization and Data Analysis.** The purified amplicons (5  $\mu$ g) were conducted using the BioPrime DNA labeling system (Invitrogen, CA, USA). Cyanine 5-ddUTP (Cy5-ddUTP) and Cyanine 3-ddUTP (Cy3-ddUTP)

(Perkin-Elmer Life Sciences, NJ, USA) fluorescent dyes were coupled to tumor (T) and normal (NT) amplicons, respectively, and cohybridized to the microarray panel. The combined tumor/normal control pair, with 8  $\mu$ g DNA, more than 180 pmol Cy5, and 150 pmol Cy3, would give strong hybridization signals. The hybridization of 4,608 spots is carried out under a 24 × 50 mm cover glass sealed tightly within a moistened hybridization chamber, GeneMachines HybChambers (Genomic Solutions, MI, USA), in a 65°C water bath from 12 to 16 h. The posthybridization washing steps are essentially those described by UltraGAPS Coated Slides instruction Manual. The hybridized slides were scanned with the GenePix 4000B scanner (Axon, CA, USA) and the acquired images were analyzed with the software GenePix Pro 4.0 (Axon, CA, USA). The microarray data was analyzed as described previously [43–46, 48, 49]. Briefly, the Cy5/Cy3 ratio and the hybridization intensity from the tumor amplicons to the hybridization intensity from the normal amplicons, from each image, are normally guided by both the average global Cy5/Cy3 ratio from each image and the Cy5/Cy3 ratios from 9 internal controls (clones without restriction cutting sites whose copy numbers remain the same in tumors and normal samples). Yellow spots (normalized Cy5/Cy3 = 1) represent equal amounts of bound DNA from each amplicon, indicating no methylation differences between tumor (T) and nontumor (NT) genomes. The analyzed data were using hierarchical clustering to classify the relationships of all genes between collected T and NT samples. A hierarchical clustering algorithm was used to investigate relationships among tumor samples. The complete linkage and the dissimilarity measure (1 minus the Pearson correlation coefficient of the log-adjusted Cy5: Cy3 ratios) were used for the analysis. The resultant dendrogram showed linked closely related colorectal tumors into a phylogenetic tree whose branch lengths represented the degree of similarity between these tumors.

**2.7. Bisulfite Sequencing and Methylation-Specific PCR (MS-PCR) for Methylation Status Analysis.** Genomic DNA (~0.5  $\mu$ g) was treated with sodium bisulfite according to the manufacturer's recommendations (EZ DNA Methylation Kit; Zymo Research, CA, USA). All selected genes methylation

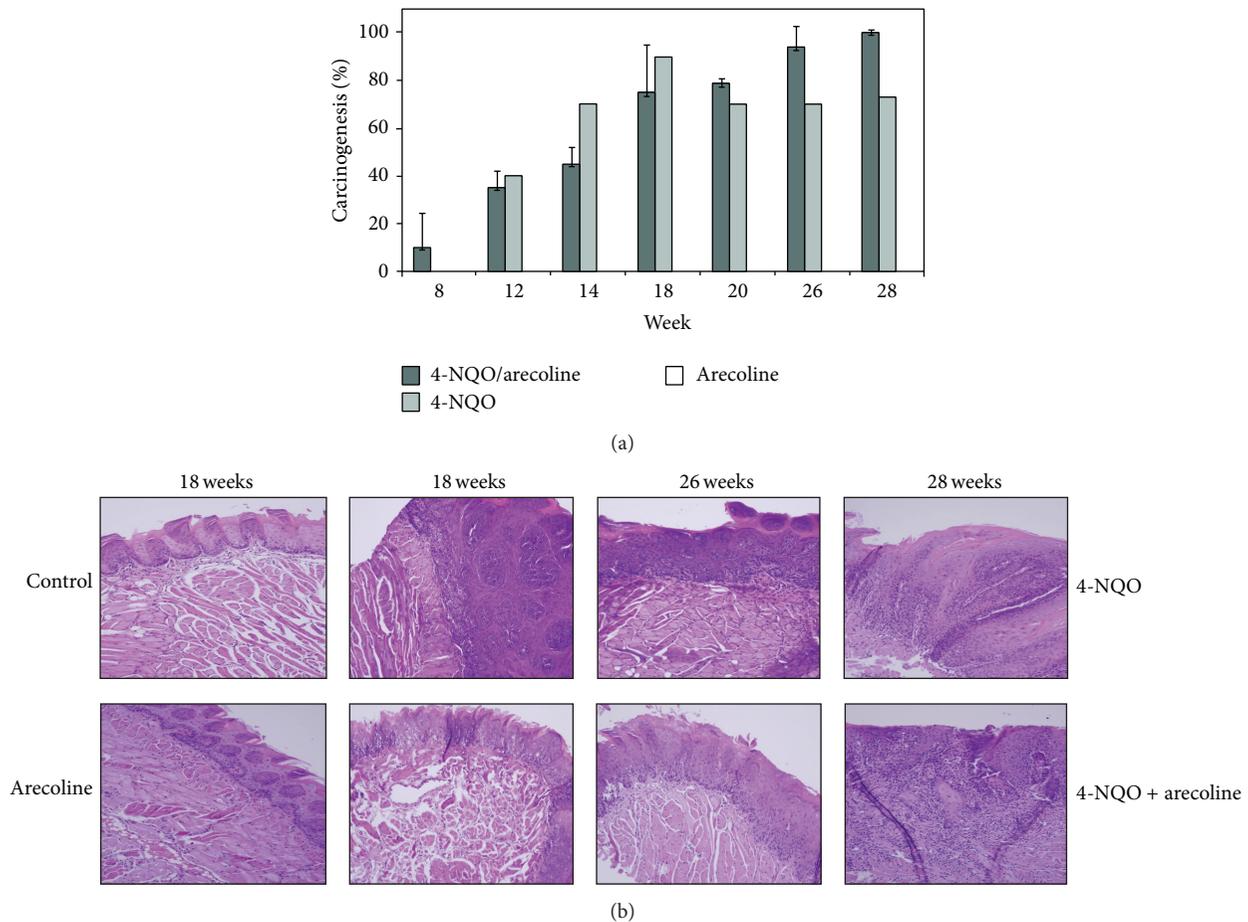


FIGURE 1: The progression of mouse model development for OSCC. (a) The ratio of carcinogenesis in mouse OSCC model. There were three treatments, 4-NQO/arecoline, 4-NQO, and arecoline. Mice were sacrificed at weeks 8, 12, 14, 18, 20, 26, and 28, respectively. The scoring criteria for mouse OSCC model are described in Section 2. (b) OSCC tongue tissues with tumors were excised, fixed, embedded, and sectioned for H&E staining. The mice that were treated with 4-NQO + arecoline would induce more serious OSCC formation than 4-NQO only and arecoline only. The order of severity was followed the time of treatment.

statuses were examined by methylation-specific PCR (MS-PCR) and sodium bisulfite genomic sequencing. The PCR reaction was as follows: 95°C for 5 min, followed by 45 cycles of 95°C, 30 sec, T<sub>m</sub> for 30 sec (Table 1), 72°C for 45 sec and ended with an extension of 72°C for 5 min and quick chill to 4°C on a Geneamp2400 PCR system (Applied Biosystems, CA, USA). For bisulfate sequencing analysis, each PCR product was subcloned into the pGEM-T Easy Vector (Promega, WI, USA) and performed 5–10 clones in each selection, respectively. Each colony was sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit and the automated ABI PRISM 3100 Genetic Analyzed (Applied Biosystems, CA, USA).

2.8. *Statistical Analysis.* Statistical analysis was performed using *t*-test to examine the association between NHOK and other cell lines. One-sided testing was used to calculate the *P*, and *P* < 0.05 was considered statistically significant.

### 3. Results

3.1. *OSCC Mouse Model Induced by Arecoline and 4-NQO.* To evaluate the efficiency of mouse model involving cotreating with arecoline and 4-NQO that mimic the etiology for OSCC tumor growth, the percentage of mice exhibiting carcinogenesis was calculated in Figure 1(a). Tumor development was assessed when treated with 4-NQO (N) and 4-NQO plus arecoline (NA) but not in arecoline. Mice were also sacrificed at 18, 26, and 28 weeks, and tongues with tumors were excised, fixed, embedded, and sectioned for H&E staining (Figure 1(b)). The H&E staining also showed that the tumor progresses were dealing with time and according to the treatments. According to the results, the incidence of tongue carcinogenesis in NA group was significantly higher than N group and arecoline group. Taken together, the treatment of NA at week 28 was much more serious than other weeks. These results suggest that arecoline promotes 4-NQO carcinogenesis in damaged oral epithelia cells.

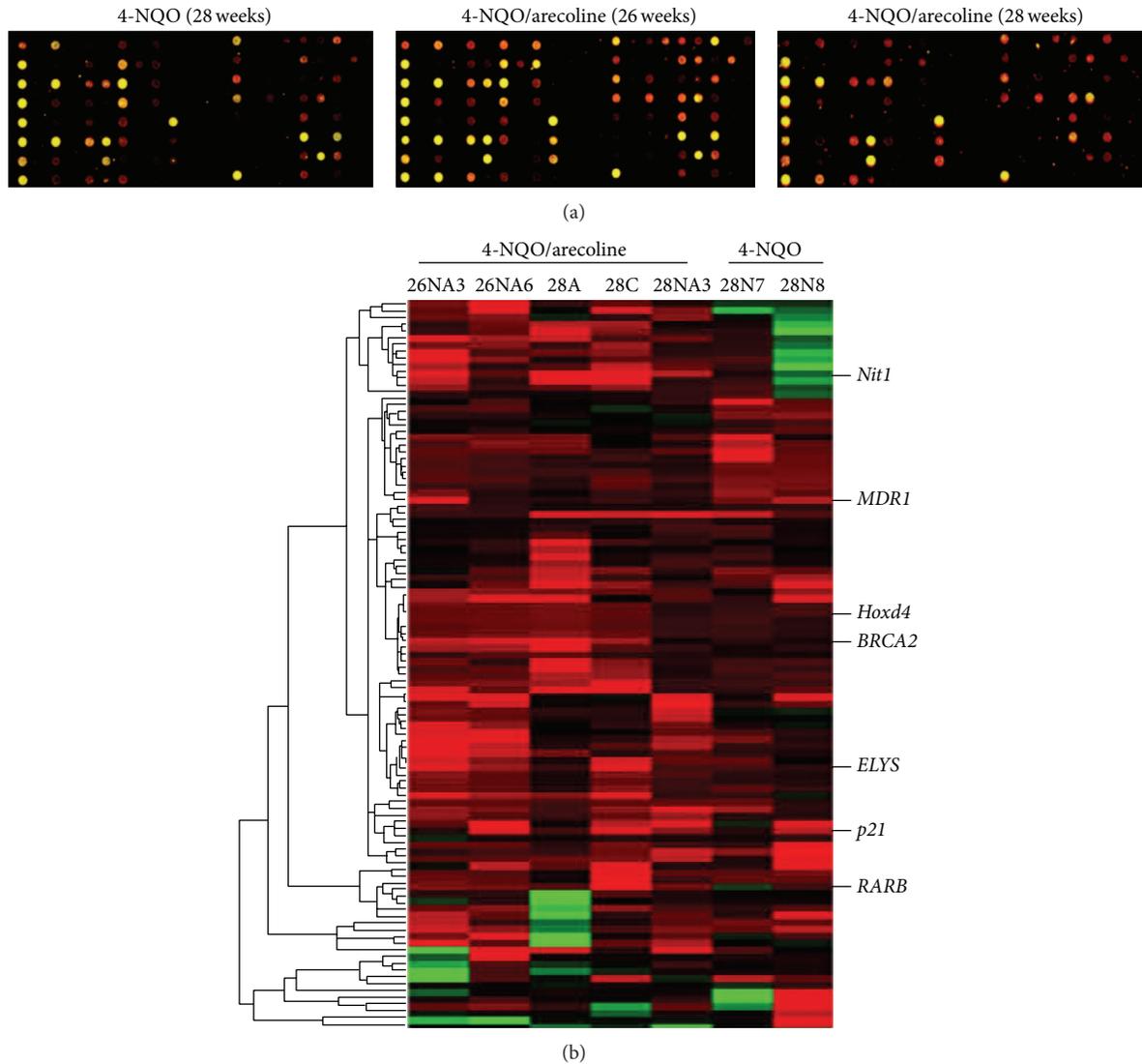


FIGURE 2: Presenting of MCGI microarray hybridization results and hierarchical clustering of methylation data of OSCC model mice. (a) MCGI microarray hybridization panel contained 4,608 duplicated CpG island tags. The expanded hybridization views showed the usefulness of the MCGI microarrays cohybridized with fluorescently labeled T (tumor part) and NT (nontumor part) with N at 28 weeks and NA at 26 and 28 weeks. Spots hybridized predominantly with tumor amplicon but not with normal amplicon would appear red and are indicative of hypermethylated CpG island loci present in the tumor genome. (b) Hierarchical clustering of N (4-NQO) and NA (4-NQO/arecoline) samples. At the top lists the 5 NA and 2 N studied. The row corresponds to each of 109 CpG island loci selected for methylation analysis. CpG islands (the normalized Cy5 : Cy3 ratios are  $\geq 2$ ) are those with hypermethylation in tumor DNA. The selected hypermethylated candidates were showed on the right sides. They are described in greater detail in Table 2.

3.2. Identification of Hypermethylation Genes from OSCC Tissues by CpG Island Microarray. According to the tumor progression percentages and H&E staining results, we selected the tongues tissues to be targets from OSCC mouse model with N and NA at 26 and 28 weeks for MCGI microarray screening. Figure 2(a) depicts representative data from the OSCC study. The expanded hybridization views showed the usefulness of the MCGI microarrays cohybridized with fluorescently labeled T (tumor part) and NT (nontumor part) with N at week 28 and NA at weeks 26 and 28, respectively. Spots hybridized predominantly with tumor amplicon, but not with nontumor amplicon, would appear

red and are indicative of hypermethylated CpG island loci, present in the tumor genome. The hybridization results showed that treatment with NA at week 28 represented much more spots that were obviously more hypermethylated than other treatments. Yellow spots (Cy5 : Cy3 = 1) represent equal amounts of bound DNA from each amplicon, an indication of no methylation differences between tumor and nontumor genomes. Selection of genes was based on the criteria described in the Materials and Methods. Figures 6(b) and 6(c) showed specific genes of selection results in hypermethylation and hypomethylation, respectively. We conducted a confirmation study to determine whether the

TABLE 2: The selected gene list of hypermethylation.

Accession number	Gene symbol	Description	Functions	Related carcinoma types
S80555	<i>RARB</i>	Retinoic acid receptor-beta	DNA binding, ligand-dependent nuclear receptor activity	Nonsmall cell lung cancer, hepatoma, bladder cancer, rectal cancer, breast cancer, head and neck squamous cancer
AF069985	<i>Nit1</i>	Nitrilase homolog I	Hydrolase activity	Oral squamous cell carcinoma (in this study)
AL355176	<i>BRCA2</i>	Breast cancer II gene	Protein binding	Bladder cancer, breast cancer, ovarian cancer
AL928664	<i>Hoxd4</i>	Homeo box D4 gene	DNA binding	Breast cancer, leukemia, neuroblastoma
AF457187	<i>p21</i>	Cyclin-dependent kinase inhibitor IA	Cyclin-dependent protein kinase inhibitor activity	Nonsmall cell lung cancer, bladder cancer, breast cancer, ovarian cancer, medulloblastoma, hepatoma
M60348	<i>MDR1</i>	Multidrug-resistance protein gene	ATP binding, ATPase activity	Pancreatic cancer, breast cancer, colorectal cancer, glioblastoma, leukemia, laryngeal cancer cell
AB081498	<i>ELYS</i>	Embryonic large molecule derived from yolk sac	DNA binding	Oral squamous cell carcinoma (in this study)

cutoff ratio ( $\geq 2$ ) could accurately identify hypermethylation. The hierarchical clustering presented the 109 gene loci of hypermethylation in the classifier in N and NA (Figure 2(b)). This methylation profile analysis has led to the identification of CpG island clusters that could evaluate many new genes correlating with OSCC progression in mouse model. These newly collected genes are shown in detail in Table 2. Upon further examination, we selected the *RARB* gene to examine in greater detail by MS-PCR bisulfite sequencing and semi-quantitative RT-PCR. The locations of CpG islands in mouse and human *RARB* genes were predicted using MethPrimer (<http://www.urogene.org/methprimer/index1.html>), respectively (Figures 3(a) and 4(a)).

**3.3. Verification of Methylated Genes by MS-PCR and Bisulfite Sequencing.** The *RARB* gene was a candidate target to verify the methylation status in these mouse OSCC tongues tissues and in human oral cancer cell lines that were also treated by 5'-aza-2'-deoxycytidine (5-aza-dC). Interestingly, *RARB* was hypermethylated in mouse OSCC and reexpression by 5-aza-dC treatment in human oral cancer cell lines. In Figure 3(b), *RARB* was hypermethylated in mouse OSCC tumor parts of N (4-NQO) and NA (4-NQO + Arecoline) compared to normal parts. Furthermore, the methylation ratio of *RARB* in NA treatment at 28 week was 60%. It was much higher than the 4-NQO treatment (20%) at 28 week. These results correlated with the microarray analysis data. *RARB* also investigated the methylation status in human oral cancer lines (Figure 4(b)). The results showed that Ca922, TW2.6, and HSC3 were hypermethylated than in NHOK. It also lost expressions in Ca922, TW2.6, and HSC3 but not in NHOK (Figure 4(c)). In Figure 4(d), the bisulfite sequencing showed the hypermethylation in TW2.6 (92.5%), Ca922 (96.3%), OC2 (90%), and HSC3 (93.8%), but in the NHOK and DOK, the normal and precancer cell lines were not methylated in bisulfite sequencing results. Taken together, these results showed that the promoter methylation of *RARB* plays the main role in OSCC progression.

**3.4. Methyltransferase Expressions in Human OSCC Cell Lines.** DNA methylation is catalyzed by the family of DNA methyltransferases (DNMT) including *Dnmt1*, *Dnmt3a*, and *Dnmt3b*. Figure 5 shows the measurement of gene expressions via real-time PCR on *Dnmt1*, *Dnmt3a*, and *Dnmt3b* in human oral cancer lines. For *Dnmt1*, there were no expression differences between NHOK and other cell lines (Figure 5(a)). However, *Dnmt3a* and *Dnmt3b* showed significantly higher expression levels in TW2.6 and Ca922 than others (Figures 5(b) and 5(c)).

## 4. Discussion

OSCC is the most common head and neck neoplasm, affecting 270,000 people worldwide each year [20, 32]. According to related research, patients who smoke, drink, and chew betel quid experience a 5.32-fold increased likelihood of death as compared to those without any oral habits [50]. In Taiwan and other Southeast Asian countries, betel quid chewing is one of the most important risk factors for oral cancer patients and associates as the main cause between betel quid chewing and oral cancer development [51]. 4-NQO is quinoline derivative and a tumorigenic compound which can induce DNA lesions. Quinone oxidoreductase is one of the major enzymes that convert 4-NQO to the more active metabolite, 3-hydroxyaminoquinoline 1-oxide [52]. This oxidoreductase can be produced from the mucosal of the sublingual in humans and mice. Arecoline is a natural alkaloid product found in the areca nut. In this study, these two compounds were used to induce oral carcinogenesis in a mouse model. When we added 4-NQO plus Arecoline to the drinking water of mice, the results showed that mice were shown to have induced carcinogenesis at 100% at week 26 and 28 (Figure 1(a)). In H&E staining, resected oral tissue at the end of week 28 (Figure 1(b)) also showed that NA and N treatment could induce squamous cell carcinoma, respectively. However, treatment with arecoline alone revealed that it did not induce tumorigenesis after 28 weeks.

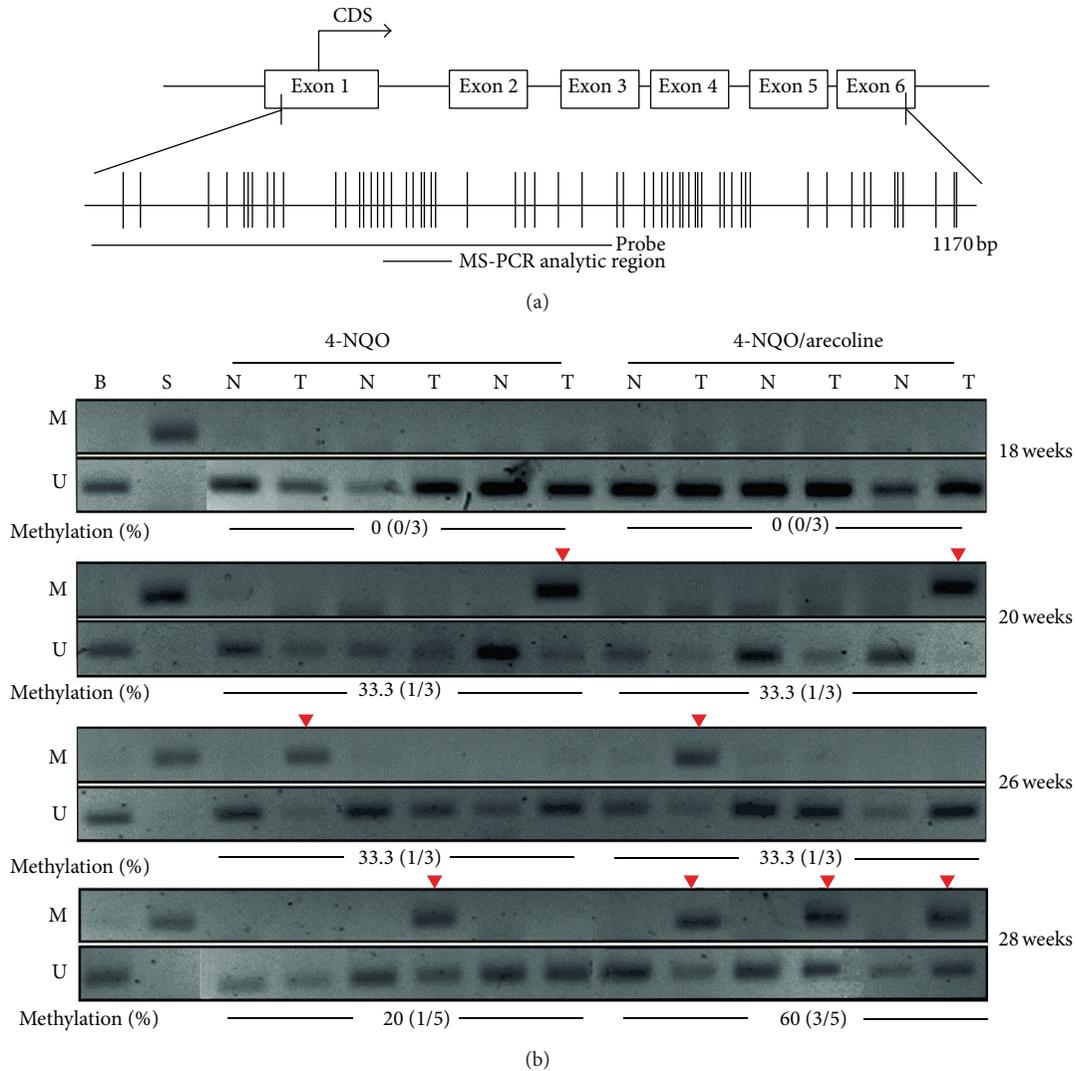


FIGURE 3: Methylation status of *RARB* gene in mouse OSCC model. (a) The CpG island diagram of *RARB* gene. The hybridized probe used to microarray located from exon one to exon three. We designed the MS-PCR primer sets located within this region. (b) The designed MS-PCR analysis results in 4-NQO and 4-NQO/arecoline at 18, 20, 26, and 28 weeks, respectively. The *RARB* gene showed hypermethylation (60%) in 4-NQO/arecoline at 28 week. M: methylated set, U: unmethylated set, B: blood DNA for methylation negative control, S: SssI treated DNA for methylation positive control, and %methylation: the percentages of methylation. The red inverted triangle showed that the methylated *RARB* gene could be amplified.

Methylation is important in the development of OSCC and many tumor suppressor genes targeted by promoter methylation will by no doubt be described in the future. The techniques used at present to detect methylation provide good sensitivity, specificity, and speed. There are many types of methylation arrays for a genome-wide approach to realize methylation profile. In this study, we applied a home-made high throughput MCGI array to analyze DNA methylation across the entire genome in the OSCC mouse tongue tissues. This home-made MCGI array not only could identify methylation profile but also could detect mRNA expression (in our previous studies). The chip was cohybridized with mouse Cot-I DNA and total RNA mixture to evaluate the quality and the exon-containing portions can be used to measure levels of gene expression. According to the previous research,

the methylation status and mRNA expression levels could be verified by MCGI array.

According to the depicted representative data from previous studies, *RARB* expression is thought to be associated with cellular sensitivity to retinoid in numerous cancer cells, including HNSCC cells, breast cancer cells, lymphocytic leukemia, and lung cancer cells [53–57]. Methylation of *RARB* was identified which had a correlation with primary oral malignant diseases [58, 59]. The methylation array of 4,608 genes (duplicate on chip) that we used in this study included the majority of genes which have previously been associated with head and neck cancer (e.g., *DAPK*, *MGMT*, and *CDHI*, etc.). However, only *RARB* in previously studied genes were shown to be positive for methylation. To explain the possibility of this discrepancy, the hybridized probes maybe

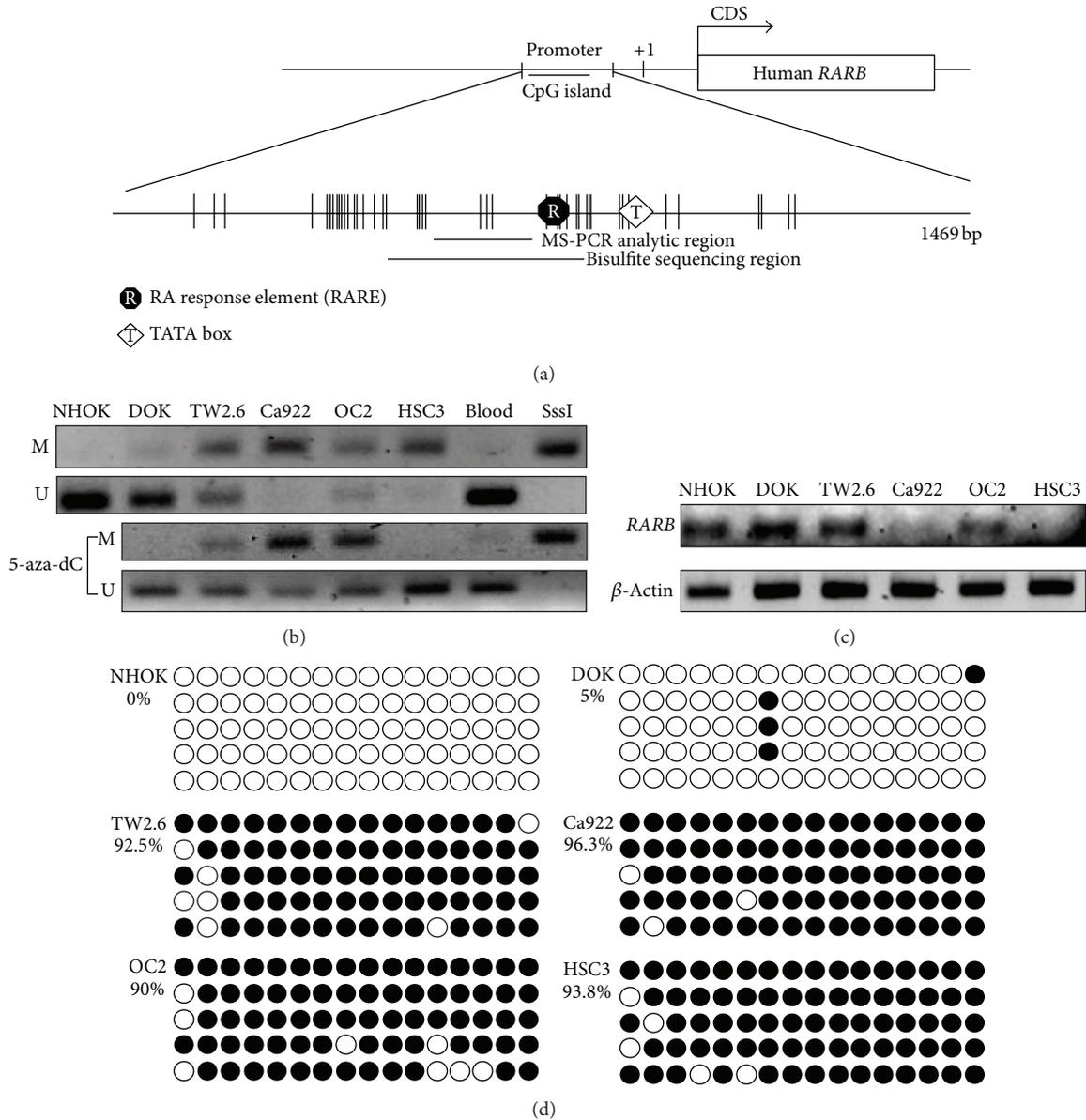


FIGURE 4: Methylation status in human oral cancer cell lines of *RARB*. (a) The schematic for CpG island presentation of *RARB*. There was a CpG island located in promoter region. We designed the MS-PCR and bisulfite sequencing primer sets in this region. (b) The *RARB* methylation status in different human oral cancer cell lines. The cell lines were also treated with 5'-aza-dC. The methylation status was recovered because of 5'-aza-dC treatment. (c) The *RARB* mRNA expression in different human oral cancer cell lines. The *RARB* was not expressed in Ca922, OC2, and HSC3 obviously. (d) The *RARB* bisulfite sequencing in different human oral cancer cell lines. There were more methylated *RARB* in TW2.6, Ca922, OC2, and HSC3 than in NHOK and DOK. The hollow circle is the unmethylated CpG site and the full circle is the methylated CpG site.

located on different regions between our array and previous methylation studies. However, we found some interesting genes and described this in greater detail in Table 2 as they were shown to have hypermethylation on the chip, but this was not reported before in OSCC.

In our study, *DNMT1* does not affect the expression in human OSCC cell lines. Nevertheless, *DNMT3a* and *DNMT3b* did affect the expressions in human OSCC cell lines (Figure 5). These results showed that *DNMT3a* and *DNMT3b*

were higher expressions in TW2.6 and Ca922, the primary oral cancer cell lines, but not in OC2 and HSC3. As compared to other tumors, no correlation was seen between DNMT upregulation and promoter hypermethylation-induced inactivation of tumor-related genes. The exact mechanisms of DNMT upregulation remain unclear, but it is suggested that aberrant DNMT activity, especially with regard to DNMT1, is due to a rapid proliferation of cancer cells because DNMT1 binds to proliferating cell nuclear antigen (PCNA) [60].

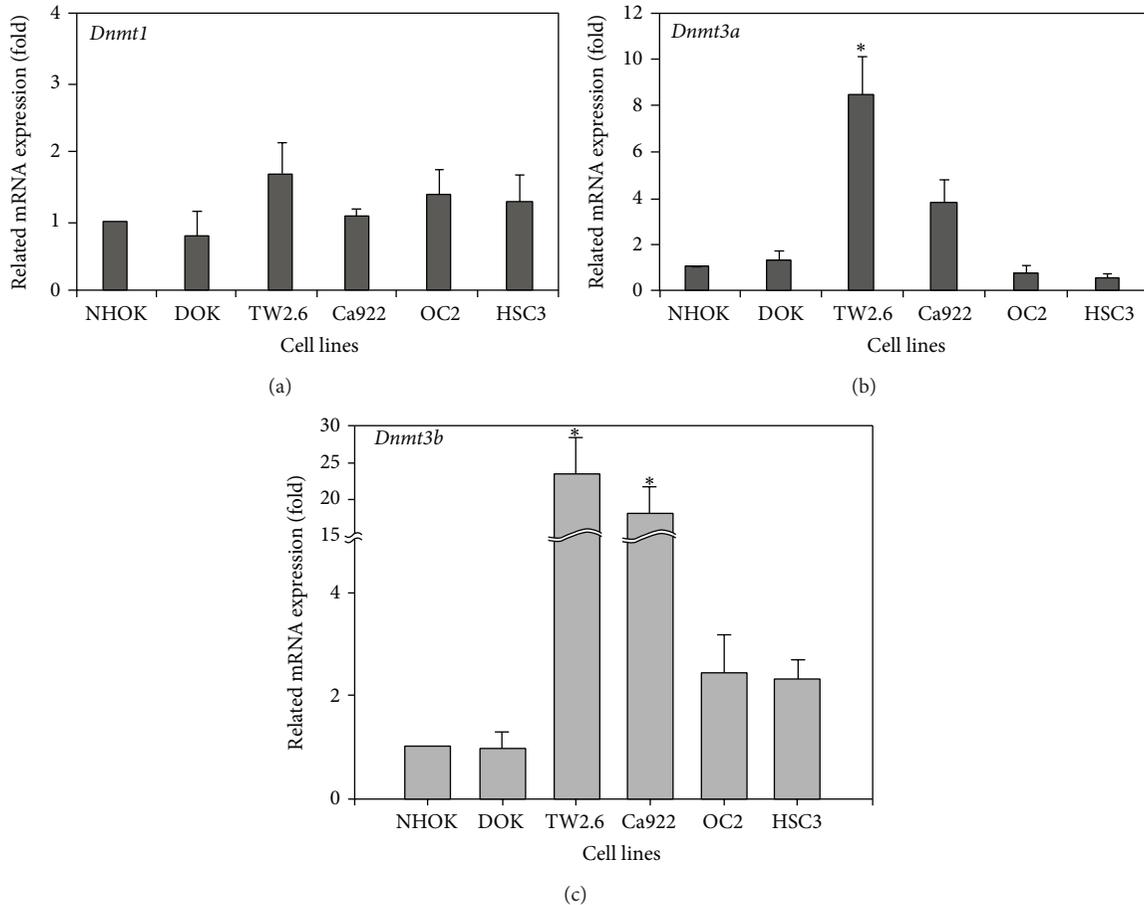


FIGURE 5: DNMTs RNA expression in human oral cancer cell lines. The measurement of DNMTs expression levels by real-time PCR was showed in (a), (b), and (c), respectively. (a) The *Dnmt1* expression was not changed in all human oral cancer cell lines. (b) and (c) TW2.6 and Ca922 were higher expression in *Dnmt3a* and *Dnmt3b* than others. The figures shown are the mean of three experiments where all of the samples were analyzed in triplicate. The start sign showed the statistically significant ( $P < 0.05$ ).

Overexpression of all the DNMTs at the mRNA level has been shown for several cancers [45, 61, 62]. However, *DNMT3a* and *DNMT3b* were the *de novo* methyltransferases. In this result (Figures 5(b) and 5(c)), we thought *DNMT3a* and *DNMT3b* were added to new methyl groups to DNA to cause DNA methylation aberrant at the early stage of OSCC progression.

We used some different approaches to deal with the sparseness of data. To begin with we used a methylation array of 4,086 genes (duplicated on chip), which provided more comprehensive data. Here we also suggest that this study could offer a basic evidence that promoter hypermethylation of RARB is correlated with the occurrence of betel-related OSCC.

Isotretinoin (13-cis-retinoic acid) is considered to have the effect of preventing second primary cancers and local or regional recurrence after head and neck cancer is treated. However, in a previous clinical trial, chemoprevention therapy involving retinoid acid does not cause significant differences in early head and neck cancers. The nonsignificant benefit result is due to the small number of patients. The prospective study rendered a small percentage of patients that

have second primary cancers and local regional recurrence, thus causing the results to be less significant even though there was trend to have reduced second primary cancers developed and less local regional recurrence in prior studies [63]. However, there was another clinical research using the isotretinoin (13-cis-retinoic acid) (50 to 100 mg per square meter of body-surface area per day) as compared with placebo, to be taken daily for 12 months. They offer significantly reduced second primary cancer development after 32 months of follow-up, and multiple second primary tumors developed in the placebo group [64]. Therefore, they suggest that isotretinoin still has the ability to prevent second primary cancers, but there was less use in preventing the primary site recurrence [64]. There was another human cohort study using an *in situ* hybridization (ISH) analysis checker with 38 pairs of surgical specimens of primary OSCC and noncancerous matched normal control to compare the cellular expression level of RARB. They found the loss of RARB in the advanced OSCCs especially when they are betel quid users [65]. This prominent result also give us an understanding that RARB is really an important factor for chemoprevention for tumor progression. Therefore, betel

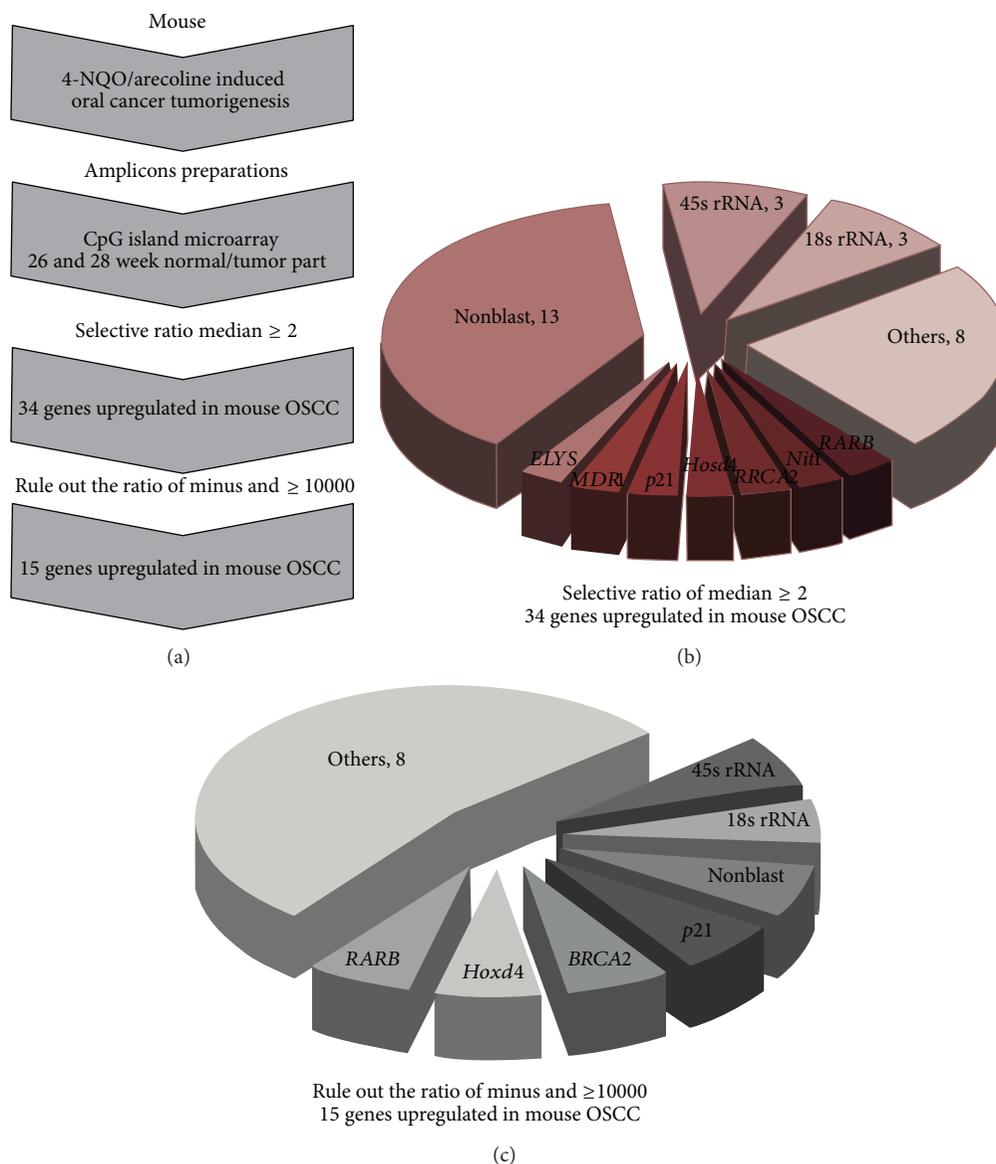


FIGURE 6: Outline of this study. (a) The flowchart of research project. (b) The selection from MCGI microarray hybridization results by the ratio higher or equal two. There were 34 genes that were hypermethylated in OSCC mouse model. (c) The selection from MCGI microarray hybridization results by ruling out the ratio was minus and higher or equal to 10,000. There were 15 genes that were hypermethylated in OSCC mouse model.

quid related hypermethylation of RARB will really increase the tumorigenesis and poor treatment outcome of oral cancer. Concerning the mechanisms of the chemoprevention function, there was another study that revealed that the retinoids could suppress basal expression of Cox-2 or EGF-mediated induction of Cox-2 in human oral squamous carcinoma cells [66]. Thus, RARB not only has cell cycle inhibition and tumor suppressant effects, but also anti-inflammatory effects that cease COX-2 related cancerization effects on the oral mucosa.

This is pilot study that talked about the detailed mechanisms of retinoid acid function affected by areca. In the future, we should consider retinoid acid use or RARB related drugs for areca users who are found to have oral tumors or

oral cancer to reduce the incidence of oral cancer and to provide a better treatment outcome.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Authors' Contribution

Zi-Lun Lai, Yung-An Tsou, and Shin-Ru Fan contributed equally to this work.

## Acknowledgments

This research was supported in part by Grants NSC-98-2313-B-005-012 and NSC-101-2313-B-005-014-MY3 from the National Science Council, Grant COA-97-6.2.1-U1(9) from the Council of Agriculture, and grant from the Ministry of Education, Taiwan, under the Aiming for the Top University plan (ATU-101-s0508). This work was supported by Grants CMU95-302, CMU97-098, and CMU97-099 from China Medical University, Taiwan (J.-C. Cheng) and DMR-103-025 from China Medical University Hospital (Y.-A. Tsou). The authors thank Dr. Dar-Bin, Shieh (National Cheng Kung University, Taiwan) to support DOK cells. The authors also thank technique support from Ms. Chia-Fan Lin.

## References

- [1] A. Jemal, R. Siegel, E. Ward et al., "Cancer statistics, 2008," *CA Cancer Journal for Clinicians*, vol. 58, no. 2, pp. 71–96, 2008.
- [2] C. Li-Ting, C. Chung-Ho, Y. Yi-Hsin, and H. Pei-Shan, "The development and validation of oral cancer staging using administrative health data," *BMC Cancer*, vol. 14, no. 1, article 380, 2014.
- [3] Y. C. Ko, Y. L. Huang, and C. H. Lee, "Betel quid chewing, cigarette smoking and alcohol consumption related to oral cancer in Taiwan," *Journal of Oral Pathology and Medicine*, vol. 24, no. 10, pp. 450–453, 1995.
- [4] W. L. Hsu, Y. C. Chien, C. J. Chiang et al., "Lifetime risk of distinct upper aerodigestive tract cancers and consumption of alcohol, betel and cigarette," *International Journal of Cancer*, vol. 135, no. 6, pp. 1480–1486, 2014.
- [5] S. H. Hung, M. C. Tsai, T. C. Liu, H. C. Lin, and S. D. Chung, "Routine endoscopy for esophageal cancer is suggestive for patients with oral, oropharyngeal and hypopharyngeal cancer," *PLoS ONE*, vol. 8, no. 8, Article ID e72097, 2013.
- [6] M. Chen, W. Huang, C. H. Chan, P. Chen, and K. Lee, "Impact of second primary esophageal or lung cancer on survival of patients with head and neck cancer," *Oral Oncology*, vol. 46, no. 4, pp. 249–254, 2010.
- [7] C. R. P. C. Gupta, "Ranks number three among ten leading causes of death," *ANNALS Academy of Medicine Singapore*, vol. 33, pp. 31S–36S, 2004.
- [8] S. Chiang, P. Chen, C. Lee et al., "Up-regulation of inflammatory signalings by areca nut extract and role of cyclooxygenase-2 -1195G>A polymorphism reveal risk of oral cancer," *Cancer Research*, vol. 68, no. 20, pp. 8489–8498, 2008.
- [9] P. Rajalalitha and S. Vali, "Molecular pathogenesis of oral submucous fibrosis—a collagen metabolic disorder," *Journal of Oral Pathology and Medicine*, vol. 34, no. 6, pp. 321–328, 2005.
- [10] A. Ariyawardana, A. D. S. Athukorala, and A. Arulanandam, "Effect of betel chewing, tobacco smoking and alcohol consumption on oral submucous fibrosis: a case-control study in Sri Lanka," *Journal of Oral Pathology and Medicine*, vol. 35, no. 4, pp. 197–201, 2006.
- [11] T. Ramaesh, B. R. R. N. Mendis, N. Ratnatunga, and R. O. Thattil, "The effect of tobacco smoking and of betel chewing with tobacco on the buccal mucosa: a cytometric analysis," *Journal of Oral Pathology and Medicine*, vol. 28, no. 9, pp. 385–388, 1999.
- [12] C. Chung, Y. Yang, T. Wang, T. Shieh, and S. Warnakulasuriya, "Oral precancerous disorders associated with areca quid chewing, smoking, and alcohol drinking in southern Taiwan," *Journal of Oral Pathology and Medicine*, vol. 34, no. 8, pp. 460–466, 2005.
- [13] X. Tang, B. Knudsen, D. Bemis, S. Tickoo, and L. J. Gudas, "Oral cavity and esophageal carcinogenesis modeled in carcinogen-treated mice," *Clinical Cancer Research*, vol. 10, no. 1, part 1, pp. 301–313, 2004.
- [14] M. Kitano, "Host genes controlling the susceptibility and resistance to squamous cell carcinoma of the tongue in a rat model," *Pathology International*, vol. 50, no. 5, pp. 353–362, 2000.
- [15] N. W. Chang, R. J. Pei, H. C. Tseng et al., "Co-treating with arecoline and 4-nitroquinoline 1-oxide to establish a mouse model mimicking oral tumorigenesis," *Chemico-Biological Interactions*, vol. 183, no. 1, pp. 231–237, 2010.
- [16] R. J. Shaw, T. Liloglou, S. N. Rogers et al., "Promoter methylation of P16, RAR $\beta$ , E-cadherin, cyclin A1 and cytoglobin in oral cancer: quantitative evaluation using pyrosequencing," *The British Journal of Cancer*, vol. 94, no. 4, pp. 561–568, 2006.
- [17] C. Y. Fan, "Epigenetic alterations in head and neck cancer: prevalence, clinical significance, and implications," *Current Oncology Reports*, vol. 6, no. 2, pp. 152–161, 2004.
- [18] M. Esteller, "Cancer epigenetics: DNA methylation and chromatin alterations in human cancer," *Advances in Experimental Medicine and Biology*, vol. 532, pp. 39–49, 2003.
- [19] P. K. Ha and J. A. Califano, "Promoter methylation and inactivation of tumour-suppressor genes in oral squamous-cell carcinoma," *The Lancet Oncology*, vol. 7, no. 1, pp. 77–82, 2006.
- [20] E. Suzuki, I. Imoto, A. Pimkhaokham et al., "PRTFDC1, a possible tumor-suppressor gene, is frequently silenced in oral squamous-cell carcinomas by aberrant promoter hypermethylation," *Oncogene*, vol. 26, no. 57, pp. 7921–7932, 2007.
- [21] C. Scully, J. K. Field, and H. Tanzawa, "Genetic aberrations in oral or head and neck squamous cell carcinoma (SCCHN): 1. Carcinogen metabolism, DNA repair and cell cycle control," *Oral Oncology*, vol. 36, no. 3, pp. 256–263, 2000.
- [22] Z. Xiong and P. W. Laird, "COBRA: A sensitive and quantitative DNA methylation assay," *Nucleic Acids Research*, vol. 25, no. 12, pp. 2532–2534, 1997.
- [23] M. Frommer, L. E. McDonald, D. S. Millar et al., "A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 5, pp. 1827–1831, 1992.
- [24] J. G. Herman, J. R. Graff, S. Myöhänen, B. D. Nelkin, and S. B. Baylin, "Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 18, pp. 9821–9826, 1996.
- [25] T. H. Huang, M. R. Perry, and D. E. Laux, "Methylation profiling of CpG islands in human breast cancer cells," *Human Molecular Genetics*, vol. 8, no. 3, pp. 459–470, 1999.
- [26] L. R. Adrien, N. F. Schlecht, N. Kawachi et al., "Classification of DNA methylation patterns in tumor cell genomes using a CpG island microarray," *Cytogenetic and Genome Research*, vol. 114, no. 1, pp. 16–23, 2006.
- [27] S. H. Wei, C. M. Chen, G. Strathdee et al., "Methylation microarray analysis of late-stage ovarian carcinomas distinguishes progression-free survival in patients and identifies candidate epigenetic markers," *Clinical Cancer Research*, vol. 8, no. 7, pp. 2246–2252, 2002.

- [28] P. S. Yan, C. Chen, H. Shi, F. Rahmatpanah, S. H. Wei, and T. H. Huang, "Applications of CpG island microarrays for high-throughput analysis of DNA methylation," *Journal of Nutrition*, vol. 132, no. 8, pp. 2430S–2434S, 2002.
- [29] C. M. Chen, H. L. Chen, T. H. Hsiau et al., "Methylation target array for rapid analysis of CpG island hypermethylation in multiple tissue genomes," *The American Journal of Pathology*, vol. 163, no. 1, pp. 37–45, 2003.
- [30] R. S. Gitan, H. Shi, C. Chen, P. S. Yan, and T. H. Huang, "Methylation-specific oligonucleotide microarray: a new potential for high-throughput methylation analysis," *Genome Research*, vol. 12, no. 1, pp. 158–164, 2002.
- [31] V. Beier, C. Mund, and J. D. Hoheisel, "Monitoring methylation changes in cancer," *Advances in biochemical engineering/biotechnology*, vol. 104, pp. 1–11, 2007.
- [32] D. M. Parkin, F. Bray, J. Ferlay, and P. Pisani, "Global cancer statistics, 2002," *CA: A Cancer Journal for Clinicians*, vol. 55, no. 2, pp. 74–108, 2005.
- [33] B. W. Neville and T. A. Day, "Oral cancer and precancerous lesions," *Ca: A Cancer Journal for Clinicians*, vol. 52, no. 4, pp. 195–215, 2002.
- [34] R. Lotan, "Retinoids in cancer chemoprevention," *The FASEB Journal*, vol. 10, no. 9, pp. 1031–1039, 1996.
- [35] D. R. Shalinsky, E. D. Bischoff, W. W. Lamph et al., "A novel retinoic acid receptor-selective retinoid, ALRT1550, has potent antitumor activity against human oral squamous carcinoma xenografts in nude mice," *Cancer Research*, vol. 57, no. 1, pp. 162–168, 1997.
- [36] F. McGregor, A. Muntoni, J. Fleming et al., "Molecular changes associated with oral dysplasia progression and acquisition of immortality: potential for its reversal by 5-azacytidine," *Cancer Research*, vol. 62, no. 16, pp. 4757–4766, 2002.
- [37] F. McGregor, E. Wagner, D. Felix, D. Soutar, K. Parkinson, and P. R. Harrison, "Inappropriate retinoic acid receptor- $\beta$  expression in oral dysplasias: correlation with acquisition of the immortal phenotype," *Cancer Research*, vol. 57, no. 18, pp. 3886–3889, 1997.
- [38] X. Xu, J. Y. Ro, J. S. Lee, D. M. Shin, W. K. Hong, and R. Lotan, "Differential expression of nuclear retinoid receptors in normal, premalignant, and malignant head and neck tissues," *Cancer Research*, vol. 54, no. 13, pp. 3580–3587, 1994.
- [39] L. Hu, D. L. Crowe, J. G. Rheinwald, P. Chambon, and L. J. Gudas, "Abnormal expression of retinoic acid receptors and keratin 19 by human oral and epidermal squamous cell carcinoma cell lines," *Cancer Research*, vol. 51, no. 15, pp. 3972–3981, 1991.
- [40] J. J. Sheu, C. H. Hua, L. Wan et al., "Functional genomic analysis identified epidermal growth factor receptor activation as the most common genetic event in oral squamous cell carcinoma," *Cancer Research*, vol. 69, no. 6, pp. 2568–2576, 2009.
- [41] K. Kozaki, I. Imoto, S. Mogi, K. Omura, and J. Inazawa, "Exploration of tumor-suppressive microRNAs silenced by DNA hypermethylation in oral cancer," *Cancer Research*, vol. 68, no. 7, pp. 2094–2105, 2008.
- [42] C. M. Chen, C. H. Wang, S. C. Wu, C. Lin, S. Lin, and W. T. K. Cheng, "Temporal and spatial expression of biologically active human factor VIII in the milk of transgenic mice driven by mammary-specific bovine  $\alpha$ -lactalbumin regulation sequences," *Transgenic Research*, vol. 11, no. 3, pp. 257–268, 2002.
- [43] A. Ahluwalia, P. Yan, J. A. Hurteau et al., "Dna methylation and ovarian cancer: I. Analysis of CpG island hypermethylation in human ovarian cancer using differential methylation hybridization," *Gynecologic Oncology*, vol. 82, no. 2, pp. 261–268, 2001.
- [44] S. H. Cross, J. A. Charlton, X. Nan, and A. P. Bird, "Purification of CpG islands using a methylated DNA binding column," *Nature Genetics*, vol. 6, no. 3, pp. 236–244, 1994.
- [45] K. P. Nephew and T. H. Huang, "Epigenetic gene silencing in cancer initiation and progression," *Cancer Letters*, vol. 190, no. 2, pp. 125–133, 2003.
- [46] P. S. Yan, M. R. Perry, D. E. Laux, A. L. Asare, C. W. Caldwell, and T. H. Huang, "CpG island arrays: an application toward deciphering epigenetic signatures of breast cancer," *Clinical Cancer Research*, vol. 6, no. 4, pp. 1432–1438, 2000.
- [47] H. Shi, P. S. Yan, C. Chen et al., "Expressed CpG island sequence tag microarray for dual screening of DNA hypermethylation and gene silencing in cancer cells," *Cancer Research*, vol. 62, no. 11, pp. 3214–3220, 2002.
- [48] P. S. Yan, C.-. Chen, H. Shi et al., "Dissecting complex epigenetic alterations in breast cancer using CpG island microarrays," *Cancer Research*, vol. 61, no. 23, pp. 8375–8380, 2001.
- [49] P. S. Yan, T. Efferth, H. L. Chen et al., "Use of CpG island microarrays to identify colorectal tumors with a high degree of concurrent methylation," *Methods*, vol. 27, no. 2, pp. 162–169, 2002.
- [50] Y. K. Chen, H. C. Huang, L. M. Lin, and C. C. Lin, "Primary oral squamous cell carcinoma: an analysis of 703 cases in southern Taiwan," *Oral Oncology*, vol. 35, no. 2, pp. 173–179, 1999.
- [51] J. H. Jeng, M. C. Chang, and L. J. Hahn, "Role of areca nut in betel quid-associated chemical carcinogenesis: current awareness and future perspectives," *Oral Oncology*, vol. 37, no. 6, pp. 477–492, 2001.
- [52] J. Tanuma, M. Hirano, Y. Hirayama et al., "Genetic predisposition to 4NQO-induced tongue carcinogenesis in the rat," *Medical Principles and Practice*, vol. 14, no. 5, pp. 297–305, 2005.
- [53] X. Zhang, Y. Liu, M. Lee, and M. Pfahl, "A specific defect in the retinoic acid response associated with human lung cancer cell lines," *Cancer Research*, vol. 54, no. 21, pp. 5663–5669, 1994.
- [54] B. Seeliger, S. Wilop, R. Osieka, O. Galm, and E. Jost, "CpG island methylation patterns in chronic lymphocytic leukemia," *Leukemia and Lymphoma*, vol. 50, no. 3, pp. 419–426, 2009.
- [55] H. Wan, N. Oridate, D. Lotan, W. K. Hong, and R. Lotan, "Overexpression of retinoic acid receptor  $\beta$  in head and neck of squamous cell carcinoma cells increases their sensitivity to retinoid-induced suppression of squamous differentiation by retinoids," *Cancer Research*, vol. 59, no. 14, pp. 3518–3526, 1999.
- [56] S. Sun, H. Wan, P. Yue, W. K. Hong, and R. Lotan, "Evidence that retinoic acid receptor  $\beta$  induction by retinoids is important for tumor cell growth inhibition," *The Journal of Biological Chemistry*, vol. 275, no. 22, pp. 17149–17153, 2000.
- [57] M. Widschwendter, G. Daxenbichler, O. Dapunt, and C. Marth, "Effects of retinoic acid and  $\gamma$ -interferon on expression of retinoic acid receptor and cellular retinoic acid-binding protein in breast cancer cells," *Cancer Research*, vol. 55, no. 10, pp. 2135–2139, 1995.
- [58] S. Maruya, J. J. Issa, R. S. Weber et al., "Differential methylation status of tumor-associated genes in head and neck squamous carcinoma: incidence and potential implications," *Clinical Cancer Research*, vol. 10, no. 11, pp. 3825–3830, 2004.
- [59] E. M. Youssef, D. Lotan, J. P. Issa et al., "Hypermethylation of the retinoic acid receptor- $\beta$ 2 gene in head and neck carcinogenesis," *Clinical Cancer Research*, vol. 10, no. 5, pp. 1733–1742, 2004.
- [60] I. Tischoff and A. Tannapfel, "DNA methylation in hepatocellular carcinoma," *World Journal of Gastroenterology*, vol. 14, no. 11, pp. 1741–1748, 2008.

- [61] K. D. Robertson, E. Uzvolgyi, G. Liang et al., "The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors," *Nucleic Acids Research*, vol. 27, no. 11, pp. 2291–2298, 1999.
- [62] S. Xie, Z. Wang, M. Okano et al., "Cloning, expression and chromosome locations of the human *DNMT3* gene family," *Gene*, vol. 236, no. 1, pp. 87–95, 1999.
- [63] F. R. Khuri, J. J. Lee, S. M. Lippman et al., "Randomized phase III trial of low-dose isotretinoin for prevention of second primary tumors in stage I and II head and neck cancer patients," *Journal of the National Cancer Institute*, vol. 98, no. 7, pp. 441–450, 2006.
- [64] W. K. Hong, S. M. Lippman, L. M. Itri et al., "Prevention of second primary tumors with isotretinoin in squamous-cell carcinoma of the head and neck," *The New England Journal of Medicine*, vol. 323, no. 12, pp. 795–801, 1990.
- [65] S. Y. Kao, H. F. Tu, K. W. Chang, C. Chang, C. Yang, and S. Lin, "The retinoic acid receptor- $\beta$  (RAR- $\beta$ ) mRNA expression in the oral squamous cell carcinoma associated with betel quid use," *Journal of Oral Pathology and Medicine*, vol. 31, no. 4, pp. 220–226, 2002.
- [66] J. R. Mestre, K. Subbaramaiah, P. G. Sacks et al., "Retinoids suppress epidermal growth factor-induced transcription of cyclooxygenase-2 in human oral squamous carcinoma cells," *Cancer Research*, vol. 57, no. 14, pp. 2890–2895, 1997.

## Research Article

# Osteoponin Promoter Controlled by DNA Methylation: Aberrant Methylation in Cloned Porcine Genome

Chih-Jie Shen,<sup>1</sup> Yung-An Tsou,<sup>1,2,3</sup> Hsiao-Ling Chen,<sup>4</sup>  
Hung-Jin Huang,<sup>5</sup> Shinn-Chih Wu,<sup>6</sup> Winston T. K. Cheng,<sup>7</sup>  
Calvin Yu-Chian Chen,<sup>2,8,9,10</sup> and Chuan-Mu Chen<sup>1,11</sup>

<sup>1</sup> Department of Life Sciences, National Chung Hsing University, Taichung 402, Taiwan

<sup>2</sup> School of Medicine, College of Medicine, China Medical University, Taichung 40402, Taiwan

<sup>3</sup> Department of Otolaryngology Head and Neck Surgery, China Medical University, Taichung 40402, Taiwan

<sup>4</sup> Department of Bioresources, Da-Yeh University, Changhwa 515, Taiwan

<sup>5</sup> Department of Chinese Pharmaceutical Sciences and Chinese Medicine Resources, College of Pharmacy, China Medical University, Taichung 40402, Taiwan

<sup>6</sup> Department of Animal Science and Technology, National Taiwan University, Taipei 106, Taiwan

<sup>7</sup> Department of Animal Science and Biotechnology, Tung Hai University, Taichung 407, Taiwan

<sup>8</sup> Department of Biomedical Informatics, Asia University, Taichung 41354, Taiwan

<sup>9</sup> Department of Medical Research, Human Genetic Center, China Medical University Hospital, Taichung 40447, Taiwan

<sup>10</sup> Research Center for Chinese Medicine & Acupuncture, China Medical University, Taichung 40402, Taiwan

<sup>11</sup> Rong Hsing Research Center for Translational Medicine and the iEGG Center, National Chung Hsing University, Taichung 402, Taiwan

Correspondence should be addressed to Calvin Yu-Chian Chen; [ycc929@mit.edu](mailto:ycc929@mit.edu)  
and Chuan-Mu Chen; [cmchen7010978@gmail.com](mailto:cmchen7010978@gmail.com)

Received 24 February 2014; Accepted 5 March 2014; Published 2 July 2014

Academic Editor: Chung Y. Hsu

Copyright © 2014 Chih-Jie Shen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cloned animals usually exhibited many defects in physical characteristics or aberrant epigenetic reprogramming, especially in some important organ development. Osteoponin (OPN) is an extracellular-matrix protein involved in heart and bone development and diseases. In this study, we investigated the correlation between *OPN* mRNA and its promoter methylation changes by the 5-aza-dc treatment in fibroblast cell and promoter assay. Aberrant methylation of porcine *OPN* was frequently found in different tissues of somatic nuclear transferred cloning pigs, and bisulfite sequence data suggested that the *OPN* promoter region -2615 to -2239 nucleotides (nt) may be a crucial regulation DNA element. In pig ear fibroblast cell culture study, the demethylation of *OPN* promoter was found in dose-dependent response of 5-aza-dc treatment and followed the *OPN* mRNA reexpression. In cloned pig study, discrepant expression pattern was identified in several cloned pig tissues, especially in brain, heart, and ear. Promoter assay data revealed that four methylated CpG sites presenting in the -2615 to -2239 nt region cause significant downregulation of *OPN* promoter activity. These data suggested that methylation in the *OPN* promoter plays a crucial role in the regulation of *OPN* expression that we found in cloned pigs genome.

## 1. Introduction

Nowadays, many of pathogenesis of diseases have been determined [1–3]. Methylation in the 5' cytosine in the CpG dinucleotides is crucial a mechanism that regulates gene expression without changing DNA sequence and can be inherited to the offspring [4]. The promoter region contains

various transcription factor binding motifs with numerous CpG dinucleotides. Some transcription factors are blocked by methylated CpG island resulting in inhibition of gene expression [5]. Somatic cell nuclear transfer (SCNT) technique is used to generate an identical genetic background offspring [6, 7]. However, SCNT cloning animals usually showed low survival rate and inappropriate methylation reprogramming

process [8]. This dilemma of SCNT animal may be caused by methylation controlled genes, such as imprinting genes [9].

*OPN* is an extracellular matrix protein and hydrophilic glycoprotein identified firstly in the bone as a sialoprotein. It contains a thrombin and transglutaminase cutting site, and the molecular weight is about 25 kDa to 75 kDa; in pig, the molecular is about 67 kDa; it contains numerous isoforms [10]. *OPN* has a hydrophobic N terminal; thus, it can be secreted out of cell membrane; the amino sequence of *OPN* is full of Asp, Thr, and Ser that can elevate the binding activity with calcium, glycosylation, and phosphorylation, respectively [11]. Thus, *OPN* plays numerous roles in many aspects, such as bone remodeling, cell migration, iNOS regulation, repairment, and leucocyte recruitment [12]. And acquired *OPN* expression has been found in a variety of cancer cell types, especially in the liver, lung, breast, prostate, colon, brain, and spleen [13, 14]. *OPN* is cleaved by MMPs protein to generate functional *OPN* that can bind to  $\alpha v \beta 3$  [15]. This integrin binding with *OPN* has influence on NF $\kappa$ B signaling transduction [16, 17]. Therefore, overexpressed *OPN* is associated with tumorigenesis, tumor invasion, and metastasis [18, 19]. Previous study suggested that overexpressed *OPN* induces the serious cardiac fibrosis [20, 21]. Thus, our cloned pigs were also surrounded by various defects in heart fibrosis and retardation of growth of bones. Therefore, this study focuses on the methylation change of *OPN* promoter that may be disrupted by inappropriate reprogramming process. Consequently, aberrant methylation of promoter could lead to aberrant expression of *OPN*. In the previous studies, *OPN* expression was induced with TSA (trichostatin A) in mouse undifferentiated mesenchymal cell line by AP1 site [22]. The TSA is a histone deacetylase inhibitor. It can lose the chromatin structure in order to let gene restore its expression. 5-aza-dc is also an analog with the same structure of cytosine without methyl group adding in the 5'C end [23].

Thus, 5-aza-dc addition leads to low methylation percentage in the CpG sites rich region. The hypomethylation status in the promoter may contribute its gene transcription activity. Porcine fetal fibroblasts in 5th passage cultures were treated with 0.5, 1.0, 2.0, and 3.0  $\mu$ M 5-aza-dc for 96 h; 5-aza-dc inhibited the growth of cell at all concentrations. 5-aza-dc induced a reduction of transcripts level in *DNMT1* and increasing expression in imprinted gene, *IGF2* [24]. Furthermore human *OPN* promoter sequence is similar to porcine in the front 400 nt of the porcine promoter. Therefore, we investigated *OPN* RNA and promoter methylation changes in the porcine ear fibroblast cell. Data showed that the elevated *OPN* expression and in 5-aza-dc treated fibroblast cell is due to the decreased methylation of *OPN* promoter. Cloned pigs samples had found extremely methylation changes, especially in the brain (99.75% upregulation), heart (11.50% down-regulation), and ear (18.03% down-regulation). Deletion analysis of the promoter region revealed 5-aza-dc induced luciferase response that was regulated by -2615 to -2239 of the *OPN* promoter. These data suggested that methylation in the *OPN* promoter plays a crucial role in the regulation of *OPN* expression. Methylation of *OPN* promoter may be an epigenetic marker of diagnosis of cancer.

## 2. Materials and Methods

**2.1. CpG Island Prediction.** The sequence of a putative CpG island in *OPN* promoter was analysed by using MethPrimer software (<http://www.urogene.org/methprimer/index1.html>).

**2.2. Cell Culture.** The porcine fibroblast cell line was grown in Dulbecco's modified Eagle's medium (DMEM; Gibco-BRL, Gaithersburg, MD, USA) supplemented with 10% fetal bovine serum (FBS, Gibco BRL) and containing 100 U/mL penicillin and streptomycin. The cells were incubated at 37°C in humidified incubator with 5% CO<sub>2</sub>.

**2.3. 5-aza-dc Demethylation Drug Treatment.** For 5-aza-dc treatment, porcine fibroblast cell in 5th passage cultures was treated with 5-aza-dc (sigma) at various concentrations, that is, 0 (control), 0.5, 1.5, and 2.0  $\mu$ M, for 72 h. Medium was changed every 24 h and then cells were collected for RNA and DNA extraction and stored at -80°C [24].

**2.4. Quantitative Real Time-PCR.** 2  $\mu$ g RNA of ear fibroblast cell was used to be transformed to cDNA. 0.5  $\mu$ L of cDNA was performed for quantitative real time-PCR with Rotor-Gene 6000 (Corbett).  $\beta$ -actin was the internal control for normalize target gene, *OPN*. The calculated gene expression fold from CT value was according to the previous study. *P* value less than 0.5 exhibited the obviously significant difference.

**2.5. Methylation Analysis by Combined Bisulfite Restriction Analysis (COBRA).** For amplification of porcine *OPN* promoter methylation analysis site, PCR was performed using 2  $\mu$ L of bisulfite-converted genomic DNA as template. The primer sets of COBRA were *OPN*-C sense 5'-TTTTTTGAGGGAGATTAGTTTTTG-3' and antisense 5'-ATTCTACTAAAATCCAACCACCC-3'. The COBRA-PCR products were purified by phenol/chloroform, followed by ethanol precipitation. The DNA was resuspended in 8.5  $\mu$ L of distilled deionized water. Purified PCR products were then digested with 10 U *Bst*UI restriction enzyme (New England Biolabs, MA, USA) at 65°C. Products were electrophoresed on 6% native acrylamide gel, stained with 200 g/mL ethidium bromide, and visualized using a Kodak 1D software.

**2.6. Methylation Specific-PCR.** Genomic DNA (0.5  $\mu$ g) was treated with sodium bisulfite according to the manufacturer's recommendations (EZ DNA Methylation Kit; Zymo research, CA, USA) and amplified with specific primers for methylated or unmethylated DNA. The primer sets of MS-PCR were *OPN*-M sense 5'-AAGCGGGGAAGGAGTTATTACGT-3', antisense 5'-TCCGACAAAACGAACGATCATAACA-3', *OPN*-U sense 5'-GAAGTGGGGAAGGAGTTATTATGT-3', and antisense 5'-CAATAACTCCAACAAAACAATC-3'. All PCR reactions were performed on PTC 200 thermocyclers (MJ Research, MA, USA) and in 25  $\mu$ L volume using the PlatinumTaq DNA polymerase system (Invitrogen, CA, USA). PCR products were separated on 1.5% agarose gels. The M-set primers contained at least three

CpG sites to distinguish the methylation status of investigated region. And U-set primers overlapping the M-set primers were used to amplify the unmethylated region.

**2.7. Plasmid Constructs.** A full length pig *OPN* promoter (-2615-luc) was amplified from wild-type pig heart tissue cDNA. This fragment was cloned into a luciferase fusion plasmid, pGL3-Enhancer vector (Promega), to generate p*OPN*-full-luc. *HindIII* and *NcoI* cutting sites were used for cloning. Three truncated forms of p*OPN* promoter were prepared by PCR using the p*OPN*-full-luc as a template and using synthesized oligonucleotides as follows: p*OPN*-full-luc: sense, 5'-AAGCTTGAATTCACCTCGTCTTTCCCTTGAGA-3', and antisense, 5'-CCATGGGCTGACAGCCTGGACCTCCCC-3'; -2239-luc: sense, 5'-AAGCTTCCTATAACTGTCTACGTTTCATATTAGAC-3', and antisense, 5'-CCATGGGCTGACAGCCTGGACCTCCCC-3'; -1505-luc: sense, 5'-AAGCTTAATTTTCATTTAAGTAACCAACTTTATATATC-3', and antisense, 5'-CCATGGGCTGACAGCCTGGACCTCCCC-3'; -495-luc: sense, 5'-AAGCTTGCCTGAACAATATAGCCTTGTCGC-3', and antisense, 5'-CCATGGGCTGACAGCCTGGACCTCCCC-3'. The sequence of constructs was confirmed by DNA sequencing. There were two-point mutation different from NCBI: one is 287A to T and the other is 957T to A.

**2.8. Transient Transfection and Luciferase Assay.** Pig ear fibroblast cells were transfected using the Lipofectamine 2000 (Invitrogen). Fibroblast cells were incubated at a density of  $8 \times 10^5$  cells into 35 mm diameter dishes. After 24 h when cell was adherent to the dishes, 3  $\mu$ g of reporter plasmid DNA was transfected for 6 h in Lipofectamine mixture (Invitrogen). 24 h after the transfection, cell lysates were collected for a luciferase assay. The luciferase activity of the cell lysates was detected by Dual-light system (Applied biosystems). The activity data was measured with PARADIGM Detection Platforms (Beckman Coulter). Luciferase activity was normalized with 1  $\mu$ g  $\beta$ -gal plasmid. All luciferase assays were carried out in triplicate.

**2.9. In Vitro Methylation of the *OPN* Promoter Region.** The *OPN* reporter construct -495-luc was methylated by incubation with *SssI* methyltransferase (New England BioLabs). The -2615-luc construct was methylated by *HhaI* and *HpaII* methyltransferase (New England BioLabs) for 16 h at 37°C. The methylation status was also verified by digested with *HhaI* and *HpaII* enzyme.

**2.10. Electrophoresis Mobile Shift Assays.** Nuclear extracts were prepared from HEK293T cells. Two probes were designed for methylation binding activity test. Two probes containing either 6-8 CpG sites or 13th CpG site in the -2615 to -2239 of the *OPN* promoter were generated by annealing two complementary oligonucleotides (*OPN* 6-8: 5'-TGCATGATCGTTCCGTCCTGCCGGAGTCACTGACGGAACCAAGACCGAGGT-3'; 5'-ACCTCGGTCTGGTTCGTCAGTGACTCCGGCAGGACGGAACGATCATGCA-3', the predicted core

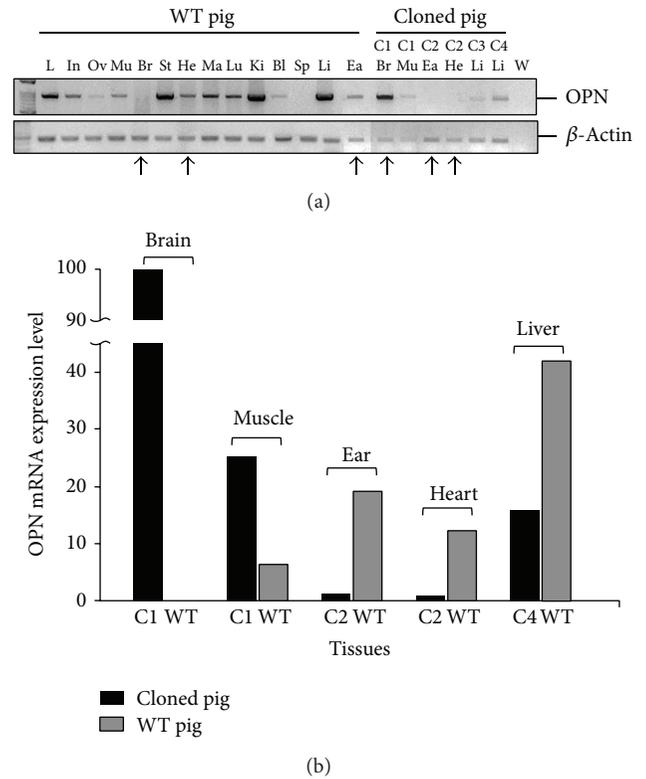


FIGURE 1: (a) (b) The semiquantitative RNA expression of *OPN*. C1-C4 indicated four different cloned pigs. The black arrows represent the contrary expression in the cloned pig tissues relative to WT tissues. Br: brain, Ea: ear, He: heart, Ki: kidney, Li: liver, Lu: lung, Mu: muscle, Sk: skin, In: intestine, Sp: spleen, Pl: placenta, Um: umbilical cord, B: blood; S: blood treated with *SssI*, and W: ddH<sub>2</sub>O.

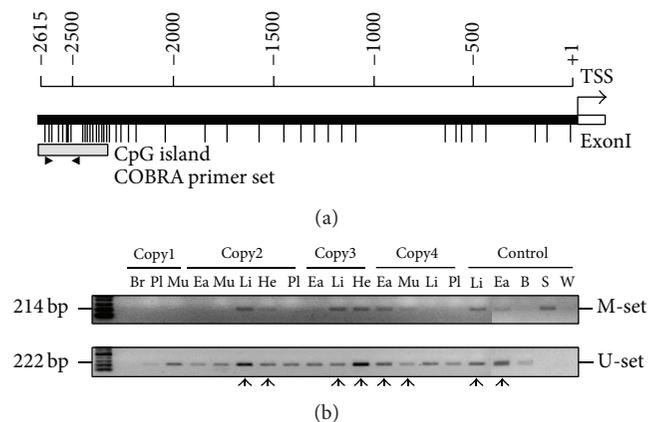


FIGURE 2: MS-PCR analysis of *OPN* promoter. (a) The distribution of CpG sites in the pig *OPN* promoter. The two arrows indicated the MS-PCR primer sets. (b) The MS-PCR results of *OPN* promoter in cloned pigs. The arrows represent both existence of methylation and unmethylation DNA element. Br: brain, Ea: ear, He: heart, Li: liver, Lu: lung, Mu: muscle, Pl: placenta, B: blood; S: blood treated with *SssI*, and W: ddH<sub>2</sub>O.

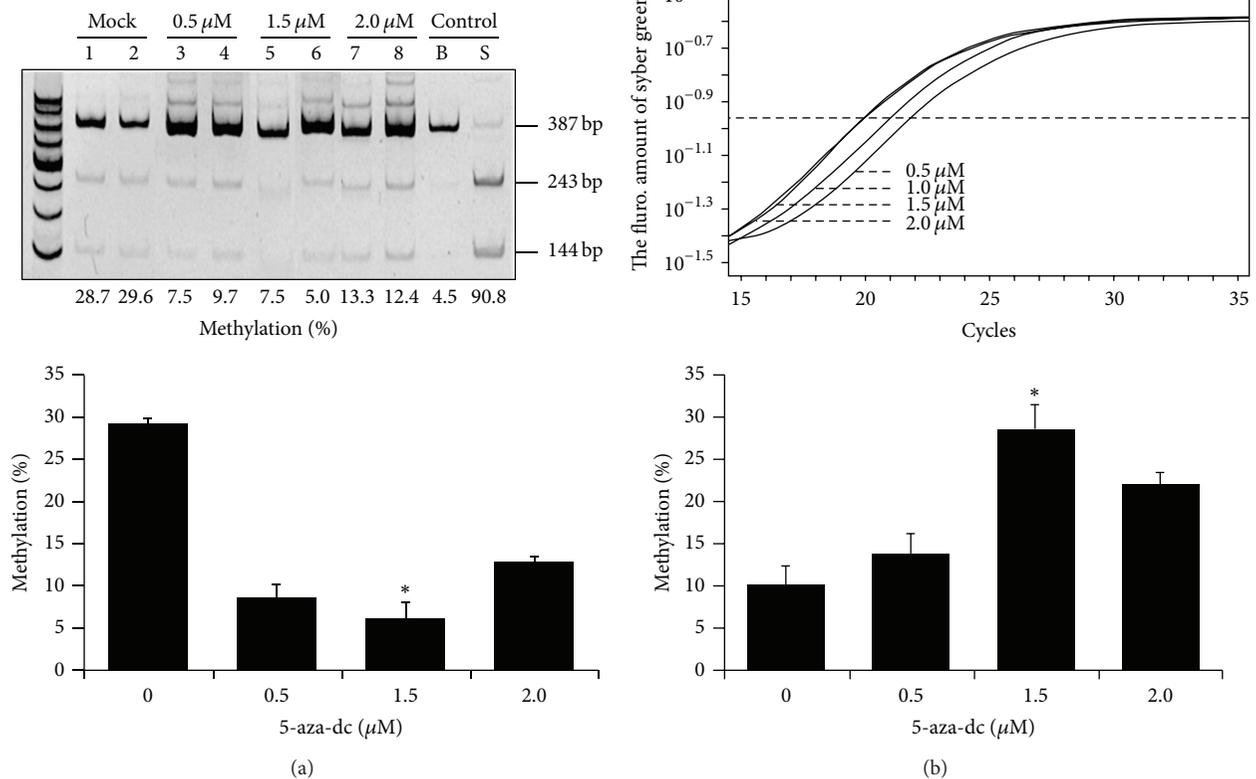


FIGURE 3: *OPN* RNA expression and DNA methylation of 5-aza-dc treated pig fibroblast cell. (a) The schematic showed the methylation percentage and mRNA expression of *OPN* promoter in different concentrations 5-aza-dc. (b) The COBRA analysis of *OPN* promoter. Below the square is the methylation percentage of *OPN* promoter methylation.

sequence of the AP1 binding site is underlined; *OPN*13th: 5'-CCTCCGTGTTCCCTGTTAATGTGT-AGCGCGTCGTTGTTGGGAAATAGTTC-3'; 5'-GAACTATTTCCCAACAACGACGCGCTACAC-ATTAACAGGGAACACGGAGG-3'; the predicted core sequence of the ADRI binding site is underlined). The transcription factor prediction software is TFSEARCH 3.0 version. The probes were labeled with  $\gamma$ -<sup>32</sup>P-ATP by using T4 kinase (Promega). Annealing probes also were methylated with SssI methyltransferase (NEB). Nuclear extracts containing 5.6 μg of the protein were preincubated in 20 μL of binding buffer (50 mM Tris-HCl (pH 8.0), 750 mM KCl, 2.5 mM EDTA, 0.5% Triton-X 100, 62.5% glycerol (v/v), and 1 mM DTT) with or without unlabeled competitor (10-fold molar excess). For supershift assay, antibody of AP1 was added to the preincubation buffer. After 10 min of preincubation on ice, the DNA probe labeled with [ $\gamma$ -<sup>32</sup>P]-ATP was added, and the mixtures were incubated at room temperature for 30 min. The reaction mixtures were resolved on 6% polyacrylamide gels. The gels were dried and subjected to PhosphorImager analysis using a Typhoon system and ImageQuant TL software (Amersham Biosciences, Sunnyvale, CA, USA).

**2.11. Molecular Modelling.** The molecular docking was then further analyzed for proving the further mechanism of our

findings; we further surveyed the interaction of AP1 (c-Jun) and TFIIB by computational biology. The AP1 (c-Jun) is possible higher spot for hypermethylation in *OPN* promoter area and provides the binding domain for RNA-polymerase II initial binding transcription factor (TFIIB) in this study. Therefore, we first utilized the Z-DOCK program to simulate the structures of c-Jun and TFIIB. After that, we further used molecular dynamics (MD) to validate the stability of c-Jun and TFIIB complex under the GROMACS 4.5.5 program [25] with charmm27 force field. The model is set in the TIP3P water modeling in 1.2 nm distance of box for water box setting. Na and Cl ions in the concentration of 0.145 M NaCl model are used for system neutralization. All bonds are fixed by linear constraint solver (LINCS) algorithm to constrain all bonds lengths in the simulation system. Newton's Law is utilized for calculating the motion of molecular dynamics as follows:

$$\frac{d^2r}{dt} = M^{-1}F. \quad (1)$$

The Particle mesh Ewald (PME) is also used for calculate the coulomb type of electrostatics. The Van der Waals (VDW) interactions are set as 1.4 nm cut-off distance for nonbound interaction. The first step is set on the 5,000 cycle steps performed in the manner of Steepest Descent algorithm for energy minimization. And then, equilibration was performed

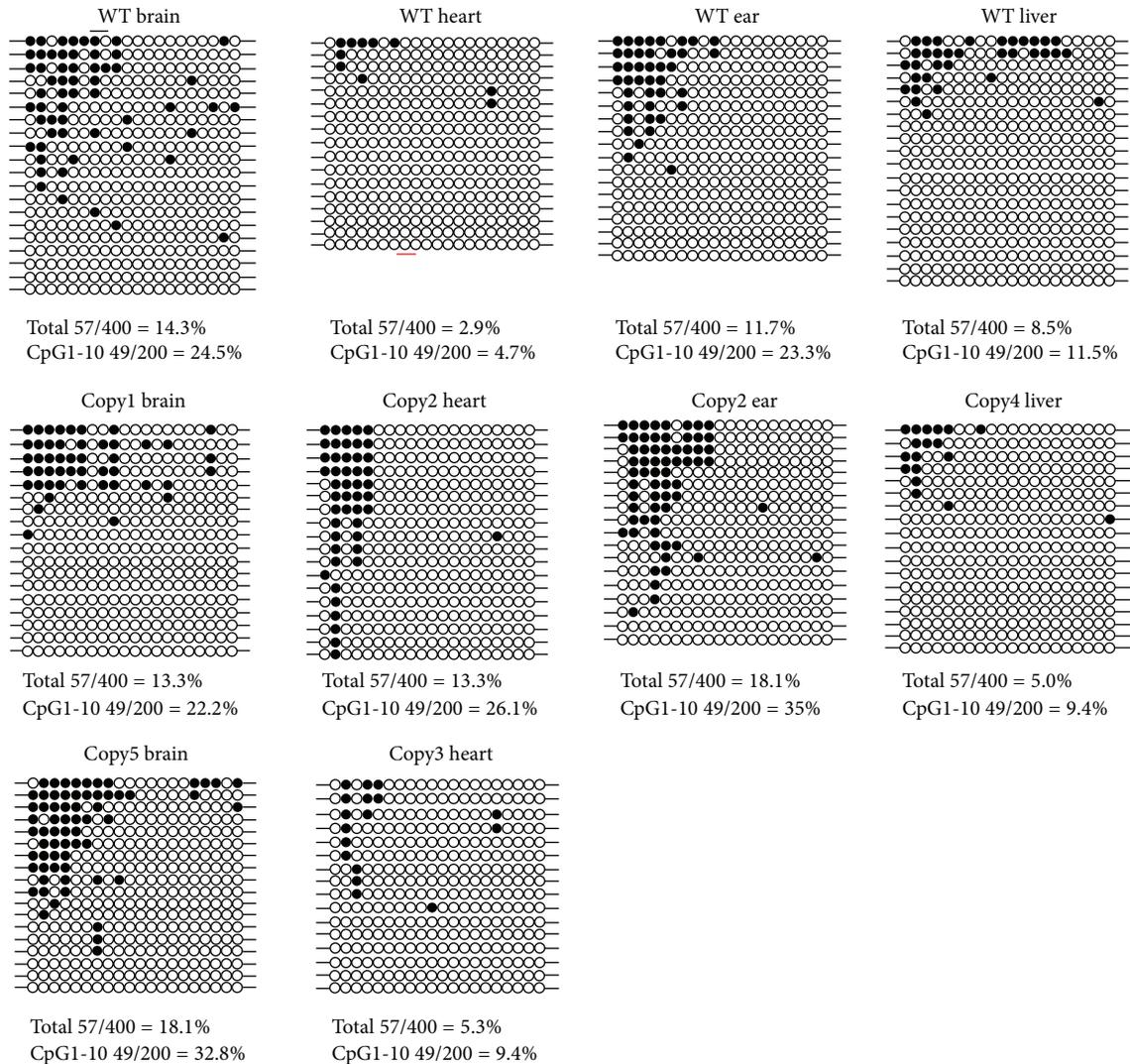


FIGURE 4: The bisulfite sequencing of *OPN* -2610~ -2400 nt upstream the promoter in cloned pigs' tissues. The closed circles represent the methylation CpG sites. The hollow circles represent the unmethylated CpG sites. The bottom number indicated the methylation percentage of each sample. The range of square showed that the region may be the methylation controlled region of *OPN* promoter.

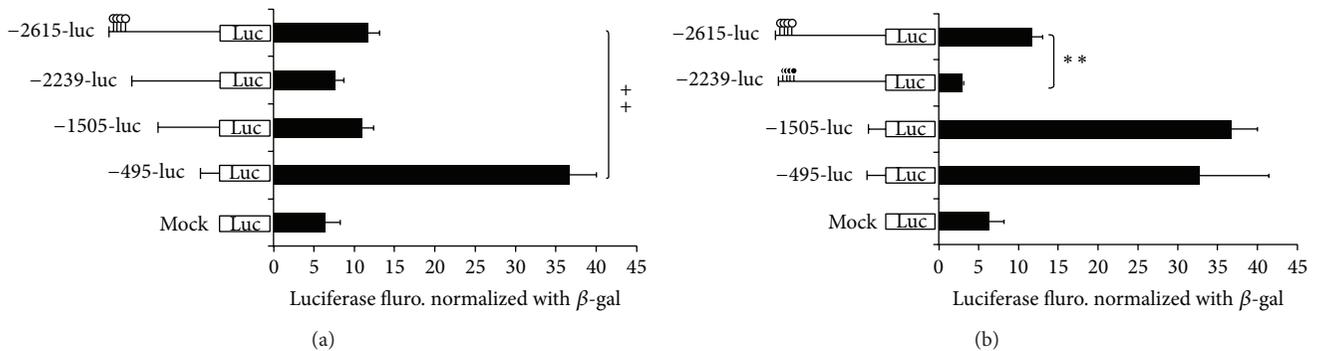


FIGURE 5: Methylation and deletion analysis of *OPN* promoter in 293T cells. The match-like bar with black circle represents the methylation CpG site; white circle of match bar indicated the unmethylation CpG sites. PGL3 vector as standard; in 293T cell line; PGL3-enhance vector as negative control; cell lysate as the background; pCMV-b-gal as internal control. The relative value is adjusted by cell lysate; -495 M-luc indicated the methylation in PGL3 backbone with *HhaI* and *HpaII* methyltransferase. (\*\* $P < 0.01$ ); -2615-luc and -2615 M-luc:  $n = 3$ ; -495-luc and -495 M-luc:  $n = 4$ . The experiments were repeated three times and the results were analyzed and presented as the mean  $\pm$  SE.

```

ttcctttgaggagaccagctctt1gAGCGAGTGTGGGAA2CGGGGAAGGAGCCCATCA3
      4      5
GTCACCTGCGTTTGCTAAAGACAACAGAGCAGAAAAAGAA4CGCTCTGCT
      6      7 8 9
TCTCTTGGCCTCGTGTTCCTGTTAATGTGTAGCGCGT7CGTTGTTGGGAA
      HhaI
ATAGTTCTCACCTGACTTTCCAAGAAATGGAGGGCCTCACAGTTGTTTGA
      10      11 12      13      14
TGGCTCGGTCATTAAATGCATGATCGTTC11CGTCTCGCGGAGTCACTGAC14
      HpaII API-like
GGAACCAGACCGAGGTCTCAGGTCCTTCTCCGAAATGCTGCCATCGTGTG
      15      16      17
GCACCTCGGAGCCATGACCGGAAGAGCCCTATGGGTATATGGTTCAGCG20
      HpaII      HhaI
CAgggtggctggactccagcagaatct
    
```

FIGURE 6: The sequence and CpG sites distribution of pig *OPN* promoter. 20 CpG sites exist in the front region of *OPN* promoter. The underline indicates the predicted API-like binding site. The gray marker indicates two methyltransferase sites of *HhaI* (GCGC) and *HpaII* (CCGG).

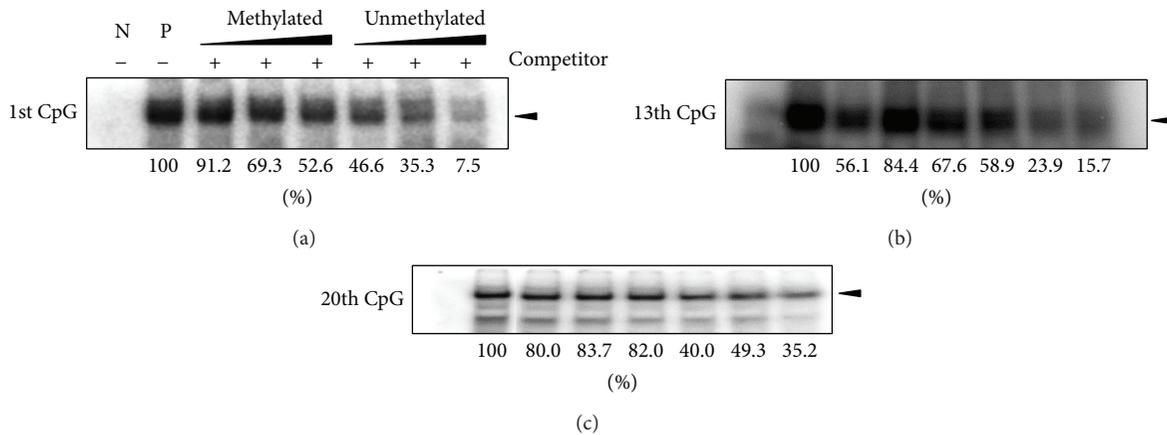


FIGURE 7: Electrophoresis mobile shift assay in porcine *OPN* promoter CpG sites. (a) Using CpG1 contained DNA element binding with SH-SY5Y nuclear extract. Methylated and unmethylated competitors were used as 2X, 5X, and 10X concentration than isotope labeled probe. (b) CpG 13th probe of *OPN* promoter binding with HEK-293 nuclear extract. (c) CpG 20th of *OPN* promoter element binding with HEK-293 nuclear extract. There was no difference between methylated or unmethylated competitors in competition experiment.

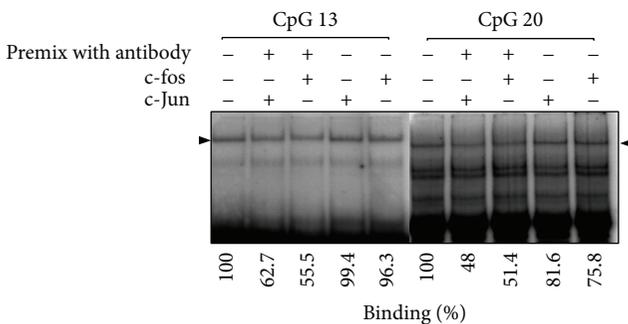


FIGURE 8: The CpG 13 and CpG 20 showed competition binding activity characteristics with c-Jun and c-Fos. The competition reduced the binding activity between transcription factors and labeled *OPN* EMSA probe. The amount of EMSA c-Jun and c-Fos antibody is 3  $\mu$ g. The EMSA probe is 30  $\mu$ g. Adding the Ab with a pre-mix way can reduce the binding effects with the transcription factors. It is suggested that the pre-mix with antibody blocks the access to its binding site (antibody-transcription factor-DNA).

with a time period of 1 ns for position restraints set under the constant temperature dynamics (NVT type) conditions. The third step is calculating the production run for 5000 ps under constant pressure and temperature dynamics (NPT type). All the MD systems are set by 310 K temperature during all simulation times. MD frames data were saved every 20 ps for all production runs.

**2.12. Molecular Dynamics Analysis.** First we survey the stability of all atoms performed by using the GROMACS 4.5.5 software though the commands of *g\_rms* and *g\_gyrate* to calculate root mean square deviation (RMSD) and radius of gyration (Rg), respectively. Secondly, we calculate the total energy for all the systems by the command of *g\_energy*. Thirdly, we further calculate root mean squared fluctuation (RMSF) for each protein residue by commands of *g\_rmsf*. Fourthly, the distance between c-Jun and TFIIB and movement analysis are calculated by the *g\_dist* program. Fifthly, the migration of dock protein (c-Jun and TFIIB) is

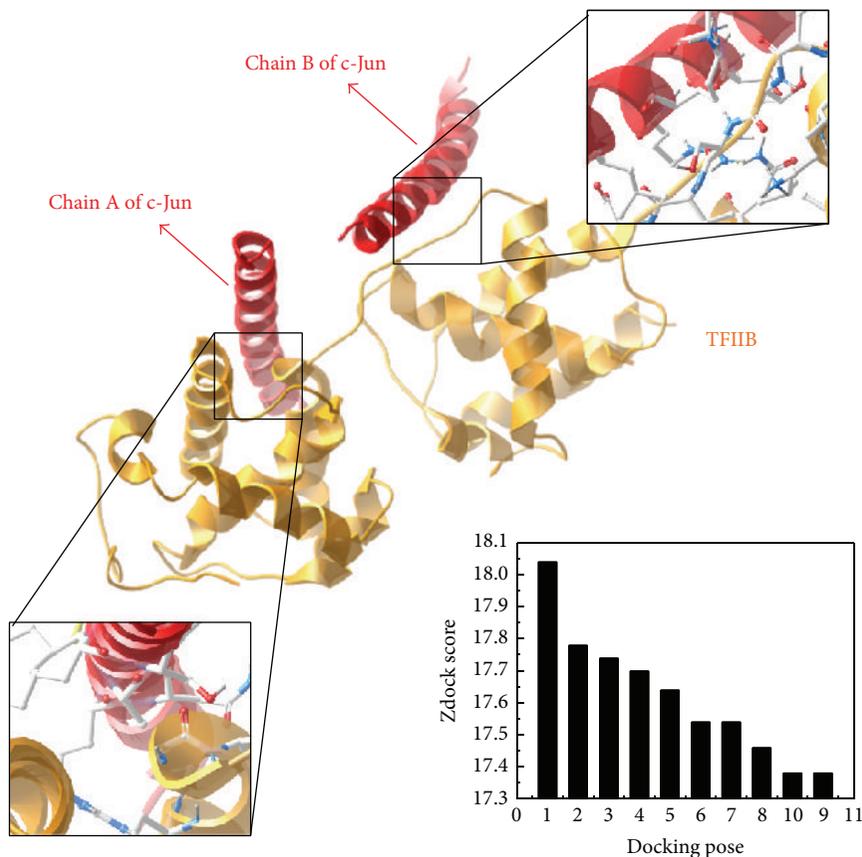


FIGURE 9: The best docking pose of c-Jun and TFIIIB with 18.04 Zdock Score. The structures of c-Jun and TFIIIB are colored in red and orange, respectively.

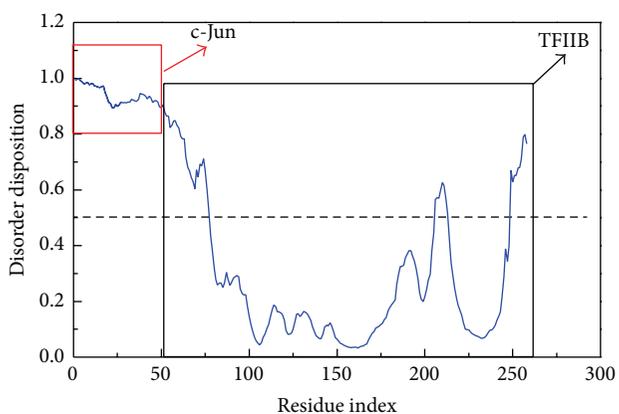


FIGURE 10: Disorder prediction of c-Jun and TFIIIB complex; the value of disorder disposition below 0.5 indicates order folding region. The sequence of c-Jun is in the region from residue index 0 to 54, and the sequence of TFIIIB is in the region from residue index 55 to 258. The folded structure of c-Jun reveals disorder; the N-terminal and C-terminal of TFIIIB structure display folded disorder.

presented by mean square displacement (MSD) under the command of `g_msd` module in GROMACS during all the simulation times. Sixthly, the `g_cluster` program is selected

for further calculation of the representative structure from all MD frames, and the representative structure is taken for further snapshot analysis. We also employed DSSP analysis and matrices of the smallest distances between each residue to investigate the stability of the protein structure. The principle component analysis (PCA) is then applied to observe the protein motion changes during all the MD frames. Finally, in order to observe the compactness between c-Jun and TFIIIB, Caver 3.0 software [26] was used to predicted space in the complex.

### 3. Results

**3.1. Distribution of Porcine OPN CpG Island.** There are dense CpG sites existing in the front of the OPN promoter region. One putative CpG island was found (CpG island size > 100, GC Percent > 50.0, Obs/Exp > 0.6) by MethPrimer program.

**3.2. Methylation and Expression Analysis of OPN in Cloned Pig Various Tissues.** Firstly, the OPN mRNA expression was investigated in the WT pig tissues and cloned pig tissues. Data showed the various expression levels in different tissues. Particularly, Copy1 brain overexpressed the OPN and Copy2 ear with no expression of OPN relative to its wild-type tissue, respectively (Figure 1). The unique aberrant

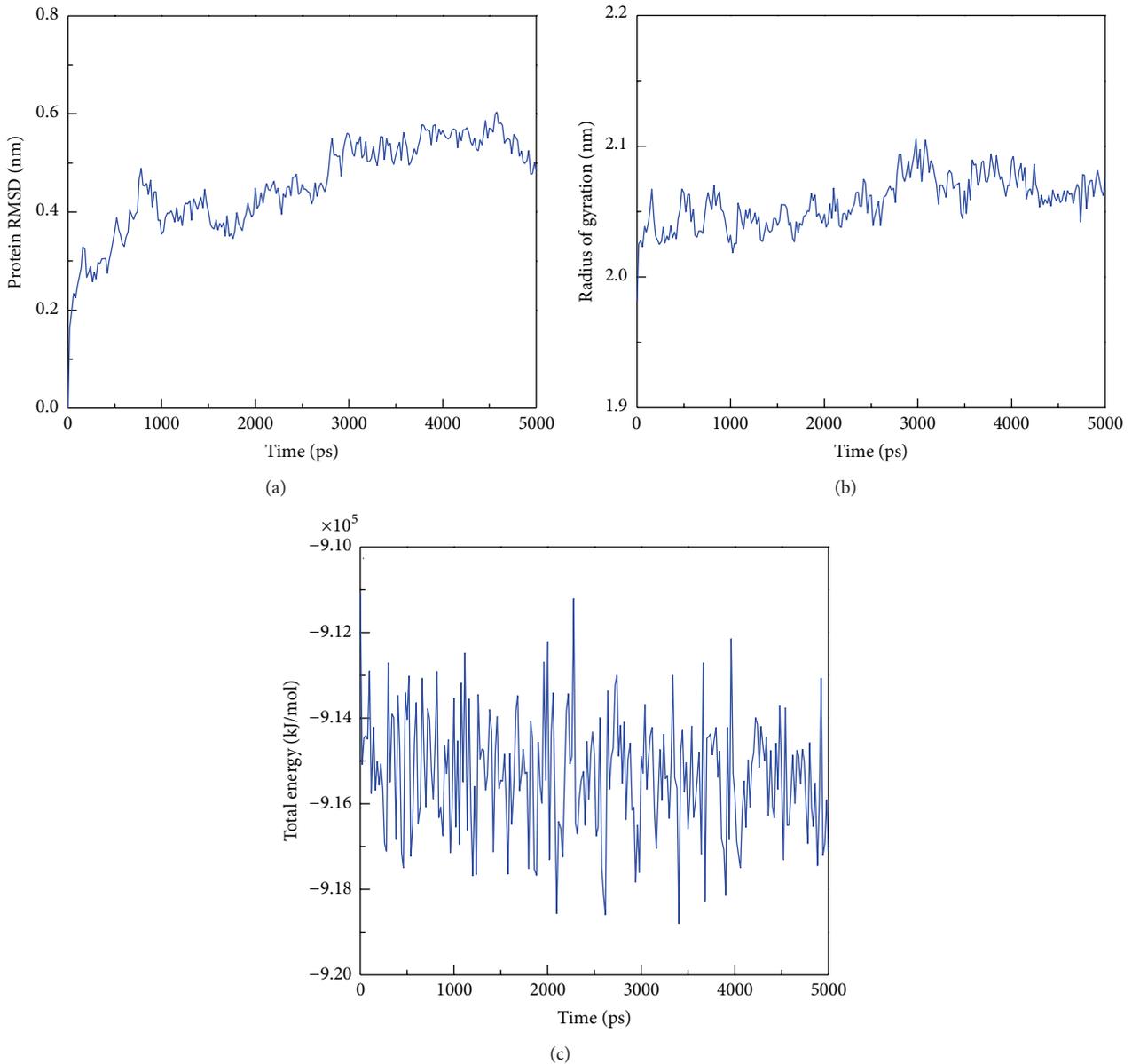


FIGURE 11: The trajectory analysis of c-Jun and TFIIB during 5000 ps simulation times. (a) RMSD values of all atoms of c-Jun and TFIIB complex; (b) radius of gyration of c-Jun and TFIIB complex for identifying the compactness of protein structure; (c) total energy of all simulated systems of c-Jun and TFIIB complex; the total energy is sum of potential energy and kinetic energy.

expression patterns exhibited the different control way of the *OPN* expression. We proposed that *OPN* expression may be a tissue-specific manner. MS-PCR primers were designed to estimate the methylation status of *OPN* promoter. Hypomethylation generally appeared in the various tissues of cloned pigs. However, there were still some tissues that showed the methylated region in *OPN* promoter (Figure 2).

**3.3. 5-aza-dc Increases *OPN* mRNA and Decreases Methylation of *OPN* Promoter in Pig Ear Fibroblast Cell.** In order to realize whether the methylated *OPN* promoter affects the

activity of *OPN* promoter. The 5-aza-dc treated porcine ear fibroblast cells showed that when the concentration of 5-aza-dc level increased, it will decrease the methylation of *OPN* promoter and restore the *OPN* RNA expression at 0.5 to 2.0  $\mu\text{M}$  (Figure 3). The results suggested that the activity of *OPN* promoter can be affected by DNA methylation directly or indirectly. COBRA assay was also used to investigate the methylation status of *OPN* promoter in WT tissues. Brain, ear, liver, and lung tissues exhibited little part methylation of the *OPN* promoter (data not show). This is thought that the methylation of *OPN* promoter in the aforementioned tissues may be involved with *OPN* transcript regulation mechanism.

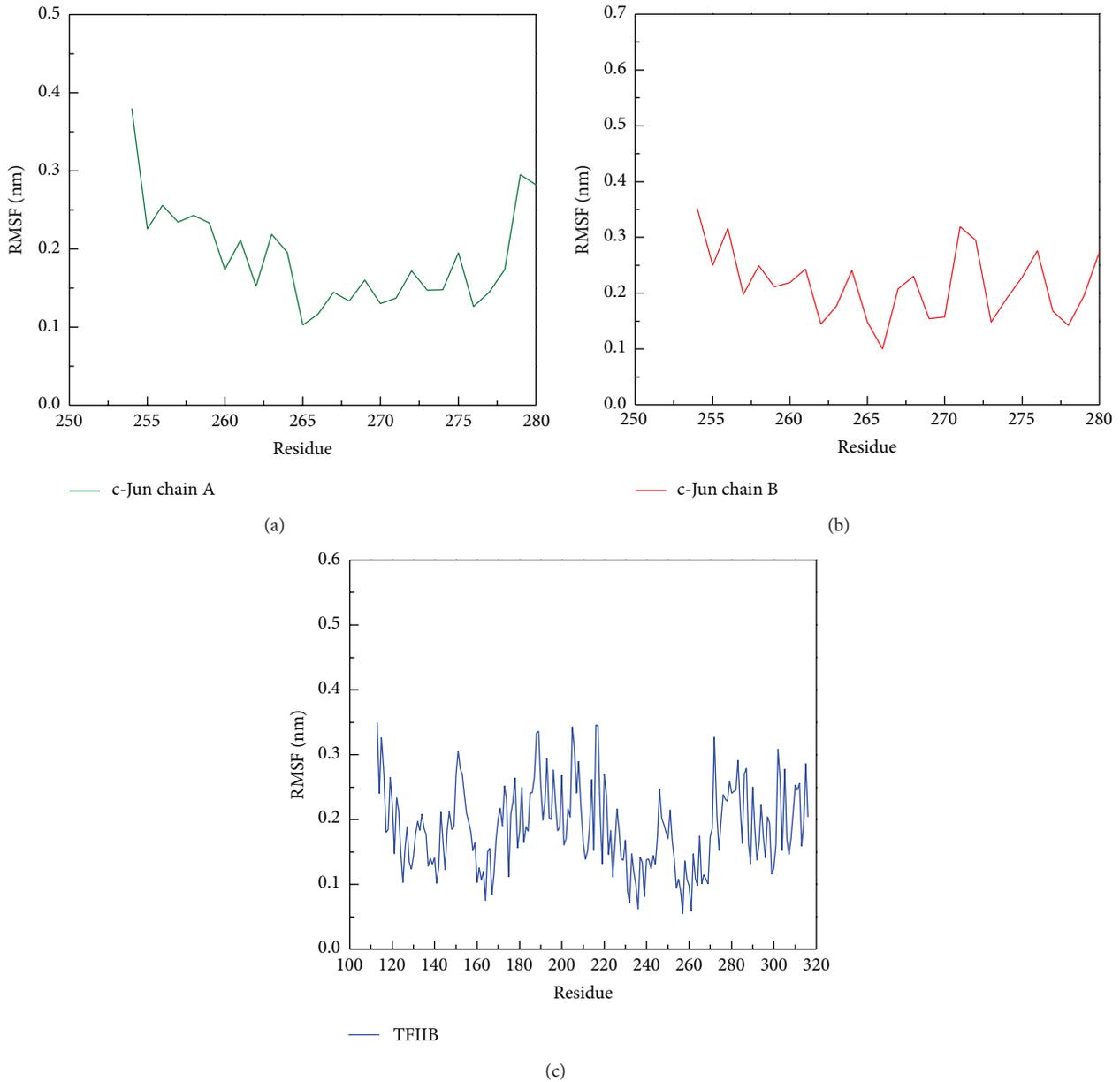


FIGURE 12: RMSF analysis of protein resides on (a) chain A of c-Jun, (b) chain B of c-Jun, and (c) TFIIB during simulation time of 5000 ps. The residue index of chain A and chain B of c-Jun is from 254 to 280, and the residue index of TFIIB is from 113 to 316. The high values of RMSF indicated the high fluctuation of residue during all simulation times.

3.4. Bisulfite Sequencing Analysis of the Whole CpG Sites Methylation Profile in Cloned Pigs. To investigate which region of the *OPN* promoter is affected by methylation in the CpG site, bisulfite sequencing was performed to dissect the methylation status of CpG sites in the *OPN* promoter. In the brain tissue, bisulfite sequencing of Copy5 and Copy1 brain exhibited the saturated status in their methylated region while WT brain exhibited fragmentary methylated CpG sites. The 18.06% methylation percentage of Copy5 brain was more than the WT brain, 14.25% (Figure 4). In the

heart tissues, bisulfite sequencing showed that the Copy2 heart had extremely hypermethylated percentage with 13.33% more than WT heart, 2.94% (Figure 4). Particularly, the data showed the inhibition of Copy2 heart mRNA (Figure 1). In the liver, the different methylation pattern also appeared in the Copy4 liver relative to WT liver. Copy2 ear with 18.05% methylation in the *OPN* analyzed region was higher than the WT ear, 11.67%; the result was proved by the experiment of *OPN* mRNA expression (Figure 1). These bisulfite sequencing results matched our previous hypothesis that methylation in

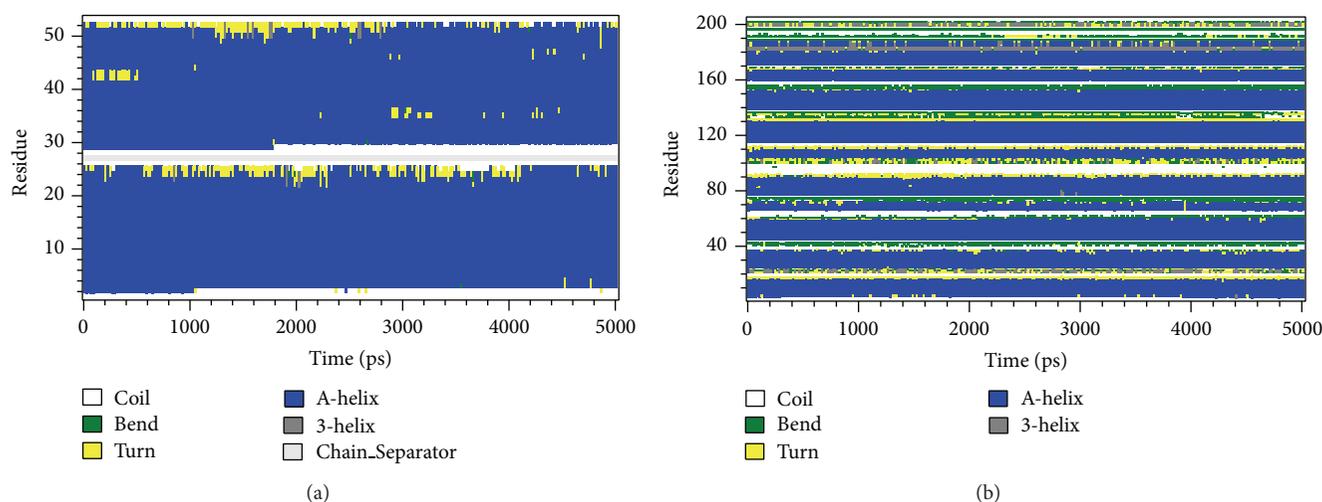


FIGURE 13: The secondary structure analysis for (a) c-Jun and (b) TFIIIB over all simulation times. Each secondary type of the structure such as Bend, Turn, alpha helix (A-Helix), and  $3_{10}$ -helix (3-helix) is colored in green, yellow, blue, and gray, respectively. The “Chain\_Separator” in (a) is used to differentiate between chain A and chain B of c-Jun.

the *OPN* promoter region regulates the activity of *OPN* promoter. Moreover, the hypermethylated *OPN* promoter may directly affect the activity of *OPN* transcription, especially in heart and ear tissues.

**3.5. Analysis of Methylation Implication of *OPN* Transcription by Promoter Assay.** The analyzed *OPN* promoter region may involved with methylated control of gene transcription. Promoter assay was designed to explain the direct inhibition of *OPN* promoter activity by methylation on the *OPN* CpG sites. Four different truncated forms of *OPN* full length (2.6 kb) were used to prove the hypothesis (Figure 5). Particularly, the truncated form 2.2 kb deleted the 377 bp promoter region (–2615~–2239 nt). This region is the analyzed region for bisulfite sequencing profile. And this region is thought to be the most possible element that regulates the *OPN* transcription. Moreover, four CpG sites were *in vitro* added to the methyl group by methyltransferase that can provide important evidence how methylation affect the *OPN* transcription. Figure 5(b) shows that the *OPN* promoter activity was significantly decreased in methylated –2615-luc plasmid. However, –2239-luc that deleted the 377 bp containing methylated characteristic DNA element leads to less inhibition of promoter activity than the methylated vector –2615 M-luc (Figures 5(a) and 5(b)). The results indicated that methylation in the front *OPN* promoter is not only decreasing the promoter activity to the basal level but also recruiting the inhibition factors to enhance the inhibition ability. In order to avoid the effects of CpG sites in PGL3-enhancer backbone, –495 M-luc that have no methylated CpG sites in the *OPN* promoter but it can be methylated in the vector backbone CpG sites compared to –495-luc. Data showed that there is no difference in the promoter activity between methylation or unmethylation in the PGL3-enhancer backbone CpG sites (Figure 5(b)). It is suggested that methylation in the critical region, such as *OPN* promoter

front end, may lead to the rearrangement of chromatin structure. Otherwise, deletion of the *OPN* promoter to the 495 bp with significant promotion of the promoter activity indicated that in the middle part of promoter DNA element may able to inhibit the activity of *OPN* promoter.

**3.6. Methylation in CpG 13th and CpG 1st of *OPN* Promoter Blocks the Binding Access of Transcription Factors.** We investigate that the methylated CpG sites in the –2615~–2239 nt of the *OPN* promoter region affect the transcription factor binding activity. Electrophoresis mobile shift assay was performed with nuclear extracts from human HEK293T and SH-SY5Y cell line. Four EMSA probes that contain the CpG 1, CpG 3-4, CpG 6-8, CpG 11-15, and CpG 19-20 were designed according to the CpG sites in our analyzed region (Figure 6). EMSA data suggested that CpG 13th and 1st sites showed methylation noncompetition phenomenon which had influence on binding with transcription factor (Figure 7). 19-20th CpG sites containing EMSA probe showed no competition ability in the methylated or unmethylated status. The premix with antibody and nuclear extracts by EMSA assay indicated that the c-Jun and c-Fos were involved in the binding to CpG sites 13 and 20 (Figure 8). However, the adding of antibody in the mixture of probe and nuclear extracts showed no significant shift bands. It is indicated that other transcription factors may also participate in the transcription activity *OPN* promoter. Thus, the c-Jun and c-Fos could be involved in the partial *OPN* transcriptional activity in a competition way. In addition, the c-Jun had higher binding affinity than c-Fos in the EMSA probe analysis. We therefore did followed computational survey to see why c-Jun affects the consequent *OPN* transcription.

**3.7. The Computational Biology Results.** After surveying the possible zone for *OPN* promoter hypermethylation, we found that the API (c-Jun) sequence frequently appeared in our

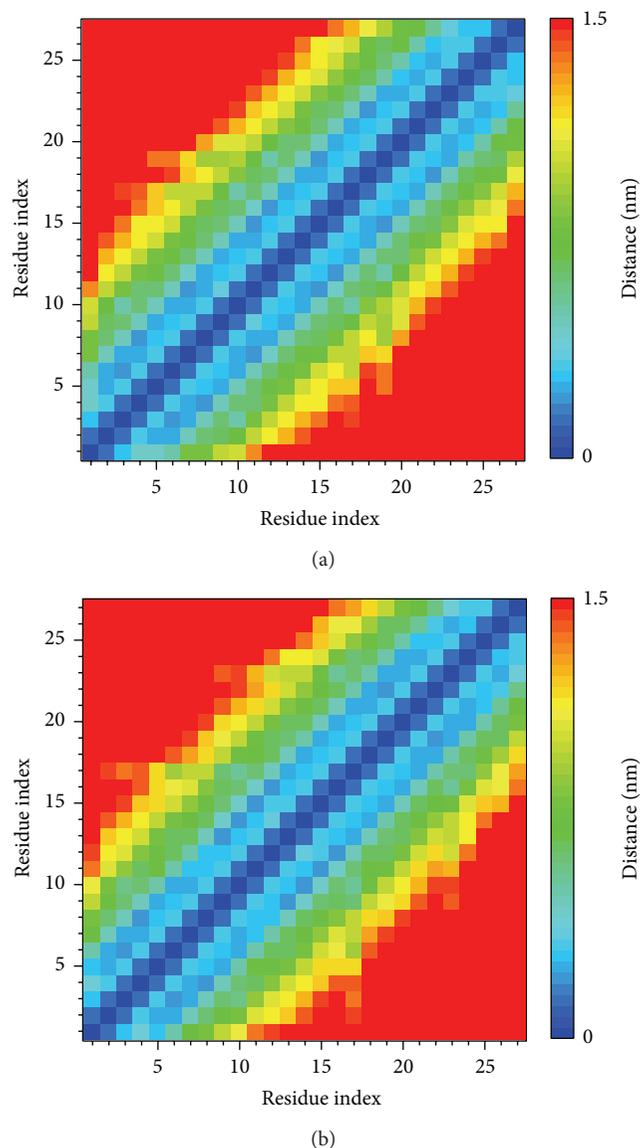


FIGURE 14: Matrices of smallest distance between each residue on (a) chain A of c-Jun and (b) chain B of c-Jun. The value of distance between residues is represented by rainbow bar, and the value of distance with longer than 1.5 nm is colored in red. The indexes of residues from 0 to 27 indicate residues from 254 to 280 on each chain of c-Jun.

molecular laboratory study. Then we further validated that the c-Jun methylation will cease the further mRNA production by inhibiting the binding of RNA-polymerase II initiation factor TFIIB. We do the further computational modelling for mechanism survey between the c-Jun and TFIIB. By the Z-DOCK analysis, we found the c-Jun and TFIIB could combine tightly (Figure 9). We chose the highest docking pose (dock score = 18.04) for further MD analysis.

Then the disorder predication was employed to observe the protein folding analysis and the result is shown in Figure 10. We found that the c-Jun has relative high disorder in folding than TFIIB; we suppose that the flexibility of c-Jun

structure could easier bound to TFIIB. This finding could be an explanation why the c-Jun bound to TFIIB by Z-DOCK program. To confirm the stability of the c-Jun and TFIIB complex, the series of molecular dynamic studies further visualize their interactions.

Protein complex RMSD analysis proved that the c-Jun and TFIIB were stable from 3000 ps to 5000 ps. In addition, we found the TFIIB are easier to be stable during the molecular dynamics (Figure 11(a)). We also found the radius of gyration tend to be stable for all simulation times with average of 2.05 nm (Figure 11(b)), suggesting that the two protein structures are compact after binding together. Figure 11(c) also shows the binding complex in a stable fluctuation and the energy of the binding complex is stable around  $-9.15 \times 10^5$  (kJ/mol).

**3.8. Stability Analysis of Residues on the Major Binding Region during MD Simulation.** To analyze the flexibility of residues on protein structure, the RMSF calculation was used to observe the flexibility of each residue; Figure 12(a) shows that the chain A of c-Jun had high frequency of fluctuation (binding site 200–210 binding residues). However, the chain B of the c-Jun has relative fewer frequency of fluctuation and the major binding region (from 228 to 240 residues) showed a less fluctuation as shown in Figure 12(b). Figure 12(c) reveals the binding regions (200–210 binding residues) for chain A of c-Jun that is more fluctuated and unstable compared to the binding regions (228–240) for chain B of c-Jun binding site in the TFIIB binding region. Figure 13 is the result for secondary structure variation calculated by DSSP analysis. Most of the main scaffold belong to alpha Helix; there are no significant changes during the whole MD simulation. All helices of the secondary structure for c-Jun and TFIIB binding remained stable during a 5,000 ps simulation time (Figure 13). We thereafter surveyed the distance between each residue of c-Jun for 5,000 ps. The variation of distances between residues in c-Jun chain is wider than the distances of residues in c-Jun chain B. Therefore, the chain B of c-Jun is more stable for TFIIB binding (Figure 14).

The hydrophobic area was then calculated by SASA in Figure 15(a); the value of hydrophobic area decreased during the last 1000 ps. This indicated the compactness of the c-Jun and TFIIB binding increased by the MD time period in our study. It is worthy to know that the distance between centrals of masses of c-Jun and TFIIB was decreasing more and more after time goes by in the 5000 ps survey (Figure 15(b)). In the migration analysis of c-Jun and TFIIB, the MSD was employed to count the migration of c-Jun and TFIIB. The c-Jun is more unstable than TFIIB during the binding interaction throughout the whole MD simulation in 5000 ps period (Figure 15(c)). Besides, we further utilized the principal component analysis (PCA) to measure all MD frames over all simulation times. The first two eigenvectors (PC1 and PC2) were shown in Figure 18; most of frames are ranged in the short range of eigenvalues  $-10$  and  $10$  in PC1 (Figure 16(a)), and arranged in eigenvalues from  $-5$  to  $5$  in PC2 (Figure 16(b)). The phase space comparing for PC1 and PC2 was shown in Figure 17; we found that each frame

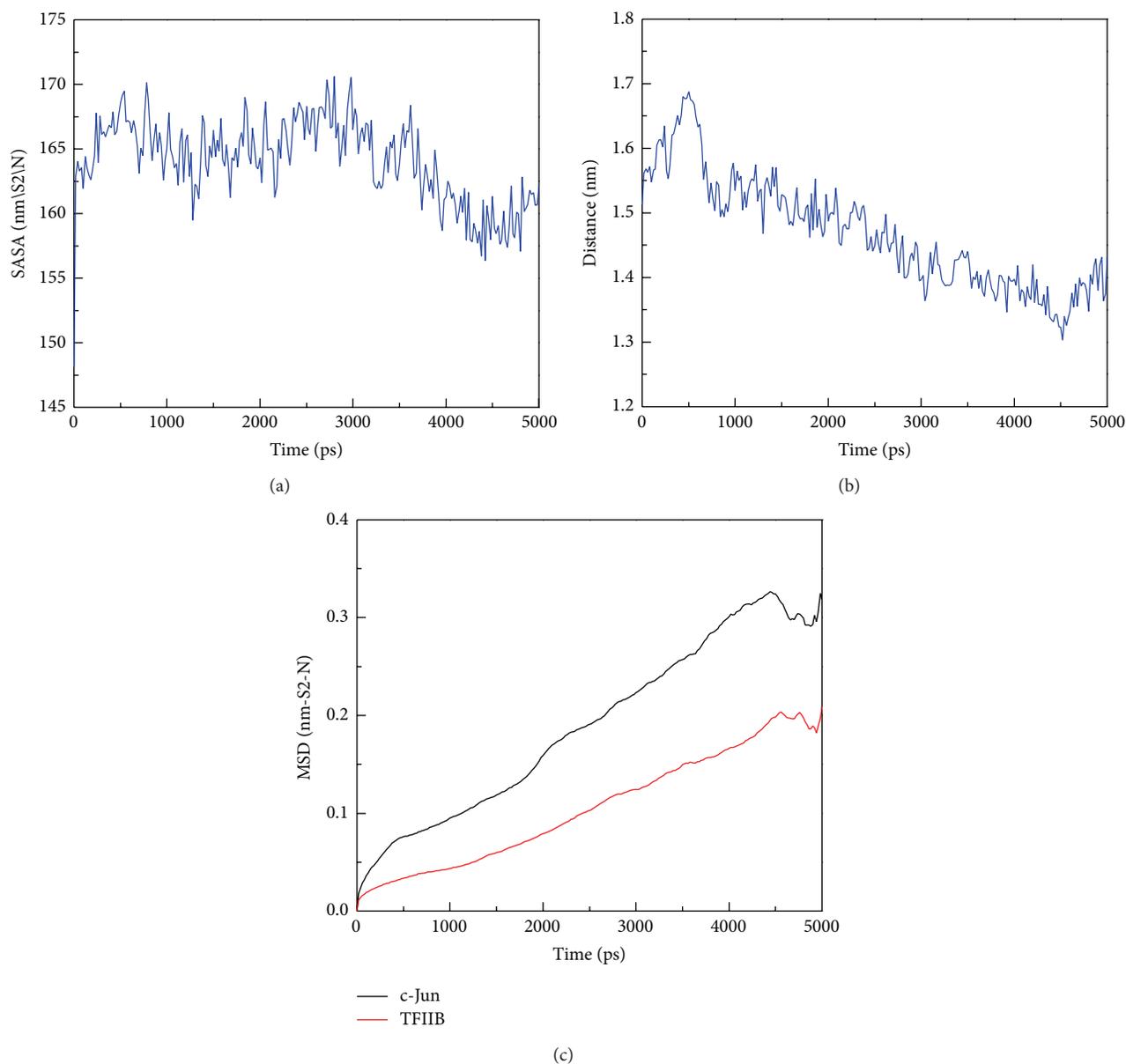
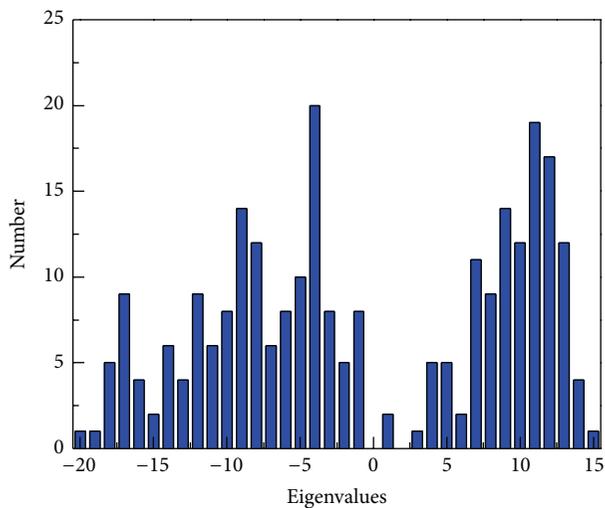


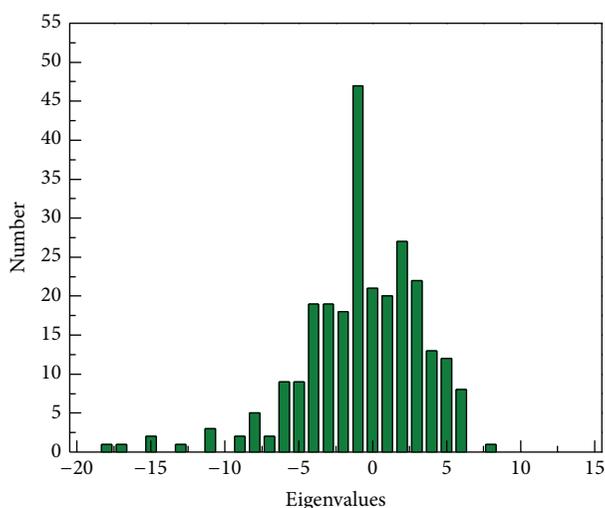
FIGURE 15: The area of solvent and protein migration analyses during simulation time of 5000 ps. (a) The total solvent accessible surface area of c-Jun and TFIIB complex; (b) the distance between the centrals of masses of c-Jun and TFIIB; (c) trajectory analysis of MSD of c-Jun and TFIIB. The high values of MSD indicated the longer distance of migration from the initial binding position.

could be grouped into two clusters. This suggests that the motion of each frame was not changed significantly over all simulation times. In order to select the most representative structure for snapshot investigation, we did cluster analysis (Figure 18). We found that the last group (cluster 14) is the predominant cluster and is also displayed in the time range from 4000 to 5000 ps; the cluster 14 also appears most predominant in frame numbers (Figure 18), and the middle structure (4260 ps) of cluster 14 is regarded as representative frame. For snapshot analysis, the comparison of initial and representative frames is shown in Figure 19; we found that

the chain A of c-Jun is more encompassed by TFIIB at 4260 ps through the inward rotation of TFIIB. This made the bindings between chain A of c-Jun and TFIIB more compacted through 0 ps to 4260 ps along with time. The elevated activation between TFIIB and chain A of c-Jun is also confirmed in RMSF analysis. Therefore, we supposed the initiated transcription factor on RNA polymerase II (TFIIB) is closed interaction to the chain A of c-Jun (AP1) from 0 ps to 4260 ps. This hypothesis was also confirmed by Figure 20. There were more spaces between chain B and TFIIB than chain A. Hence, we could see that the TFIIB acts more close to



(a)



(b)

FIGURE 16: The number of MD frames of the first two eigenvectors (PC1 and PC2) by PCA analysis during simulation time of 5000 ps. The higher range of eigenvalue denotes the wider motion of protein structure over all simulation times.

the chain A of c-Jun (Figure 20). Overall, we presume that the initiation of *OPN* transcription started from TFIIB binding to chain A of c-Jun.

#### 4. Discussion

Previous study has shown that DNA element (GGGT-CATATGGTTCA) located in osteopontin promoter -2245 to -2259 nt can be regulated by vitamin D3 [27]. This DNA regulation region can easily be affected by the change of calcium concentration. The promoter region of porcine *OPN* was analyzed in transcription factor binding sites except the region -2615 to -2239. Interestingly, this region of porcine *OPN* promoter is rich in CpG sites compared to human, mouse, and bovine genome. Sakata also proved that

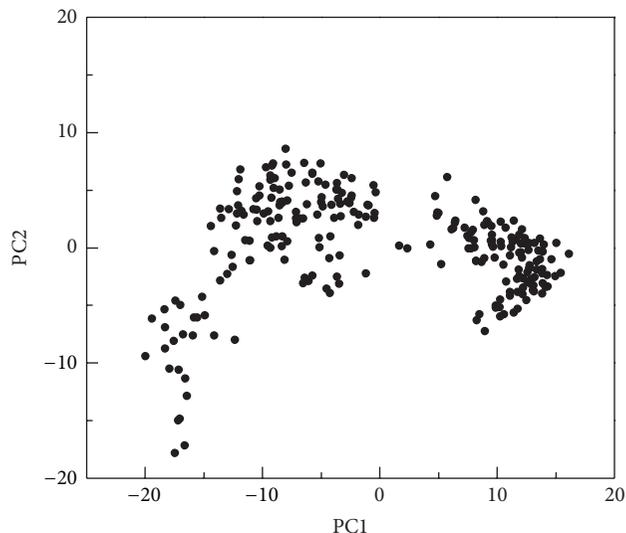


FIGURE 17: Phase space analysis of comparing the first two eigenvectors (PC1 and PC2) for principle component analysis. The eigenvalues of the two eigenvectors are projected into one phase space; small motion of protein structure could be grouped into clusters.

*OPN* promoter transcription activity is regulated by some specific DNA modification mechanism of rearrangement of chromatin structure [22].

In the present study, four cloned pigs were surrounded by many defects. For example, Copy1 pig had a retardation of limb bone growth. Copy2 heart organ showed a pericarditis and copy3 heart had valvular heart disease. This physiology defects appeared aberrant development, especially in bone or heart, may involved in the initially fetus stage with inappropriate organ differentiate. Thus, our data suggested that the consequent result in aberrant *OPN* expression or incompletely epigenetic modification in *OPN* promoter (Figure 5). These aberrant molecular data of *OPN* are correlated with the defects of bone and heart in cloned pigs. Semiquantitative PCR of *OPN* mRNA showed that discrepant expression pattern was identified in several cloned pig tissues, especially in brain (99.75% up-regulation), heart (11.5% downregulation), and ear (18.03% downregulation) (Figure 2). *OPN* mRNA has different expression in brain development in different embryonic stages [28]. The overexpression of *OPN* in brain tissue may cause some unexpected brain damage or neuron development. *OPN* can induce myocardial fibrosis and repair tissue after inflammation. Lacking *OPN* will cause faulty wound healing after myocardial infarction [29, 30]. Silent expression of *OPN* in cloned pig heart tissue may also be the main cause of heart disease.

Recent studies indicated that *OPN* gene expression may be affected by treatment of TSA (a histone deacetylase inhibitor). The results applied that *OPN* promoter could be regulated by epigenetic mechanism [22]. In this study, we investigated *OPN* methylation profile after 5-aza-dc treatment. The results indicated that mRNA expression of *OPN* is

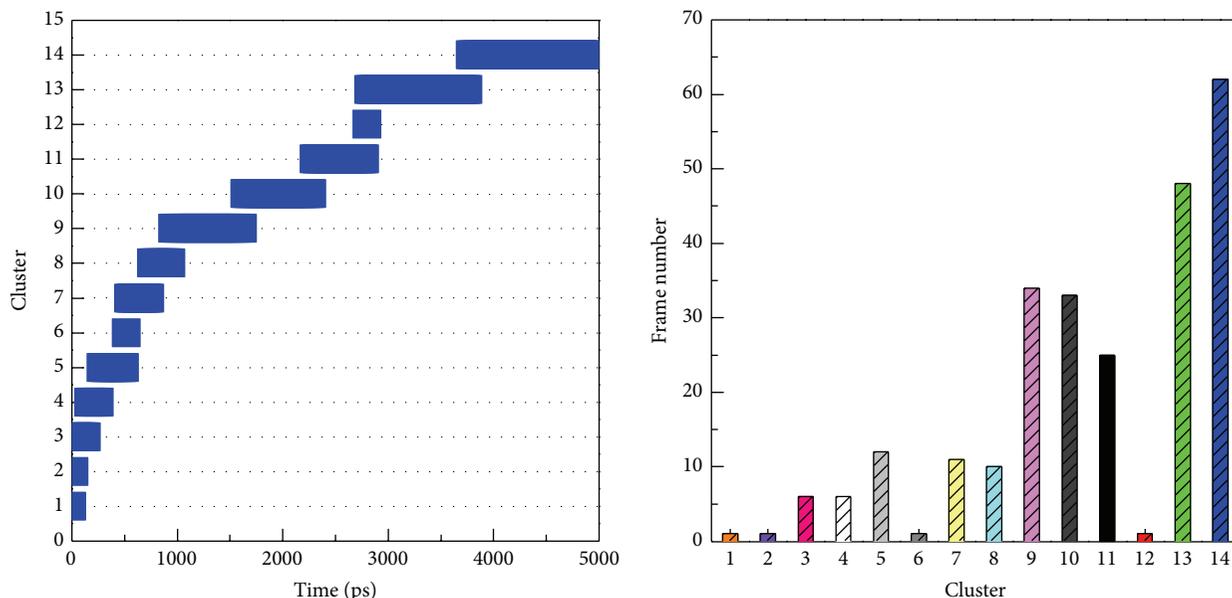


FIGURE 18: Cluster analysis of all MD frames of c-Jun and TFIIIB complex during simulation time of 5 ns for identifying representative structure. All MD frames were grouped into fourteen clusters by linkage method; the RMSD cut-off distance between each neighbor frame is 0.14 nm. The most predominant group is the cluster 14, which is displaced in the region of simulation time from 4000 to 5000 ps, and the MD frames in cluster 14 are the most number among all clusters. The middle frame of cluster 14 is displaced in simulation time of 4260 ps.

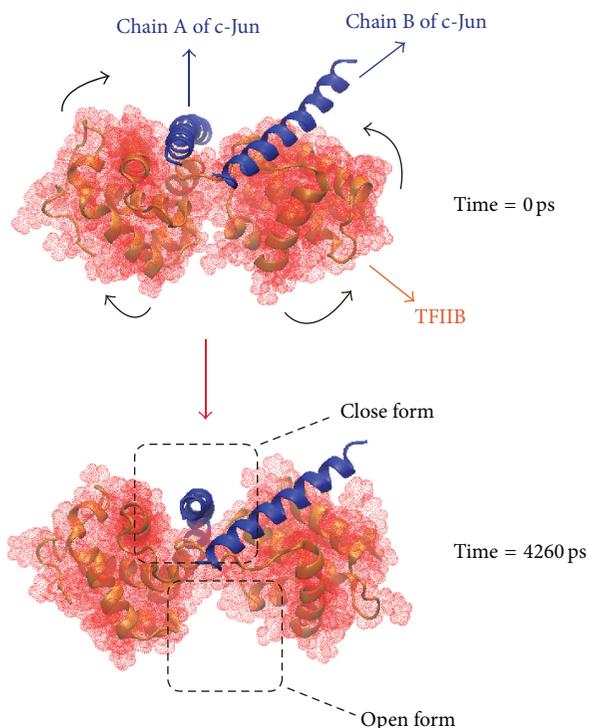


FIGURE 19: The structural comparison between the first frame (0 ps) and representative structure (4260 ps). The structures of c-Jun and TFIIIB are represented by blue ribbon and red solid phase, respectively. The chain A of c-Jun was surrounded more compactly by TFIIIB at 4260 ps by the structural inward rotation to make more compactness between TFIIIB and c-Jun through 0 ps to 4260 ps.

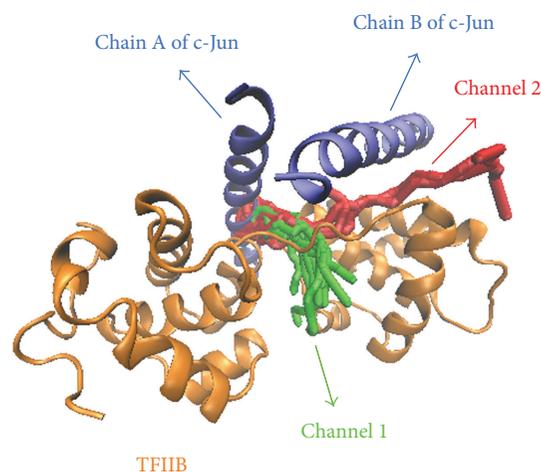


FIGURE 20: The possible space prediction between c-Jun and TFIIIB among all simulation times. The predicted channels are colored in red and green. The structures of c-Jun and TFIIIB are colored in blue and orange, respectively. Each possible space is represented by channels; each channel was generated by Caver 3.0 program.

directly affected by adding methyltransferase inhibitor 5-aza-dc (Figure 4). COBRA was performed to study the methylation of *OPN* front end promoter in wild-type and cloned pig different tissues. Sodium bisulfite sequencing analysis also revealed that the methylation of CpG sites concentrated in front of the 20 CpG sites in front of *OPN* promoter (Figure 5). Discrepancy methylation in this promoter region

also happened in brain, heart, ear, and liver tissues between wild-type and cloned pigs (Figure 5). It revealed that DNA methylation of *OPN* promoter may be involved with regulation of expression of *OPN* mRNA. In order to characterise which promoter DNA element is important, four constructs of *OPN* promoter (−2615-luc, −2239-luc, −1505-luc, and −495-luc) were used for analysis. Obvious downregulation in methylated −2615-luc (*Hpa*II and *Hha*I methyltransferase) was observed. Compared to the deletion of this control region (~390 nt) in front of *OPN* promoter, the decreasing level of promoter activity is not as obvious as −2615 M-luc construct (Figure 7). It means that methylation in the front of *OPN* promoter caused some silent mechanism that make chromosome structure more compact or block some promotion transcription factors. The EMSA data indicated that 13th CpG site of our analyzed region could bind to AP1 transcription factor and binding activity is affected by methylation in this CpG site (Figure 7). Taken together, all these findings correlated with DNA methylation in tissue- or cell-specific gene expression. *OPN* promoter region was densely methylated in some low expression (Figures 2, 4, and 5).

Our data revealed that DNA methylation of CpG sites in *OPN* promoter was the main mechanism through specific transcription factor that makes the tissue-specific expression. In previous study, AP1-like binding site (TGAGCGA) was identified as a methylated insulator region in human blastoma cell line [31]. Analyzed region in front of porcine *OPN* promoter showed that CpG 1st binding site contained the specific binding site sequence. While in our interesting DNA regulation region range from −2615 to −2239 nt of *OPN* promoter also exhibited little block access in the competition of probe. It is suggested that CpG 1st and CpG 13th play an important role in methylation controlled mechanism to regulate gene expression. We finally utilize Z-dock program [32] to analyze the interaction between AP1 (PDB code: 1JNM) and RNA polymerase II initial transcription factor (TFIIB) (PDB code: 1VOL) [33] to see if they had stable binding. From the docking result of Z-dock (Figure 8), we proved the AP1 is significantly bound to TFIIB. We also found the AP1 can autoregulate the HDAC-1 in promoter region and lead to significant higher degrees of hypermethylation in the *OPN* promoter region and cause AP1 to be hypermethylated consequently ceasing the *OPN* mRNA expression [34].

Further mechanical studies by the computational biology also pointed out that the DNA sequence for hypermethylation of *OPN* promoter binding sites is c-Jun. The chain A of c-Jun could be encompassed more tightly by inward rotational structure change of TFIIB during the MD process (Figure 19). Therefore, we found c-Jun had crucial role for interaction of initiating transcription by RNA polymerase II. The methylation of c-Jun leads to of hyper-condense helix structural change and makes transcription termination which stops *OPN* mRNA production. Therefore, the MD docking results reconfirm the c-Jun partake the crucial roles in consequent *OPN* transcription that matches our wet laboratory studies. We suppose this will cause the problems in the embryonic

development and lead to threatened conditions. Therefore, adjusting *OPN* promoter c-Jun (AP1) methylation will affect transcription binding and could be the treatment for genetic developing errors in the future.

In conclusion, aberrant methylation of porcine *OPN* gene was frequently found in different tissues of somatic nuclear transferred cloning pigs and bisulfite sequence data suggested that the *OPN* promoter region of −2615 to −2239 nt may be a crucial regulation DNA element. In pig ear fibroblast cell culture study, the demethylation of *OPN* promoter was found in dose-dependent response of 5-aza-dc treatment and followed the *OPN* mRNA reexpression. In cloned pig study, discrepant expression pattern was identified in several cloned pig tissues, especially in brain (99.75% up-regulation), heart (11.5% down-regulation), and ear (18.03% down-regulation). Promoter assay data revealed that four methylated CpG sites presenting in the −2615 to −2239 nt region cause significant downregulation (approximately 75%) of *OPN* promoter activity ( $P < 0.001$ ). EMSA data also suggested that CpG 13th and 1st sites showed methylation noncompetition phenomenon which had influence on binding with transcription factor.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Authors' Contribution

Chih-Jie Shen, Yung-An Tsou, and Hsiao-Ling Chen contributed equally to this work.

## Acknowledgments

This research was supported by Grants NSC-95-2313-B-005-030-MY3, NSC-101-2313-B-005-014-MY3, 101-2314-B-039-013-MY3, NSC102-2325-B039-001, and NSC102-2221-E-468-027- from the National Science Council and was partly supported by the Ministry of Education, Taiwan, under the aiming top university plan (ATU-101-s0508). The authors would like to thank their colleagues (Drs. Tung-Chou Tsai, Cheng-Wei Lai, and Zi-Lun Lai) in the Molecular Embryology & DNA Methylation Laboratory for their help with discussions and technical issues. This study is also supported in part by Taiwan Department of Health Clinical Trial and Research Center of Excellence (DOH102-TD-B-111-004), Taiwan Department of Health Cancer Research Center of Excellence (MOHW103-TD-B-111-03), and CMU under the Aim for Top University Plan of the Ministry of Education, Taiwan.

## References

- [1] Y. M. Leung, Y. M. Wong, S. W. Chen et al., "Down-regulation of voltage-gated  $Ca^{2+}$  channels in  $Ca^{2+}$  store-depleted rat insulinoma RINm5F cells," *BioMedicine*, vol. 3, no. 3, pp. 130–139, 2013.

- [2] M.-C. Lin, S.-Y. Tsai, F.-Y. Wang et al., "Leptin induces cell invasion and the upregulation of matrilysin in human colon cancer cells," *BioMedicine*, vol. 3, no. 4, pp. 174–180, 2013.
- [3] Y.-M. Chang, B. K. Velmurugan, W. W. Kuo, and et al, "Inhibitory effect of alpinate *Oxyphyllae fructus* extracts on Ang II-induced cardiac pathological remodeling-related pathways in H9c2 cardiomyoblast cells," *BioMedicine*, vol. 3, no. 4, pp. 148–152, 2013.
- [4] T. H. Bestor, "The DNA methyltransferases of mammals," *Human Molecular Genetics*, vol. 9, no. 16, pp. 2395–2402, 2000.
- [5] P. H. Tate and A. P. Bird, "Effects of DNA methylation on DNA-binding proteins and gene expression," *Current Opinion in Genetics and Development*, vol. 3, no. 2, pp. 226–231, 1993.
- [6] W. H. Eyestone and K. H. Campbell, "Nuclear transfer from somatic cells: applications in farm animal species," *Journal of Reproduction and Fertility*, vol. 54, pp. 489–497, 1999.
- [7] T. Wakayama, A. C. F. Perry, M. Zuccotti, K. R. Johnson, and R. Yanagimachi, "Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei," *Nature*, vol. 394, no. 6691, pp. 369–374, 1998.
- [8] T. Wakayama and R. Yanagimachi, "Mouse cloning with nucleus donor cells of different age and type," *Molecular Reproduction and Development*, vol. 58, no. 4, pp. 376–383, 2001.
- [9] J. Ohgane, T. Wakayama, Y. Kogo et al., "DNA methylation variation in cloned mice," *Genesis*, vol. 30, no. 2, pp. 45–50, 2001.
- [10] G. A. Johnson, R. C. Burghardt, F. W. Bazer, and T. E. Spencer, "Osteopontin: roles in implantation and placentation," *Biology of Reproduction*, vol. 69, no. 5, pp. 1458–1471, 2003.
- [11] G. F. Weber, S. Zawaideh, S. Hikita, V. A. Kumar, H. Cantor, and S. Ashkar, "Phosphorylation-dependent interaction of osteopontin with its receptors regulates macrophage migration and activation," *Journal of Leukocyte Biology*, vol. 72, no. 4, pp. 752–761, 2002.
- [12] M. Mazzali, T. Kipari, V. Ophascharoensuk, J. A. Wesson, R. Johnson, and J. Hughes, "Osteopontin—a molecule for all seasons," *QJM*, vol. 95, no. 1, pp. 3–13, 2002.
- [13] N. S. Fedarko, A. Jain, A. Karadag, M. R. van Eman, and L. W. Fisher, "Elevated serum bone sialoprotein and osteopontin in colon, breast, prostate, and lung cancer," *Clinical Cancer Research*, vol. 7, no. 12, pp. 4060–4066, 2001.
- [14] G. Chakraborty, S. Jain, and G. C. Kundu, "Osteopontin promotes vascular endothelial growth factor-dependent breast tumor growth and angiogenesis via autocrine and paracrine mechanisms," *Cancer Research*, vol. 68, no. 1, pp. 152–161, 2008.
- [15] R. Agnihotri, H. C. Crawford, H. Haro, L. M. Matrisian, M. C. Havrda, and L. Liaw, "Osteopontin, a novel substrate for matrix metalloproteinase-3 (stromelysin-1) and matrix metalloproteinase-7 (matrilysin)," *The Journal of Biological Chemistry*, vol. 276, no. 30, pp. 28261–28267, 2001.
- [16] D. Wang, S. Yamamoto, N. Hijiya, E. N. Benveniste, and C. L. Gladson, "Transcriptional regulation of the human osteopontin promoter: Functional analysis and DNA-protein interactions," *Oncogene*, vol. 19, no. 50, pp. 5801–5809, 2000.
- [17] H. Rangaswami, A. Bulbule, and G. C. Kundu, "Osteopontin: role in cell signaling and cancer progression," *Trends in Cell Biology*, vol. 16, no. 2, pp. 79–87, 2006.
- [18] H. Rangaswami, A. Bulbule, and G. C. Kundu, "Nuclear factor-inducing kinase plays a crucial role in osteopontin-induced MAPK/IKK $\alpha$  kinase-dependent nuclear factor  $\kappa$ B-mediated promatrix metalloproteinase-9 activation," *Journal of Biological Chemistry*, vol. 279, no. 37, pp. 38921–38935, 2004.
- [19] A. B. Tuck, D. M. Arsenault, F. P. O'Malley et al., "Osteopontin induces increased invasiveness and plasminogen activator expression of human mammary epithelial cells," *Oncogene*, vol. 18, no. 29, pp. 4237–4246, 1999.
- [20] H. Okamoto, "Osteopontin and cardiovascular system," *Molecular and Cellular Biochemistry*, vol. 300, no. 1-2, pp. 1–7, 2007.
- [21] F. Schoensiegel, R. Bekeredjian, A. Schrewe et al., "Atrial natriuretic peptide and osteopontin are useful markers of cardiac disorders in mice," *Comparative Medicine*, vol. 57, no. 6, pp. 546–553, 2007.
- [22] R. Sakata, S. Minami, Y. Sowa, M. Yoshida, and T. Tamaki, "Trichostatin A activates the osteopontin gene promoter through API site," *Biochemical and Biophysical Research Communications*, vol. 315, no. 4, pp. 959–963, 2004.
- [23] K. H. Choi, H. Basma, J. Singh, and P. W. Cheng, "Activation of CMV promoter-controlled glycosyltransferase and  $\beta$ -galactosidase glycogenes by butyrate, trichostatin A, and 5-aza-2'-deoxycytidine," *Glycoconjugate Journal*, vol. 22, no. 1-2, pp. 63–69, 2005.
- [24] B. M. Kumar, H.-F. Jin, J.-G. Kim et al., "DNA methylation levels in porcine fetal fibroblasts induced by an inhibitor of methylation, 5-azacytidine," *Cell and Tissue Research*, vol. 325, no. 3, pp. 445–454, 2006.
- [25] S. Pronk, S. Páll, R. Schulz et al., "GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit," *Bioinformatics*, vol. 29, no. 7, pp. 845–854, 2013.
- [26] E. Chovancova, A. Pavelka, P. Benes et al., "CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures," *PLoS Computational Biology*, vol. 8, no. 10, Article ID e1002708, 2012.
- [27] Q. Zhang, J. L. Wrana, and J. Sodek, "Characterization of the promoter region of the porcine opn (osteopontin, secreted phosphoprotein 1) gene. Identification of positive and negative regulatory elements and a "silent" second promoter," *European Journal of Biochemistry*, vol. 207, no. 2, pp. 649–659, 1992.
- [28] M. Y. Lee, J. S. Choi, S. W. Lim, J. H. Cha, M. H. Chun, and J. W. Chung, "Expression of osteopontin mRNA in developing rat brainstem and cerebellum," *Cell and Tissue Research*, vol. 306, no. 2, pp. 179–185, 2001.
- [29] L. Liaw, D. E. Birk, C. B. Ballas, J. S. Whitsitt, J. M. Davidson, and B. L. Hogan, "Altered wound healing in mice lacking a functional osteopontin gene (spp1)," *The Journal of Clinical Investigation*, vol. 101, no. 7, pp. 1468–1478, 1998.
- [30] N. A. Trueblood, Z. Xie, C. Communal et al., "Exaggerated left ventricular dilation and reduced collagen deposition after myocardial infarction in mice lacking osteopontin," *Circulation Research*, vol. 88, no. 10, pp. 1080–1087, 2001.
- [31] M. Fujimoto, R. Kitazawa, S. Maeda, and S. Kitazawa, "Methylation adjacent to negatively regulating AP-1 site reactivates TrkA gene expression during cancer progression," *Oncogene*, vol. 24, no. 32, pp. 5108–5118, 2005.
- [32] Accelrys, *Discovery Studio Client v2.5*, Accelrys, San Diego, Calif, USA, 2009.
- [33] D. B. Nikolov, H. Chen, E. D. Halay et al., "Crystal structure of a TFIIB-TBP-TATA-element ternary complex," *Nature*, vol. 377, no. 6545, pp. 119–128, 1995.
- [34] P. Sharma, S. Kumar, and G. C. Kundu, "Transcriptional regulation of human osteopontin promoter by histone deacetylase inhibitor, trichostatin A in cervical cancer cells," *Molecular Cancer*, vol. 9, article 178, 2010.

## Research Article

# Muscle-Type Specific Autophosphorylation of CaMKII Isoforms after Paced Contractions

Wouter Eilers,<sup>1,2</sup> Wouter Gevers,<sup>3,4</sup> Daniëlle van Overbeek,<sup>2</sup> Arnold de Haan,<sup>1,2</sup>  
Richard T. Jaspers,<sup>2</sup> Peter A. Hilbers,<sup>3</sup> Natal van Riel,<sup>3</sup> and Martin Flück<sup>1,4</sup>

<sup>1</sup> School of HealthCare Science, Institute for Biomedical Research into Human Movement and Health, Manchester Metropolitan University, Manchester M1 5GD, UK

<sup>2</sup> Laboratory for Myology, Faculty of Human Movement Sciences, MOVE Research Institute Amsterdam, VU University Amsterdam, 1081 BT Amsterdam, The Netherlands

<sup>3</sup> Department of Biomedical Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands

<sup>4</sup> Laboratory for Muscle Plasticity, Department of Orthopedics, University of Zurich, Balgrist University Hospital, Forchstrasse 340, 8008 Zurich, Switzerland

Correspondence should be addressed to Martin Flück; [mflueck@research.balgrist.ch](mailto:mflueck@research.balgrist.ch)

Received 6 May 2014; Accepted 27 May 2014; Published 26 June 2014

Academic Editor: Calvin Yu-Chian Chen

Copyright © 2014 Wouter Eilers et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We explored to what extent isoforms of the regulator of excitation-contraction and excitation-transcription coupling, calcium/calmodulin protein kinase II (CaMKII) contribute to the specificity of myocellular calcium sensing between muscle types and whether concentration transients in its autophosphorylation can be simulated. CaMKII autophosphorylation at Thr287 was assessed in three muscle compartments of the rat after slow or fast motor unit-type stimulation and was compared against a computational model (CaMuZclE) coupling myocellular calcium dynamics with CaMKII Thr287 phosphorylation. Qualitative differences existed between fast- (*gastrocnemius medialis*) and slow-type muscle (*soleus*) for the expression pattern of CaMKII isoforms. Phospho-Thr287 content of  $\delta$ A CaMKII, associated with nuclear functions, demonstrated a transient and compartment-specific increase after excitation, which contrasted to the delayed autophosphorylation of the sarcoplasmic reticulum-associated  $\beta$ M CaMKII. In soleus muscle, excitation-induced  $\delta$ A CaMKII autophosphorylation demonstrated frequency dependence ( $P = 0.02$ ). In the glycolytic compartment of *gastrocnemius medialis*, CaMKII autophosphorylation after excitation was blunted. *In silico* assessment emphasized the importance of mitochondrial calcium buffer capacity for excitation-induced CaMKII autophosphorylation but did not predict its isoform specificity. The findings expose that CaMKII autophosphorylation with paced contractions is regulated in an isoform and muscle type-specific fashion and highlight properties emerging for phenotype-specific regulation of CaMKII.

## 1. Introduction

Myocellular calcium is an important second messenger of muscle regulation. This role is mediated by increases in sarcoplasmic calcium concentration with functional recruitment of muscle fibers following motoneuron-induced muscle excitation (i.e., recruitment; [1]). The elevated calcium initiates sarcomere shortening, that is, excitation-contraction coupling, and produces downstream effects on gene regulation, that is, excitation-transcription coupling [2–5]. Three types of motor units can be distinguished based on their contractile and metabolic characteristics, that is: a fast-fatigable, a

fast-fatigue-resistant, and a slow-fatigue-resistant type [1, 6]. Thereby it is observed that slow and fast contractile characteristics of motor units are reflected by differences in the (electric) firing pattern of the innervating motoneuron and concentration differences in the rise of myocellular calcium with excitation [7]. Experimental studies have established a specific preference of motoneuron activation frequencies for the three contractile types of motor units [1, 8]. Slow fatigue-resistant motor units are preferentially recruited by low-frequency depolarisation of motor units. In contrast, recruitment of fast-type motor units often occurs at higher stimulation frequencies [9, 10]. Furthermore, the elevated

content of the organelles of excitation-contraction coupling, t-tubuli and sarcoplasmic reticulum, in fast-type fibres is associated with larger increases in myocellular calcium concentration during recruitment [11].

Along with the calcium/calmodulin-dependent phosphatase calcineurin [7, 14] changes in the activity of calcium/calmodulin dependent kinase II (CaMKII) are a potential interface explaining the downstream effects of elevated myocellular calcium on calcium channel activity and gene regulation [7, 12, 14] and indicated differences between muscle fiber types [7]. To determine the relevance of CaMKII for the regulation of calcium channel activity and gene expression, it is important to understand the regulation of CaMKII activity. CaMKII is a multimeric phosphotransfer enzyme that assembles from different isoforms [15]. Its phosphotransfer activity is induced following rises in intracellular calcium after autophosphorylation of Thr287 in its autoinhibitory domain, enhancing its phosphotransfer activity and rendering it calcium-independent through a subsequent conformational change [15]. Thr287 phosphorylation of CaMKII and a concomitant increase in calcium/calmodulin-independent phosphotransfer activity can be observed within seconds after onset of muscle contraction [12]. Based on model studies of cardiac and brain isoforms, this increase in CaMKII autophosphorylation is believed to rapidly fall within seconds to baseline levels [16, 17]. This is, however, not what can be observed in skeletal muscle where autonomous CaMKII activity can remain specifically elevated after prolonged repetition of contraction, such as running type exercise [18].

Four isoforms of CaMKII (i.e.,  $\beta$ M,  $\delta$ A,  $\delta$ D, and  $\gamma$ B) have been reported to exist in skeletal muscle [12, 13]. *In vitro* measurements point out characteristic differences in calcium sensitivity of activation between CaMKII isoforms [19, 20]. The biochemical observations are in line with cellular investigations that allocate different functions for the various CaMKII isoforms. For instance, the  $\beta$ M isoform has been found to associate with sarcoplasmic reticulum hinting that it possibly operates in the regulation of calcium channels involved in muscle contraction and relaxation [13]. Conversely, the  $\delta$ D/ $\gamma$ B CaMKII, and possibly  $\delta$ A, isoforms are associated with nuclear functions [21–23]. *In vitro* studies of CaMKII revealed that CaMKII autophosphorylation is subject to discrete regulation by the frequency and amplitude of calcium transients, which differs between CaMKII isoforms [24]. Studies in hippocampal cultures identified that Thr286-phosphorylation of the neuron-specific CaMKII isoform  $\alpha$  acts as a frequency/number decoder for sensory input [25]. Computational modeling of CaMKII activation by postsynaptic calcium pointed out that autophosphorylation of  $\beta$  CaMKII demonstrates a greater response range than  $\alpha$  CaMKII [26]. The extent to which isoform-specific CaMKII autophosphorylation is implicated in the response of muscle fiber types to functional recruitment is not understood.

Towards this end we characterised the pattern of CaMKII isoform expression and CaMKII autophosphorylation following electrically paced contractions in phenotypically distinct motor compartments of the rat using muscle-tendon preparation *in situ*. Specifically we asked whether autophosphorylation of the different isoforms of CaMKII is subject to

the same regulation by motoneuron excitation and whether this would differ between three distinct compartments of the *triceps surae* muscle group, representing the three types of motor units, when excitation is paced with slow-type and fast-type electric stimulation protocols [9, 10]. Emphasis was put on the question of whether the reported contribution of mitochondria to calcium buffer capacity would exert an influence on Thr287 content of CaMKII [27]. We then asked whether we can reveal mechanistic understanding of the properties of excitation-induced CaMKII regulation between muscle types and isoforms by confronting experimentally observed “on-” and “off-rates” of CaMKII Thr287-phosphorylation with results from a newly assembled computational model.

## 2. Materials and Methods

**2.1. Design.** Four experimental protocols were run to pace contraction of the *triceps surae* muscle group of one leg of rats with a slow and fast motor unit-targeted protocol [8, 9]. Three muscle compartments of the *triceps surae* were rapidly isolated and subjected to the measure of CaMKII Thr287 phosphorylation with biochemical means. The nonstimulated muscles of the contralateral side served as controls for Protocols 1 and 2. Collaterally, CaMKII Thr287 phosphorylation was assessed using an *in silico* model combining myocellular calcium dynamics with CaMKII isoform autophosphorylation in different muscle types.

**2.2. Animals.** Three-month-old female Wistar rats (Harlan Laboratories) were anaesthetized by intraperitoneal injection of 1.2 mL/100 gram body weight of 12.5% urethane. Ear and foot reflexes were tested to check whether the animal was sufficiently anaesthetized. Subsequently, injections of 0.3–0.5 mL, up to a maximum of 1.5 mL, were given every 10 minutes afterwards until reflexes had disappeared. Rats were kept on a heated pad ( $37 \pm 0.5^\circ\text{C}$ ) to prevent hypothermia. Experiments were carried out with approval of the local Animal Experiments Committee at the MOVE Research Institute Amsterdam, VU University Amsterdam. 12 and 20 animals entered the experimental Protocols 1 and 2 (mean body weight of 191–230 grams) and Protocols 3 and 4 (mean body weight of 205–220 grams), respectively.

**2.3. Muscle-Tendon Preparation.** Hind limbs were shaved and the skin was removed, after which *gastrocnemius medialis* and *soleus* muscles were exposed and mechanically isolated by removing as much as possible the myofascial connections to surrounding muscles. Blood supply to and nerve innervations of *m. gastrocnemius medialis* and *m. soleus* were kept intact. The calcaneus was cut from the talus, while still attached to the Achilles tendon. The sciatic nerve was dissected free, proximally severed, and electrically stimulated through an electrode by different protocols being controlled by a computer.

*Protocol 1* (150 Hz protocol). *Gastrocnemius medialis* and *soleus* muscles of the right leg were kept below slack length (without determination of optimal length) and electrically

stimulated via the sciatic nerve with a train of 100 rectangular electric pulses of 50 microseconds duration and 3 mA amplitude delivered at 150 Hz. Muscles of the left leg served as nonstimulated controls. The experiment was performed at a room temperature of 23°C.

*Protocol 2* (10 Hz protocol). *Gastrocnemius medialis* and *soleus* muscles were electrically stimulated as described for Protocol 1 but with the modification of a single train of 100 pulses at 10 Hz.

*Protocol 3* (tetanic contraction at optimal length). 3 tetanic contractions with 2 minutes of rest in between were imposed on *gastrocnemius medialis* muscle via the sciatic nerve with a modification of the 150 Hz-protocol where the train consisted of 30, rather than 100, electric pulses of 50 microseconds duration and 3 mA amplitude, at 150 Hz. The temperature of the *gastrocnemius* muscle was kept at 35°C using an envelope into which warm water vapor was sprayed.

*Protocol 4* (24 tetanic contractions). A 2 minute protocol of 24 tetanic contractions (1 contraction every 5 seconds) with the same duration and frequency as described for Protocol 3 was imposed. The experiment was performed at 35°C. This protocol resulted in a decrease in maximal tetanic force of approximately 30%.

**2.4. Sampling.** After completion of the protocol, the proximal (i.e., oxidative) and distal (i.e., glycolytic) portion of *gastrocnemius medialis* muscle [28] and *soleus* muscle (in Protocols 1 and 2 only) was dissected as rapidly as possible and snap-frozen in liquid nitrogen. The first muscle was typically sampled after 1 minute and the second after 2 minutes. In half of the experiments, *soleus* muscle was sampled first, and in the other half *gastrocnemius medialis* muscle was sampled first. Subsequently, the nonstimulated left muscles were dissected and snap-frozen. Muscles subjected to Protocol 4 were subjected to sampling over a time-course covering 10 and 60 minutes after contraction to estimate the rate of CaMKII dephosphorylation. Rats were subsequently euthanized by intracardial injection of Euthasol while fully anaesthetized. Muscles were stored at -80°C until used for western blot analysis.

**2.5. Biochemical Analysis of CaMKII Thr287 Phosphorylation.** 25 µm thick cryosections were prepared from the frozen muscles, pooled and homogenized on ice using a Polytron homogeniser (Kinematica AG, Luzern, Switzerland) in ice-cold RIPA buffer (50 mM TRIS-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1% v/v Nonidet P40 substitute, and 0.25% w/v sodium deoxycholate, including freshly added protease/phosphatase inhibitors: 1 mM NaF, 1 mM Na<sub>3</sub>VO<sub>4</sub>, 0.1 mM PMSF, 1 µg/mL leupeptin, 0.2 µg/mL pepstatin, and 0.1 µg/mL aprotinin. Chemicals were obtained from Sigma-Aldrich (Poole, United Kingdom) unless stated otherwise. Crude homogenates were aspirated 5–10-times through a 0.8 mm syringe needle and stored at -80°C until being further processed. An aliquot of the aspirated homogenate was taken for determination of protein concentration with

the bicinchoninic acid protein assay (Pierce, Rockford, IL, USA).

Protein levels of total CaMKII, phospho-Thr287-CaMKII, and cytochrome-c oxidase subunit IV (COXIV) were analyzed by western blotting followed by immunodetection. Homogenates were denatured by addition of sample buffer (final concentration: 50 mM TRIS-HCl (pH 6.8), 2% w/v SDS, 10% v/v glycerol, and 2% β-mercaptoethanol) and 5 minutes heating at 95°C. 20 µg of protein per lane was separated with SDS-PAGE on a 15% acrylamide (Biorad) gel and transferred overnight in ice-cold buffer onto a nitrocellulose membrane (GE Healthcare Life Sciences, Little Chalfont, United Kingdom). Membranes were stained with Ponceau S solution to confirm equal protein loading and transfer and then subjected to immunodetection. After blocking, 2-hour incubations were carried out with primary antibody against pan-CaMKII (BD Bioscience, no. 611292, dilution: 1/2500), phospho-Thr287-CaMKII (Cell Signalling Technology, no. 3361, dilution: 1/1000), or COXIV (Cell Signalling Technology, no. 4850, dilution 1/2000) in TTBS (20 mM TRIS-HCl (pH 7.5), 0.9% w/v NaCl, 0.05% Tween-20) with 5% milk (pan-CaMKII, COXIV) or 5% bovine serum albumin (phospho-Thr287-CaMKII) as blocking agent. After serial washes in TTBS, membranes were incubated with species-specific horseradish peroxidase-conjugated secondary antibodies and signal detected enhanced chemiluminescence (Pierce, Rockford, IL, USA) and recorded with a ChemiDoc XRS system (Biorad, Hemel Hempstead, United Kingdom).

Samples from the oxidative and glycolytic *m. gastrocnemius medialis* and *m. soleus* were analyzed on separate immunoblots. For each muscle compartment, two sample pairs of stimulated and nonstimulated contralateral control muscles from Protocols 1 and 2, respectively, were analyzed on a same blot. Signal intensity of the protein band of interest was quantified using Quantity One version 4.6.8 (Biorad), background-corrected, and normalized to the average of CaMKII proteins signals from all lanes on the blot. Then for each blot, the normalized signals were related to the total CaMKII signal for the respective resting (i.e., “nonstimulated”) muscle and data pooled between experiments. Finally the values were related to the mean of the “nonstimulated” samples for the respective experimental protocol and subjected to statistical analysis.

The specificity of CaMKII and phospho-Thr287-CaMKII detection was ensured in control experiments monitoring calcium/calmodulin-inducible phosphorylation in “cold” *in vitro* kinase assays as described [18]. In brief, 5 µL of total muscle homogenate was suspended in a reaction mix containing calcium/calmodulin (Enzo Life Sciences, Exeter, United Kingdom) (total volume: 50 µL; 0.1 mM ATP, 10 mM HEPES (pH 7.4), 5 mM MgCl<sub>2</sub>, 0.1% Tween-20, and 0.5 mM CaCl<sub>2</sub> + 1 µM calmodulin or 5 mM EGTA) for 30 minutes at 30°C. The reaction was stopped by the addition of 16.7 µL 4x sample buffer. Samples were heated to 95°C and proteins were separated with SDS-PAGE on a 7.5% (or 15%) acrylamide (Biorad) gel and subjected to western blotting as described. Based on data from [13], we identified the detected CaMKII bands as βM (72 kDa), δA (60 kDa), and a combination of δD and γB (58 kDa), the latter two of

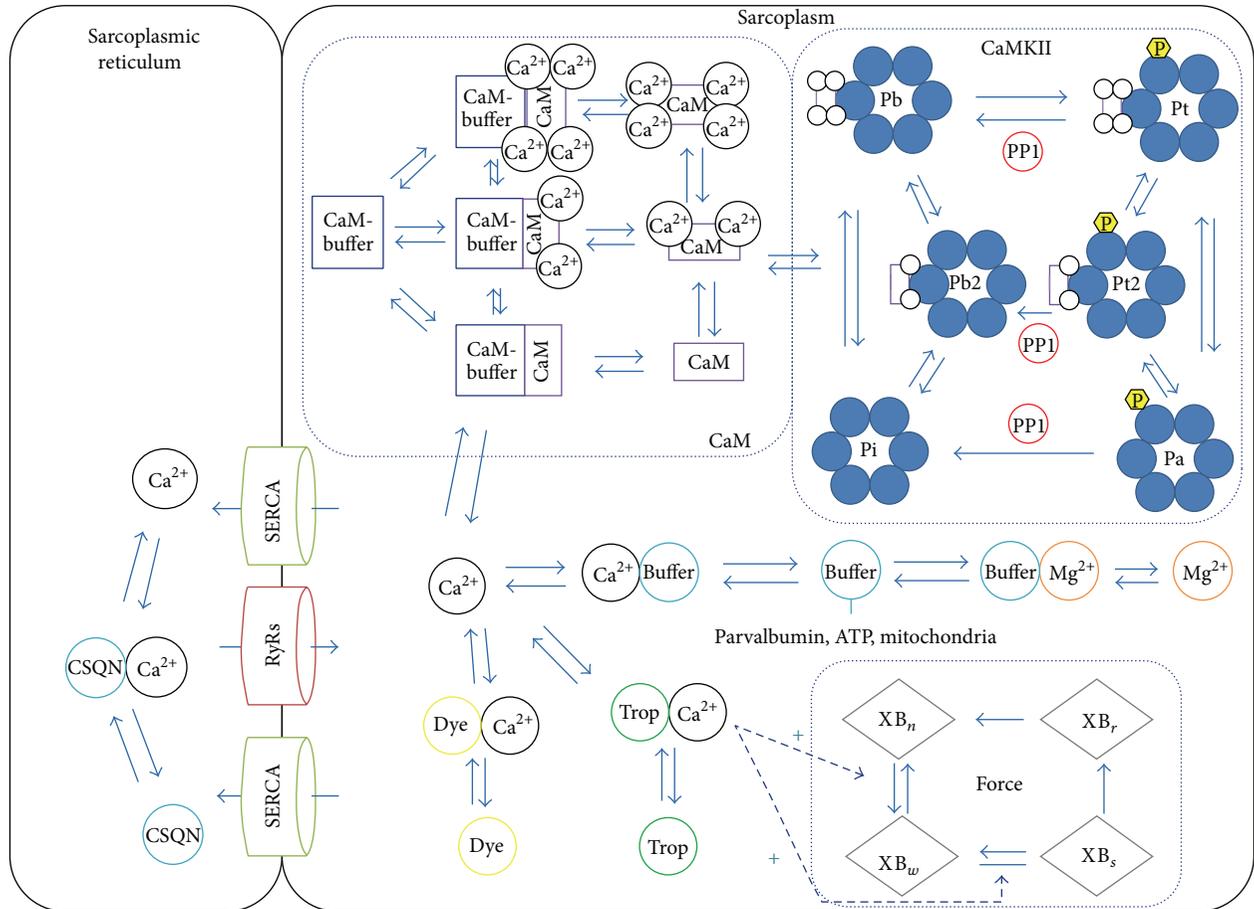


FIGURE 1: Schematic representation of the components of the CaMuZcIE model to simulate calcium-driven Thr287 phosphorylation of CaMKII in skeletal muscle. The model includes calcium flow through the sarcoplasmic reticulum and sarcoplasm. Calcium is bound to the buffer calsequestrin in the SR and is transported by RyRs channel and SERCA pump between the SR and sarcoplasm. Within the sarcoplasm calcium binds to the buffers ATP and parvalbumin, mitochondria, and troponin C, initiating sarcomere shortening and force production. Furthermore, calcium binds to the calcium sensor calmodulin, which in turn activates the phosphotransferase CaMKII, initiating several pathways. A comprehensive list of all abbreviations can be found in the Appendices.

which could not be separated in all gels. Therefore, these two CaMKII isoforms were assessed as one band.

**2.6. Model Development.** We set out to develop a computational model to predict calcium driven changes in Thr 287 phosphorylation of CaMKII (termed CaMuZcIE for CaMKII muscle model from Zurich-Eindhoven). The model is based on a spatiotemporal model of calcium dynamics in the half sarcomere of fast-twitch muscle [29] and a biochemical model describing CaMKII activation [16] which was reduced to a lumped version by removing the spatial component.

The model is based on the interactions between the chemical species inside a sarcomere, consisting of the ions calcium, magnesium and potassium, the calcium buffers calsequestrin, parvalbumin, troponin, adenosine triphosphate (ATP), and mitochondria that act as a calcium buffer, the proteins calmodulin (CaM) and calcium/calmodulin-dependent kinase (CaMKII), CaM buffers, and a calcium-binding dye. Both the sarcoplasmic reticulum and the

sarcoplasm are modelled, including the pumps facilitating transport of calcium between the compartments, Ryanodine receptors (RyR), and sarco/endoplasmic reticulum calcium-ATPase pumps (SERCA). Furthermore, the development of contractile force created by contraction is modelled. Reactions are described in terms of changes in the concentration (fluxes) of respective species using coupled ordinary differential equations (ODEs) based on a set of initial concentrations and parameters of the reaction rates. The system of ODEs resulting from these sets is then solved using a numerical integration algorithm, in this case the Matlab (The Mathworks) built-in solver `ode15s` using an implicit integration scheme with numerical differentiation formulas. The calculation takes around 20 seconds to simulate 10 minutes of experimental time using an Intel core i7 processor (3.4 GHz). Memory usage is negligible (<100 MB). The model was calibrated by measures on sarcoplasmic calcium based on the calcium-binding dye. A schematic representation of all the model components and fluxes can be seen in Figure 1. Sensitivity analysis was performed using the multiparametric

TABLE 1: Relative influences of the model input parameters on the transient in phosphoThr287-CaMKII.

Parameter	Relative influence
kbta	0.32
kPP1	0.24
CaMax	0.22
KmPP1	0.20
CaPump	0.11
kbi	0.06
kbtc	0.06
kTrop1	0.04
kMitCa2	0.04
kDye2	0.03
kMitCa1	0.03
kATP2	0.03
kCaM42	0.03
kParvCa1	0.03
kParvCa2	0.03
kMitMg1	0.03
kDye1	0.03
kATP1	0.03
kbtb	0.03
kCSQN1	0.02
kCaM0Boff	0.02
KdPump	0.02
kMitMg2	0.02
kCaM20	0.02
kTrop2	0.02
kCSQN2	0.02
kParvMg1	0.02
kParvMg2	0.02

Values were calculated as described under multiparametric sensitivity analysis in Section 2.6 “Model Development” of the “Materials and Methods” section. A higher value indicates a larger influence. The most influential parameters are the speed of auto- and dephosphorylation (i.e., kbta, kPPi) and the activity of the calcium pumps (i.e., CaMax, CaPump). For abbreviations consult Tables 2 and 3.

sensitivity analysis (MPSA) method, describing the speed of the calcium pumps, total calsequestrin concentration, and total CaMKII concentration as most influential parameters (Table 1). Further details on the model equations can be found in the Appendices and Tables 2–6. MPSA was conducted by uniform sampling of 4000 parameter sets from an interval from 75% to 125% of the default parameter set using Latin hypercube sampling, ensuring the sets cover the complete parameter space. Stimulation with a pulse train of 100 pulses at 150 Hz was used (Protocol 1), as this describes a full and complete stimulation. For each parameter set the time course of phospho287-CaMKII was compared to the output with the default set. A sum of squared differences was used as measure. Subsequently, sensitivity values were calculated as described [30].

Interdependent parameters were excluded as these variables change together with the variables they are dependent

on. Exponentials of the description of the RyR activity were excluded as well as the RyR activity is regulated by varying the parameter CaMax. The remaining parameters are ordered with respect to their influence on the trace of phosphorylated CaMKII, as calculated via the MPSA method. The MPSA results were used to perform uncertainty analysis [31]. Poisson distributions were fitted to the set of output transients resulting from MPSA and were used to draw 95% confidence intervals. For clarity, only the trace of the  $\beta_M$  CaMKII isoform is presented; the traces of the other isoforms display similar behavior.

**2.7. Statistics.** All statistical tests were performed using Statistica 10.0 (Statsoft Inc., Tulsa, OK, USA). The effect of stimulation with Protocols 1 and 2 on total and phospho-Thr287-CaMKII levels was tested with repeated-measures ANOVA on the factor “stimulation” [stimulated, rest] with a post hoc test of Fisher. Additional factors were “CaMKII isoform” [ $\beta_M$ ,  $\delta A$ ,  $\delta D/\gamma B$ ], ‘stimulation frequency’ [10 Hz, 150 Hz], and muscle compartment [glycolytic *gastrocnemius medialis*, oxidative *gastrocnemius medialis*, and *soleus*]. These were run separately or in combination. The effect of stimulation with Protocols 3 and 4 on phospho-Thr287-CaMKII levels was tested with ANOVAs on the factor “stimulation” [rest, after 3 tetanic contractions, after 24 tetanic contraction]  $\times$  “muscle compartment” [glycolytic *gastrocnemius medialis*, oxidative *gastrocnemius medialis*]  $\times$  “CaMKII isoform” [ $\beta_M$ ,  $\delta A$ ,  $\delta d/\gamma B$ ] with a post hoc test of Fisher. Two-sided post hoc tests were carried out to analyze the effect of electric stimulation on phospho-Thr287-CaMKII.

### 3. Results

**3.1. CaMKII Isoforms in Skeletal Muscle.** Four CaMKII isoforms, that is,  $\beta_M$ ,  $\delta A$ , and  $\delta d/\gamma B$  were identified in rat *gastrocnemius* muscle in accordance with Rose [12] and Bayer ([13]; Figures 2(a) and 2(b)). *In vitro* autokinase assays confirmed the specificity of detecting CaMKII phosphorylation at Thr287.

The abundance of CaMKII isoforms did not differ between the glycolytic and oxidative compartment of *gastrocnemius medialis* muscle (Figures 2(c), 2(e)/2(f)). The phosphorylation levels of CaMKII at Thr287 were however 2-fold higher in the glycolytic compared to the oxidative compartment of *gastrocnemius medialis* muscle ( $P \leq 0.01$ ; Figure 2(d)). In the oxidative muscle, *m. soleus*, lower levels of the  $\delta d/\gamma B$  CaMKII isoform and two bands at the height of  $\beta_M$  were identified (Figure 2(c)).

**3.2. CaMKII Phosphorylation in Fast-Type Muscle In Silico Is Graded by the Content of Mitochondria.** Calculations were run using CaMuZcIE to assess CaMKII-Thr287 phosphorylation in the two compartments of the fast-type *gastrocnemius medialis* muscle after pacing the muscle with a fast-type, 150 Hz protocol. The results of the calculations for these two *gastrocnemius* compartments are shown in Figure 3.

The *in silico* experiment demonstrated a further pronounced increase in Thr287 phosphorylation of CaMKII in

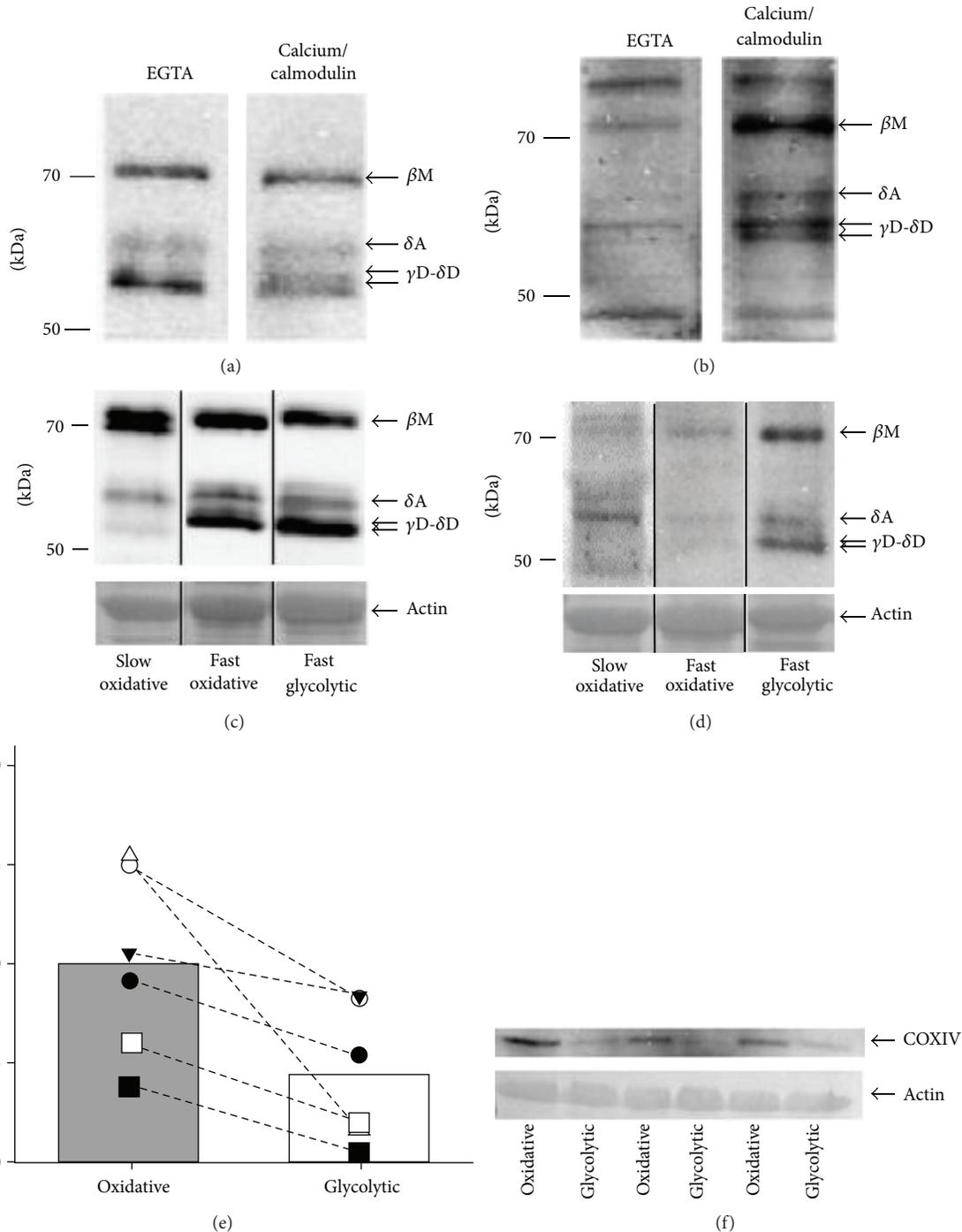


FIGURE 2: Calcium/calmodulin-dependent kinase II isoforms in rat skeletal muscle. Total homogenate of a *gastrocnemius* muscle from a rat was prepared and subjected to *in vitro* kinase reactions in the presence of EGTA or calcium/calmodulin and was subjected to immunoblotting with a pan-CaMKII (a) and phospho-Thr287 specific antibody (b). CaMKII isoforms were then assigned based on a calcium/calmodulin-inducible phospho-Thr287 signal and a detection of similar sized bands according to the nomenclature established by Rose [12] and Bayer [13]. (c)(d) CaMKII isoforms (c) and Thr287 phosphorylated CaMKII (d) in equal protein amount in total homogenate of fatigue resistant (i.e., oxidative) and fatigable (i.e., glycolytic) compartments of the fast *gastrocnemius medialis* muscle and the slow fatigue resistant *soleus* muscle. Lanes originating from different parts of SDS-PAGE gels are separated by black lines. A second band at the height of the  $\beta$ M isoform is detected in *soleus* muscle. (e) Graph displays COXIV levels in the oxidative and glycolytic compartments of *gastrocnemius medialis* as determined by western blotting followed by immunodetection. Bars represent mean COXIV levels and symbols represent the levels in individual sample pairs. Symbols being connected by a stippled line reflect intra-animal pairs. The *P*-level of the difference between red and white GM is indicated (paired *t*-test). (f) Example immunoblot showing the detection of COXIV protein and the actin loading control in the oxidative and glycolytic compartments of *gastrocnemius medialis*.

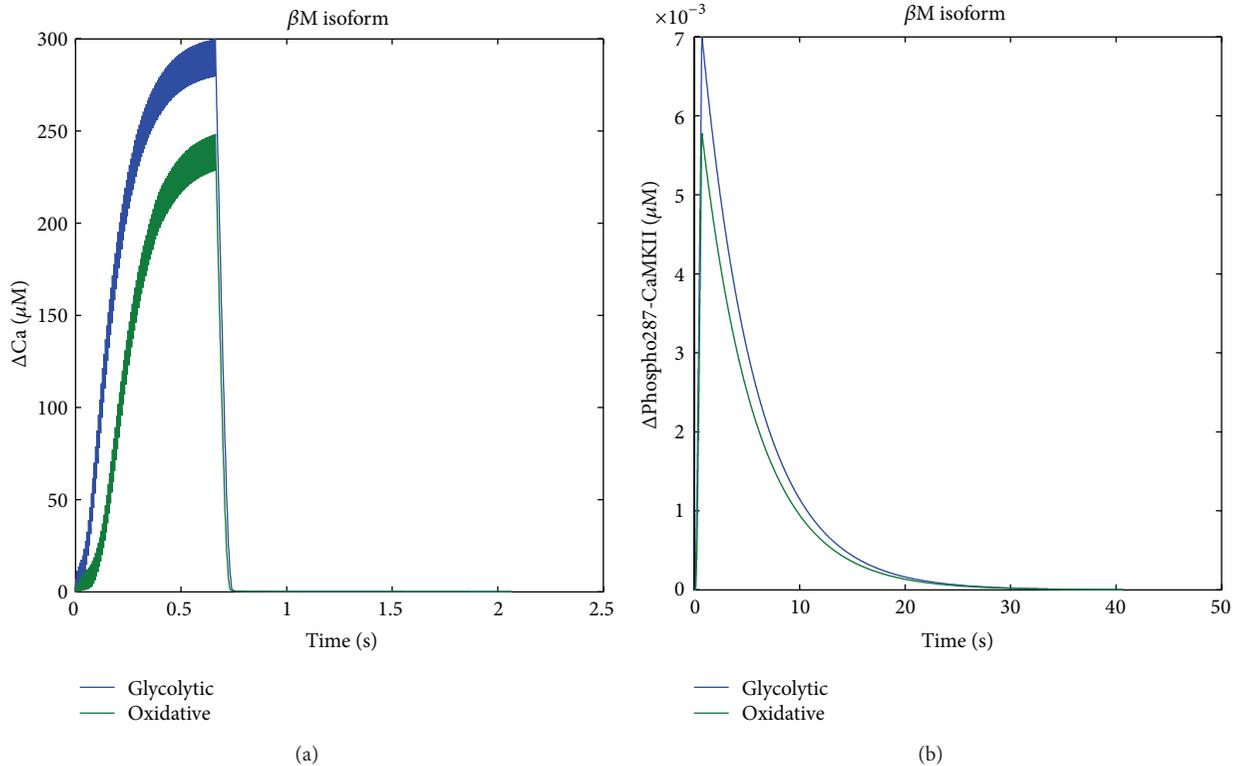


FIGURE 3: Thr287 phosphorylation of CaMKII in fast muscle compartment after paced contractions *in silico*. (a)(b) Line graphs displaying the level changes of calcium (a) and CaMKII-Thr287 (b) in the glycolytic and oxidative compartments of *gastrocnemius* muscle after an input signal of 100 pulses at 150 Hz.

the glycolytic compared to the oxidative compartment of the *gastrocnemius medialis* muscle after paced contractions. Essentially, the same result arose when the calculations were run for individual CaMKII isoforms, yet the degree of peak Thr287 phosphorylation differed in the order  $\beta M > \delta D/\gamma B > \delta A$  (data not shown).

Multiparametric sensitivity analysis was carried out to assess the influence of the model parameters on the simulation of calcium-induced Thr287 phosphorylation of CaMKII (Figure 4). Table 1 shows these results in order of decreasing influence of the assessed parameters.

**3.3. CaMKII Phosphorylation after Paced Contractions in Glycolytic and Oxidative Compartment of Fast-Type Muscle In Situ.** Contraction of *gastrocnemius medialis* muscle was stimulated by applying electric pulses at 10 Hz or 150 Hz to the sciatic nerve (Protocols 1 and 2). The biochemical characterisation identified an increase in Thr287-phosphorylation of all combined CaMKII isoforms in the oxidative compartment of *gastrocnemius medialis* muscle ( $P = 0.002$ ). The analysis of individual isoforms demonstrated increased phospho-Thr287 levels of  $\delta A$  CaMKII ( $P = 0.02$ ), but not  $\beta M$  ( $P = 0.07$ ) and  $\delta D/\gamma B$  CaMKII ( $P = 0.17$ ), in the oxidative muscle compartment (Figures 5(a) and 5(b)). The levels of neither CaMKII isoform differed between resting and 150 Hz

stimulated in the oxidative compartment (all  $P$ -values  $> 0.30$ ; data not shown).

In the glycolytic compartment, Thr287 phosphorylation of CaMKII tended to be less affected than in the oxidative portion of *gastrocnemius medialis* after stimulation ( $P = 0.055$ , for the interaction of "stimulation"  $\times$  "muscle compartment"). In the glycolytic portion none of the CaMKII isoforms demonstrated regulation after stimulation (Figure 5(c)). No interaction effect was identified between stimulation with the slow versus the fast-type protocol in either portion of *gastrocnemius medialis* muscle for Thr287 phosphorylation.

**3.4. Frequency Dependent  $\delta A$  CaMKII Isoform Phosphorylation in the Slow Oxidative Compartment.** We assessed CaMKII isoform phosphorylation in a third muscle of the *triceps surae* complex with a slow-oxidative phenotype, that is, the *m. soleus*, after stimulation with the slow-type (10 Hz) and fast-type (150 Hz) protocol, respectively (Figure 6(a)). The characterisation identified a trend for an interaction between CaMKII isoform and stimulation frequency ( $P = 0.08$ ). When assessing phosphorylation separately per CaMKII isoform we identify a frequency-dependent Thr287 phosphorylation of the  $\delta A$ -CaMKII isoform ( $P = 0.017$ ; Figure 6(b)). Phospho-Thr287 levels of this isoform were selectively downregulated after electric stimulation with the

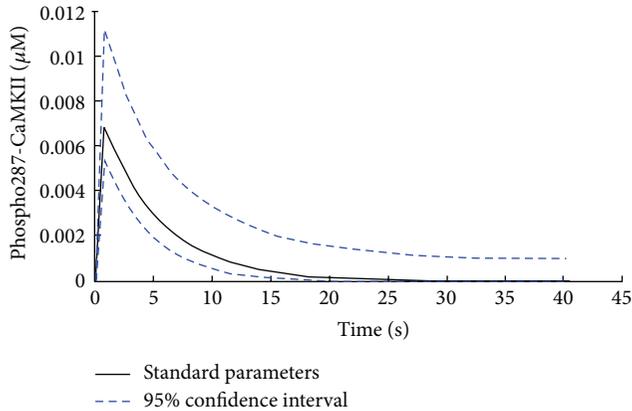


FIGURE 4: Phospho-Thr287-transient of  $\beta$ M CaMKII with 95% confidence intervals. Poisson distribution fitted to the set of output transients obtained by 4000 Latin hypercube-sampled input parameters sets from 75% to 125% of default values. The stimulation protocol with a pulse train of 100 pulses at 150 Hz was used and the model was set up for fast glycolytic muscle. The black line indicates the output transient using the default parameter set. The stippled blue lines indicate the upper and lower 2.5% confidence limit of the Poisson distribution.

fast-type protocol, but it was increased with the slow-type protocol (Figure 6(c)).

**3.5. Muscle Specific CaMKII Activation by Repeated Tetanic Contractions of Gastrocnemius Muscle.** We investigated whether CaMKII activation in glycolytic and oxidative compartment of fast-twitch muscle would be affected with further repetitions of tetanic contractions. The computational model predicted a rapid increase in CaMKII phosphorylation for all isoforms after repeated tetanic contractions in the oxidative portion of the *gastrocnemius* muscle which followed the order  $\beta$ M >  $\delta$ D/ $\gamma$ B >  $\delta$ A. Thr287 phosphorylation had early maxima after 1-2 contractions before falling to near resting values (Figures 7(a) and 7(b)).

Except for the  $\delta$ A isoform, the simulated biphasic relationship between phospho-Thr287 and the number of contraction was not what we found experimentally *in situ*. There was a significant interaction effect of the stimulation and compartment ( $P = 0.02$ ) for Thr287 phosphorylation of CaMKII *in situ*. Experiments *in situ* reproduced the model observations on the sensitivity of phospho-Thr287 content of CaMKII in the oxidative compartment of *gastrocnemius medialis* and pointed out isoform-specific phosphorylation in this muscle. Thr287 phosphorylation of the  $\delta$ A CaMKII isoform remained increased after three tetanic contractions but ceased after 24 contractions (Figure 7(g)). By contrast Thr287 content of the  $\beta$ M isoform was first significantly increased after 24 tetanic contractions (Figure 7(e)).

## 4. Discussion

The contribution of muscle contractions to activities of daily living varies from single short duration, for the purpose of

acceleration, to continuous contractions being repeated for minutes to hours at a given load until fatigue occurs. While the calcium-induced regulation of myocellular ATPases that produce contractile output is relatively well understood [32], knowledge of calcium-induced regulation of phosphotransferase activities is limited [7, 14]. Towards this end we established a new computational model, CaMuZcE, to investigate whether changes in Thr287 phosphorylation of the calcium-receptive phosphotransferase, CaMKII, following electrically paced contractions could be predicted. CaMKII has been shown to encode information provided by the frequency and amplitude of calcium transients [24] and to regulate excitation-contraction and excitation-transcription coupling [33, 34]. It is unclear which CaMKII isoforms mediate these effects in skeletal muscle but their autophosphorylation pattern in response to fast or slow motor unit-type stimulation might provide clues about CaMKII isoform-specific functions. Therefore, emphasis was put on the elucidation of the response of different isoforms in phenotypically distinct muscles and the influence of slow/fast motor unit stimulation frequencies.

A possible limitation of our approach was that contraction-induced phosphorylation of downstream targets of CaMKII, was not assessed. This experiment is indicated for subsequent studies on CaMKII-regulated signal transduction because Thr287 phosphorylation of CaMKII and *in vitro* measured phosphotransferase activity of CaMKII appears not to be correlated [12]. However, as demonstrated by the same authors, Thr287 phosphorylation of combined CaMKII isoforms increases 1.5-fold in *gastrocnemius* muscle after five repeated tetanic contractions over 10 seconds *in situ* with a further 5-fold increase after 3 minutes of repeated contractions. Our biochemical measurements on the effect of tetanic stimulation over a similar duration (i.e., 0.66 seconds to 2 minutes) are compatible with these results (Figure 5(b)). However, we did not see such large effects on phospho-Thr287 content of CaMKII after one to twenty tetanic contractions and we identify that the  $\delta$ A and  $\beta$ M CaMKII isoforms follow a different time course of phosphorylation (Figures 5 and 7). The latter finding indicates a different contribution of CaMKII isoforms to the reported response of combined phospho-Thr287 content of CaMKII with paced contractions *in situ*. The contrasting magnitude in the level increases for the phospho-Thr287 content of CaMKII isoforms respective to Rose [12] is possibly explained by differences in the stimulation protocol and the procedure of harvesting the muscle. In our case, the latter was carried out to separate the three major compartments of the ankle extensor group *triceps surae*. This sampling scheme allowed us to investigate and identify previously unresolved differences in CaMKII phosphorylation between muscle types, that is, the distal (i.e., glycolytic) and proximal (i.e., oxidative) *m. gastrocnemius medialis* and the slow oxidative *m. soleus*. Our biochemical measures identified that phospho-Thr287 content of  $\delta$ A and  $\beta$ M CaMKII is increased after externally paced tetanic contractions in the oxidative *gastrocnemius medialis* muscle but is not significantly affected in the glycolytic compartment of the same muscle (Figures 5 and 7). The observed difference in

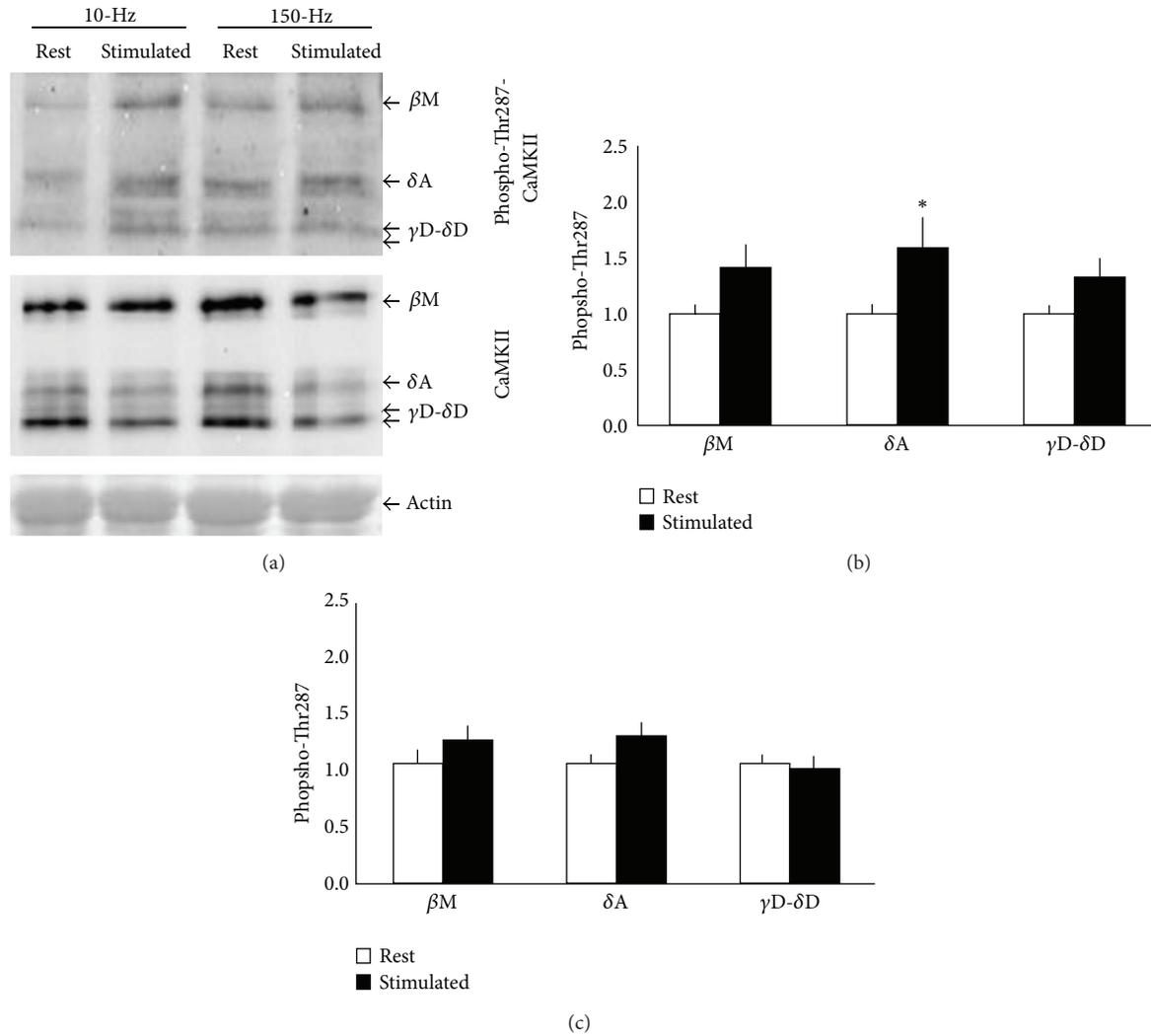


FIGURE 5: Thr287 phosphorylation of CaMKII in glycolytic and oxidative fast muscle compartment after paced contractions *in situ*. (a) Immunoblot showing the phospho-Thr287-CaMKII signal in stimulated and resting control *gastrocnemius medialis* muscle (oxidative compartment) after stimulation of the sciatic nerve with a train of 100 pulses of a slow (i.e., 10 Hz) or a fast (i.e., 150 Hz) motor unit-targeted protocol [9]. (b)(c) Bar graphs visualizing the mean + SE of changes in phospho-Thr287-CaMKII levels with stimulation of the oxidative (b) and glycolytic (c) compartment of *gastrocnemius medialis*. Data reflect combined values from the stimulation 10-Hz and 150-Hz protocol. \*denotes  $P < 0.05$  versus rest (two-sided paired  $t$ -test,  $n = 12-14$ ).

sensitivity of contraction-induced CaMKII phosphorylation between the studied muscle compartments suggests that the anatomical specialisation of muscle fibres is an important factor in the posttranslational regulation of CaMKII.

The role of fibre specialisation for CaMKII isoform autophosphorylation is further corroborated by higher phospho-Thr287 levels of CaMKII at rest between the oxidative and glycolytic portion of *gastrocnemius medialis* (Figure 2). The mechanism behind this baseline difference awaits further exploration. Interestingly, however, the fast-fatigable white *vastus lateralis* muscle of rats responds to repeated voluntary running with an increase in autonomous CaMKII activity [18]. Autonomous phosphotransfer activity of CaMKII reflects increased Thr287 phosphorylation and a subsequent conformation change in the CaMKII enzyme [15]. This suggests a possible contribution of fibre recruitment

during cage activity to the baseline differences in phospho287 content of CaMKII between the oxidative and glycolytic compartment of *gastrocnemius medialis* muscle.

Our investigation demonstrates that autophosphorylation of CaMKII isoforms in skeletal muscle depends on the duration of muscle activity. Interestingly, phospho-Thr287 content of the  $\delta$ A isoform was transiently increased and faded after twenty-four tetanic contractions of *gastrocnemius medialis* muscle, when phosphorylation of the major CaMKII isoform of skeletal muscle,  $\beta$ M, increased (Figures 7(e)/7(g)). The findings suggest that the  $\beta$ M CaMKII isoform of skeletal muscle does, like the  $\alpha$ CaMKII isoform in neurons, act as a decoder of neuronal input [25].  $\beta$ M CaMKII regulates calcium release from the sarcoplasmic reticulum [7, 14] whereas the  $\delta$ A CaMKII isoform has been demonstrated to affect nuclear processes and modify cardiac growth [22, 23].

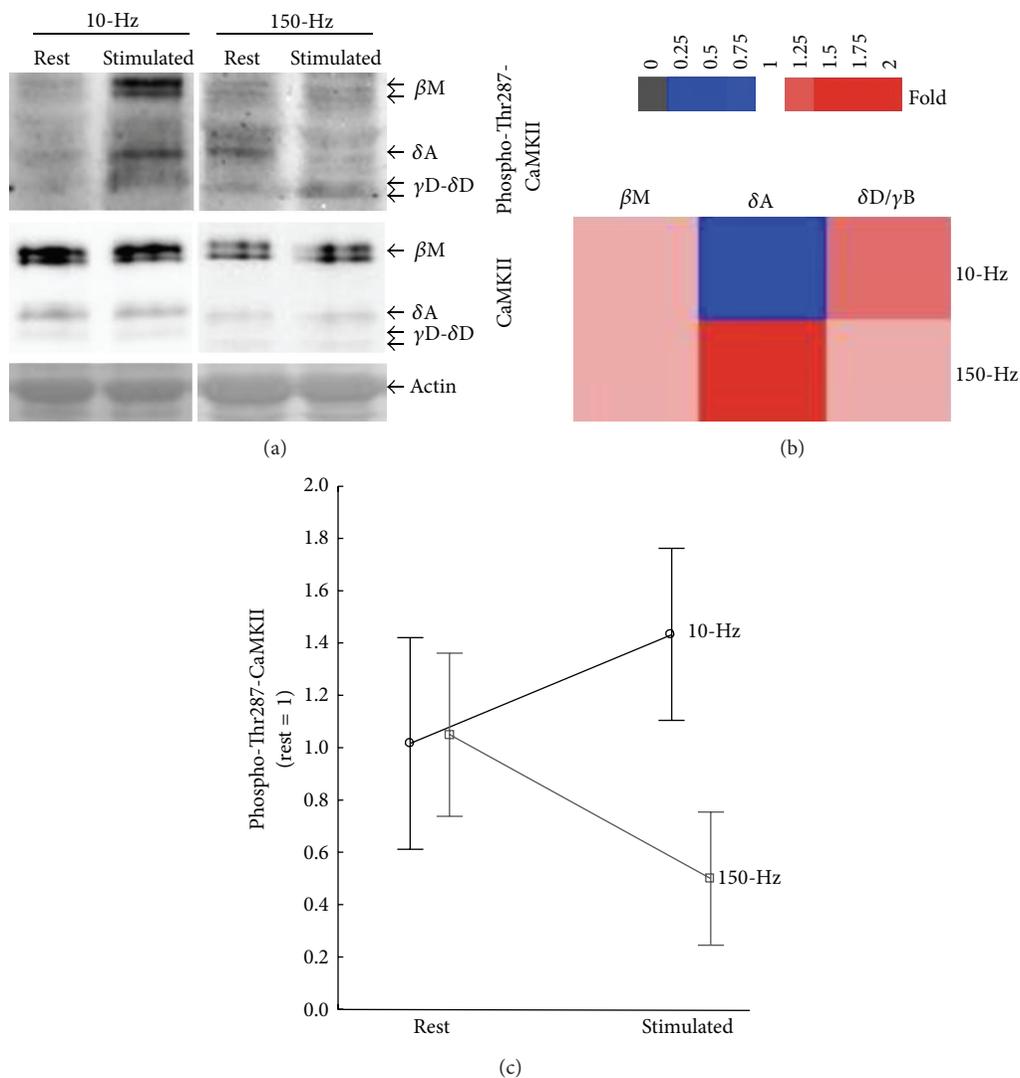


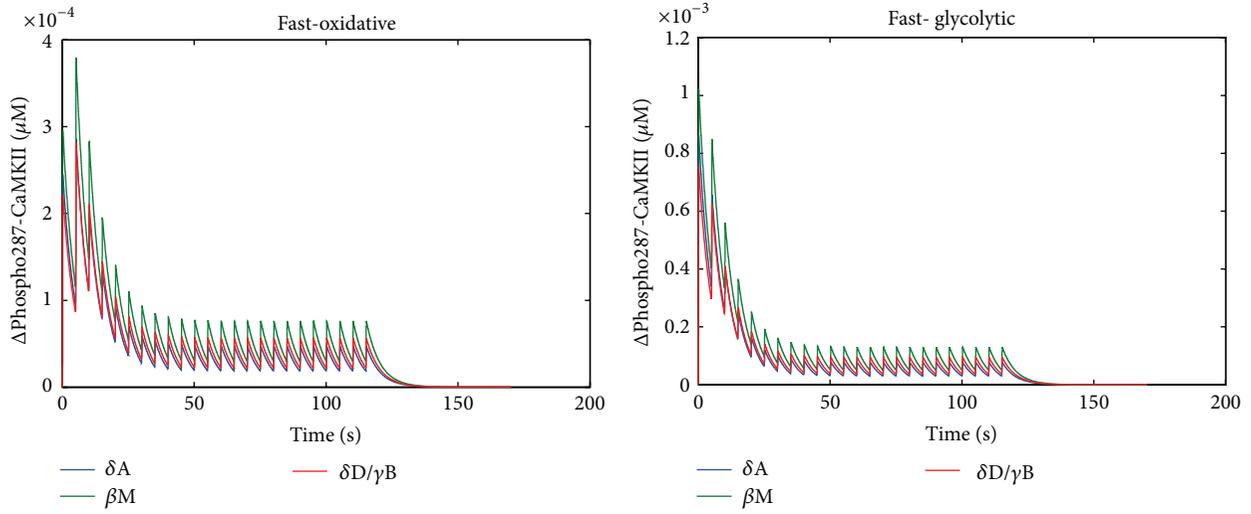
FIGURE 6: Frequency dependent CaMKII isoform autophosphorylation in slow oxidative muscle after paced contractions *in situ*. (a) Example of Thr 287 phosphorylation of CaMKII isoforms in the stimulated (stim) *soleus* muscle and its contralateral control (rest) after electric stimulation with 150 Hz (Protocol 1) and 10 Hz (Protocol 2). The position of the respective bands is indicated by an arrow. Please note the presence of a second band at the height of the  $\beta$ M isoform in stimulated muscle. (b) Composite figure visualising the fold changes in CaMKII-Thr287 content of CaMKII isoforms for slow and fast motor unit-targeted stimulation for CaMKII isoforms in slow oxidative *soleus* muscle in a colour code. (c) Line graph resolving the interaction effect of stimulation frequency on Thr287 phosphorylation of the  $\delta$ A-CaMKII isoform. Repeated ANOVA with post hoc test of Fisher,  $n = 6-10$ .

Collectively the findings highlight a possible contribution of isoform-specific CaMKII autophosphorylation to excitation-induced signalling in skeletal muscle.

The observed Thr287 phosphorylation of  $\beta$ M CaMKII *in situ* (Figure 2), and its association with muscle fatigue (as indicated by a 30% decrease in maximal tetanic force), is in line with the proposition by Tavi [14] on a summation of CaMKII autophosphorylation with repeated contractions. The latter observation was however not diversified for CaMKII isoforms and verified with biochemical analysis. The mechanistic basis of this effect is therefore unclear. Our report points out that autophosphorylation of the  $\beta$ M isoform rises

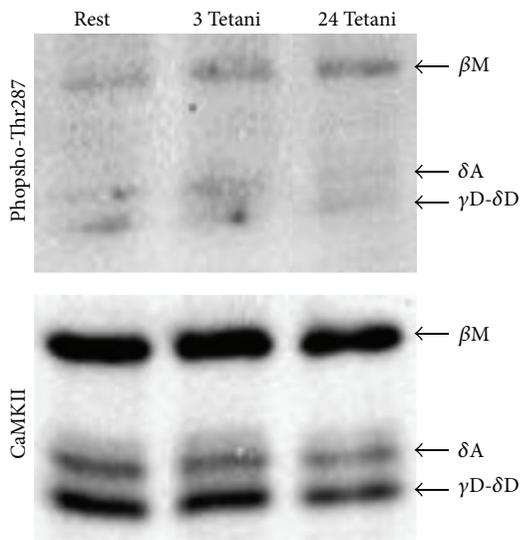
slower than  $\delta$ A CaMKII isoform after repeated contractions *in situ* (Figures 7(e)/7(g)). This is intriguing because  $\beta$ M localizes to the sarcoplasmic reticulum from which intracellular calcium is released [13] and because sensitivity analysis identifies that the most influential parameter for the rate of CaMKII autophosphorylation,  $k_{\text{bta}}$  [20], is considerably higher for the  $\beta$ M than the  $\delta$ A CaMKII isoform (Table 1).

By contrast, the calmodulin affinity rate constants, which are derived from the parameter  $k_{\text{bi}}$ , are of lesser importance for our model (0.06 versus 0.32, Table 1), indicating that instead of the calmodulin affinity, the rate constant of autophosphorylation is the rate-determining step. At

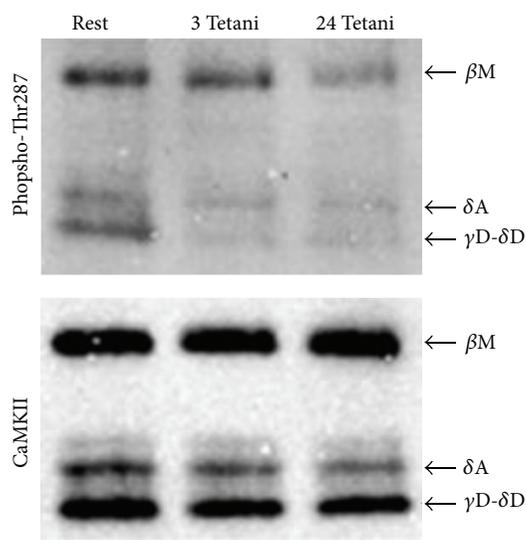


(a)

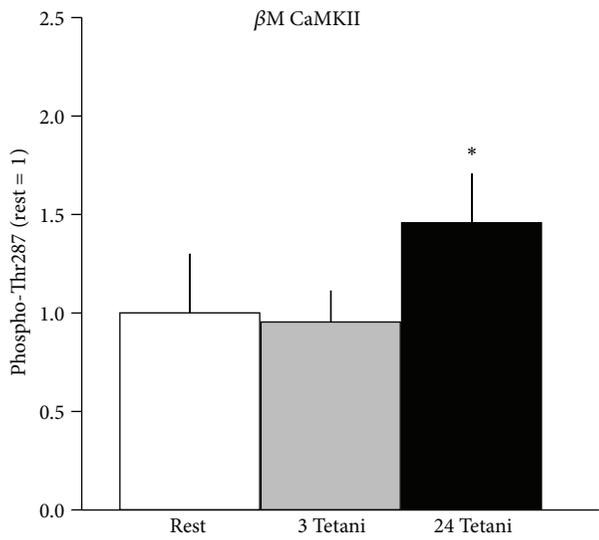
(b)



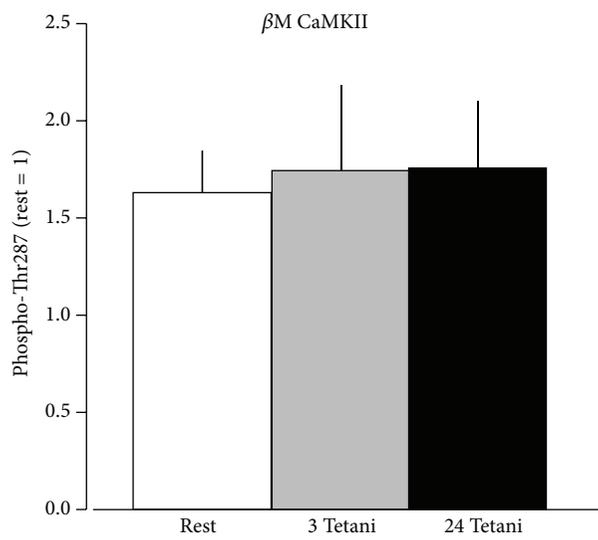
(c)



(d)



(e)



(f)

FIGURE 7: Continued.

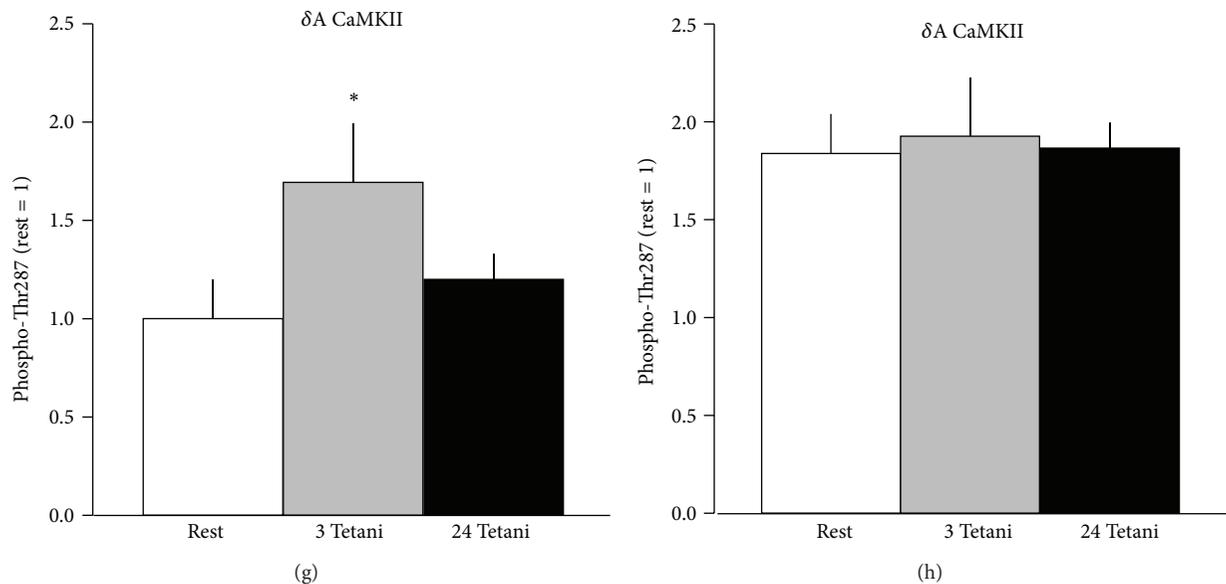


FIGURE 7: Isoform specific CaMKII autophosphorylation with repeated contractions *in situ*. (a)(b) Line graph of the calculated changes in Thr287 phosphorylation of CaMKII isoforms in the oxidative (a) and glycolytic (b) compartment of *gastrocnemius medialis* muscle as a function of the number of tetanic contractions as paced with repeated trains of 30 pulses of 150 Hz with 4.8 seconds of rest between trains. (c)–(h) Representative immunoblots (c)(d) and bar graphs of mean (e)–(h) of Thr287 phosphorylated CaMKII isoforms in the oxidative (c) (e) (g) and glycolytic (d) (f) (h) compartment of *gastrocnemius medialis* muscle after 3 and 24 repeated tetanic contractions stimulated via the sciatic nerve with trains of 30 pulses at 150 Hz. \* denotes  $P < 0.05$  versus rest (ANOVA with post hoc test of Fisher,  $n = 6$ ).

higher values of this rate constant of autophosphorylation, however, the calmodulin affinity to CaMKII becomes a more important parameter as binding to calmodulin needs to occur before CaMKII autophosphorylation can take place [15]. Both parameters need to be elevated to facilitate a higher total rate of autophosphorylation. Accordingly, this influence is specifically more pronounced for the  $\beta M$  isoform due to its higher rate constant [20]. We therefore speculate that the *in situ* observed pattern of  $\beta M$  CaMKII isoform autophosphorylation (Figures 7(e)/7(g)) reflects differences in the threshold of calmodulin-dependent autophosphorylation for CaMKII isoforms, or calmodulin availability, rather than kinetic constants of autophosphorylation as characterized *in vitro* [20, 24].

A major finding in this regard was that Thr287 phosphorylation of the  $\delta A$  CaMKII isoform is highly sensitive to excitation and differentiates the response of the fatigue-resistant *soleus* muscle between slow and fast motor unit-type stimulation (Figures 5 and 6; [1]). Interestingly, the increased phospho-Thr287 content of  $\delta A$  CaMKII was only observed with the stimulation frequency corresponding to the “natural” recruitment frequency of the motoneuron in this muscle, that is, 10 Hz; [1, 8]. This supports the idea that elevated  $\delta A$  CaMKII Thr287 phosphorylation in oxidative muscle types depends on the recruitment pattern of motor units. As the *soleus* muscle is also stimulated by the 150 Hz protocol our findings bear the notion that it is the high frequency tetanic excitation that leads to a dephosphorylation of CaMKII  $\delta A$ .

In the course of our investigation, we have established a new computational model that connects calcium dynamics to downstream activation of the regulator of excitation-contraction coupling, CaMKII, in muscle fibers. We used this model to predict possible CaMKII isoform differences in level alterations of Thr287 phosphorylation between stimulation protocols and motor unit types. The phosphatase concentration *in silico* was calibrated by taking the measured decay of Thr287 phosphorylation of CaMKII into account. Of interest in this regard is the fact that *in situ* we observed only a relatively moderate reduction in the content of CaMKII-Thr 287 phosphorylation during recovery from 24 tetanic contractions (data not shown). These values indicate that  $V_{max}$  or concentration of the phosphatase activity dephosphorylating CaMKII is considerably lower than the value used in the first computational model of CaMKII phosphorylation [16], approximating the initially published values [17]. Accordingly the phosphatases PP1 would be far less efficient to dephosphorylate Thr287 on CaMKII in rat skeletal muscle compared to other tissues.

The output of our *in silico* model emphasises that differences in mitochondrial content as seen for the glycolytic and oxidative *gastrocnemius medialis* muscle bring about different alterations in Thr287 phosphorylation of CaMKII with excitation. The model explains this difference in terms of lower increases in free calcium in the sarcoplasm of oxidative muscle fibers due to the calcium buffer capacity of mitochondria [27]. The model also predicted that Thr287 phosphorylation of CaMKII would fall after an initial peak

fade with repeated contraction (Figure 3), a prediction that was confirmed by the observed transient increase in phospho-Thr287 content of  $\delta A$  CaMKII (Figure 7(g)). Our approach can serve as an example as to how *in silico* data can be used to plan experiments and interpret biological data.

However, the results from our modelling did not predict the *in situ* observed differences in the time course of Thr287 phosphorylation between CaMKII isoforms and the discrete influence of the muscle phenotype. For instance modelling predicted a similar transient increase in phospho-Thr287 content of  $\beta M$  and  $\delta A$  CaMKII after tetanic contractions but which differed from the data gathered *in situ* (Figure 3 versus Figures 7(a)/7(b)). As well, the predictions on the larger increase in Thr287-phosphorylation in the glycolytic relative to the oxidative compartment of *m. gastrocnemius medialis* differed from the measured data (Figure 3). This disconnection between *in silico* and *in situ* data emphasises that further experimental input from measured parameters, specifically those addressing influential model parameters (Table 1) and their relationship to baseline values and threshold of calcium-induced activation, and  $\beta M$  and  $\delta A$  isoform localization, is necessary to refine the model.

## 5. Conclusions

Our findings provide the first evidence that CaMKII isoforms serve as muscle and frequency-specific sensors of muscle excitation. This indicates that aside from the characteristics of a contractile protocol, isoform and muscle type-specific activation of CaMKII must be taken into consideration when interpreting the physiological activation of reactions downstream of CaMKII with muscle contraction *in vivo*. Results from our computational investigations offer a first consolidation of the experimental observations and indicate that the boundary conditions for the modelling of CaMKII regulation need to be adjusted to reveal a mechanistic explanation of CaMKII autophosphorylation in function of calcium dynamics.

## Appendices

### A. Fluxes

*A.1. Pumps of the Sarcoplasmic Reticulum.* The RyRs are modeled using an empirical function, derived from Baylor [36]:

$$J_{\text{RyR}} = \left( [\text{Ca}]_{\text{SR}} - [\text{Ca}]_{\text{Myo}} \right)^{e^{-t/\tau_2}} * \text{CaMax} * \left( 1 - e^{-t/\tau_1} \right)^{\text{power1}}. \quad (\text{A.1})$$

It is based on the opening of the channels inside the RyRs, which are regulated by electrical signals coming from the nerves going via the sarcolemma. Parameters involved are the reference calcium flow, the two characteristic opening times of the channels, and the exponential coefficients describing the weight of the opening times. Furthermore, the flux is made dependent on the difference in calcium concentration to incorporate a diffusion component.

The SERCA pumps are modeled as described in [37]:

$$J_{\text{SERCA}} = \frac{\text{CaPump} * [\text{Ca}]_{\text{Myo}}^2}{[\text{Ca}]_{\text{Myo}}^2 + K_{d,\text{pump}}^2}. \quad (\text{A.2})$$

The rate is second order, indicating cooperativity, as the pump transports two ions at the same time for one ATP molecule. Other involved parameters are the reference pumping rate and the dissociation constant of calcium binding.

*A.2. Calcium Buffers.* The change in concentration of the buffers calsequestrin, parvalbumin, troponin, ATP, and the mitochondria is modeled using mass-action kinetics. Calsequestrin is a low-affinity, high-capacity buffer inside the SR that stores calcium. Parvalbumin is a high capacity calcium buffer that also binds magnesium, therefore making it a slow onset calcium buffer. Mitochondria also bind calcium and magnesium, with the same parameters as parvalbumin. ATP can buffer calcium as well as magnesium; but using a reduced reaction this can be approximated by only modeling the calcium binding, as changes in total free Mg concentration are small [36]. It is a low-affinity, rapid acting calcium buffer. Calcium binds to troponin to induce contraction. These buffers are modeled using a second order reaction

$$J_{\text{buf-Ca}} = k_{\text{on}} * [\text{Ca}]_{\text{Myo}} * [\text{buf}]_{\text{free}} - k_{\text{off}} * [\text{buf}]_{\text{Ca}}, \quad (\text{A.3})$$

with  $k_{\text{on}}$  and  $k_{\text{off}}$  the rate constants of binding and release of the complex, respectively, for each buffer,  $[\text{Ca}]_{\text{Myo}}$  the calcium concentration in the sarcoplasm,  $[\text{buf}]_{\text{free}}$  the free buffer concentration, and  $[\text{buf}]_{\text{Ca}}$  the concentration of buffer bound to calcium. As parvalbumin also binds to magnesium, a second reaction describes the reaction with magnesium

$$J_{\text{Parv-Mg}} = k_{\text{on,Mg}} * [\text{Mg}]_{\text{Myo}} * [\text{Parv}]_{\text{free}} - k_{\text{off,Mg}} * [\text{Parv}]_{\text{Mg}}. \quad (\text{A.4})$$

The fluxes describe the change in concentration of the buffers bound to calcium or magnesium. The concentrations of all the other species can then be deduced from these fluxes.

*A.3. CaM and CaMKII.* The CaM and CaMKII reactions are modeled as described by Saucerman [16]. CaM is a calcium sensor that can bind 4 calcium molecules, two to the N-terminal EF hand and two to the C-terminal EF hand of the amino-acid chain. The binding kinetics can be reduced to a three state model to reduce parameters, as binding at each location is cooperative. Furthermore, each state also exists in a buffered form. This model is included in the solver by first calculating the reaction rates, where for each state transition the rate is calculated by taking the binding rate and subtracting the release rate (A.5). These rates are described by mass action kinetics, resulting in a third order reaction as 2 calcium ions bind to one CaM protein at the same time. The reaction rates are then combined into fluxes for each of the six CaM and CaM-buffer states (A.5).

Next, CaMKII reactions are modeled in an equal fashion. CaMKII exists in 6 different states, with 4 of these states

TABLE 2: Initial concentrations.

Name	Value	Unit	Description	Reference
CaMKIItot	1	[ $\mu\text{M}$ ]	Total CaMKII concentration	<a href="http://dare.ubv.vu.nl/handle/1871/40249">http://dare.ubv.vu.nl/handle/1871/40249</a>
CaMfreeM	0.15	[ $\mu\text{M}$ ]	Total CaM concentration	<a href="http://dare.ubv.vu.nl/handle/1871/40249">http://dare.ubv.vu.nl/handle/1871/40249</a>
BfreeM	20	[ $\mu\text{M}$ ]	Total CaM buffer concentration	Saucerman [16]
PPItot	0.15	[ $\mu\text{M}$ ]	PP1 concentration	Saucerman [16]
Mg	1000	[ $\mu\text{M}$ ]	Magnesium concentration sarcoplasm	Saucerman [16]
K	160000	[ $\mu\text{M}$ ]	Potassium concentration sarcoplasm	Saucerman [16]
CaSRfree	1000	[ $\mu\text{M}$ ]	Initial calcium concentration in sarcoplasmic reticulum	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
CSQNT	60000	[ $\mu\text{M}$ ]	Total calsequestrin concentration	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
Ca	0.1	[ $\mu\text{M}$ ]	Initial calcium concentration sarcoplasm	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
TropT	205	[ $\mu\text{M}$ ]	Total troponin C concentration	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
ParvT	1500	[ $\mu\text{M}$ ]	Total parvalbumin concentration	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
MitT	100	[ $\mu\text{M}$ ]	Total mitochondria concentration	[35]
ATPT	8000	[ $\mu\text{M}$ ]	Total APT concentration	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
DyeT	81	[ $\mu\text{M}$ ]	Total dye concentration	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
Pb2R	0	[ $\mu\text{M}$ ]	Initial CaMKII Pb2 in state	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
PbR	0	[ $\mu\text{M}$ ]	Initial CaMKII Pb in state	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
PtR	0	[ $\mu\text{M}$ ]	Initial CaMKII Pt in state	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
Pt2R	0	[ $\mu\text{M}$ ]	Initial CaMKII Pt2 in state	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
PaR	0	[ $\mu\text{M}$ ]	Initial CaMKII PaR in state	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
PiR	1	[ $\mu\text{M}$ ]	Initial CaMKII PiR in state	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
$\text{XB}_n$	1	[—]	Initial fraction of total action/myosin chains in $\text{XB}_n$ state	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
$\text{XB}_w$	0	[—]	Initial fraction of total action/myosin chains in $\text{XB}_w$ state	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
$\text{XB}_s$	0	[—]	Initial fraction of total action/myosin chains in $\text{XB}_s$ (force) state	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
$\text{XB}_r$	0	[—]	Initial fraction of total action/myosin chains in $\text{XB}_r$ state	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>

together forming the active fraction of CaMKII (Pb, Pt, Pa, Pt2) and 3 of these four states phosphorylated at the threonine 287 site (Pt, Pa, Pt2). Switching between these states is facilitated by either (de) phosphorylation, Ca2CaM binding and release, or Ca4CaM binding and release. Again, the reaction rates are calculated for each state transition (A.5), resulting in either second or third order reactions. Next, the reaction rates are combined into fluxes for each state (A.5). Consider

$$R02 = k_{\text{CaM}02} * [\text{Ca}]_{\text{Myo}}^2 * [\text{CaM}] - k_{\text{CaM}20} * [\text{Ca}_2 - \text{CaM}],$$

$$R24 = k_{\text{CaM}24} * [\text{Ca}]_{\text{Myo}}^2 * [\text{Ca}_2 - \text{CaM}]$$

$$- k_{\text{CaM}42} * [\text{Ca}_4 - \text{CaM}],$$

$$R02B = k_{\text{CaM}02B} * [\text{Ca}]_{\text{Myo}}^2 * [\text{CaM} - \text{Buffer}]$$

$$- k_{\text{CaM}20B} * [\text{CaM} - \text{Buffer} - \text{Ca}_2],$$

$$R24B = k_{\text{CaM}24B} * [\text{Ca}]_{\text{Myo}}^2 * [\text{CaM} - \text{Buffer} - \text{Ca}_2]$$

$$- k_{\text{CaM}42B} * [\text{CaM} - \text{Buffer} - \text{Ca}_4],$$

$$R2B = k_{\text{CaM}2\text{Bon}} * [\text{Ca}_2 - \text{CaM}] * [\text{Buffer}]$$

$$- k_{\text{CaM}2\text{Boff}} * [\text{CaM} - \text{Buffer} - \text{Ca}_2],$$

$$R4B = k_{\text{CaM}4\text{Bon}} * [\text{CaM} - \text{Buffer}] * [\text{Buffer}]$$

$$- k_{\text{CaM}4\text{Boff}} * [\text{CaM} - \text{Buffer} - \text{Ca}_4],$$

$$RCKib = k_{ib} * [\text{Ca}_4 - \text{CaM}] * [\text{Pi}] - k_{bi} * [\text{Pb}],$$

$$RCKib2 = k_{ib2} * [\text{Ca}_2 - \text{CaM}] * [\text{Pi}] - k_{b2i} * [\text{Pb}2],$$

$$RCKb2b = k_{b24} * [\text{Ca}]_{\text{Myo}}^2 * [\text{Pb}2] - k_{b42} * [\text{Pb}],$$

$$RCKbt = k_{bt} * [\text{Pb}] - \frac{k_{\text{PP1}} * [\text{PP1}] * [\text{Pt}]}{K_{m,\text{PP1}} + [\text{CaMKII}]_{\text{tot}} * [\text{Pt}]},$$

$$RCKtt2 = k_{t42} * [\text{Pt}] - k_{t24} * [\text{Ca}]_{\text{Myo}}^2 * [\text{Pt}2],$$

$$RCKt2b2 = \frac{k_{\text{PP1}} * [\text{PP1}] * [\text{Pt}2]}{K_{m,\text{PP1}} + [\text{CaMKII}]_{\text{tot}} * [\text{Pt}2]},$$

$$RCKta = k_{ta} * [\text{Pt}] - k_{at} * [\text{Ca}_4 - \text{CaM}] * [\text{Pa}],$$

TABLE 3: Rate constants.

Name	Value	Unit	Description	Reference
CaMax	25	$[\mu\text{M}^{-1}]$	$\text{Ca}^{2+}$ release rate	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
Power1	5	[—]	RyR kinetic parameter	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
tau1	0.609	[s]	RyR kinetic parameter	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
Power2	3	[—]	RyR kinetic parameter	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
tau2	0.774	[s]	RyR kinetic parameter	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
CaPump	20	$[\mu\text{M} * \text{ms}^{-1}]$	Maximum SERCA pumping	<a href="http://dare.uvu.vu.nl/handle/1871/40249">http://dare.uvu.vu.nl/handle/1871/40249</a>
KdPump	1	$[\mu\text{M}]$	Dissociation constant calcium binding to SERCA	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kAtp1	0.054	$[\mu\text{M}^{-1} * \text{ms}^{-1}]$	On rate reaction $\text{Ca}^{2+}$ -ATP	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kAtp2	120	$[\text{ms}^{-1}]$	Off rate reaction $\text{Ca}^{2+}$ -ATP	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kTrop1	0.08	$[\mu\text{M}^{-1} * \text{ms}^{-1}]$	On rate reaction $\text{Ca}^{2+}$ -troponin C	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kTrop2	0.32	$[\text{ms}^{-1}]$	Off rate reaction $\text{Ca}^{2+}$ -troponin C	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kParvCa1	0.37	$[\mu\text{M}^{-1} * \text{ms}^{-1}]$	On rate reaction $\text{Ca}^{2+}$ -parvalbumin	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kParvCa2	0.003	$[\text{ms}^{-1}]$	Off rate reaction $\text{Ca}^{2+}$ -parvalbumin	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kParvMg1	$1.05E - 04$	$[\mu\text{M}^{-1} * \text{ms}^{-1}]$	On rate reaction $\text{Mg}^{2+}$ -parvalbumin	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kParvMg2	0.012	$[\text{ms}^{-1}]$	Off rate reaction $\text{Mg}^{2+}$ -parvalbumin	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kMitCa1	0.37	$[\mu\text{M}^{-1} * \text{ms}^{-1}]$	On rate reaction $\text{Ca}^{2+}$ -mitochondria	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kMitCa2	0.003	$[\text{ms}^{-1}]$	Off rate reaction $\text{Ca}^{2+}$ -mitochondria	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kMitMg1	$1.05E - 04$	$[\mu\text{M}^{-1} * \text{ms}^{-1}]$	On rate reaction $\text{Mg}^{2+}$ -mitochondria	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kMitMg2	0.012	$[\text{ms}^{-1}]$	Off rate reaction $\text{Mg}^{2+}$ -mitochondria	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kCsqn1	0.1	$[\mu\text{M}^{-1} * \text{ms}^{-1}]$	On rate reaction $\text{Ca}^{2+}$ -calsequestrin	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kCsqn2	100	$[\text{ms}^{-1}]$	Off rate reaction $\text{Ca}^{2+}$ -calsequestrin	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kDye1	0.0864	$[\mu\text{M}^{-1} * \text{ms}^{-1}]$	On rate reaction $\text{Ca}^{2+}$ -Dye	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kDye2	6.05	$[\text{ms}^{-1}]$	Off rate reaction $\text{Ca}^{2+}$ -Dye	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kCam20	0.01	$[\text{ms}^{-1}]$	Ca dissociation from CaM (C-terminal)	Saucerman [16]
kCam42	0.5	$[\text{ms}^{-1}]$	2 Ca dissociation from CaM (N-terminal)	Saucerman [16]
kCam0Boff	$1.40E - 06$	$[\text{ms}^{-1}]$	CaM dissociation from buffer	Saucerman [16]
kPP1	0.0017	$[\text{ms}^{-1}]$	Thr287 dephosphorylation	Saucerman [16]
KmPP1	11	$[\mu\text{M}]$	Km for Thr287 dephosphorylation	Saucerman [16]
kbi	0.0013	$[\text{ms}^{-1}]$	Ca4CaM dissociation from Pb	Gaertner [20]
kbta	0.018	[—]	Polynomial factor autophosphorylation	Gaertner [20]
kbtb	0.015	[—]	Polynomial factor autophosphorylation	Gaertner [20]
kbtc	0.033	[—]	Polynomial factor autophosphorylation	Gaertner [20]

TABLE 3: Continued.

Name	Value	Unit	Description	Reference
KF50	8.4	[ $\mu\text{M}$ ]	Dissociation constant Ca-troponin C from $\text{XB}_n$ state	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
NF2	1.68	[—]	Power constant for Ca-troponin C influence on kf1	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
KF502	27	[ $\mu\text{M}$ ]	Dissociation constant Ca-troponin C from $\text{XB}_w$ state	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
NF2	8	[—]	Power constant for Ca-troponin C influence on kf2	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kFm1s	0.0045	[ $\text{ms}^{-1}$ ]	On rate reaction $\text{XB}_n\text{-XB}_w$	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kFm1	0.045	[ $\text{ms}^{-1}$ ]	Off rate reaction $\text{XB}_n\text{-XB}_w$	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kF0	$1.00E - 03$	[ $\text{ms}^{-1}$ ]	Rate second reaction $\text{XB}_n\text{-XB}_w$	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kFg1	0.0169	[ $\text{ms}^{-1}$ ]	Rate reaction $\text{XB}_s\text{-XB}_r$	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kFg2	0.0337	[ $\text{ms}^{-1}$ ]	Rate reaction $\text{XB}_r\text{-XB}_n$	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kFm2	0.13	[ $\text{ms}^{-1}$ ]	On rate reaction $\text{XB}_w\text{-XB}_s$	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kFm2s	1	[ $\text{ms}^{-1}$ ]	Off rate reaction $\text{XB}_w\text{-XB}_s$	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>

TABLE 4: CaMKII isoforms.

Name	Value	Unit	Description	Reference
$\delta A$				
kbi	0.0013	[ $\text{ms}^{-1}$ ]	Ca4CaM dissociation from Pb	Gaertner [20]
kbta	0.018	[—]	Polynomial factor autophosphorylation	Gaertner [20]
kbtb	0.015	[—]	Polynomial factor autophosphorylation	Gaertner [20]
kbtc	0.033	[—]	Polynomial factor autophosphorylation	Gaertner [20]
$\beta M$				
kbi	0.00054	[ $\text{ms}^{-1}$ ]	Ca4CaM dissociation from Pb	Gaertner [20]
kbta	0.043	[—]	Polynomial factor autophosphorylation	Gaertner [20]
kbtb	0.0062	[—]	Polynomial factor autophosphorylation	Gaertner [20]
kbtc	0.019	[—]	Polynomial factor autophosphorylation	Gaertner [20]
$\delta d$				
kbi	0.00015	[ $\text{ms}^{-1}$ ]	Ca4CaM dissociation from Pb	Gaertner [20]
kbta	0.0053	[—]	Polynomial factor autophosphorylation	Gaertner [20]
kbtb	0.00092	[—]	Polynomial factor autophosphorylation	Gaertner [20]
kbtc	0.066	[—]	Polynomial factor autophosphorylation	Gaertner [20]
$\gamma B$				
kbi	0.00070	[ $\text{ms}^{-1}$ ]	Ca4CaM dissociation from Pb	Gaertner [20]
kbta	0.055	[—]	Polynomial factor autophosphorylation	Gaertner [20]
kbtb	0.0074	[—]	Polynomial factor autophosphorylation	Gaertner [20]
kbtc	0.015	[—]	Polynomial factor autophosphorylation	Gaertner [20]

$$RCKt2a = k_{t2a} * [\text{Pt2}] - k_{at2} * [\text{Ca}_2 - \text{CaM}] * [\text{Pa}],$$

$$RCKai = \frac{k_{PP1} * [\text{PP1}] * [\text{Pa}]}{K_{m,PP1} + [\text{CaMKII}]_{\text{tot}} * [\text{Pa}]},$$

$$J_{\text{CaM}} = -R02 - R0B,$$

$$J_{\text{Ca}_2\text{-CaM}} = R02 - R24 - R2B + RCKt2a - RCKib2,$$

$$J_{\text{Ca}_4\text{-CaM}} = R24 - R4B + RCKta - RCKib,$$

$$J_{\text{CaM-Buffer}} = R0B - R02B,$$

$$J_{\text{CaM-Buffer-Ca}_2} = R02B - R24B + R2B,$$

$$J_{\text{CaM-Buffer-Ca}_4} = R24B + R4B,$$

$$J_{\text{Pi}} = -RCKib - RCKib2 + RCKai,$$

TABLE 5: Slow glycolytic parameter changes compared to fast glycolytic.

Name	Value	Unit	Description	Reference
Ca	0.15	[ $\mu\text{M}$ ]	Initial calcium concentration sarcoplasm	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
TropT	102	[ $\mu\text{M}$ ]	Total troponin C concentration	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
ParvT	0	[ $\mu\text{M}$ ]	Total parvalbumin concentration	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
MitT	300	[ $\mu\text{M}$ ]	Total parvalbumin concentration	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
ATPT	5000	[ $\mu\text{M}$ ]	Total APT concentration	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
CaMax	8	[ $\mu\text{M}^{-1}$ ]	$\text{Ca}^{2+}$ release rate	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
Power1	6.71	[—]	RyR kinetic parameter	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
tau1	1.06	[s]	RyR kinetic parameter	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
Power2	2.98	[—]	RyR kinetic parameter	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
tau2	0.98	[s]	RyR kinetic parameter	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
CaPump	1	[ $\mu\text{M} * \text{ms}^{-1}$ ]	Maximum SERCA pumping	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>
kTrop2	0.16	[ $\text{ms}^{-1}$ ]	Off rate reaction $\text{Ca}^{2+}$ -troponin C	<a href="http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/">http://www.tue.nl/en/publication/ep/p/d/ep-uid/247565/</a>

TABLE 6: Fast oxidative parameter changes compared to fast glycolytic.

Name	Value	Unit	Description	Reference
MitT	250	[ $\mu\text{M}$ ]	Total concentration of mitochondria	<a href="http://dare.ubvu.vu.nl/handle/1871/40249">http://dare.ubvu.vu.nl/handle/1871/40249</a>

$$J_{Pb} = RCKib + RCKb2b - RCKbt,$$

$$J_{Pb2} = RCKib2 - RCKb2b + RCKt2b2,$$

$$J_{Pt} = RCKbt - RCKtt2 - RCKta,$$

$$J_{Pt2} = RCKtt2 - RCKt2b2 - RCKt2a,$$

$$J_{Pa} = -RCKai + RCKta + RCKt2a.$$

(A.5)

Some of the rate constants depend on the concentration of the ions magnesium and potassium. Phosphorylation is caused by the protein phosphorylating itself (termed autophosphorylation), where dephosphorylation is facilitated by the protein PPI.

**A.4. Force.** The force is modeled as described by Groenendaal [29], which compared different models for coupling the troponin concentration to force production by the myofibrils. It is described by a four state model of the actin and tropomyosin chains. Calcium bound troponin C (ca-TropC)

facilitates the transformation between states. Without ca-TropC all the myosin chains are in the  $\text{XB}_n$  state. When ca-TropC is formed, the myosin chains go through a series of states called the cross bridge cycle, thereby inducing contraction of the muscle fibers and force production.

The cycle is modeled using rate constants, with some dependent on the concentration of ca-TropC:

$$J_{\text{XB}_n} = -k1 ([\text{Trop} - \text{Ca}]) * \text{XB}_n - k_{F,0} * \text{XB}_n + k_{F,-1} * \text{XB}_w + k_{Fg1} * \text{XB}_r,$$

$$J_{\text{XB}_w} = k1 ([\text{Trop} - \text{Ca}]) * \text{XB}_n + k_{F,0} * \text{XB}_n - k_{F,-1} * \text{XB}_w - k2 ([\text{Trop} - \text{Ca}]) * \text{XB}_w + k_{F,-2} * \text{XB}_s,$$

$$J_{\text{XB}_s} = k2 ([\text{Trop} - \text{Ca}]) * \text{XB}_w - k_{F,-2} * \text{XB}_s - k_{F,g1} * \text{XB}_s,$$

$$J_{\text{XB}_r} = -k_{(F,g2)\text{XB}_r} + k_{F,g1} * \text{XB}_s,$$

$$k_{F,-1} * \left( \frac{[\text{Trop} - \text{Ca}]}{K_{F,50}} \right)^{\text{NF}} = k1 ([\text{Trop} - \text{Ca}]),$$

$$\begin{aligned}
k_{F,-2} * \left( \frac{[\text{Trop} - \text{Ca}]}{K_{F,502}} \right)^{\text{NF2}} \\
= k_2 ([\text{Trop} - \text{Ca}]).
\end{aligned}
\tag{A.6}$$

The force is described by the fraction of myosin chains in the  $\text{XB}_s$  state. This model is not able to produce an actual quantitative measure of force, as this is beyond the scope of this research. The force output of this model is therefore normalized between zero and one to compare it with experimental data.

**A.5. Contractile Phenotypes and CaMKII Isoforms.** The model is then further expanded to describe the different CaMKII isoforms  $\beta\text{M}$ ,  $\delta\text{A}$ ,  $\delta\text{d}$ , and  $\gamma\text{B}$ . Although the functions of the different isoforms are generally equal, they reside in different compartments in the cell and have a different structure and molecular weight, resulting in different reactions constants.

Furthermore, three different twitch types of muscle fibers are modeled, fast glycolytic, slow glycolytic, and fast oxidative. Compared to fast glycolytic, slow glycolytic fibers have less active RyRs and less active SERCA pumps, resulting in a lower baseline of calcium concentration in the sarcoplasm. Furthermore, the fibers have a lower concentration of the calcium buffers ATP and parvalbumin. Troponin C has a lower amount of binding sites, which is modeled as a lower concentration. Fast oxidative uses the same parameters as fast glycolytic but has a higher amount of mitochondria compared to the other fiber types.

**A.6. Equilibration.** Before the model can be run the total amounts of all derivative species need to be calculated; as for most species only the total amount is known. If the amount of free calcium and free buffer inside the sarcoplasm in equilibrium is known, the concentrations of calcium bound to buffer can be calculated by setting the fluxes of calcium binding and release equal to each other (A.7). If only the total amount of buffer is known, the Michaelis-Menten equilibrium approximation can be used (A.7). For the competitive binding of calcium and magnesium to parvalbumin or mitochondria (A.7) can be used:

$$[\text{buf}]_{\text{Ca}} * k_2 = [\text{buf}]_{\text{free}} * [\text{Ca}]_{\text{Myo}} * k_1,$$

$$[\text{buf}]_{\text{Ca}} = \frac{[\text{buf}]_{\text{free}} * [\text{Ca}]_{\text{Myo}} * k_1}{k_2},$$

$$[\text{buf}]_{\text{Ca}} = \frac{[\text{buf}]_{\text{free}} * [\text{Ca}]_{\text{Myo}}}{(k_1/k_2) + [\text{Ca}]_{\text{Myo}}},$$

$$[\text{buf}]_{\text{tot}} = [\text{buf}]_{\text{free}} + [\text{buf}]_{\text{Ca}} + [\text{buf}]_{\text{Mg}},$$

$$\begin{aligned}
[\text{buf}]_{\text{tot}} = [\text{buf}]_{\text{free}} + \frac{[\text{buf}]_{\text{free}} * [\text{Ca}]_{\text{Myo}}}{k_2/k_1} \\
+ \frac{[\text{buf}]_{\text{free}} * [\text{Mg}]}{k_2/k_1},
\end{aligned}$$

$$[\text{buf}]_{\text{tot}} = [\text{buf}]_{\text{free}} * \left( 1 + \frac{[\text{Ca}]_{\text{Myo}}}{k_2/k_1} + \frac{[\text{Mg}]}{k_2/k_1} \right),$$

$$[\text{buf}]_{\text{free}} = \frac{[\text{buf}]_{\text{tot}}}{1 + ([\text{Ca}]_{\text{Myo}} / (k_2/k_1)) + ([\text{Mg}] / (k_2/k_1))}. \tag{A.7}$$

## B. Model Parameters

For more details see Tables 2, 3, 4, 5, and 6.

## Supplementary Data

Program lines and source codes will be made available upon request.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Authors' Contribution

Wouter Eilers and Wouter Gevers have contributed equally to this work. Conception and design of research was done by Wouter Eilers, Martin Flück, Richard T. Jaspers, Arnold de Haan, and Natal van Riel. Wouter Eilers, WH, and Daniëlle van Overbeek performed experiments. Wouter Eilers, Wouter Gevers, Martin Flück, and Natal van Riel analyzed data. Wouter Eilers, Wouter Gevers, Richard T. Jaspers, Martin Flück, Natal van Riel, Peter A. Hilbers, and Arnold de Haan interpreted results of experiments. Wouter Eilers, Wouter Gevers, and Martin Flück prepared figures. Martin Flück and Natal van Riel drafted the paper. Wouter Eilers, Wouter Gevers, Martin Flück, and Natal van Riel edited and revised paper. Martin Flück, Peter A. Hilbers, and Arnold de Haan provided financial support. All the authors mentioned above approved the final version of the paper.

## Acknowledgments

Proficient technical assistance of Tinelines Busé-Pot and Guus Baan for the *in situ* contraction experiments is greatly acknowledged. This work was supported by a start-up grant by Manchester Metropolitan University (Martin Flück) and the University of Zurich (Martin Flück) and Erasmus Mundus (Wouter Gevers).

## References

- [1] R. Hennig and T. Lomo, "Firing patterns of motor units in normal rats," *Nature*, vol. 314, no. 6007, pp. 164–166, 1985.
- [2] M. E. Anderson, "Connections count: excitation-contraction meets excitation-transcription coupling," *Circulation Research*, vol. 86, no. 7, pp. 717–719, 2000.
- [3] S. Schiaffino and A. L. Serrano, "Calcineurin signaling and neural control of skeletal muscle fiber type and size," *Trends in Pharmacological Sciences*, vol. 23, no. 12, pp. 569–575, 2002.
- [4] M. W. Berchtold, H. Brinkmeier, and M. Müntener, "Calcium ion in skeletal muscle: its crucial role for muscle function, plasticity, and disease," *Physiological Reviews*, vol. 80, no. 3, pp. 1215–1265, 2000.
- [5] M. Brini, T. Cali, D. Ottolini, and E. Carafoli, "Intracellular calcium homeostasis and signaling," in *Metalloids and the Cell*, vol. 12 of *Metal Ions in Life Sciences*, pp. 119–168, 2013.
- [6] S. Schiaffino and C. Reggiani, "Fiber types in mammalian skeletal muscles," *Physiological Reviews*, vol. 91, no. 4, pp. 1447–1531, 2011.
- [7] E. R. Chin, "Role of  $\text{Ca}^{2+}$ /calmodulin-dependent kinases in skeletal muscle plasticity," *Journal of Applied Physiology*, vol. 99, no. 2, pp. 414–423, 2005.
- [8] D. Kernell, O. Eerbeek, B. A. Verhey, and Y. Donselaar, "Effects of physiological amounts of high- and low-rate chronic stimulation on fast-twitch muscle of the cat hindlimb. I. Speed- and force-related properties," *Journal of Neurophysiology*, vol. 58, no. 3, pp. 598–613, 1987.
- [9] A. de Haan, C. J. de Ruiter, A. Lind, and A. J. Sargeant, "Age-related changes in force and efficiency in rat skeletal muscle," *Acta Physiologica Scandinavica*, vol. 147, no. 4, pp. 347–355, 1993.
- [10] K. Gundersen, E. Leberer, T. Lomo, D. Pette, and R. S. Staron, "Fibre types, calcium-sequestering proteins and metabolic enzymes in denervated and chronically stimulated muscles of the rat," *Journal of Physiology*, vol. 398, pp. 177–189, 1988.
- [11] B. R. Eisenberg, D. J. Dix, Z. W. Lin, and M. P. Wenderoth, "Relationship of membrane systems in muscle to isomyosin content," *Canadian Journal of Physiology and Pharmacology*, vol. 65, no. 4, pp. 598–605, 1987.
- [12] A. J. Rose, T. J. Alsted, J. B. Kobberø, and E. A. Richter, "Regulation and function of  $\text{Ca}^{2+}$ -calmodulin-dependent protein kinase II of fast-twitch rat skeletal muscle," *Journal of Physiology*, vol. 580, no. 3, pp. 993–1005, 2007.
- [13] K.-U. Bayer, K. Harbers, and H. Schulman, " $\alpha$ KAP is an anchoring protein for a novel CaM kinase II isoform in skeletal muscle," *EMBO Journal*, vol. 17, no. 19, pp. 5598–5605, 1998.
- [14] P. Tavi and H. Westerblad, "The role of in vivo  $\text{Ca}^{2+}$  signals acting on  $\text{Ca}^{2+}$ -calmodulin-dependent proteins for skeletal muscle plasticity," *Journal of Physiology*, vol. 589, no. 21, pp. 5021–5031, 2011.
- [15] A. Hudmon and H. Schulman, "Structure-function of the multifunctional  $\text{Ca}^{2+}$ /calmodulin-dependent protein kinase II," *Biochemical Journal*, vol. 364, no. 3, pp. 593–611, 2002.
- [16] J. J. Saucerman and D. M. Bers, "Calmodulin mediates differential sensitivity of CaMKII and calcineurin to local  $\text{Ca}^{2+}$  in cardiac myocytes," *Biophysical Journal*, vol. 95, no. 10, pp. 4597–4612, 2008.
- [17] J. M. Bradshaw, Y. Kubota, T. Meyer, and H. Schulman, "An ultrasensitive  $\text{Ca}^{2+}$ /calmodulin-dependent protein kinase II-protein phosphatase 1 switch facilitates specificity in postsynaptic calcium signaling," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 18, pp. 10512–10517, 2003.
- [18] M. Flück, M. N. Waxham, M. T. Hamilton, and F. W. Booth, "Skeletal muscle  $\text{Ca}^{2+}$ -independent kinase activity increases during either hypertrophy or running," *Journal of Applied Physiology*, vol. 88, no. 1, pp. 352–358, 2000.
- [19] G. Dupont and A. Goldbeter, "CaM kinase II as frequency decoder of  $\text{Ca}^{2+}$  oscillations," *BioEssays*, vol. 20, no. 8, pp. 607–610, 1998.
- [20] T. R. Gaertner, S. J. Kolodziej, D. Wang et al., "Comparative analyses of the three-dimensional structures and enzymatic properties of  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  isoforms of  $\text{Ca}^{2+}$ -calmodulin-dependent protein kinase II," *Journal of Biological Chemistry*, vol. 279, no. 13, pp. 12484–12494, 2004.
- [21] L. S. Maier, " $\text{Ca}^{2+}$ /calmodulin-dependent protein kinase II (CaMKII) in the heart," *Advances in Experimental Medicine and Biology*, vol. 740, pp. 685–702, 2012.
- [22] L. S. Maier, "CaMKII $\delta$  overexpression in hypertrophy and heart failure: cellular consequences for excitation-contraction coupling," *Brazilian Journal of Medical and Biological Research*, vol. 38, no. 9, pp. 1293–1302, 2005.
- [23] C. Li, X. Cai, H. Sun et al., "The  $\delta$ A isoform of calmodulin kinase II mediates pathological cardiac hypertrophy by interfering with the HDAC4-MEF2 signaling pathway," *Biochemical and Biophysical Research Communications*, vol. 409, no. 1, pp. 125–130, 2011.
- [24] P. De Koninck and H. Schulman, "Sensitivity of CaM kinase II to the frequency of  $\text{Ca}^{2+}$  oscillations," *Science*, vol. 279, no. 5348, pp. 227–230, 1998.
- [25] H. Fujii, M. Inoue, H. Okuno et al., "Nonlinear decoding and asymmetric representation of neuronal input information by CaMKII $\alpha$  and calcineurin," *Cell Reports*, vol. 3, no. 4, pp. 978–987, 2013.
- [26] L. Li, M. I. Stefan, and N. Le Novère, "Calcium input frequency, duration and amplitude differentially modulate the relative activation of calcineurin and CaMKII," *PLoS ONE*, vol. 7, no. 9, Article ID e43810, 2012.
- [27] S. Carroll, P. Nicotera, and D. Pette, "Calcium transients in single fibers of low-frequency stimulated fast-twitch muscle of rat," *American Journal of Physiology: Cell Physiology*, vol. 277, no. 6, pp. C1122–C1129, 1999.
- [28] C. J. De Ruiter, A. De Haan, and A. J. Sargeant, "Physiological characteristics of two extreme muscle compartments in gastrocnemius medialis of the anaesthetized rat," *Acta Physiologica Scandinavica*, vol. 153, no. 4, pp. 313–324, 1995.
- [29] W. Groenendaal, J. A. L. Jeneson, P. J. Verhoog et al., "Computational modelling identifies the impact of subtle anatomical variations between amphibian and mammalian skeletal muscle on spatiotemporal calcium dynamics," *IET Systems Biology*, vol. 2, no. 6, pp. 411–422, 2008.
- [30] Z. Zi, K.-H. Cho, M.-H. Sung, X. Xia, J. Zheng, and Z. Sun, "In silico identification of the key components and steps in IFN- $\gamma$  induced JAK-STAT signaling pathway," *FEBS Letters*, vol. 579, no. 5, pp. 1101–1108, 2005.
- [31] J. Vanlier, C. A. Tiemann, P. A. J. Hilbers, and N. A. W. van Riel, "Parameter uncertainty in biochemical models described by ordinary differential equations," *Mathematical Biosciences*, vol. 246, no. 2, pp. 305–314, 2013.
- [32] R. Bottinelli, "Functional heterogeneity of mammalian single muscle fibres: do myosin isoforms tell the whole story?" *Pflugers Archiv European Journal of Physiology*, vol. 443, no. 1, pp. 6–17, 2001.

- [33] N. Narayanan and A. Xu, "Phosphorylation and regulation of the  $\text{Ca}^{2+}$ -pumping ATPase in cardiac sarcoplasmic reticulum by calcium/calmodulin-dependent protein kinase," *Basic Research in Cardiology*, vol. 92, supplement 1, pp. 25–35, 1997.
- [34] X. Mu, L. D. Brown, Y. Liu, and M. F. Schneider, "Roles of the calcineurin and CaMK signaling pathways in fast-to-slow fiber type transformation of cultured adult mouse skeletal muscle fibers," *Physiological Genomics*, vol. 30, no. 3, pp. 300–312, 2007.
- [35] M. Philippi and A. H. Sillau, "Oxidative capacity distribution in skeletal muscle fibers of the rat," *Journal of Experimental Biology*, vol. 189, pp. 1–11, 1994.
- [36] S. M. Baylor and S. Hollingworth, "Model of sarcomeric  $\text{Ca}^{2+}$  movements, including ATP  $\text{Ca}^{2+}$  binding and diffusion, during activation of frog skeletal muscle," *Journal of General Physiology*, vol. 112, no. 3, pp. 297–316, 1998.
- [37] T. J. Hund and Y. Rudy, "Rate dependence and regulation of action potential and calcium transient in a canine cardiac ventricular cell model," *Circulation*, vol. 110, no. 20, pp. 3168–3174, 2004.

## Research Article

# CELLCOUNTER: Novel Open-Source Software for Counting Cell Migration and Invasion *In Vitro*

Xiaoni Li,<sup>1</sup> Hongshun Yang,<sup>2</sup> Hailiang Huang,<sup>3,4</sup> and Tao Zhu<sup>1</sup>

<sup>1</sup> Hefei National Laboratory for Physical Sciences at Microscale and School of Life Sciences, USTC, Hefei, Anhui, China

<sup>2</sup> School of Information Science and Technology, USTC, Hefei, Anhui, China

<sup>3</sup> Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA

<sup>4</sup> Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

Correspondence should be addressed to Hailiang Huang; [hhuang@atgu.mgh.harvard.edu](mailto:hhuang@atgu.mgh.harvard.edu) and Tao Zhu; [zhut@ustc.edu.cn](mailto:zhut@ustc.edu.cn)

Received 16 April 2014; Revised 7 June 2014; Accepted 7 June 2014; Published 26 June 2014

Academic Editor: Calvin Yu-Chian Chen

Copyright © 2014 Xiaoni Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Transwell Boyden chamber based migration/invasion assay is a simple and extensively used approach for the characterization of cell motility *in vitro*. Cell motility is quantified by counting the number of cells that pass through the filter membrane. The counting is usually performed manually, which is laborious and error prone. We have therefore developed CELLCOUNTER, an application that is capable of recognizing and counting the total number of cells through an intuitive graphical user interface. The counting can be performed in batch, and the counting results can be visualized and further curated manually. CELLCOUNTER will be helpful in streamlining the experimental process and improving the reliability of the data acquisition.

## 1. Introduction

Cell migration is the movement of cells from one location to another generally in response to and toward specific external chemical signals. Cell invasion is similar to cell migration, except that it requires the cell to migrate through an extracellular matrix or basement membrane barrier by enzymatically degrading the barrier. Cell migration/invasion is central to many physiological and pathological processes such as embryonic development, wound repair, and tumor metastasis [1–3].

Transwell Boyden chamber [4] based cell migration/invasion assay is a simple and extensively used approach for the quantitation of cell motility *in vitro* [3, 5]. The number of cells that pass through the filter membrane from Boyden chamber is usually counted manually from the inverted microscopic images. Such images may contain hundreds of cells and manually counting the number of them is not a trivial work. Existing image analysis programs, for example, CELLPROFILER and IMAGEJ [6–8], require pipeline/macro/plugin files that are specific to the cell/assay types, which is not

available yet for the Transwell assays. Although it is possible to create a CELLPROFILER/IMAGEJ pipeline/macro/plugin for the Transwell assays, an independent program taking into account the specific characteristics of the assay can most likely perform better [9]. Another related application [10], although available for some migration assays, requires fluorescently stained cells and does not work for the Transwell assays.

We have therefore developed CELLCOUNTER, an application specialized in automatically counting the number of cells in Transwell assays. This application supports all major image formats (JPEG, PNG, and GIF) and offers an intuitive graphical user interface (GUI). The default parameters in CELLCOUNTER work well for most users, while the more advanced users have the option to set the parameters of their choice to further improve the counting accuracy (Figure 1).

CELLCOUNTER is capable of processing hundreds of images in a batch and exporting the results to a plain text file for statistical analysis. The recognized cells will be marked on the assay images for curation purpose. Users may manually add or remove recognized cells through the GUI.

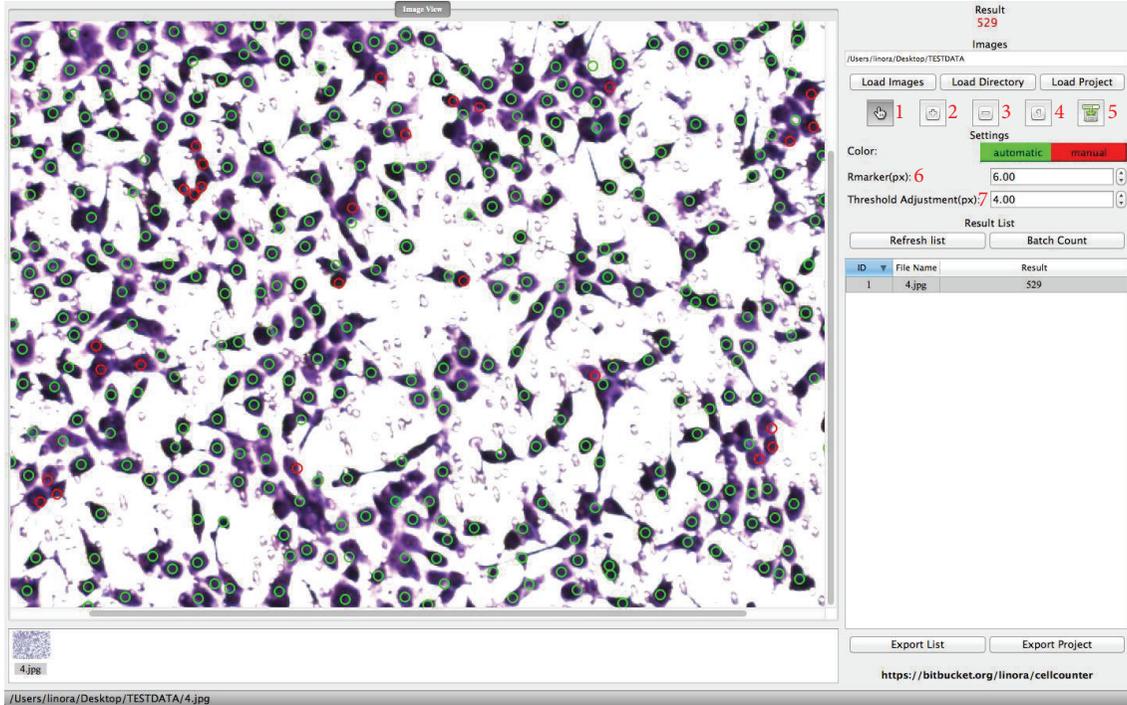


FIGURE 1: The graphical user interface of CELLCOUNTER. Green circles indicate automatically recognized cells and red circles indicate manually labeled cells. Functions for buttons/inputs are (1) manual counting/curation; (2) zooming in; (3) zooming out; (4) restoring to the original image size; (5) saving the counting results (including images); (6) setting the radius of the label (for visualization only); (7) setting the radius of small wells (in px, see the main text for more details, default recommended).

The counting process, including the automatic recognition and manual curations, can be saved to a disk file and loaded back to resume the work. CELLCOUNTER will prove to be useful in streamlining the analysis of Transwell assays, reducing human errors, and improving the reliability of the assay quantitation.

## 2. Materials and Methods

**2.1. Cell Lines.** 20 images from 6 human cancer cell lines were used in this study (Table 1): 17 images are from breast cancer cell lines (1 x T47D, 13 x MCF-7, 2 x MDA-MB-231, and 1 x MDA-MB-435s), 2 images are from hepatic carcinoma (Hep3B), and 1 image is from prostatic cancer (pc-3). All cell lines were obtained from the American Type Culture Collection (Rockville, MD, USA).

**2.2. Transwell Assay.** Assays were performed in BioCoat Matrigel invasion chambers (Corning Costar, Acton, MA) as described previously [11–14]. The assay images were captured using a Nikon camera with 5 M pixels.

**2.3. Biology Stimulation.** Five images (out of the 20 total images) are from stimulated cell (4 x MCF-7 and 1 x MDA-MB-231). We stimulated the cells with 20 ng/mL epidermal growth factor (EGF) when the confluency reached 95%. For

the non-stimulated cells, we added ethyl alcohol as a control compound. After the stimulation we incubated the cells for 4 hours.

**2.4. Existing Software.** CELLPROFILER and IMAGEJ are versatile and flexible image processing software for scientific purposes. For comparison purpose, we used a naïve setup and choice of parameters.

CELLPROFILER uses a customized pipeline, including input and analysis modules, to process images. The input modules in our pipeline are “Images,” “Metadata,” “NameAndType,” and “Groups.” The analysis modules are “ColorToGray,” “ApplyThreshold,” and “IdentifyPrimaryObjects.” There are many tunable parameters in each module. We chose the parameters that make the most sense to us. The project configuration file used in this study (CellProfiler.cproproj) with all parameters is available to download at <https://bitbucket.org/linora/cellcounter/downloads>.

For IMAGEJ, after loading the images, we first adjusted the color threshold using the default thresholding method (Image → Adjust → Color Threshold → select). We then analyzed the images using Analyze → Analyze particle → summarize.

**2.5. Implementation.** CELLCOUNTER was developed using the C++ programming language. Some image processing

TABLE 1: Images used in this study. Cells were stained with 0.1% crystal violet for 20 minutes for the deep staining group and 8 minutes for the light staining group. High density images refer to those having over 400 cells and low density images refer to those having less than 200 cells. Biology stimulation of cells was discussed in Section 2.

ID	Cell line	Stimulation	Stain	Density	Number of cells		
					Expert	CELLCOUNTER	Beginner
1	T47D	No	—	High	931	957	955
2	MCF-7	No	Deep	High	792	748	538
3	MDA-MB-231	No	—	High	475	479	413
4	pc-3	No	—	—	523	507	390
5	MCF-7	—	Deep	—	314	322	263
6	MCF-7	Yes	Deep	—	359	364	300
7	MDA-MB-231	Yes	—	High	566	572	366
8	Hep3B	—	Light	—	284	279	249
9	MDA-MB-435s	—	—	—	464	461	430
10	Hep3B	—	—	—	186	178	144
11	MCF-7	—	Light	Low	165	159	—
12	MCF-7	Yes	—	—	403	403	—
13	MCF-7	No	Deep	—	417	395	—
14	MCF-7	Yes	Light	—	189	187	—
15	MCF-7	—	Light	Low	94	98	—
16	MCF-7	Yes	—	Low	120	118	—
17	MCF-7	—	—	Low	139	140	—
18	MCF-7	—	Light	—	8	8	—
19	MCF-7	—	Deep	High	554	553	—
20	MCF-7	—	—	Low	191	192	—

algorithms were implemented with the help of Open Source Computer Vision Library [15], and the graphical user interface was designed using Nokia Qt (a cross platform application and UI framework). CELLCOUNTER is compatible with all major platforms including Windows (XP or newer), Mac OSX (10.6+), and Linux (Kernel 3.x). Assays images for CELLCOUNTER have to have an aspect ratio of 4:3 (default in most digital cameras) and a minimal resolution of 1280 px by 960 px. Images that have the correct aspect ratio but in higher resolutions will be resized to 1280 px by 960 px. Resizing to a lower resolution reduces the use of memory and CPU time in analyzing images. We have verified that the resolution of 1280 px by 960 px is high enough to retain all the necessary details for cell counting.

In CELLCOUNTER, the original assay image is first converted to a grayscale image, and further partitioned into cell areas and background areas using an adaptive threshold. Areas that have grayscale below the threshold (close to white)

are background and areas that have grayscale above the threshold (close to black) are cells. The threshold is established by using the maximum entropy method [16]. A small value can be added to the threshold (threshold adjustment) to fine-tune the counting accuracy. Users may adjust this value in GUI and use the visual feedback to choose the optimal value. Alternately, the default value works very well in most circumstances. Standard image processing procedures, such as contrast enhancement, smoothing, eroding, and dilating [17], are also performed to remove noises in the image.

Due to the design of the assay, small wells in the equipment can be captured in the images and counted falsely as cells. We solve this problem by further partitioning the cell areas into small wells and true cells. Another adaptive threshold is established by applying the OSTU method [18] to only the cell areas identified in the previous step. Because the small wells are generally darker than the true

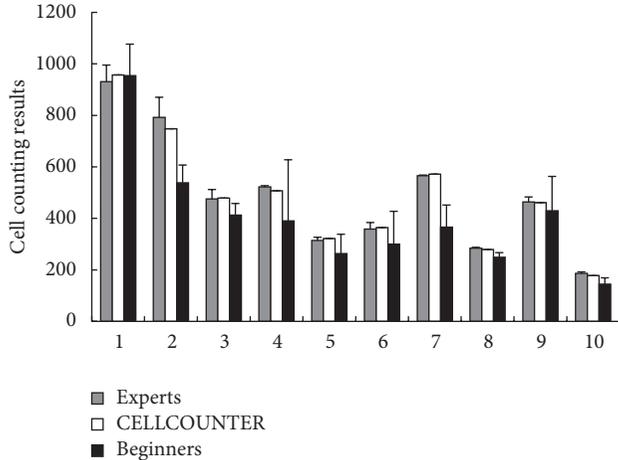


FIGURE 2: Cell counting results for 10 assay images from 3 experts, CELLCOUNTER, and 3 beginners. Error bars are standard deviations (only available for human counting results).

cells, applying this adaptive threshold, followed by image smoothing, eroding, and dilating, can successfully remove the small wells from the cell areas.

One cell area may contain multiple overlapping cells. We count the number of cells in a particular cell area by using the radius of its maximum inscribed circle. If the radius is smaller than the empirical threshold of the cell radius  $r$  (default to 6 pixels) and greater than the empirical threshold for noise (default to 4 pixels), we count this cell area as a single cell. If the radius is greater than the empirical threshold ( $r$ ), we count the number of cells in this cell area as

$$h \times \left[ \frac{L}{2r} + a \right] \times \left[ \frac{W}{2r} + a \right], \quad (1)$$

in which  $L$  and  $W$  are the length and width of the minimum bounding rectangle,  $h$  is the number of layers that the cells stack, and  $a$  is a parameter to account for cells at the boundary of the rectangle.  $h$  and  $a$  are unknown and have been estimated using the human counted cell numbers as the training dataset ( $h \approx 4.0$  and  $a \approx 0.5$ ). The total number of cells in the image is then the summation of cell numbers in each cell area.

### 3. Results and Discussion

The user interface and a counting example using CELLCOUNTER is provided in Figure 1. We used CELLCOUNTER to analyze 10 assay images (randomly selected from the 20 images in Section 2). The cell numbers counted by the experts (3 researchers that are proficient in cell counting) are used as the gold standard and compared with the cell numbers counted by CELLCOUNTER (default parameters, without manual curation) and beginners (3 researchers with basic trainings in cell counting).

As shown in Figure 2, we found that the cell numbers counted by CELLCOUNTER are statistically the same as

the numbers counted by experts (2-way ANOVA  $P$  value = 0.91). In contrast, the beginners gave slightly smaller cell counts (2-way ANOVA  $P$  value = 0.04), and these cell counts have a significantly higher standard deviation (Wilcoxon test  $P$  value = 0.004) than the cell counts from the experts. Therefore, we conclude that CELLCOUNTER is able to perform accurate cell counting and improve the stability of the counting results. Additional tests using unpublished images have confirmed this conclusion (results not shown).

Systematic biases can lead to false positive findings. For example, if software systematically reports a larger (comparing with the true) number of cells if they are stained for a longer time, and if the person who stains the case samples tends to stain them longer, we might falsely claim a significant case/control difference. Therefore, we determined CELLCOUNTER for systematic biases towards the staining, the density of cells, and the biology stimulation. We found no systematic bias in all 3 conditions (Table 1). The  $P$  values are 0.42 for deep versus light stains, 0.91 for high versus low cell densities, and 0.96 for stimulated versus nonstimulated cells.

CELLPROFILER and IMAGEJ are powerful scientific image processing software. They are versatile and flexible but also have steep learning curves and require efforts to set up the right pipeline/parameters. Using a naïve setup as discussed in Section 2, we noticed that both methods perform worse comparing with CELLCOUNTER (Figure 3). The correlation coefficients ( $R^2$ ) for CELLPROFILER and IMAGEJ are 0.49 and 0.12, respectively, while the  $R^2$  for CELLCOUNTER reached almost 1. CELLCOUNTER performs better because it was designed for images from these assays, while CELLPROFILER and IMAGEJ have more general purposes. We are aware that more sophisticated setup or choices of parameters for CELLPROFILER and IMAGEJ may improve their performance. However, we argue that our choices are representative of average users with basic training and knowledge in image processing.

### 4. Conclusion

We have developed CELLCOUNTER, a program that features an intuitive graphical user interface to count the number of cells in Transwell Boyden chamber based migration/invasion assays. This program allows high-throughput analysis of a large number of assay images. The counted cells are visibly marked on the assay images and can be manually curated. The accuracy of the counting results has been validated using expert counted cell numbers as the gold standard.

CELLCOUNTER significantly simplifies the data acquisition in the Transwell assays, reduces human errors, and improves the stability of counting results. It will prove to be a helpful tool in the study of cell invasion and metastasis *in vitro*.

CELLCOUNTER is currently available for Windows, Mac OSX, and Linux platforms and can be downloaded from <https://bitbucket.org/linora/cellcounter/downloads>.

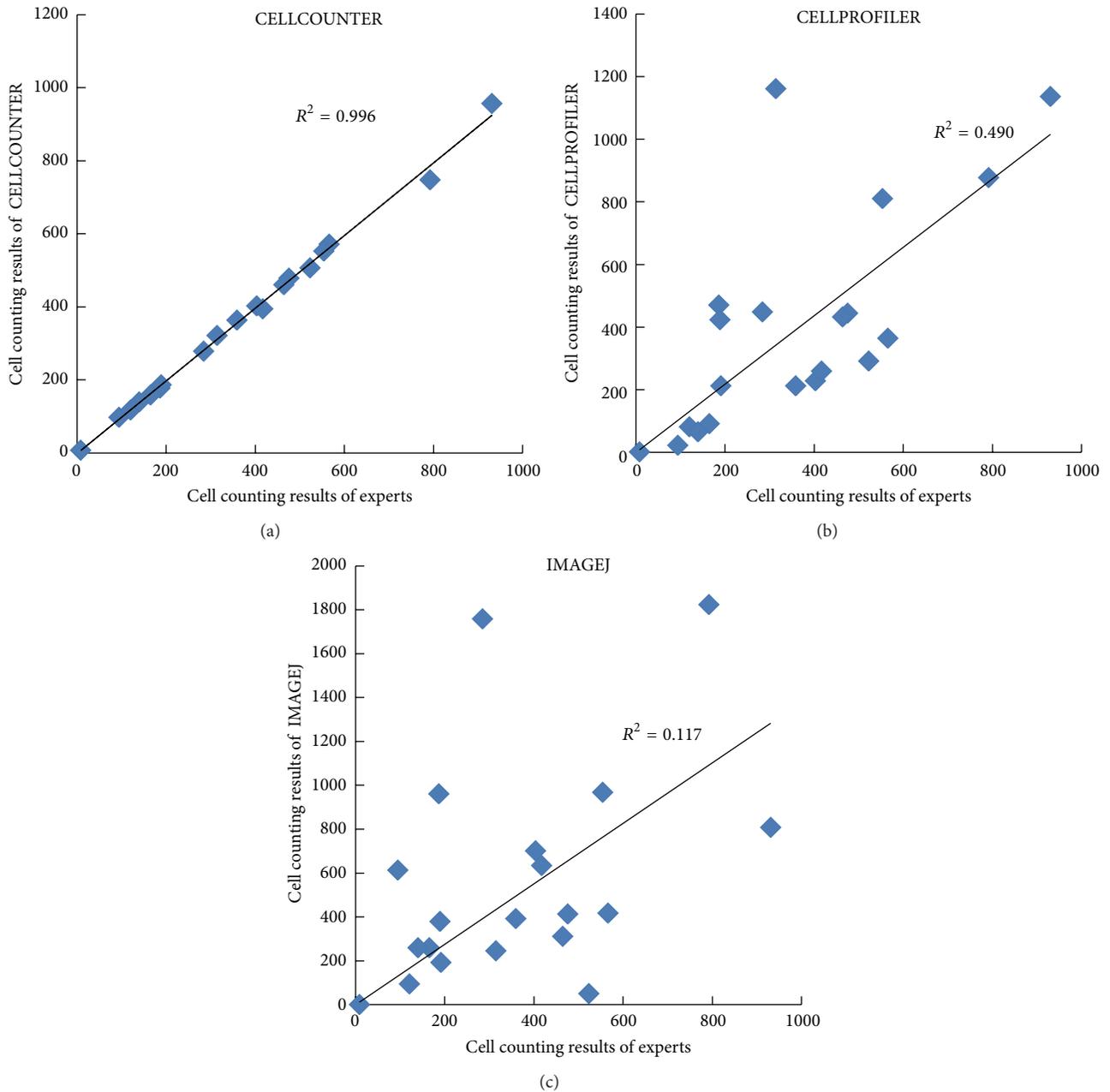


FIGURE 3: Counting results from CELLCOUNTER (a), CELLPROFILER (b), and IMAGEJ (c) comparing with human counting results.

**Conflict of Interests**

The authors declare that there is no conflict of interests.

**Acknowledgments**

The authors thank Dr. Zhiming Pi for helpful advice to improve this software. They would also like to thank the members of their laboratories for contributing to the development of the software and this paper. This work was

supported by the National Key Scientific Program of China (2012CB934002, 2010CB912804) and National Natural Science Foundation of China (81272925).

**References**

[1] A. J. Ridley, M. A. Schwartz, K. Burridge et al., "Cell migration: integrating signals from front to back," *Science*, vol. 302, no. 5651, pp. 1704–1709, 2003.

- [2] S. A. Courtneidge, "Cell migration and invasion in human disease: the Tks adaptor proteins," *Biochemical Society Transactions*, vol. 40, no. 1, pp. 129–132, 2012.
- [3] E. A. Goncharova, D. A. Goncharov, and V. P. Krymskaya, "Assays for in vitro monitoring of human airway smooth muscle (ASM) and human pulmonary arterial vascular smooth muscle (VSM) cell migration," *Nature Protocols*, vol. 1, no. 6, pp. 2933–2939, 2006.
- [4] H. Chen, "Boyden chamber assay," *Methods in Molecular Biology*, vol. 294, pp. 15–22, 2005.
- [5] X. H. Liu, K. H. Lu, K. M. Wang et al., "MicroRNA-196a promotes non-small cell lung cancer cell proliferation and invasion through targeting HOXA5," *BMC Cancer*, vol. 12, article 348, 2012.
- [6] A. E. Carpenter, T. R. Jones, M. R. Lamprecht et al., "CellProfiler: image analysis software for identifying and quantifying cell phenotypes," *Genome Biology*, vol. 7, no. 10, article R100, 2006.
- [7] M. R. Lamprecht, D. M. Sabatini, and A. E. Carpenter, "Cell-Profiler: free, versatile software for automated biological image analysis," *BioTechniques*, vol. 42, no. 1, pp. 71–75, 2007.
- [8] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, "NIH Image to ImageJ: 25 years of image analysis," *Nature Methods*, vol. 9, no. 7, pp. 671–675, 2012.
- [9] N. N. Kachouie, L. Kang, and A. Khademhosseini, "Arraycount, an algorithm for automatic cell counting in microwell arrays," *BioTechniques*, vol. 47, no. 3, pp. 1–100, 2009.
- [10] B. K. Al-Khazraji, P. J. Medeiros, N. M. Novielli, and D. N. Jackson, "An automated cell-counting algorithm for fluorescently-stained cells in migration assays," *Biological Procedures Online*, vol. 13, no. 1, article 9, pp. 1–9, 2011.
- [11] S. Mukhina, H. C. Mertani, K. Guo, K. Lee, P. D. Gluckman, and P. E. Lobie, "Phenotypic conversion of human mammary carcinoma cells by autocrine human growth hormone," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 42, pp. 15166–15171, 2004.
- [12] X. Li, X. Liu, W. Xu et al., "C-MYC-regulated miR-23a/24-2/27a cluster promotes mammary carcinoma cell invasion and hepatic metastasis by targeting sprouty2," *Journal of Biological Chemistry*, vol. 288, no. 25, pp. 18121–18133, 2013.
- [13] P. Qian, Z. Zuo, Z. Wu et al., "Pivotal role of reduced let-7g expression in breast cancer invasion and metastasis," *Cancer Research*, vol. 71, no. 20, pp. 6463–6474, 2011.
- [14] J. Chen, Y. Yao, C. Gong et al., "CCL18 from tumor-associated macrophages promotes breast cancer metastasis via PITPNM3," *Cancer Cell*, vol. 19, no. 4, pp. 541–555, 2011.
- [15] G. Bradski, "The OpenCV library," *Dr Dobbs Journal*, vol. 25, no. 11, pp. 120–126, 2000.
- [16] S. F. Gull and J. Skilling, "Maximum-entropy method in image-processing," *IEE Proceedings F: Communications, Radar and Signal Processing*, vol. 131, no. 6, pp. 646–659, 1984.
- [17] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, PWS, Pacific Grove, Calif, USA, 2nd edition, 1999.
- [18] H. Ng, "Automatic thresholding for defect detection," *Pattern Recognition Letters*, vol. 27, no. 14, pp. 1644–1649, 2006.

## Research Article

# Treatment of Cardiovascular Disease by Traditional Chinese Medicine against Pregnane X Receptor

Kuen-Bao Chen,<sup>1,2,3</sup> Hsin-Yi Chen,<sup>2</sup> Kuan-Chung Chen,<sup>4</sup> and Calvin Yu-Chian Chen<sup>1,2,5,6</sup>

<sup>1</sup> School of Medicine, College of Medicine, China Medical University, Taichung 40402, Taiwan

<sup>2</sup> Department of Biomedical Informatics, Asia University, Taichung 41354, Taiwan

<sup>3</sup> Department of Anesthesiology, China Medical University Hospital, Taichung 40447, Taiwan

<sup>4</sup> School of Pharmacy, China Medical University, Taichung 40402, Taiwan

<sup>5</sup> Research Center for Chinese Medicine & Acupuncture, China Medical University, Taichung 40402, Taiwan

<sup>6</sup> Human Genetic Center, Department of Medical Research, China Medical University Hospital, Taichung, Taiwan

Correspondence should be addressed to Calvin Yu-Chian Chen; [ycc929@MIT.edu](mailto:ycc929@MIT.edu)

Received 19 February 2014; Accepted 5 March 2014; Published 22 June 2014

Academic Editor: Chung Y. Hsu

Copyright © 2014 Kuen-Bao Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recently, cardiovascular disease, also known as loop circulatory system diseases or disorders, is one of the serious diseases including heart disease, stroke, atherosclerosis, myocardial infarction, hypertension, hypotension, and thrombosis. Human pregnane X receptor, PXR, plays a crucial role in exogenous and endobiotic metabolism for rabbit, rat, mouse, and human. The PXR activation can protect the blood vessels from damage of hazardous substances. In this study we aim to investigate the potent lead compounds as PXR receptor agonist against cardiovascular disease. To improve drug development of TCM compounds, we aim to investigate the potent lead compounds as PXR agonists from the TCM compounds in TCM Database@Taiwan. The top three TCM compounds, bis(4-hydroxybenzyl) ether mono- $\beta$ -D-glucopyranoside (BEMG), ixerisoid, and tangshenoside II, have displayed higher potent binding affinities than the positive control, PNU-142721, in the docking simulation. After MD simulations, which can optimize the result of docking simulation and validate the stability of H-bonds between each ligand and PXR protein under dynamic conditions, top TCM compounds, BEMG and tangshenoside II, maintain most of interactions with PXR protein, which keep the ligand binding stable in the binding domain. Hence, we propose BEMG and tangshenoside II as potential lead compounds for further study in drug development process with the PXR protein.

## 1. Introduction

Recently, cardiovascular disease, also known as loop circulatory system diseases or disorders, is one of the serious diseases including heart disease, stroke, atherosclerosis, myocardial infarction, hypertension, hypotension, and thrombosis. It is the top leading cause of death in the United States and most European countries. More than 83.6 million Americans have the cardiovascular problems; the patients of cardiovascular disease in other Western countries are also growing yearly [1]. Family history, obesity, latent diseases, such as diabetes, gout and kidney disease, and bad habits, diet, the environment of toxic substances, and drugs are the risk factors for cardiovascular disease [2–4]. Environmental

pollution and chemicals also promote the occurrence of blood vessel function disorders and cardiovascular diseases. We should consider how to regulate and protect the blood vessels [5].

Nowadays, many distinct mechanisms of diseases have been identified [6, 7] to determine the potential target proteins for drug design against each disease [8–11]. Human pregnane X receptor, PXR, plays a crucial role in exogenous metabolism for rabbit, rat, mouse, and human [2–5]. Some studies indicate that PXR also plays an important role in endobiotic metabolism for rabbit, rat, mouse, and human [12–22]. Activated PXR binds to response elements in the promoters and upregulates the transcription of Phases I and II drug-metabolizing enzymes, for example, glutathione

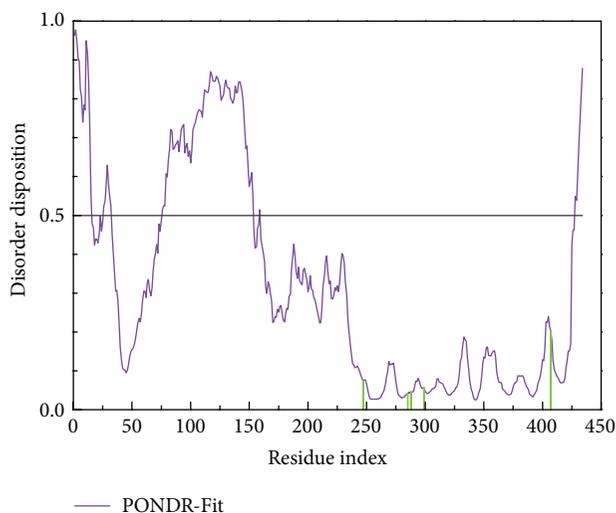


FIGURE 1: Disordered disposition predicted by PONDR-Fit.

S-transferases (GSTs) and cytochrome P450 (CYP)s, and transporters, for example, multidrug resistance protein 1 (MDR1) [12, 13]. It provides a mechanism for the blood vessels to protect itself and the underlying tissue under exogenous and endobiotic insults [14].

The human pregnane X receptor, PXR (NR1I2, also known as PAR or SXR), is a key transcription factor gene expression and regulation of CYP3A. It is combined by DNA binding domain (DBD) and ligand binding domain (LBD) [19–21]. It is composed of three  $\alpha$ -helices and five  $\beta$ -folds formed around globular ligand binding cavity [22]. PXR can be activated by variant ligands, including drug [12], endogenous compounds [12, 23], and environmental contaminants [24]. PXR has a similar protective effect in the vessel and in liver, which can stop the liquid, the solute, and the cells in the vessel wall. Therefore, the PXR activation can protect the blood vessels from damage of hazardous substances. In this study, we aim to investigate the potent lead compounds as PXR receptor agonist against cardiovascular disease.

Recently, *in silico* researches have been broadly used in the drug design [25–29]. Many compounds extracted from traditional Chinese medicine (TCM) had been determined as potential lead compounds for many different diseases, such as stroke [30–32], tumors [33–36], inflammation [37], metabolic syndrome [38–40], viral infection [41, 42], and some disorders [43–45]. As structural disordered amino acids in the binding domain of protein may affect the ligand binding with target protein and induce side effect [46, 47], the disordered amino acids of PXR protein were predicted before virtual screening. For TCM compounds filtered by virtual screening, the interactions of the docking poses in the docking simulation may be modified under dynamic conditions. We employed the molecular dynamics (MD) simulations to validate the stability

of each docking pose. In addition, the biological activities of potential TCM candidates were predicted by three distinct models.

## 2. Materials and Methods

**2.1. Data Collection.** The X-ray crystallography structure of the human pregnane X receptor (PXR) was downloaded from RCSB Protein Data Bank with PDB ID 3R8D [48]. The disordered amino acids of PXR protein were predicted using PONDR-Fit [49] protocol with the sequence of PXR protein from Swiss-Prot (UniProtKB: O75469). The PXR protein has protonated the final structure of protein with Chemistry at HARvard Macromolecular Mechanics (CHARMM) force field [50] and removed crystal water using Prepare Protein module in Discovery Studio 2.5 (DS 2.5). The binding domain was defined by the volume of the cocrystallized anti-HIV drug, PNU-142721. TCM compounds from TCM Database@Taiwan [51] have protonated the final structure and have been filtered by Lipinski et al.'s Rule of Five [52] using Prepare Ligand module in DS 2.5.

**2.2. Docking Simulation.** The prepared TCM compounds have been docked in the binding domain of PXR protein using LigandFit protocol [53] in DS 2.5 which docks ligands into the binding domain using a shape filter and Monte-Carlo ligand conformation generation and then optionally minimized with CHARMM force field [50] and rejected the similar poses by the clustering of saved docking pose. The consensus scores were calculated using the properties of -PLP1, -PLP2, -PMF, -PMF04, dock score, Jain, LigScore1 Dreiding, LigScore2 Dreiding, ligand internal energy, Ludi 1, Ludi 2, and Ludi 3.

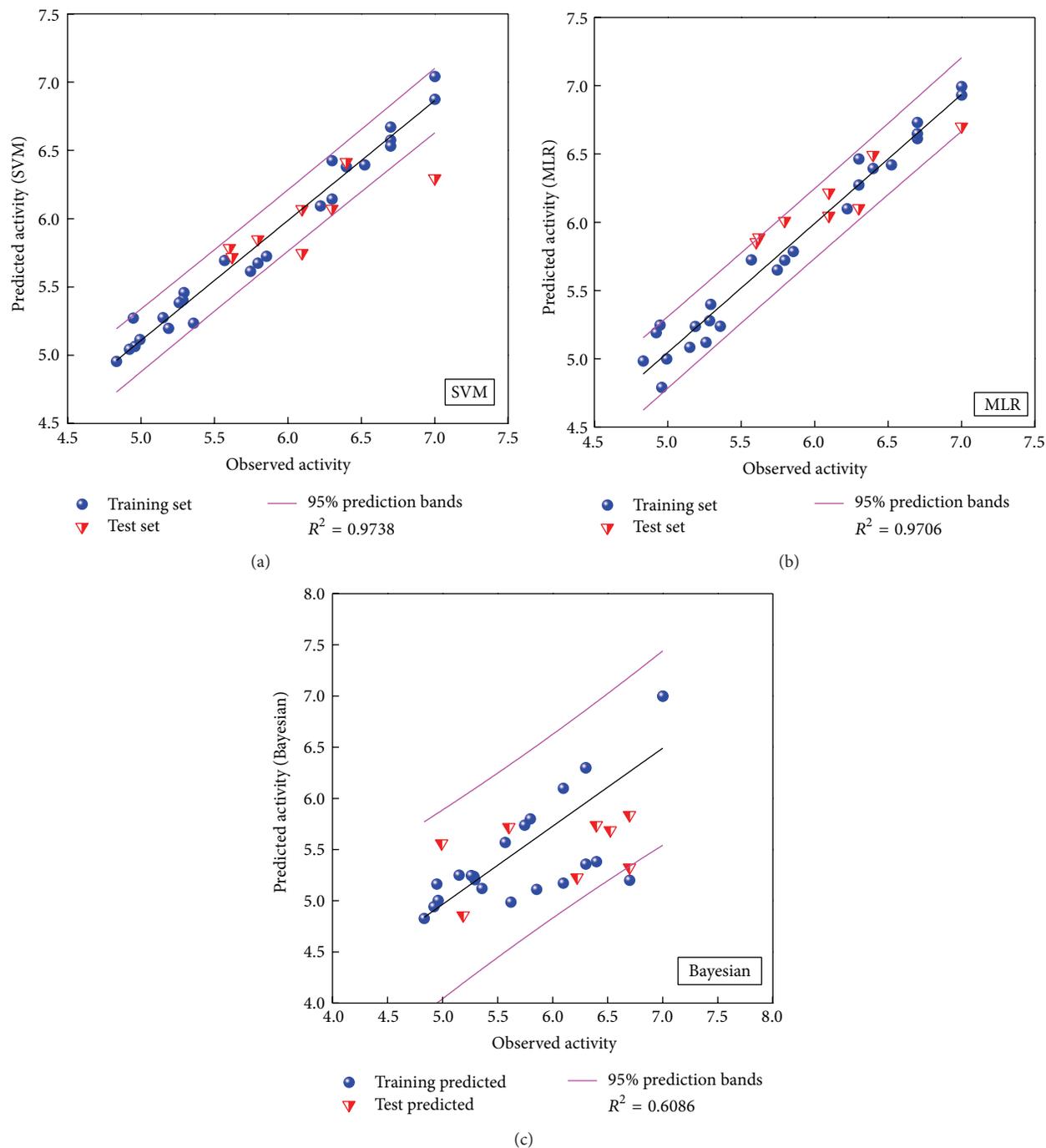


FIGURE 2: Comparative plots of observed versus predicted activity for (a) SVM, (b) MLR, and (c) BNT models.

**2.3. Biological Activity Prediction.** Three distinct prediction models, multiple linear regression (MLR), support vector machine (SVM), and Bayes network toolbox (BNT) models, were employed to predict the biological activity for the TCM compounds using the  $pEC_{50}$  ( $\log(1/EC_{50})$ ) value of 25 compounds out of 33 PXR agonists [54]. The suitable molecular descriptors for constructing the prediction models

were selected using genetic function approximation module [55] in DS 2.5, and the protocol estimates the fitness of individual model using square correlation coefficient ( $R^2$ ). The prediction models have also been validated by cross validation test. In addition, MLR and BNT models were performed using MATLAB, and SVM model was performed using LibSVM developed by Chang and Lin [56].

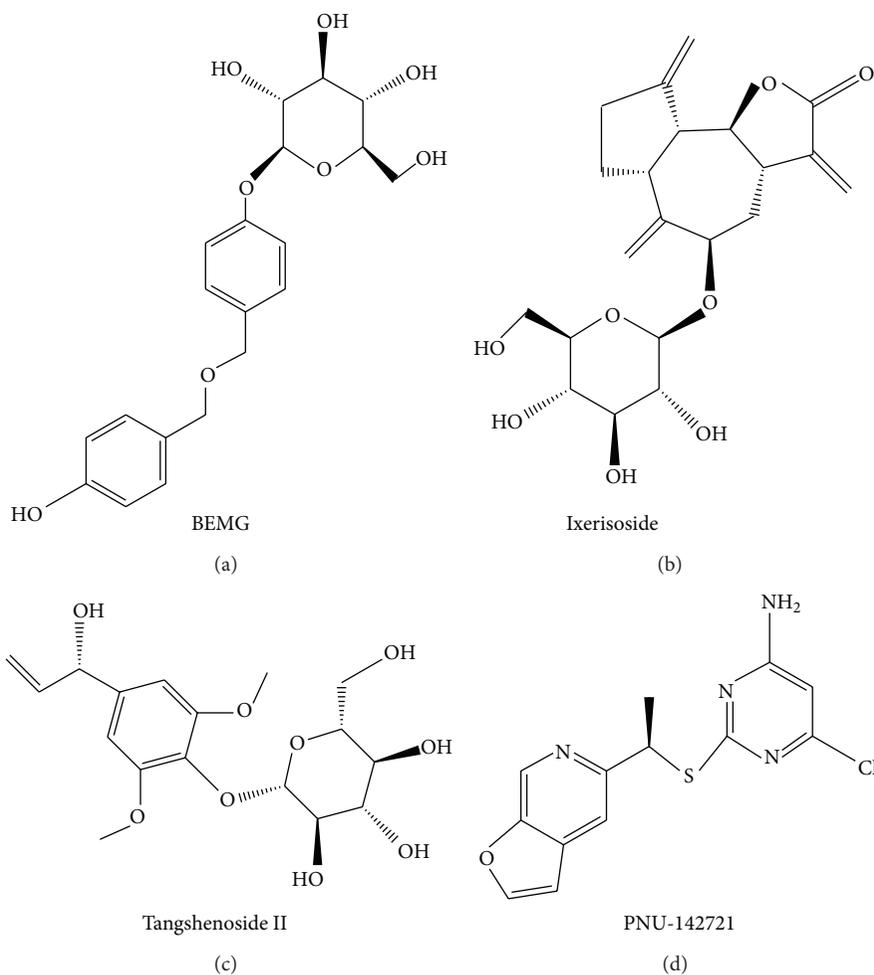


FIGURE 3: Chemical scaffold of control and the top three candidates: (a) bis(4-hydroxybenzyl) ether mono-beta-D-glucopyranoside (BEMG), (b) ixerisoside, (c) tangshenoside II, and (d) PNU-142721.

**2.4. Molecular Dynamics (MD) Simulation.** For each docking pose in the dock simulation, the protein-ligand complex has been simulated under dynamic conditions with classical molecular dynamics theory using Gromacs 4.5.5 [57]. The topology and parameters for PXR protein with CHARMM27 force field and each ligand were provided using pdb2gmx protocol in Gromacs and SwissParam program [58], respectively. A cubic box is performed with the box edge approximate 1.2 nm from the molecules periphery and solvated using TIP3P water model neutralized by 0.145 M NaCl model using Gromacs. Then the steepest descent [59] was employed to remove bad van der Waals contacts with a maximum of 5,000 steps. In equilibration section, the position-restrained molecular dynamics simulation was employed using linear constraint algorithm, NVT equilibration, Berendsen weak thermal coupling method, and particle mesh Ewald method.

A total of 40 ns production simulation with time step in unit of 2 fs was performed using particle mesh Ewald (PME) option and NPT ensembles. A series of protocols in Gromacs, such as *g\_rms*, *g\_gyrate*, *g\_msd*, *g\_sas*, *g\_energy*, *g\_rmsf*, and *do\_dssp*, was employed to analyze the MD trajectories.

### 3. Results and Discussion

**3.1. Disordered Protein Prediction.** The disordered disposition for the sequence of PXR protein from Swiss-Prot (UniProtKB: O75469) predicted by PONDR-Fit was illustrated in Figure 1. As the residues in the binding domain do not lie in the disordered region, the binding domain of PXR protein has a stable structure in protein folding.

**3.2. Biological Activity Prediction.** GFA (genetic functional analysis) protocol in DS 2.5 was employed with 204 descriptors to determine the ten optimum molecular descriptors for constructing prediction models with 25 compounds of training set. The selected descriptors were ES\_Sum\_dNH, ES\_Sum\_ssNH, ES\_Sum\_sssN, ES\_Count\_aaCH, ES\_Count\_ssNH, Num\_RingBonds, Molecular\_PolarSASA, IAC\_Total, Jurs\_DPSA\_3, and Jurs\_PPSA\_1. According to these selected descriptors, the functional formula of MLR model was constructed as follows:

$$\text{pEC}_{50} = -1.24629 - 0.44990 \times \text{ES\_Sum\_dNH} + 1.29360 \times \text{ES\_Sum\_ssNH}$$



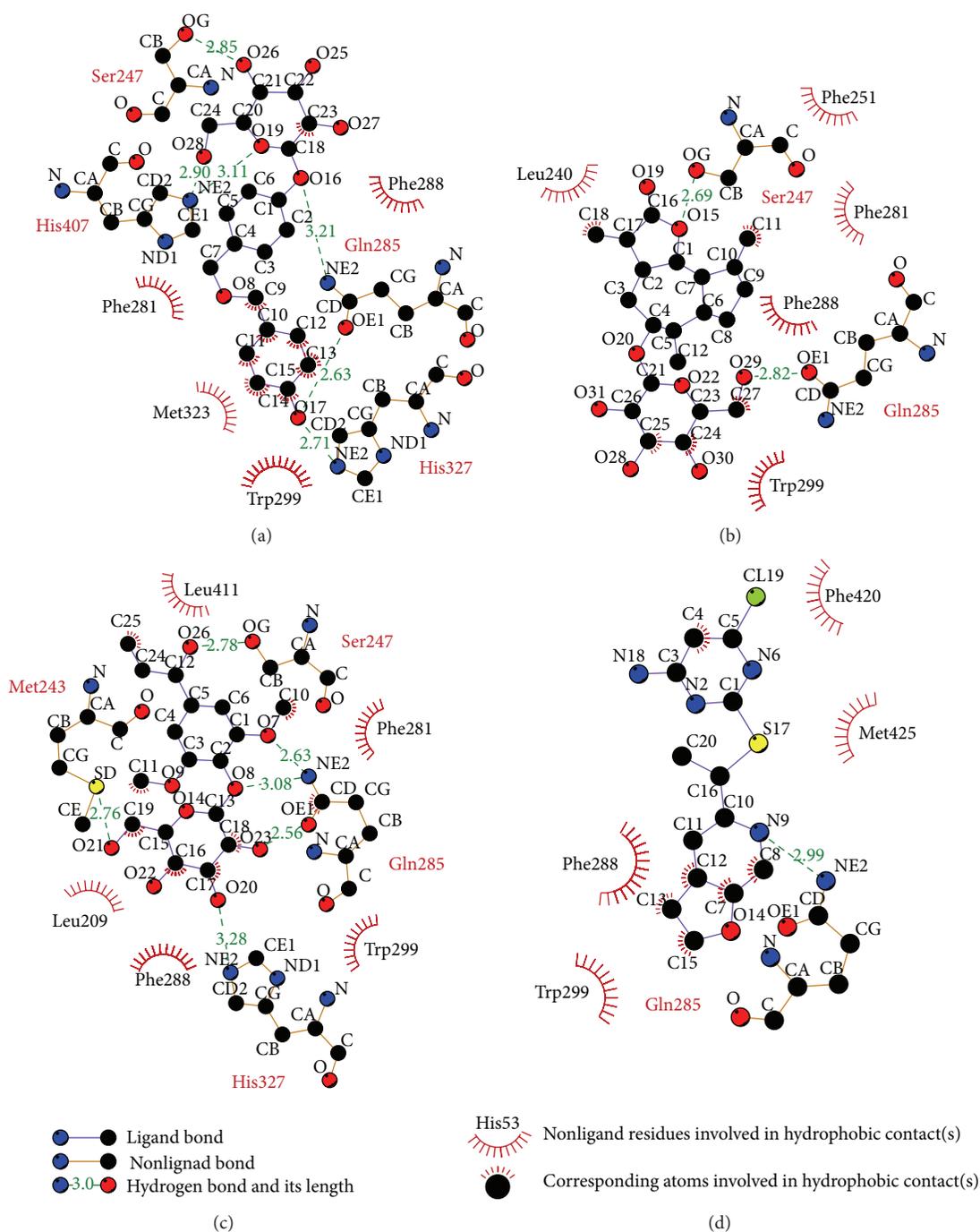


FIGURE 5: Docking pose of PXR complex with (a) BEMG, (b) ixeriside, (c) tangshenoside II, and (d) PNU-142721 drawn by LigPlot program.

rank the top 20 TCM compounds. For the top three TCM compounds, bis(4-hydroxybenzyl) ether mono- $\beta$ -D-glucopyranoside (BEMG), ixeriside, and tangshenoside II, BEMG was extracted from *Gastrodia elata* [60], which have been indicated the effect of reducing blood pressure, increasing the heart, cerebral blood flow, and reducing cerebral vascular resistance [61, 62]. Ixeriside was extracted from *Cichorium intybus* [63], which can improve diabetes [64] and clear toxins in the liver [65]. Tangshenoside II

was extracted from root of *Codonopsis tangshen* [66], which has excitatory effects for nervous system, and can enhance the body resistance; expansion of peripheral vascular and blood pressure, and inhibit the pressor effect of epinephrine, regulate gastrointestinal motility, anti-ulcer, inhibition of gastric acid secretion, reducing the activity of pepsin, raise leukocyte level declined after chemotherapy and radiation. The chemical scaffold top TCM compounds and PNU-142721 are illustrated in Figure 3. According to the docking poses

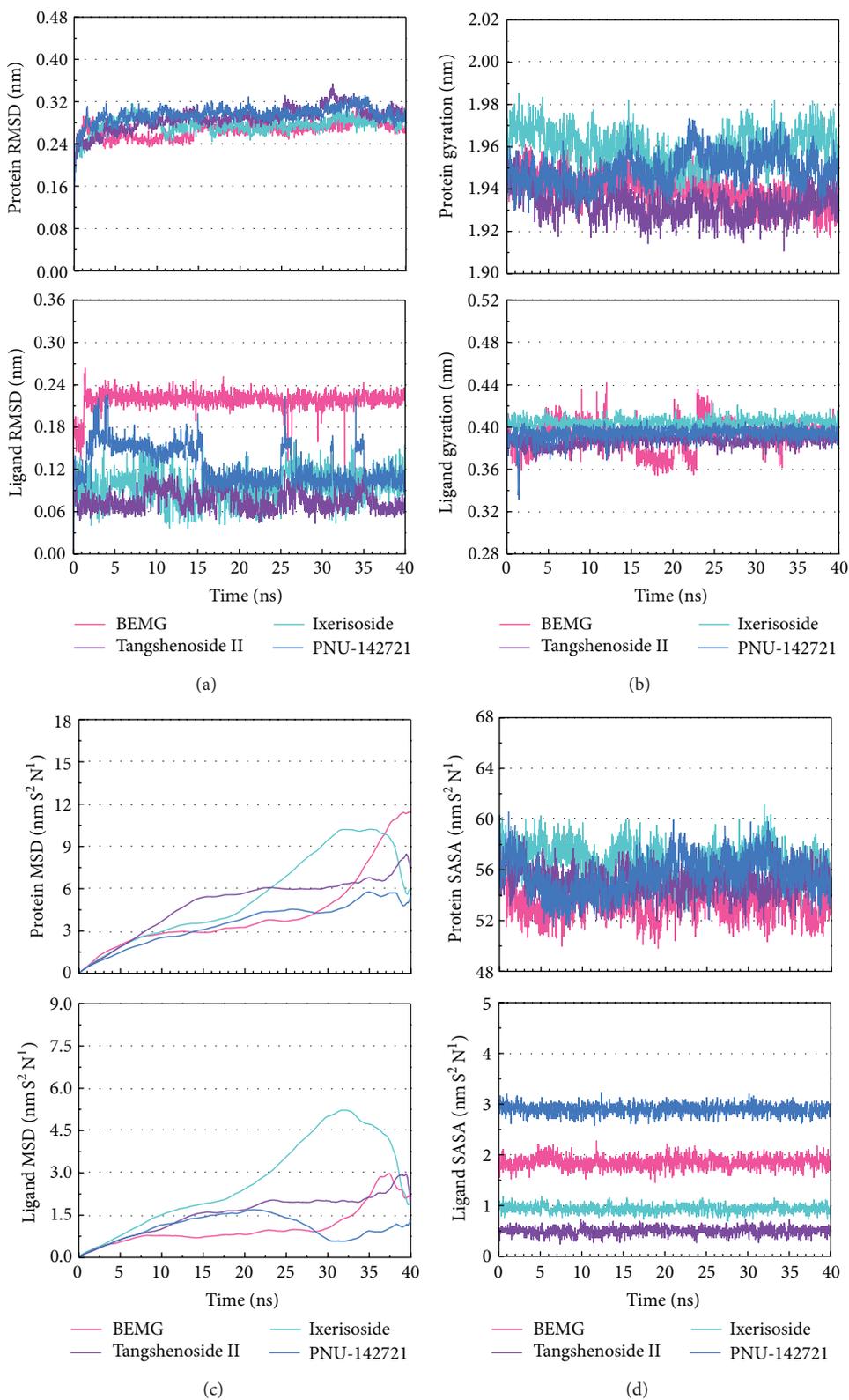


FIGURE 6: Analysis of MD trajectories generated by Gromacs: (a) root-mean-square deviations (RMSDs), (b) radii of gyration, (c) mean square deviation (MSD), and (d) total solvent accessible surface area (SASA).

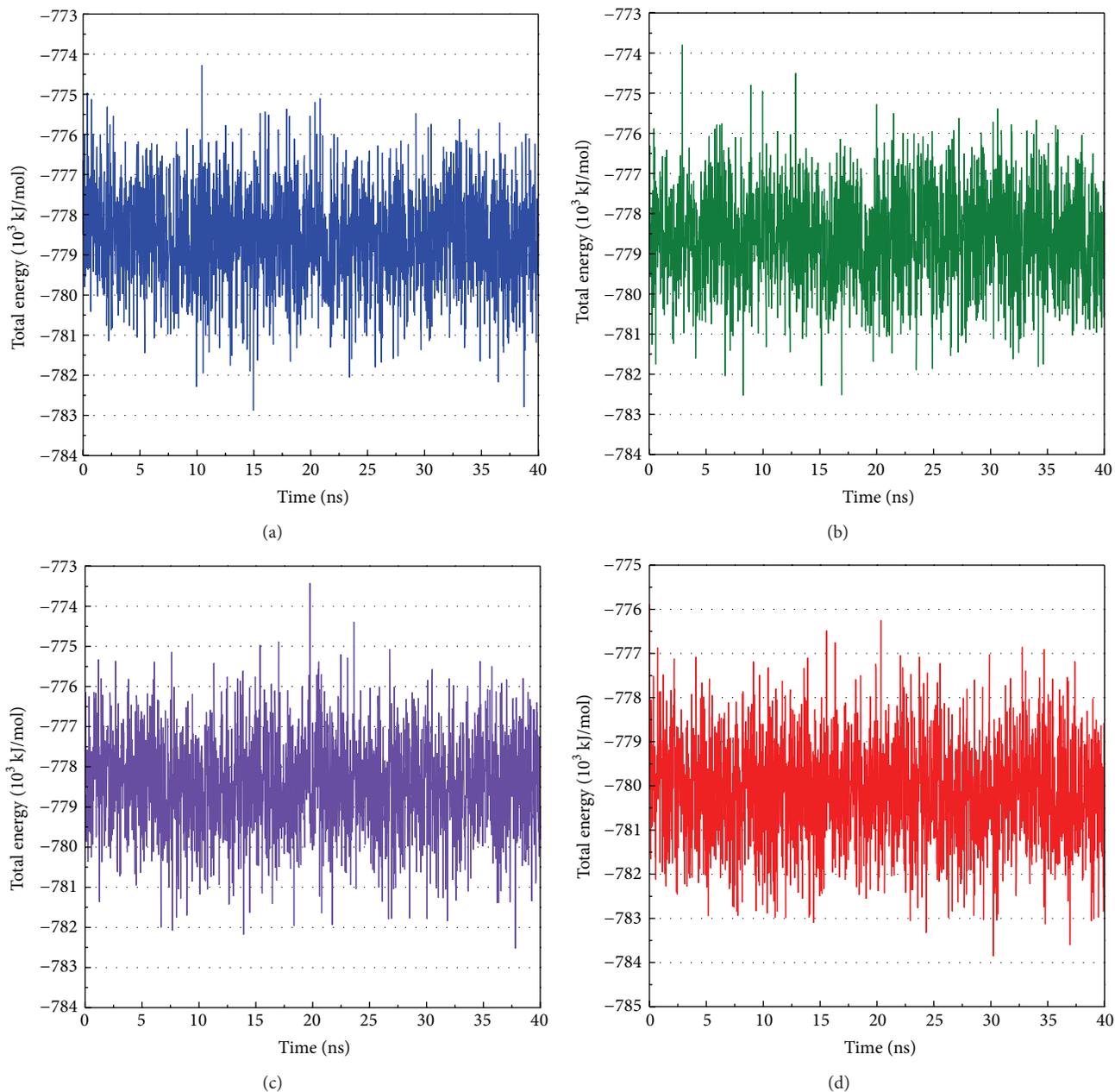


FIGURE 7: Total energy of PXR complex with (a) BEMG, (b) ixeriside, (c) tangshenoside II, and (d) PNU-142721.

in Figures 4 and 5, the top three candidate compounds and control have hydrogen bonds (H-bonds) with the common amino acid Gln285 exist. The top three candidate compounds have H-bonds with Ser247. In addition, BEMG still produces hydrogen bonds with His327 and His407 and generates  $\pi$  bond with His407 and Trp299. Tangshenoside II will produce additional hydrogen bond with Met243, as well as PNU-142721 will produce  $\pi$  bond with Phe288. Figure 5 illustrates the hydrophobic contacts between each compound and residues in the binding domain. The top three candidate compounds and control have hydrophobic contacts with common residues Phe288 and Trp299, and all TCM

compounds have hydrophobic contacts with residue Phe281. The docking results indicate that the top three TCM candidate compounds have higher binding affinities than control. In addition, they have H-bonds with key residues Ser247 and Gln285 and hydrophobic contacts with key residues Trp299 and Phe288.

**3.4. Molecular Dynamics Simulation.** MD simulation was employed to validate the stability of interactions between PXR protein and each compound. Root-mean-square deviation (RMSD) illustrated the atomic fluctuations during MD simulation in Figure 6(a). Protein RMSD displays the

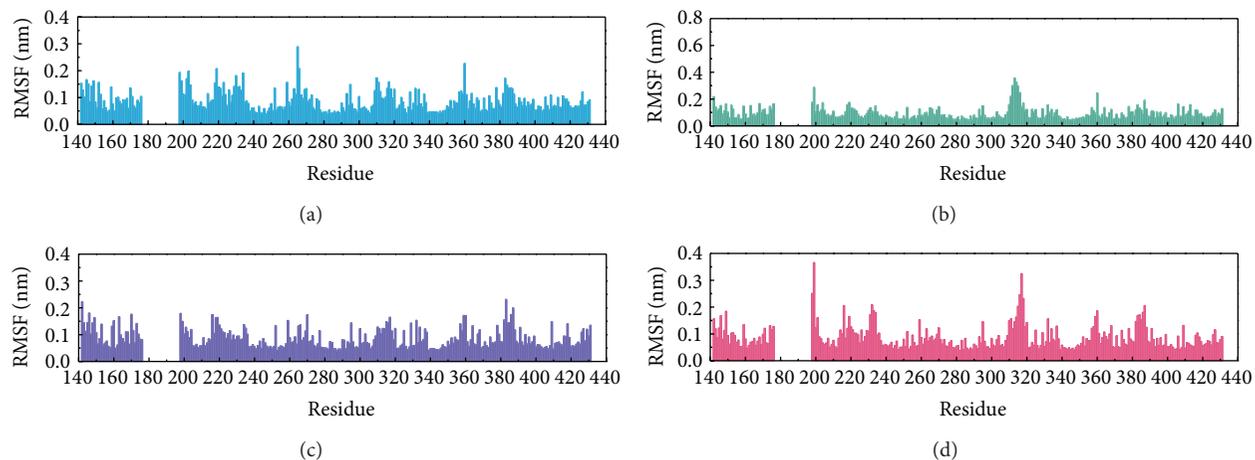


FIGURE 8: Root mean square fluctuation (RMSF) for residues in PXR complex with (a) BEMG, (b) ixerisoside, (c) tangshenoside II, and (d) PNU-142721 over 35–40 ns MD simulation.

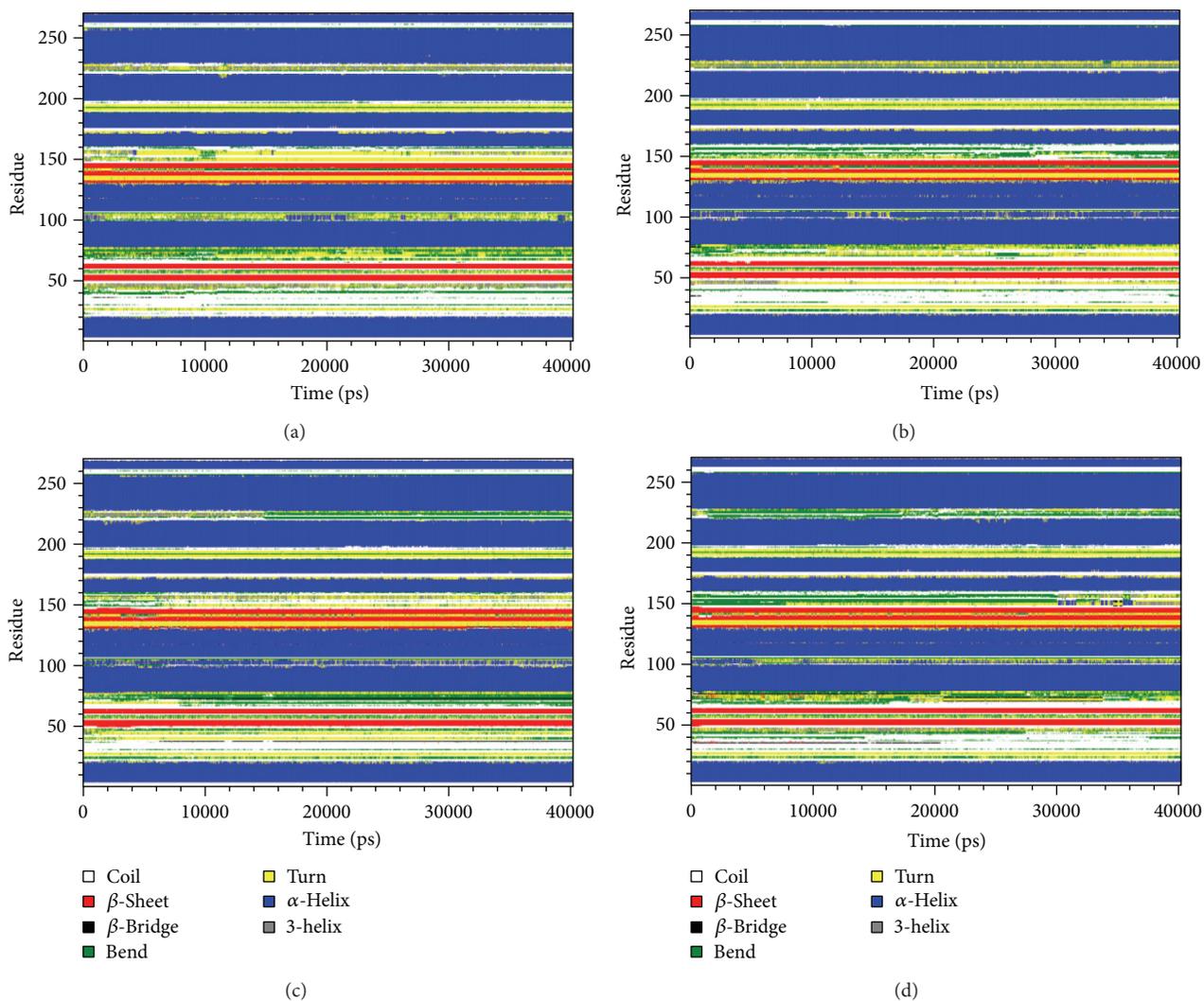


FIGURE 9: Changes of secondary structure in the PXR complex with (a) BEMG, (b) ixerisoside, (c) tangshenoside II, and (d) PNU-142721 during MD simulation.

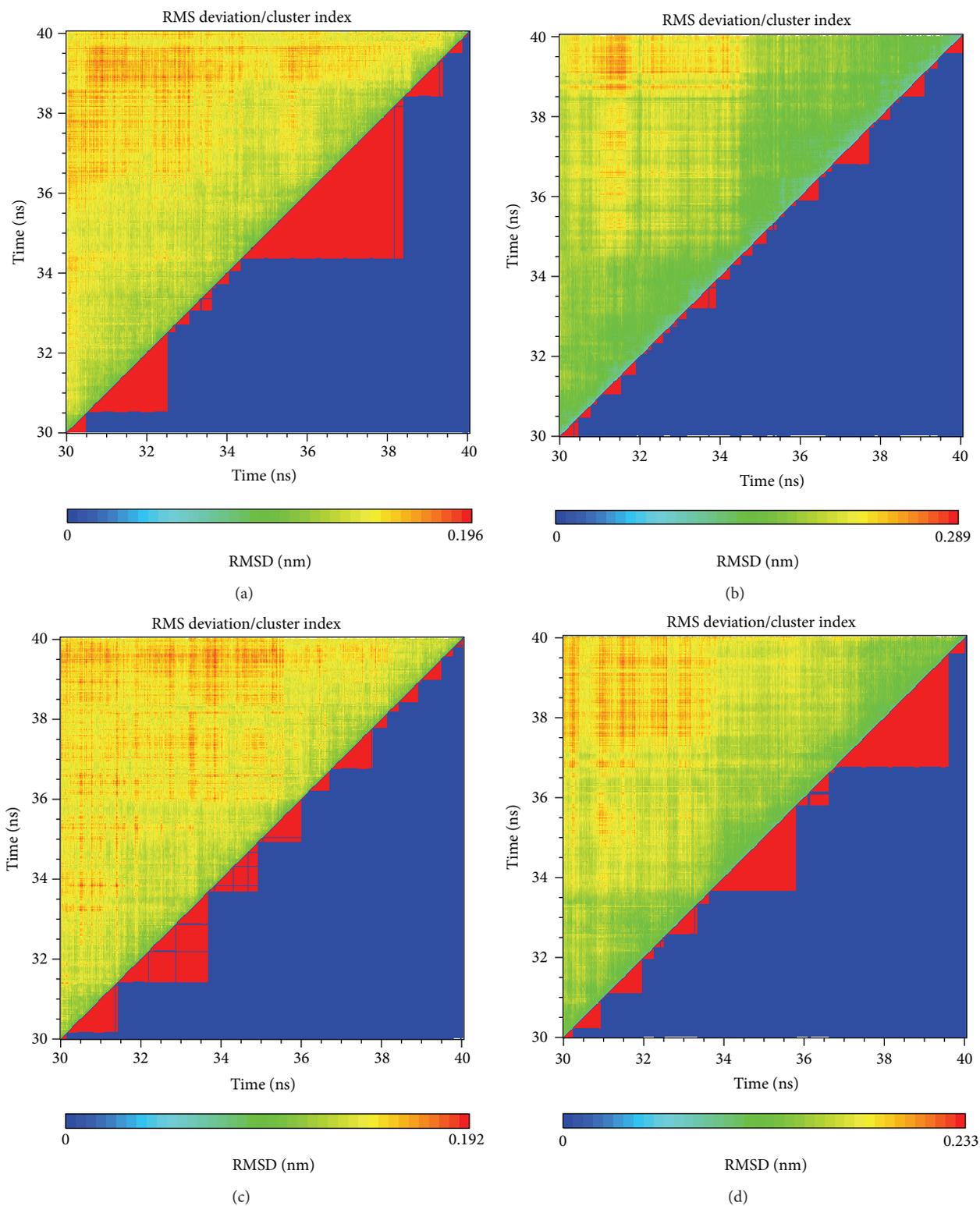


FIGURE 10: RMSD matrix and clustering diagram of MD conformations over 30–40 ns for PXR complex with (a) BEMG, (b) ixerisidide, (c) tangshenaside II, and (d) PNU-142721. Clusters were calculated using a cutoff of 0.1 nm.

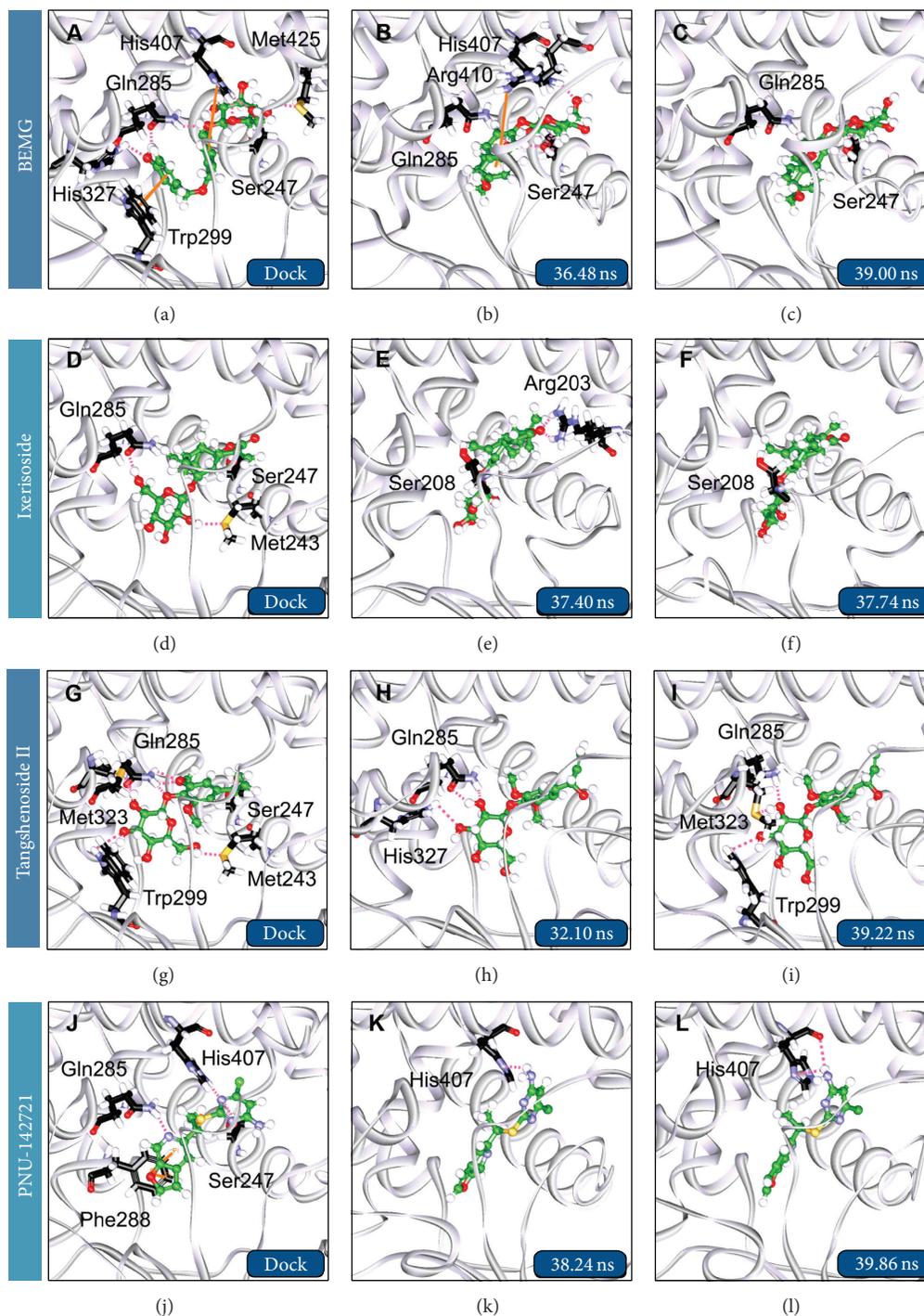


FIGURE 11: Snapshots of docking pose in docking and MD simulation for PXR complex with (a)–(c) BEMG, (d)–(f) ixeriside, (g)–(i) tangshenoside II, and (j)–(l) PNU-142721.

changes in the protein structure of PXR induced by the TCM candidates and control, which are tended to stabilize after MD simulation. For the ligand RMSD in Figure 6(a), the value of BMEG tends to stabilize after 2 ns of MD simulation at approximately 0.21 nm. For the other TCM candidates and control, the ligand RMSD also tends to stabilize after 20 ns of MD simulation. The variation of radii of gyration

for protein and each ligand in Figure 6(b) indicates that each compound may not lead to significant variation to PXR protein under dynamics condition. The slope of the MSD showed in Figure 6(c) indicates that ixeriside induces larger diffusion changes than others, which has an increase the slope after 20 ns. The variation of solvent accessible surface area (SASA) of PXR protein and each ligand in the complexes

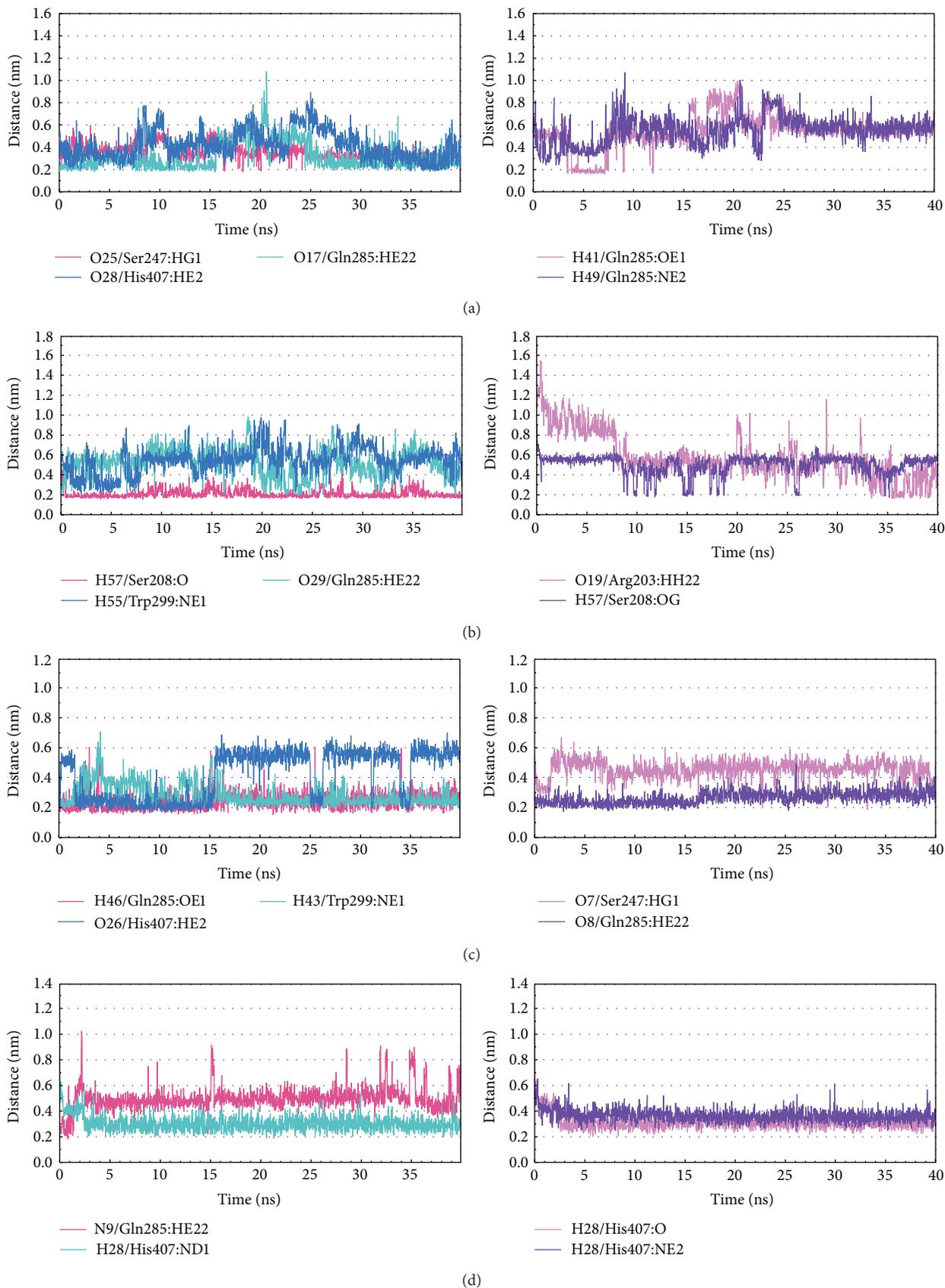


FIGURE 12: Distance variation of H-bonds for PXR complex with (a) BEMG, (b) ixeriside, (c) tangshenoside II, and (d) PNU-142721 during MD simulation.

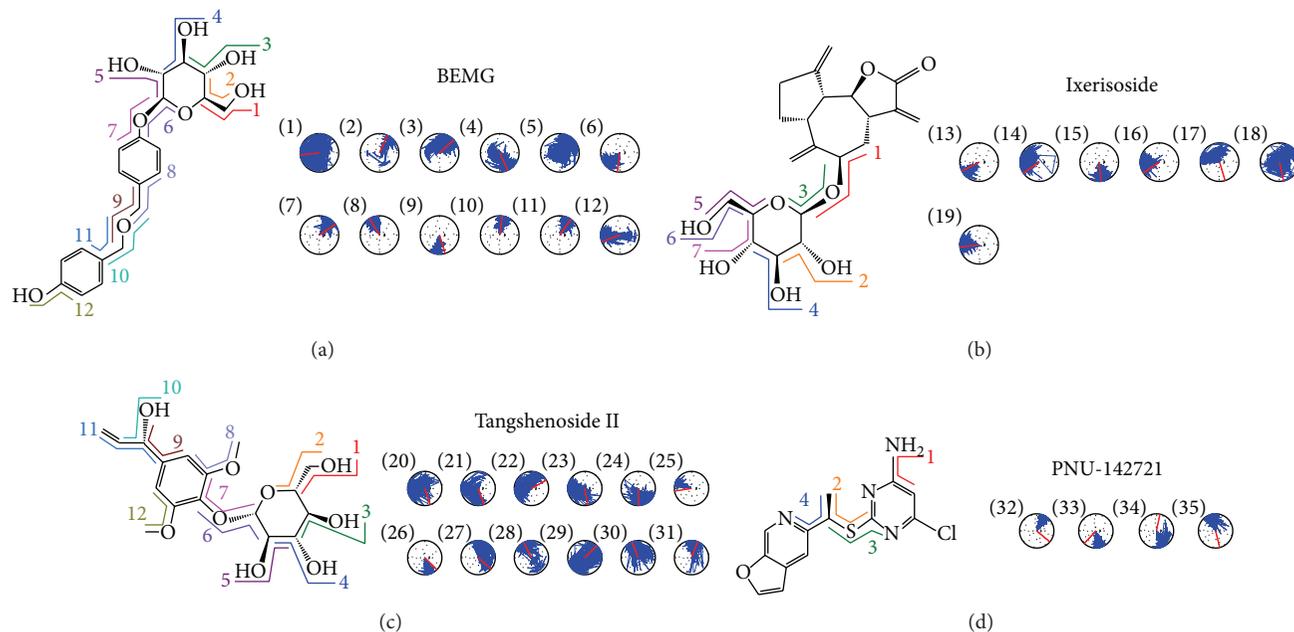


FIGURE 13: Variation of ligand torsion angles for each PXR complex during 40 ns of MD simulation.

over 40 ns of MD simulation is illustrated in Figure 6(d). It shows that there is no significant change in both protein SASA and ligand SASA. The averages of ligand SASA of BEMG, ixeriside, tangshenoside II, and PNU-142721 are 1.85482 nm<sup>2</sup>/NS<sup>2</sup>, 0.937577 nm<sup>2</sup>/NS<sup>2</sup>, 0.499383 nm<sup>2</sup>/NS<sup>2</sup>, and 2.896435 nm<sup>2</sup>/NS<sup>2</sup>, respectively. For the variation of total energy of each protein complex displayed in Figure 7, there is also no significant change under dynamic conditions. Figure 8 displays the root mean square fluctuation (RMSF) of each residue in each PXR protein complex. The key residues in docking simulation, which are Ser247, Gln285, Phe288, Trp299, and His407, have less flexibility under dynamic conditions. Figure 9 displays the change of secondary structure of PXR protein in each complex. There is no significant change in the secondary structure of PXR protein for each protein complex.

The representative structures of PXR protein complexes after MD simulation were identified by the RMSD values and graphical depiction of the clusters analysis with a RMSD cutoff of 0.1 nm during 30–40 ns of MD simulation (Figure 10). The docking poses in docking simulation and two representative structures after MD simulation for each PXR protein complex are illustrated in Figure 11. For BEMG, it maintains the H-bonds with Gln285 and Ser247. Ixeriside forms the H-bond with Ser208 instead of the H-bonds in docking simulation. Tangshenoside II also has stable H-bonds with Gln285, Trp299, and Met323 after MD simulation, as PNU-142721 maintains H-bonds with His407. To discuss the stabilities of H-bonds under dynamics condition, the H-bond occupancy for key residues of PXR protein and variation of each H-bond over 40 ns of MD simulation are displayed in Table 2 and Figure 12, respectively. BEMG has the stable H-bonds with Ser247, Gln285, and His407 after 30 ns of MD simulation. Ixeriside has stable H-bonds

with Ser208 and forms an H-bond with Arg203 instead of Gln285 and Trp299. Tangshenoside II has stable H-bonds with Gln285 and Trp299 and loses the H-bond with Ser247 after 2 ns of MD simulation. For control, PNU-142721, has stable H-bonds with His407. Figure 13 illustrates the variation of torsion angles in each ligand over 40 ns of MD simulation. The variation of each torsion angle supports the result of distance variation of H-bonds in Figure 12.

#### 4. Conclusion

This study aims to investigate the potent TCM candidates for PXR protein. The top three TCM compounds, BEMG, ixeriside, and tangshenoside II, have displayed higher potent binding affinities than the positive control, PNU-142721, in the docking simulation. According to the docking results, the top three candidate compounds and control has hydrophobic contacts with common residues Phe288 and Trp299, and all TCM compounds have hydrophobic contacts with residue Phe281. The docking results indicate that the top three TCM candidate compounds have higher binding affinities than control. In addition, they have H-bonds with key residues Ser247 and Gln285 and hydrophobic contacts with key residues Trp299 and Phe288. The MD simulations are performed to optimize the result of docking simulation and validate the stability of H-bonds between each ligand and PXR protein under dynamic conditions. For the MD simulation, the top three TCM compounds maintain most of interactions with PXR protein, which keep the ligand binding stable in the binding domain. In addition, they have potential bioactivities predicted by the three distinct models. Hence, we propose BEMG and tangshenoside II as potential lead compounds for further study in drug development process with the PXR protein.

TABLE 1: Docking results and predicted pEC<sub>50</sub> for top TCM compounds and PNU-142721.

Name	CS*	Dock score	H-bond forming residues	H-bond quantity	Predicted activity*		
					SVM*	MLR*	BNT*
Bis(4-hydroxybenzyl) ether mono-beta-D-glucopyranoside	11.00	100.59	Ser247, Gln285, His327, His407	6	5.17	2.92	5.29
Ixerisoideside	10.00	103.44	Ser247, Gln285	2	6.34	4.22	5.23
Tangshenoside II	8.00	105.556	Ser247, Gln285, Met243	6	6.51	3.77	4.89
Ruine	8.00	104.085	Ser247, His407	4	5.24	2.73	5.22
Crotalaburnine	9.00	100.181	Ser247, Gln285	3	5.62	1.70	4.50
Dihydroferulic acid [3-(4-hydroxy-3-methoxyphenyl) propionic acid]	9.00	98.854	Gln285, His327, Met425	4	6.11	2.89	4.44
Corchoionoside C	8.00	99.075	Ser247, Met425	2	6.36	3.10	4.64
Beta-D-glucosyl-columbianetin	8.00	98.283	Gln285	1	6.05	3.96	5.24
Ethyl rosmarinat	8.00	97.721	His327, His407	2	5.91	2.94	4.63
Persicarin	7.00	102.239	Ser247, Gln285	2	7.24	9.68	5.83
6beta,7beta,16beta,17-Tetrahydroxy-ent-kauranoic acid	7.00	100.892	Ser247, Gln285, His407	5	6.43	4.22	4.96
Androsin	7.00	99.939	Gln285, His327, His407	5	6.35	3.47	4.62
Baihuaqianhuoside	7.00	98.004	Ser247, Gln285, Met243	3	6.08	2.77	4.51
Eleutheroside B	7.00	97.823	Gln285	3	6.58	3.94	4.93
Androsin	6.00	102.747	Gln285, His327, His407	4	6.28	3.22	4.55
4-Hydroxy-3-methoxy-acetophenone-4-O-beta-D-glucopyranoside	6.00	99.008	Gln285, His407	3	6.35	3.47	4.62
Nortrachelogenin	6.00	98.94	Gln285, Met425	2	5.14	1.75	4.64
3-Methoxy-4-beta-D-glucopyranosyloxypropiofenone	5.00	98.645	Ser247, Gln285, His327	4	6.08	2.77	4.51
Azelaic acid	3.00	102.435	Ser247, His327, His407	4	7.16	5.44	3.63
Sulfoorientalol D	3.00	100.402	Gln285, His407	3	6.59	3.90	4.42
PNU-142721*	0	46.172	Ser247, Gln285, His407	1	5.72	0.54	4.10

PNU-142721: control.

CS: consensus score.

SVM: support vector machine.

MLR: multiple linear regression.

BNT: Bayesian network.

Predicted activity:  $-\log(\text{activity})$ , where activity = % transactivation of PXR receptor  $\times 10 \mu\text{mol/L}$ .

## Conflict of Interests

The authors declared that there is no conflict of interests.

## Authors' Contribution

Kuen-Bao Chen, Hsin-Yi Chen, and Kuan-Chung Chen contributed equally to this paper.

## Acknowledgments

The research was supported by Grants from the National Science Council of Taiwan (NSC102-2325-B039-001 and NSC102-2221-E-468-027-), Asia University (ASIA100-CMU-2, ASIA101-CMU-2, and 102-ASIA-07), and China Medical University Hospital (DMR-103-058, DMR-103-001, and DMR-103-096). This study is also supported in part by Taiwan

TABLE 2: H-bond occupancy for key residues of PXR protein complex with the top three candidates and PNU-142721 over 40 ns molecular dynamics simulation.

Ligand	H-bond	Ligand atom	Amino acid	Distance (nm)			Occupancy (%)
				Max.	Min.	Average	
BEMG	1	O25	Ser247:HG1	0.62	0.18	0.36	12.85%
	2	O26	Ser247:HG1	0.90	0.19	0.49	7.60%
	3	O17	Ser247:HG1	0.91	0.18	0.48	4.40%
	4	O27	Ser247:HG1	0.86	0.19	0.47	4.00%
	5	O17	Gln285:HE22	1.08	0.17	0.32	59.90%
	6	O27	Gln285:HE22	0.89	0.18	0.36	31.45%
	7	H41	Gln285:OE1	1.00	0.16	0.54	9.75%
	8	H49	Gln285:NE2	1.07	0.24	0.55	1.95%
	9	O28	His407:HE2	0.89	0.18	0.42	18.95%
	10	O19	His407:HE2	0.88	0.20	0.56	4.20%
	11	O25	His407:HE2	0.61	0.21	0.40	5.75%
	12	H52	His407:ND1	0.78	0.20	0.47	1.80%
Ixeriside	1	H57	Ser208:O	0.57	0.15	0.21	95.70%
	2	O19	Arg203:HH22	1.55	0.16	0.58	7.15%
	3	H57	Ser208:OG	0.90	0.17	0.50	6.85%
	4	O19	Arg203:HH12	1.66	0.16	0.77	2.45%
	5	O29	Gln285:HE22	0.99	0.18	0.52	6.45%
	6	H55	Gln285:OE1	0.99	0.18	0.60	3.45%
	7	O22	Gln285:HE22	0.91	0.21	0.65	0.50%
	8	H55	Trp299:NE1	0.98	0.21	0.55	5.50%
	9	H56	Trp299:NE1	0.69	0.23	0.45	1.55%
Tangshenoside II	1	H46	Gln285:OE1	0.61	0.15	0.25	84.20%
	2	O8	Gln285:HE22	0.50	0.17	0.26	78.75%
	3	O23	Gln285:HE22	0.51	0.17	0.26	76.70%
	4	O9	Gln285:HE22	0.57	0.20	0.38	13.40%
	5	O7	Gln285:HE22	0.75	0.23	0.47	1.00%
	6	O7	Ser247:HG1	0.67	0.19	0.46	1.10%
	7	O26	Ser247:HG1	0.85	0.19	0.59	0.30%
	8	H43	Trp299:NE1	0.71	0.19	0.28	71.05%
	9	H45	Trp299:NE1	0.69	0.22	0.44	6.50%
	10	O26	His407:HE2	0.70	0.17	0.42	38.05%
PNU-142721	1	N9	Gln285:HE22	1.03	0.17	0.50	2.40%
	2	H28	His407:ND1	0.82	0.19	0.30	54.55%
	3	H28	His407:O	1.05	0.20	0.31	44.55%
	4	H28	His407:NE2	0.65	0.24	0.37	5.60%

H-bond occupancy cutoff: 0.3 nm.

Department of Health Clinical Trial and Research Center of Excellence (DOH102-TD-B-111-004), Taiwan Department of Health Cancer Research Center of Excellence (MOHW103-TD-B-111-03), and CMU under the Aim for Top University Plan of the Ministry of Education, Taiwan.

## References

- [1] A. S. Go, D. Mozaffarian, V. L. Roger et al., "Heart disease and stroke statistics—2013 update: a report from the American Heart Association," *Circulation*, vol. 127, no. 1, Article ID e6-e245, 2013.
- [2] C. A. Ihunnah, M. Jiang, and W. Xie, "Nuclear receptor PXR, transcriptional circuits and metabolic relevance," *Biochimica et Biophysica Acta: Molecular Basis of Disease*, vol. 1812, no. 8, pp. 956–963, 2011.
- [3] K. E. Swales and D. Bishop-Bailey, "The potential use of the pregnane X receptor in cardiovascular therapy," *Expert Review of Cardiovascular Therapy*, vol. 10, no. 9, pp. 1079–1082, 2012.
- [4] A. R. Pries and W. M. Kuebler, "Normal endothelium," *Handbook of Experimental Pharmacology*, no. 176, pp. 1–40, 2006.
- [5] B. Hennig, G. Reiterer, Z. Majkova, E. Oesterling, P. Meerarani, and M. Toborek, "Modification of environmental toxicity by nutrients: implications in atherosclerosis," *Cardiovascular Toxicology*, vol. 5, no. 2, pp. 153–160, 2005.

- [6] Y. M. Leung, K. L. Wong, S. W. Chen et al., "Down-regulation of voltage-gated Ca<sup>2+</sup> channels in Ca<sup>2+</sup> store-depleted rat insulinoma RINm5F cells," *BioMedicine*, vol. 3, no. 3, pp. 130–139, 2013.
- [7] Y. Jiang, X. Li, W. Yang et al., "PKM2 regulates chromosome segregation and mitosis progression of tumor cells," *Molecular Cell*, vol. 53, no. 1, pp. 75–87, 2014.
- [8] V. Janssens and J. Goris, "Protein phosphatase 2A: a highly regulated family of serine/threonine phosphatases implicated in cell growth and signalling," *The Biochemical Journal*, vol. 353, no. 3, pp. 417–439, 2001.
- [9] M. A. Leissring, E. Malito, S. Hedouin et al., "Designed inhibitors of insulin-degrading enzyme regulate the catabolism and activity of insulin," *PLoS ONE*, vol. 5, no. 5, Article ID e10504, 2010.
- [10] K.-P. Su, "Inflammation in psychopathology of depression: clinical, biological, and therapeutic implications," *BioMedicine*, vol. 2, no. 2, pp. 68–74, 2012.
- [11] C.-L. Jao, S.-L. Huang, and K.-C. Hsu, "Angiotensin I-converting enzyme inhibitory peptides: inhibition mode, bioavailability, and antihypertensive effects," *BioMedicine*, vol. 2, no. 4, pp. 130–136, 2012.
- [12] S. A. Kliewer, B. Goodwin, and T. M. Willson, "The nuclear pregnane X receptor: a key regulator of xenobiotic metabolism," *Endocrine Reviews*, vol. 23, no. 5, pp. 687–702, 2002.
- [13] A. Geick, M. Eichelbaum, and O. Burk, "Nuclear receptor response elements mediate induction of intestinal MDR1 by rifampin," *Journal of Biological Chemistry*, vol. 276, no. 18, pp. 14581–14587, 2001.
- [14] K. E. Swales, R. Moore, N. J. Truss et al., "Pregnane X receptor regulates drug metabolism and transport in the vasculature and protects from oxidative stress," *Cardiovascular Research*, vol. 93, no. 4, pp. 674–681, 2012.
- [15] G. Bertilsson, J. Heidrich, K. Svensson et al., "Identification of a human nuclear receptor defines a new signaling pathway for CYP3A induction," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 21, pp. 12208–12213, 1998.
- [16] S. A. Kliewer, J. T. Moore, L. Wade et al., "An orphan nuclear receptor activated by pregnanes defines a novel steroid signaling pathway," *Cell*, vol. 92, no. 1, pp. 73–82, 1998.
- [17] B. Blumberg, W. Sabbagh Jr., H. Juguilon et al., "SXR, a novel steroid and xenobiotic-sensing nuclear receptor," *Genes and Development*, vol. 12, no. 20, pp. 3195–3205, 1998.
- [18] L. M. Tompkins and A. D. Wallace, "Mechanisms of cytochrome P450 induction," *Journal of Biochemical and Molecular Toxicology*, vol. 21, no. 4, pp. 176–181, 2007.
- [19] R. E. Watkins, P. R. Davis-Searles, M. H. Lambert, and M. R. Redinbo, "Coactivator binding promotes the specific interaction between ligand and the pregnane X receptor," *Journal of Molecular Biology*, vol. 331, no. 4, pp. 815–828, 2003.
- [20] R. E. Watkins, G. B. Wisely, L. B. Moore et al., "The human nuclear xenobiotic receptor PXR: structural determinants of directed promiscuity," *Science*, vol. 292, no. 5525, pp. 2329–2333, 2001.
- [21] R. E. Watkins, S. M. Noble, and M. R. Redinbo, "Structural insights into the promiscuity and function of the human pregnane X receptor," *Current Opinion in Drug Discovery and Development*, vol. 5, no. 1, pp. 150–158, 2002.
- [22] J. Orans, D. G. Teotico, and M. R. Redinbo, "The nuclear xenobiotic receptor pregnane X receptor: recent insights and new challenges," *Molecular Endocrinology*, vol. 19, no. 12, pp. 2891–2900, 2005.
- [23] S. D. Shenoy, T. A. Spencer, N. A. Mercer-Haines et al., "Induction of CYP3A by 2,3-oxidosqualene:lanosterol cyclase inhibitors is mediated by an endogenous squalene metabolite in primary cultured rat hepatocytes," *Molecular Pharmacology*, vol. 65, no. 5, pp. 1302–1312, 2004.
- [24] E. G. Schuetz, C. Brimer, and J. D. Schuetz, "Environmental xenobiotics and the antihormones cyproterone acetate and spironolactone use the nuclear hormone pregnenolone X receptor to activate the CYP3A23 hormone response element," *Molecular Pharmacology*, vol. 54, no. 6, pp. 1113–1117, 1998.
- [25] C. Y.-C. Chen, "A novel integrated framework and improved methodology of computer-aided drug design," *Current Topics in Medicinal Chemistry*, vol. 13, no. 9, pp. 965–988, 2013.
- [26] W. I. Tou, S.-S. Chang, C.-C. Lee, and C. Y.-C. Chen, "Drug design for neuropathic pain regulation from traditional Chinese medicine," *Scientific Reports*, vol. 3, p. 844, 2013.
- [27] K. C. Chen, Y. R. Jian, M. F. Sun et al., "Investigation of silent information regulator 1 (Sirt1) agonists from Traditional Chinese Medicine," *Journal of Biomolecular Structure & Dynamics*, vol. 31, no. 11, pp. 1207–1218, 2013.
- [28] H.-J. Huang, H. W. Yu, C.-Y. Chen et al., "Current developments of computer-aided drug design," *Journal of the Taiwan Institute of Chemical Engineers*, vol. 41, no. 6, pp. 623–635, 2010.
- [29] C. Y.-C. Chen, "Weighted equation and rules—a novel concept for evaluating protein-ligand interaction," *Journal of Biomolecular Structure and Dynamics*, vol. 27, no. 3, pp. 271–282, 2009.
- [30] K.-C. Chen and C. Yu-Chian Chen, "Stroke prevention by traditional Chinese medicine? A genetic algorithm, support vector machine and molecular dynamics approach," *Soft Matter*, vol. 7, no. 8, pp. 4001–4008, 2011.
- [31] K.-C. Chen, K.-W. Chang, H.-Y. Chen, and C. Y.-C. Chen, "Traditional Chinese medicine, a solution for reducing dual stroke risk factors at once?" *Molecular BioSystems*, vol. 7, no. 9, pp. 2711–2719, 2011.
- [32] T.-T. Chang, K.-C. Chen, K.-W. Chang et al., "In silico pharmacology suggests ginger extracts may reduce stroke risks," *Molecular BioSystems*, vol. 7, no. 9, pp. 2702–2710, 2011.
- [33] Y. A. Tsou, K. C. Chen, S. S. Chang, Y. R. Wen, and C. Y. Chen, "A possible strategy against head and neck cancer: in silico investigation of three-in-one inhibitors," *Journal of Biomolecular Structure & Dynamics*, vol. 31, no. 12, pp. 1358–1369, 2013.
- [34] Y.-A. Tsou, K.-C. Chen, H.-C. Lin, S.-S. Chang, and C. Y.-C. Chen, "Uroporphyrinogen decarboxylase as a potential target for specific components of Traditional Chinese Medicine: a virtual screening and molecular dynamics study," *PLoS ONE*, vol. 7, no. 11, Article ID e50087, 2012.
- [35] S.-C. Yang, S.-S. Chang, H.-Y. Chen, and C. Y.-C. Chen, "Identification of potent EGFR inhibitors from TCM Database@Taiwan," *PLoS Computational Biology*, vol. 7, no. 10, Article ID e1002189, 2011.
- [36] C.-Y. Chen and C. Y.-C. Chen, "Insights into designing the dual-targeted HER2/HSP90 inhibitors," *Journal of Molecular Graphics and Modelling*, vol. 29, no. 1, pp. 21–31, 2010.
- [37] K.-C. Chen, M.-F. Sun, S.-C. Yang et al., "Investigation into potent inflammation inhibitors from Traditional Chinese Medicine," *Chemical Biology and Drug Design*, vol. 78, no. 4, pp. 679–688, 2011.

- [38] K. C. Chen, S. S. Chang, F. J. Tsai, and C. Y. Chen, "Han ethnicity-specific type 2 diabetic treatment from traditional Chinese medicine?" *Journal of Biomolecular Structure & Dynamics*, vol. 31, no. 11, pp. 1219–1235, 2013.
- [39] H.-J. Huang, K.-J. Lee, H. W. Yu, H.-Y. Chen, F.-J. Tsai, and C. Y.-C. Chen, "A novel strategy for designing the selective PPAR agonist by the "Sum of activity" model," *Journal of Biomolecular Structure and Dynamics*, vol. 28, no. 2, pp. 187–200, 2010.
- [40] K.-C. Chen, S.-S. Chang, H.-J. Huang, T.-L. Lin, Y.-J. Wu, and C. Y.-C. Chen, "Three-in-one agonists for PPAR- $\alpha$ , PPAR- $\gamma$ , and PPAR-d from traditional Chinese medicine," *Journal of Biomolecular Structure and Dynamics*, vol. 30, no. 6, pp. 662–683, 2012.
- [41] S.-S. Chang, H.-J. Huang, and C. Y.-C. Chen, "Two birds with one stone? Possible dual-targeting H1N1 inhibitors from traditional Chinese medicine," *PLoS Computational Biology*, vol. 7, no. 12, Article ID e1002315, 2011.
- [42] H. J. Huang, Y. R. Jian, and C. Y. Chen, "Traditional Chinese medicine application in HIV: an in silico study," *Journal of Biomolecular Structure & Dynamics*, vol. 32, no. 1, pp. 1–12, 2014.
- [43] H.-C. Tang and C. Y.-C. Chen -C, "Investigation of the novel lead of melanocortin 1 receptor for pigmentary disorders," *Evidence-Based Complementary and Alternative Medicine*, vol. 2014, Article ID 254678, 13 pages, 2014.
- [44] H.-J. Huang, C.-C. Lee, and C. Y.-C. Chen, "Pharmacological chaperone design for reducing risk factor of Parkinson's disease from traditional Chinese medicine," *Evidence-Based Complementary and Alternative Medicine*, vol. 2014, Article ID 830490, 12 pages, 2014.
- [45] H. Y. Chen, S. S. Chang, Y. C. Chan, and C. Y. Chen, "Discovery of novel insomnia leads from screening traditional Chinese medicine database," *Journal of Biomolecular Structure & Dynamics*, vol. 32, no. 5, pp. 776–791, 2014.
- [46] W. I. Tou and C. Y. Chen, "May disordered protein cause serious drug side effect?" *Drug Discovery Today*, vol. 19, no. 4, pp. 367–372, 2013.
- [47] C. Y.-C. Chen and W. I. Tou, "How to design a drug for the disordered proteins?" *Drug Discovery Today*, vol. 18, no. 19–20, pp. 910–915, 2013.
- [48] Y. Cheng and M. R. Redinbo, "Activation of the human nuclear xenobiotic receptor PXR by the reverse transcriptase-targeted anti-HIV drug PNU-142721," *Protein Science*, vol. 20, no. 10, pp. 1713–1719, 2011.
- [49] B. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker, and V. N. Uversky, "PONDR-FIT: a meta-predictor of intrinsically disordered amino acids," *Biochimica et Biophysica Acta: Proteins and Proteomics*, vol. 1804, no. 4, pp. 996–1010, 2010.
- [50] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson et al., "CHARMM: a program for macromolecular energy minimization and dynamics calculations," *Journal of Computational Chemistry*, vol. 4, pp. 187–217, 1983.
- [51] C. Y.-C. Chen, "TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico," *PLoS ONE*, vol. 6, no. 1, Article ID e15939, 2011.
- [52] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings," *Advanced Drug Delivery Reviews*, vol. 46, no. 1–3, pp. 3–26, 2001.
- [53] C. M. Venkatachalam, X. Jiang, T. Oldfield, and M. Waldman, "LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites," *Journal of Molecular Graphics and Modelling*, vol. 21, no. 4, pp. 289–307, 2003.
- [54] K. Zimmermann, M. D. Wittman, M. G. Saulnier et al., "SAR of PXR transactivation in benzimidazole-based IGF-1R kinase inhibitors," *Bioorganic and Medicinal Chemistry Letters*, vol. 20, no. 5, pp. 1744–1748, 2010.
- [55] D. Rogers and A. J. Hopfinger, "Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships," *Journal of Chemical Information and Computer Sciences*, vol. 34, no. 4, pp. 854–866, 1994.
- [56] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, article 27, 2011.
- [57] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, "GRGMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation," *Journal of Chemical Theory and Computation*, vol. 4, no. 3, pp. 435–447, 2008.
- [58] V. Zoete, M. A. Cuendet, A. Grosdidier, and O. Michielin, "SwissParam: a fast force field generation tool for small organic molecules," *Journal of Computational Chemistry*, vol. 32, no. 11, pp. 2359–2368, 2011.
- [59] R. Fletcher, *Optimization*, Academic Press, New York, NY, USA, 1969.
- [60] L. G. Jian Yin, *Modern Study of Chinese Drugs*, Xueyuan Press, Beijing, China, 1993, (Chinese).
- [61] X. Xiong, X. Yang, W. Liu et al., "Banxia baizhu tianma decoction for essential hypertension: a systematic review of randomized controlled trials," *Evidence-based Complementary and Alternative Medicine*, vol. 2012, Article ID 271462, 10 pages, 2012.
- [62] H. W. Zhang, J. Tong, G. Zhou, H. Jia, and J. Y. Jiang, "Tianma Gouteng Yin Formula for treating primary hypertension," *Cochrane Database of Systematic Reviews*, vol. 6, Article ID CD008166, 2012.
- [63] W. Kisiel and K. Zielińska, "Guaianolides from Cichorium intybus and structure revision of Cichorium sesquiterpene lactones," *Phytochemistry*, vol. 57, no. 4, pp. 523–527, 2001.
- [64] P. N. Pushparaj, H. K. Low, J. Manikandan, B. K. H. Tan, and C. H. Tan, "Anti-diabetic effects of Cichorium intybus in streptozotocin-induced diabetic rats," *Journal of Ethnopharmacology*, vol. 111, no. 2, pp. 430–434, 2007.
- [65] A. J. Atta, T. A. Elkoly, S. M. Mouneir, G. Kamel, N. A. Alwabel, and S. Zaher, "Hepatoprotective effect of methanol extracts of Zingiber officinale and Cichorium intybus," *Indian Journal of Pharmaceutical Sciences*, vol. 72, no. 5, pp. 564–570, 2010.
- [66] D. Song, G.-X. Chou, G.-Y. Zhong, and Z.-T. Wang, "Two new phenylpropanoid derivatives from codonopsis tangshen OLIV," *Helvetica Chimica Acta*, vol. 91, no. 10, pp. 1984–1988, 2008.

## Research Article

# Evaluation of Correlation of Cell Cycle Proteins and Ki-67 Interaction in Paranasal Sinus Inverted Papilloma Prognosis and Squamous Cell Carcinoma Transformation

Yung-An Tsou,<sup>1,2,3</sup> Hung-Jin Huang,<sup>4</sup> Tang-Chuan Wang,<sup>1</sup> Chih-Jaan Tai,<sup>2</sup> Chuan-Mu Chen,<sup>3</sup> and Calvin Yu-Chian Chen<sup>2,5,6,7</sup>

<sup>1</sup> Department of Otolaryngology Head and Neck Surgery, China Medical University, Taichung 40402, Taiwan

<sup>2</sup> Department of Medicine, School of Medicine, China Medical University, Taichung 40402, Taiwan

<sup>3</sup> Department of Life Sciences, Agricultural Biotechnology Center, National Chung Hsing University, Taichung 402, Taiwan

<sup>4</sup> Department of Chinese Pharmaceutical Sciences and Chinese Medicine Resources, College of Pharmacy, China Medical University, Taichung 40402, Taiwan

<sup>5</sup> Department of Biomedical Informatics, Asia University, Taichung 41354, Taiwan

<sup>6</sup> Computational and Systems Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>7</sup> Research Center for Chinese Medicine and Acupuncture, China Medical University, Taichung 40402, Taiwan

Correspondence should be addressed to Chuan-Mu Chen; [cmchen7010978@gmail.com](mailto:cmchen7010978@gmail.com) and Calvin Yu-Chian Chen; [ycc929@MIT.edu](mailto:ycc929@MIT.edu)

Received 22 February 2014; Revised 5 March 2014; Accepted 5 March 2014; Published 12 June 2014

Academic Editor: Chung Y. Hsu

Copyright © 2014 Yung-An Tsou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The recurrent sinonasal inverted papilloma (IP) could be transformed to sinonasal squamous cell carcinoma. We use protein expression patterns by immunohistochemical method to see whether the expression of p53, p16, p21, and p27 belongs to cell-cycle-regulators and PCNA (proliferating cell nuclear antigen) and Ki-67 the proliferation markers in sixty patients with sinonasal inverted papilloma, and 10 of them with squamous cell carcinoma transformation. Significantly elevated levels of Ki-67, p27, and PCNA in IP with squamous cell carcinoma transformation of sinonasal tract compared with inverted papilloma were revealed. No variation of p16, p21, PLUNC (palate, lung, and nasal epithelium clone protein) and p53 expression was correlated to sinonasal IP malignant transformation by multivariate survey. However, we found elevated PLUNC expression in IPs with multiple recurrences. Finally, we found that PCNA, p27 may interact with CDK1 which promote IP cell proliferation and correlate to sinonasal squamous cell carcinoma. Ki-67 could work throughout the cell cycles to cause malignant transformation. In conclusion, this is a first study showing the correlation of Ki-67, PCNA interacted with CDK1 might lead to malignant transformation. Elevated PLUNC expression in the sinonasal IPs was related to multiple recurrences in human.

## 1. Introduction

The inverted papilloma (IP) is a type of tumor in which surface epithelial cells grow downward into the underlying supportive tissue. The bladder, renal pelvis, ureter, urethra, nose, and paranasal sinuses are all possible areas for IP occurrence [1]. Epistaxis and nasal obstruction with facial pain or headache attacked when the nose or sinuses mucosa bear the IP [2]. Although the IP originating from the out-lining respiratory membrane belongs to a benign epithelial

neoplasm, the local invasiveness, higher recurrence rate, and malignant transformation make it difficult to treat [3]. The malignant transformation rate is 5–10% and many of them are synchronous presenting with squamous cell carcinoma [4]. PCNA (proliferating cell nuclear antigen) acts as an antigen expression in the cell nuclei in the phase of DNA synthesis during cell cycle and considers the cancer prognosis, but the relationship to IP is still controversial [5].

Cyclin and cyclin dependent kinase (CDK) partake the important roles in cell cycle during proliferation and affect

different cell cycle phases [6–9]. The CDK1 is a very crucial initiator for cell proliferation and malignant transformation factor [10]. On the contrary, the p21, p27, and CDK inhibitors, limited CDKs and arrest the cell cycle [8]. The p21 is an inhibitor of G1 cell phase CDKs which restrain cells entry into S phase. The p27 could also bind to CDKs and act as CDKI as p21. The p27 interacts with cyclin E-CDK2, cyclin A-CDK2, and cyclin D1-CDK4 complexes, affecting cell proliferation and apoptosis [6, 7].

The proliferation marker Ki-67 antigen, detected with monoclonal antibody MIB-1, is expressed in all G1, S, G2, and M cell phases except G0 [8]. Ki-67 expression also correlated to the tumor behavior, pathologic tumor grade, and early recurrence in various carcinomas [11–15].

Concerning the IPs transformed or synchronous with cancers, we also do the p16, p53, Ki-67, and PLUNC (palate, lung, and nasal epithelium clone protein) IHC study. The p16 is a tumor suppressor protein decelerating G1 to S cell phase and prevent of malignant transformation [8]. PLUNC is an innate immune material that has an anticancer effect for nasopharyngeal cancer but no studies revealed its relevance to sinonasal IPs [16].

PCNA, p53, p21, p27, and Ki-67 surveys were done to see whether they were correlating to tumor extents and the treatment outcome of sinonasal IPs [8].

The aim of this study is to investigate the roles of cell-cycle-regulators p53, p21, p27, proliferation marker Ki-67, p16, PCNA, and innate immune material PLUNC in recurrence and malignant transformation for sinonasal IPs. We had also surveyed whether possible predicted factors of Ki-67, PCNA, and p27 correlate to CDKs by computational simulation finally to give a possible explanation to the mechanism of Ki-67, PCNA, and p27 to the CDKs in IPs prognosis.

## 2. Materials and Methods

From Department of China Medical University Hospital from January 2000 to June 2010, 60 cases of sinonasal IPs and 10 cases of IPs with squamous cell carcinoma transformation were collected and reviewed from the medical records in a retrospective manner. All of the tissues fixed in 10% formalin and prepared in paraffin were used for IHC studies by avidin-biotin-peroxidase complex method. Mouse monoclonal antibodies (mAb), anti-p16 (Neomarkers, DCS-50, 200 mg/L), anti-p21 (Neomarkers, MS 387-P, 200 mg/L), anti-p27 (Neomarkers, MS 256-P, 200 mg/L), anti-p53 (Neomarkers, RM 9105-S), Ki-67 (Neomarkers RM 9106-S) PCNA, and PLUNC (Santa Cruz Biotechnology, Santa Cruz, CA, USA) were used for immunohistochemistry through the streptavidin-biotin peroxidase method [11]. Xylene was used for deparaffinizing all of the tissue sections and then rehydrated by alcohol series, and finally saturated in distilled water. Then the tissue was shifted into phosphate-buffered saline with adding 0.3% solution of hydrogen peroxidase to block the endogenous peroxidase activity at room temperature for 10 min, then Tris buffer rinsed thereafter. The sections were then boiled by a microwave oven for 15 min in citrate buffer solution (10 mmol/L; pH, 6.0). Primary antibodies p53, p21, p27, p16,

PCNA, Ki-67, and PLUNC was applied one by one for 60 min at room temperature [11].

Then we add linking antibody and streptavidin peroxidase complex (DAKO LSAB Kit, K-0675; Carpinteria, CA) for 15 min at room temperature. The 0.05% diaminobenzidine tetrahydrochloride (DAB) was used for 15 min for tissue stain and finally washed twice by the Tris buffer [8, 11].

The sections were counterstained with Mayer's hematoxylin after washing by deionized water. The immunostained nuclei were quantified in each case. All counting was performed under a standard light microscope in 1000x field to evaluate positive nuclei/total number of cells. Ten fields or at least 500 cells were counted on each section. Tumor sections were considered negative if staining was absent or present in <10% of tumor cells. A score of 1+ was given when 10–30% of the cells were positive to the reaction. A score of 2+ was given when 30–50% of the cells were positive to the reaction. A score of 3+ was given when >50% of the cells were positive to the reaction, respectively (Tables 1 and 2) [17]. Statistical significance was analyzed using the Pearson's chi-squared test or Fisher exact test for univariate analysis and multiple logistic regression test was used for multivariate analysis. Results were considered statistically significant when the *P* value was <0.05.

The protein-protein docking was carried out by ZDOCK program [18] for analyzing the three possible prognostic factors for malignant transformation bound to CDK1. To render the possible mechanism to what we found in this study, we calculate the interaction of the Ki-67, p27, and PCNA to CDK1 by computational biology. We further utilized GROMACS 4.5.5 program [19] to observe the stability of the complexes after ZDOCK binding predications. The environment of MD system was set in the TIP3P water modeling with 1.2 nm distance water box which contained Na and Cl ions in the concentration of 0.145 M NaCl for system neutralization. First, we set 5,000-cycle steps in the steepest descent algorithm for energy minimization. Secondly, the constant temperature dynamics (NVT type) conditions were employed to provide MD simulation for equilibration and performed in 1 ns time period. In the final step, the constant pressure and temperature dynamics (NPT type) were set for the production run in 5000 ps time period. The temperature of the system during the simulation process was set as 310 K. For trajectory analysis, we employed software GROMACS 4.5.5 to count the root mean square deviation (RMSD) and radius of gyration (Rg), respectively. Series of MD conformation data was surveyed every 20 ps of all production runs.

## 3. Results

There were a total of 55 males and 15 females included in this study. Sixty of them are IPs and the other 10 are sinonasal IPs collected with malignant transformation. The age is  $45.64 \pm 14.45$  years old ranging from 25 to 78 years old. The stains revealed significantly elevated levels of PLUNC, but decreased levels of Ki-67, p53, p21, and p27 in patients with multiple recurrence of IPs (Table 1). We also found

TABLE 1: The correlation of IHC results with multiple recurrence in univariant by Pearson Xi square and Logistic multivariant regression analysis.

	>3 times surgery		Sum	P	pp
	0	1			
PCNA					
0	5	5	10	0.03	0.144
1	20	5	25		
2	15	10	25		
3	10	0	10		
Ki-67					
0	25	15	40	0.05	0.01
1	10	5	15		
2	10	0	10		
3	5	0	5		
P53					
0	10	15	25	0.0001	0.002
1	10	0	10		
2	10	0	10		
3	20	5	25		
P16					
0	20	5	25	0.04	1.000
1	10	10	20		
2	20	5	25		
3	0	0	0		
P21					
0	35	20	55	0.002	0.01
1	10	0	10		
2	5	0	5		
3	0	0	0		
P27					
0	10	10	20	0.02	0.04
1	15	0	15		
2	15	0	15		
3	20	0	20		
PLUNC					
0	10	0	0	0.02	0.002
1	5	0	0		
2	15	5	5		
3	20	5	15		

TABLE 2: The correlation of IHC results with malignant transformation in univariant by Pearson Xi square and Logistic multivariant regression analysis.

	Malignant transformation		Sum	P	pp
	0	1			
PCNA					
0	10	10	10	0.001	0.0001
1	25	25	25		
2	25	25	25		
3	0	10	10		
Ki-67					
0	40	0	40	0.001	0.0001
1	10	5	15		
2	10	0	10		
3	0	5	5		
P53					
0	25	0	25	0.001	0.187
1	5	5	10		
2	10	0	10		
3	20	5	25		
P16					
0	20	5	25	0.097	1.000
1	20	0	20		
2	20	5	25		
3	0	0	0		
P21					
0	50	5	55	0.002	0.214
1	5	5	10		
2	5	0	5		
3	0	0	0		
P27					
0	10	10	20	0.001	0.001
1	15	0	15		
2	15	0	15		
3	20	0	20		
PLUNC					
0	10	0	10	0.224	0.250
1	5	0	5		
2	15	5	20		
3	30	5	35		

the elevated PCNA, Ki-67, and p27 in the sinonasal IPs with squamous cell carcinoma transformation compared with sinonasal IPs alone with no synchronous malignancy (Table 2). The elevated PLUNC expression is correlated to multiple sinus surgery in the patients with sinonasal IPs. However, the PLUNC expression level is not correlated to the malignant transformation of IPs to SCC. We also showed the IHC expression of different levels for PCNA, Ki-67, p27, and PLUNC in Figures 1, 2, 3, and 4. The preoperative MRI or CT

scan were all perform for all the patients as in Figures 5 and 6.

The elevated Ki-67 immunohistochemical staining is found in both sinonasal IPs with multiple recurrences and malignant transformation in our univariant and multivariant analysis by Pearson's chi-squared test and multiple logistic regression test as showed in Tables 1 and 2 and Figure 2(a). Therefore, high Ki-67 index could be considered an important factor for prognosis and malignant predicted

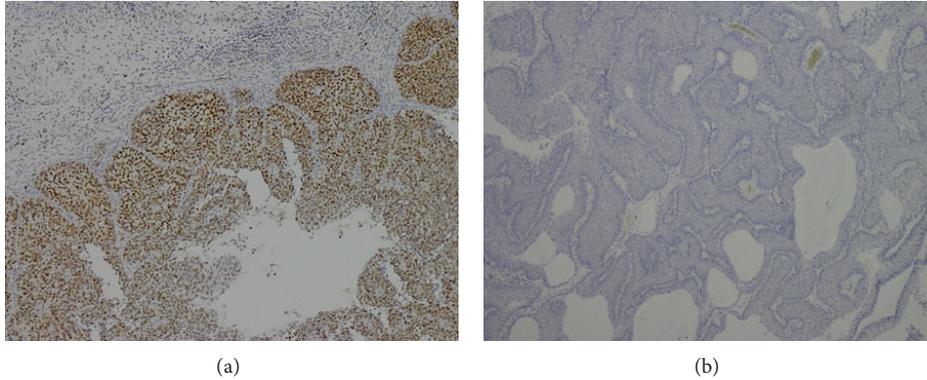


FIGURE 1: The PCNA IHC (a) expression +++ and (b) 0 in IPs.

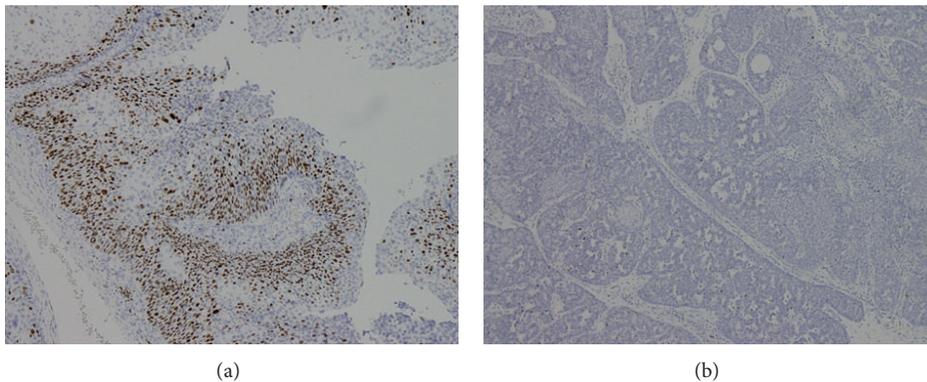


FIGURE 2: The Ki67 IHC (a) expression +++ and (b) 0 in IPs.

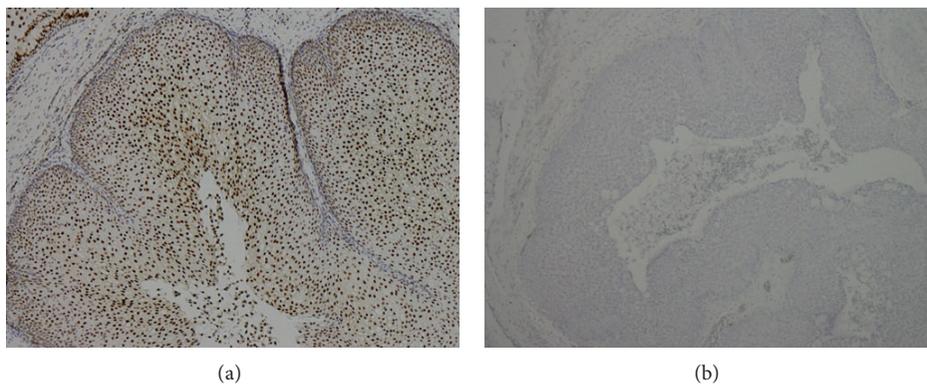


FIGURE 3: The p27 IHC (a) expression +++ and (b) 0 in IPs.

markers. We could even combine the survey of the significantly increased Ki-67 and PCNA IHC expression level and the lower p27 level to predict the higher malignant transformation trend in patients with sinonasal IPs in our survey.

We consequently did some docking analyses with final molecular dynamic studies between what we found the prognosis factors of PCNA and CDK1, Ki-67 and CDK1, and p27 and CDK1 which all showed stable docking in Figure 7 and their docking score is also showed in by ZDOCK program. The ZDOCK generated the top 10 docking poses

of CDK1 and Ki-67, CDK1-p27, and CDK1-PCNA. We chose the best docking score for analyzing the stability of protein-protein interaction. The Ki-67, p27, and PCNA to the CDK1 binding score are 20.92, 21.72, and 24.32, respectively by the ZDOCK program. The key residues are Gly9, Ser10, Ile11, Leu12, Lys13, Lys14, and Val15 which were the major binding domains for both Ki-67 and p27 to the CDK1 shown in red ribbon (Figure 7). We further did MD analysis for these key residues to see the interaction of these three proteins to the CDK1. After the MD simulation of CDK1 complex with Ki-67, p27, and PCNA, we performed the RMSD analysis

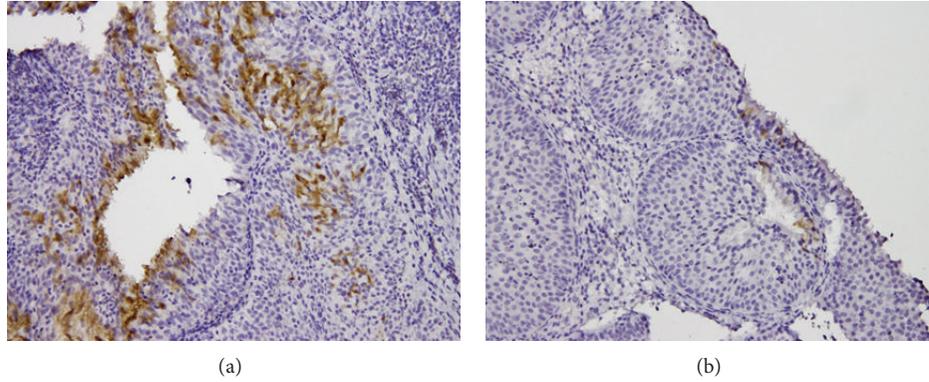


FIGURE 4: The PLUNC IHC (a) expression +++ and (b) 0 in IPs.

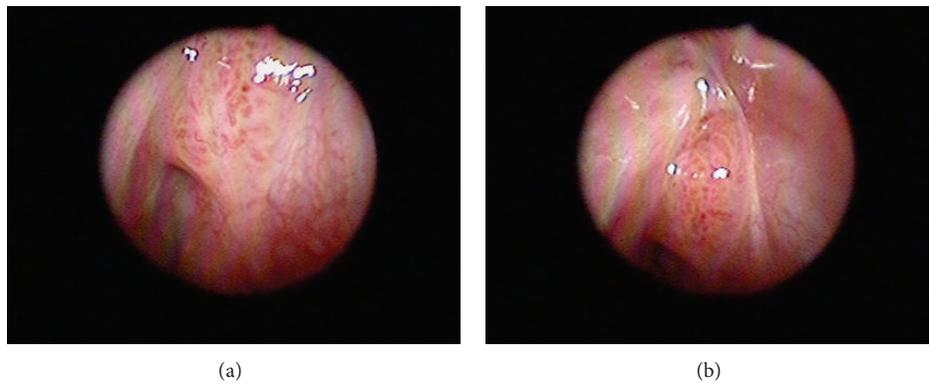


FIGURE 5: Sinoscopic view of sinonasal inverted papilloma with malignancy transformation.



FIGURE 6: Sinus MRI (coronary view) and sinus CT scan (axial view) of sinonasal inverted papilloma with malignant transformation.

for 5000 ps simulation times. All the three proteins (Ki-67, p27, and PCNA) are revealed to have stable fluctuation after 1000 ps simulation time (Figure 8).

In Figure 9, we calculated the proteins radius gyration for 5000 ps simulation time period. The three complexes tend to low values of radius of gyration and stable fluctuation over

all MD simulation, which indicate the compacted complexes between each of the protein structure. The SASA (area of solvent) method was used for hydrophobic nature of the three complexes, these three became stable fluctuation and also showed stable hydrophobic change between each protein-protein interactions (Figure 10). Besides the total

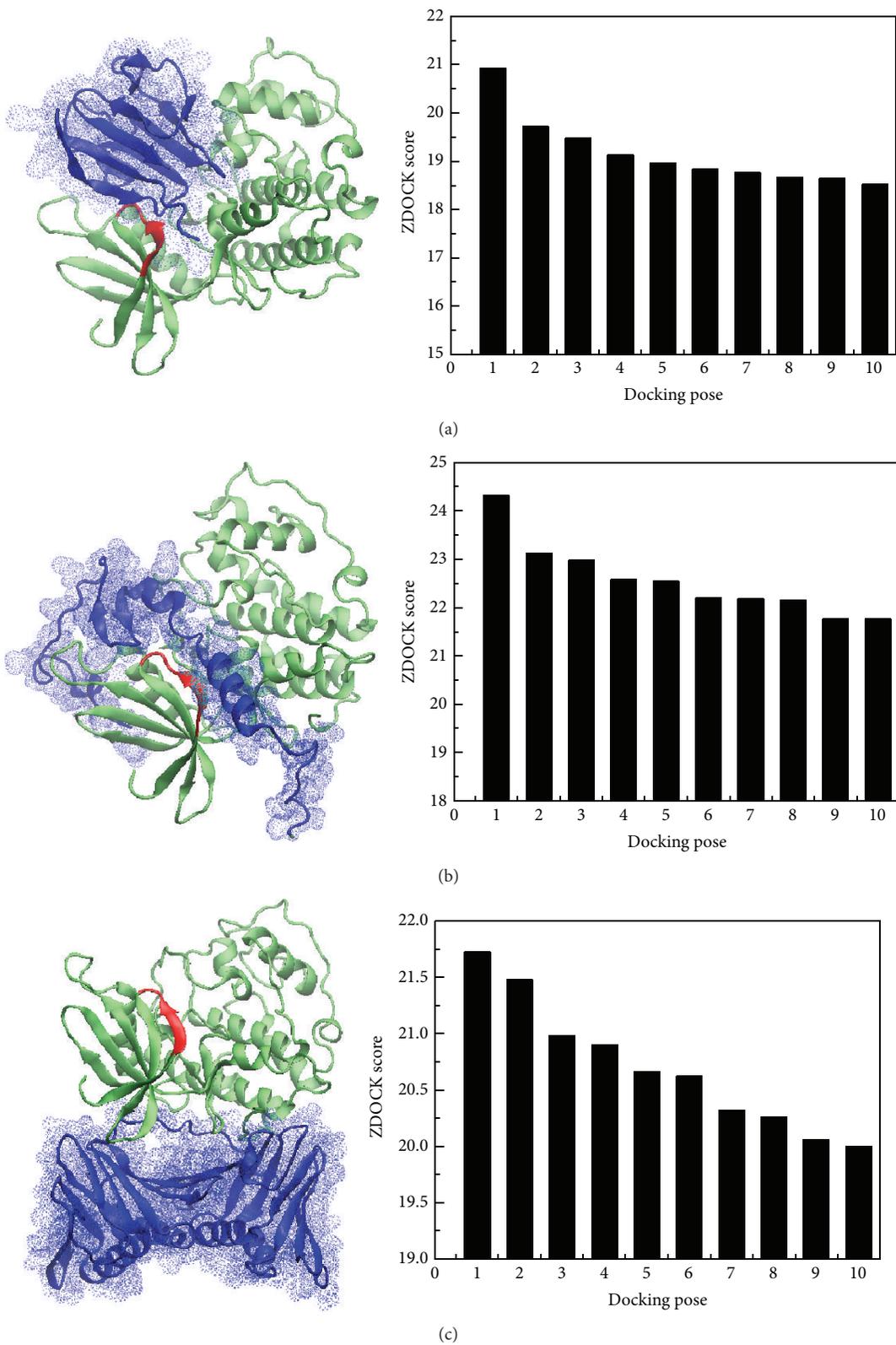


FIGURE 7: The best docking poses of CDK1 (green) with target protein (blue): (a) Ki-67, (b) p27, and (c) PCNA. The top ten protein-protein complexes with ZDOCK scores were generated by ZDOCK program. The highest ZDOCK score of Ki-67, p27, and PCNA are 20.92, 21.72, and 24.32, respectively. The key binding residues of Ki-67 and p27 are colored in red, and the key residues include Gly9, Ser10, Ile11, Leu12, Lys13, Lys14, and Val15.

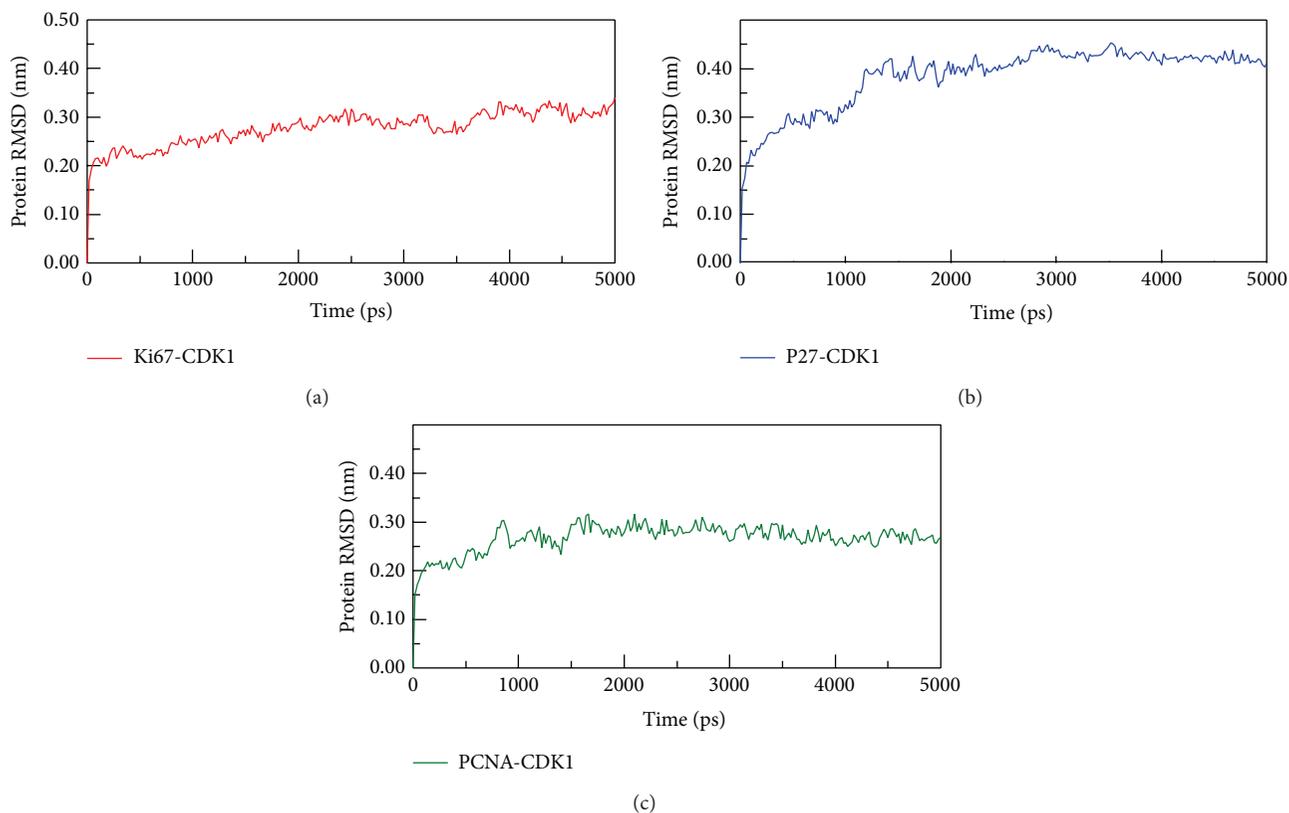


FIGURE 8: The RMSD analysis of all atoms of CDK1 complexes with (a) Ki-67, (b) p27, and (c) PCNA during 5000 ps simulation times. All complexes tend to stable fluctuation after simulation time of 1000 ps.

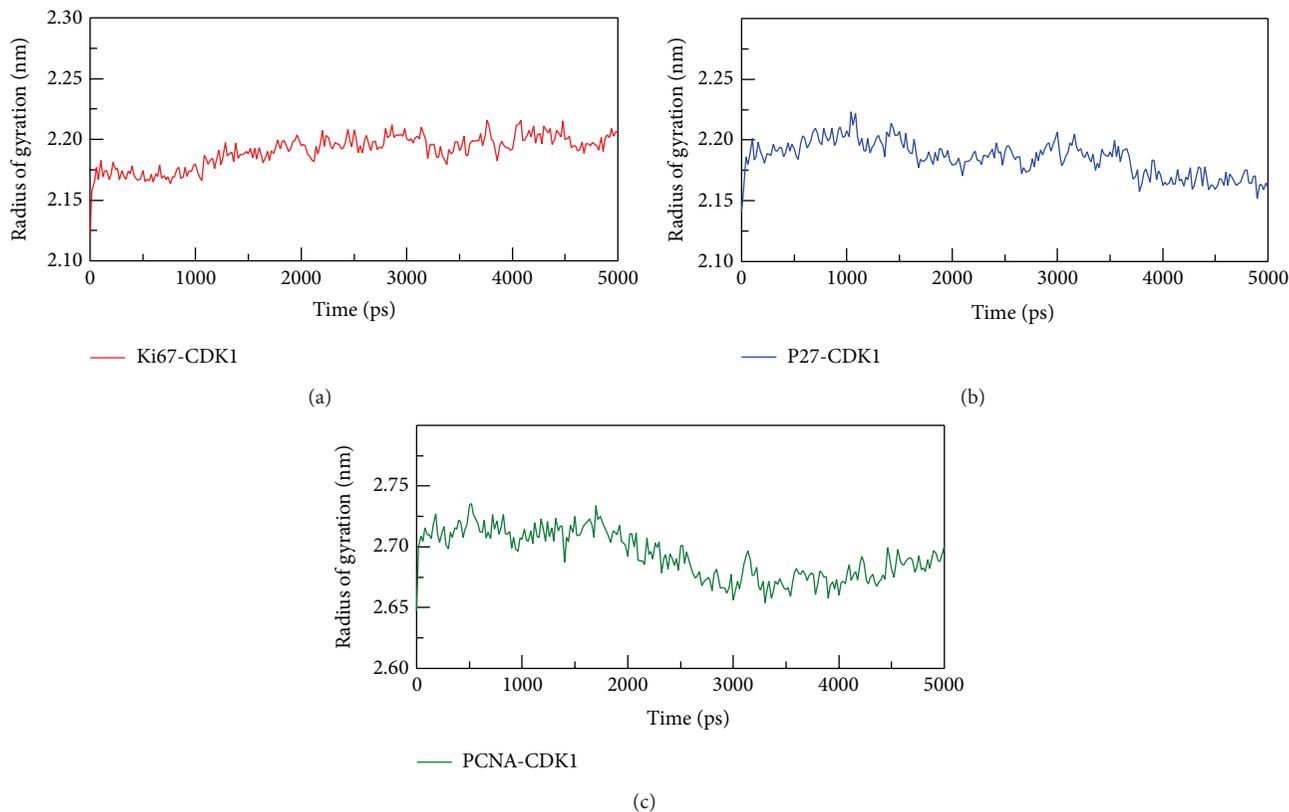


FIGURE 9: The radius of gyration of CDK1 complexes with (a) Ki-67, (b) p27, and (c) PCNA during 5000 ps simulation times. The low values of radius of gyration indicate the compacted complexes between two protein structures.

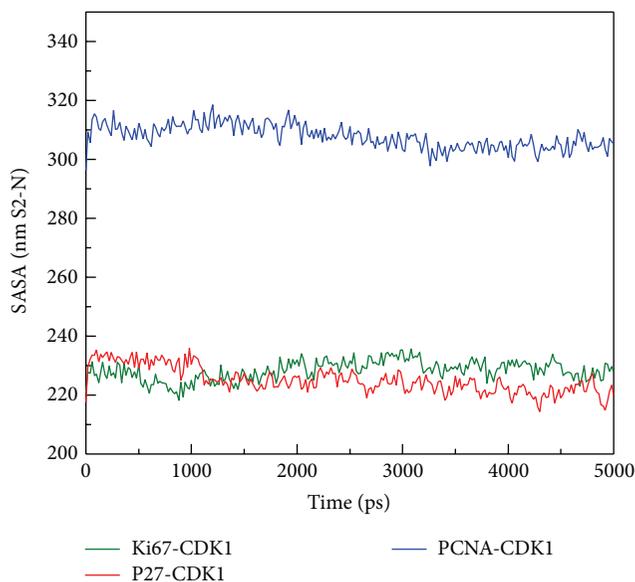


FIGURE 10: The SASA (area of solvent) analysis of all complexes conformation for hydrophobic definition during 5000 ps; the stable fluctuation indicated no distinct change between protein-protein interactions.

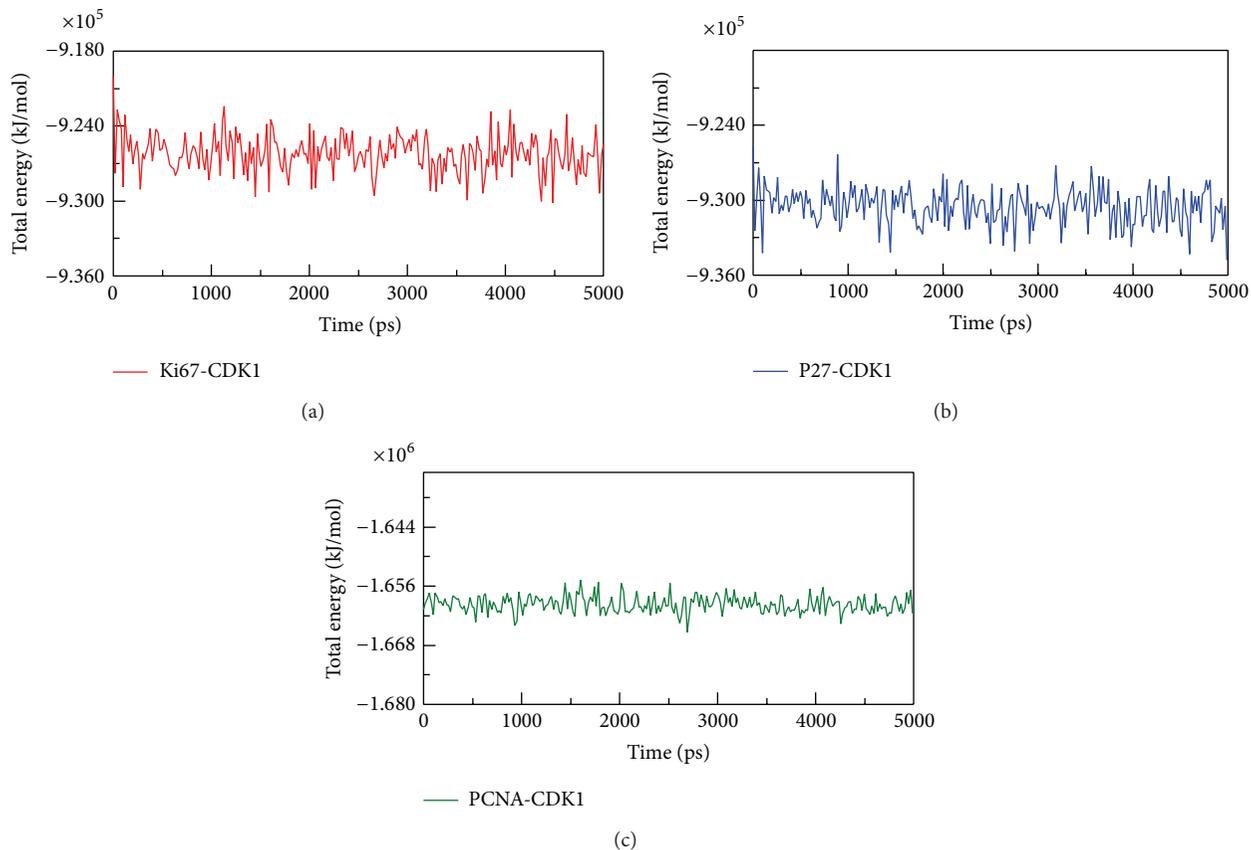
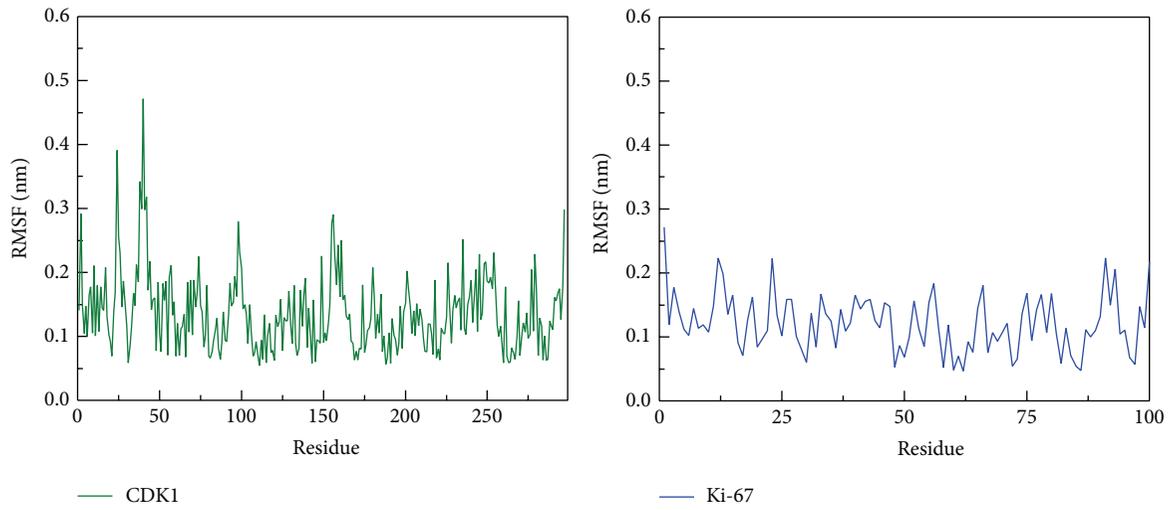
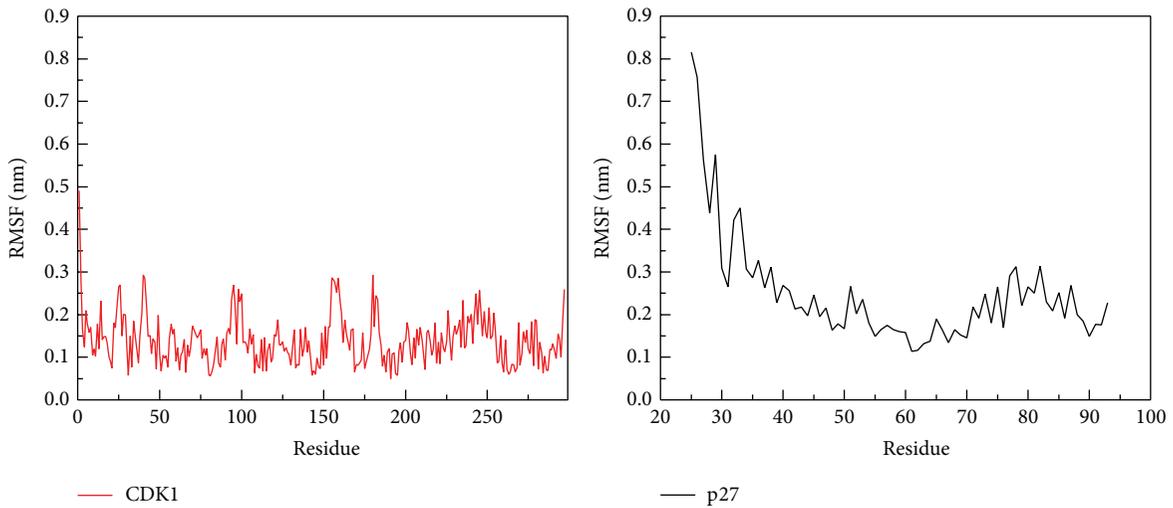


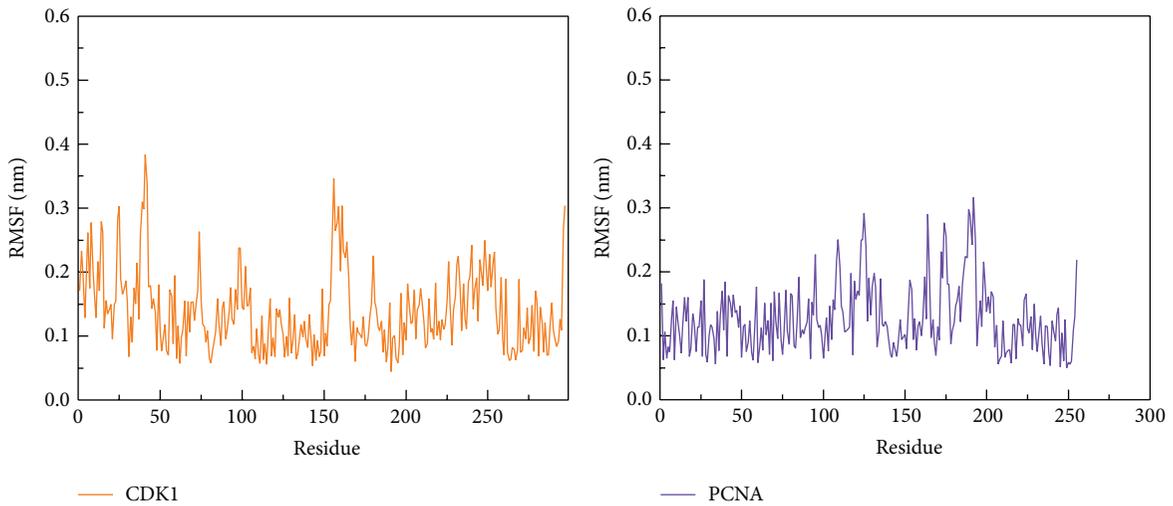
FIGURE 11: Total energy calculation of MD systems of (a) Ki-67, (b) p27 and (c) PCNA during 5000 ps simulation times, each of average fluctuations are  $-9.27 \times 10^5$ ,  $-9.30 \times 10^5$ , and  $-1.66 \times 10^6$ , respectively.



(a)



(b)



(c)

FIGURE 12: RMSF analysis of protein residues on (a) CDK1 and Ki-67, (b) CDK1 and p27, and (c) CDK1 and PCNA during simulation time of 5000 ps. The high value of RMSF fluctuation denotes strong variation of protein structure over all MD simulation.

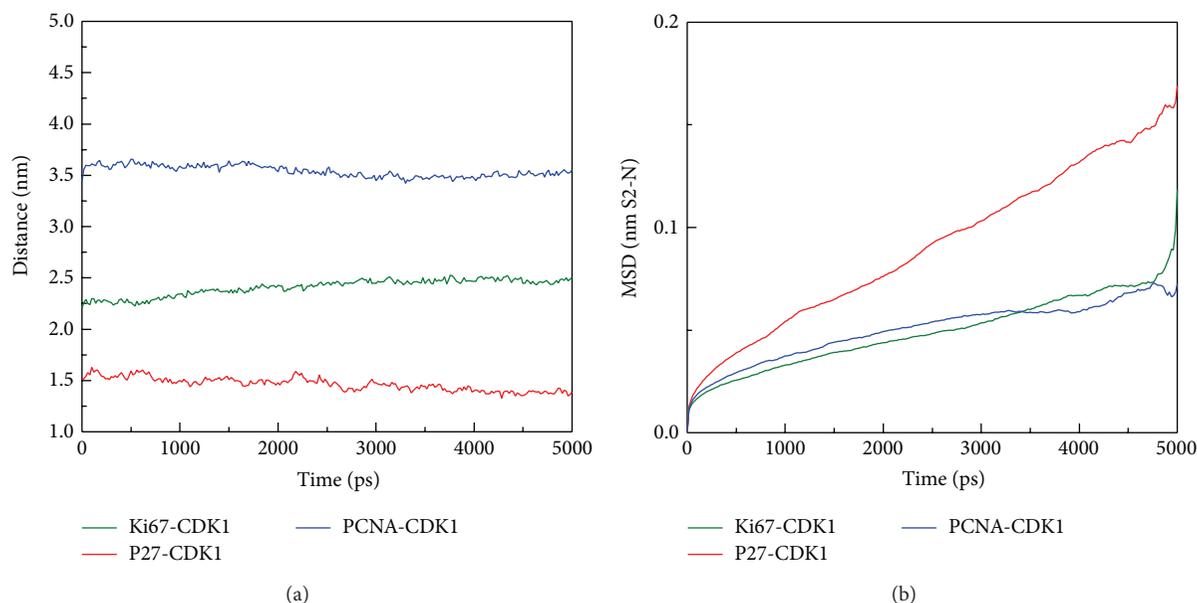


FIGURE 13: The migration analysis of complexes during simulation time of 5000 ps: (a) the distance between CDK1 and the docked proteins over all MD time and (b) the mean square displacement (MSD) analysis for all CDK1 complexes; the high value of MSD denotes the high distance of protein migration from initial position.

energy calculation of MD systems for each of the three complexes during 5000 ps simulation times, they all had stable energy variation in the range of  $-9.27 \times 10^5$ ,  $-9.30 \times 10^5$ , and  $-1.66 \times 10^6$ , respectively (Figure 11). The result of energy analysis reveals that all simulation systems are stable during 5000 ps. In residues fluctuation analysis, we measure RMSF value of all residues for the three complexes (Figure 12). In RMSF validation of CDK1-Ki67 complex, there are significant fluctuations observed on residues from 38 to 43 of CDK1, which has great change during the MD simulation (Figure 12(a)). There was also significant variation observed in the residues from 25 to 40 on p27 protein structure during the MD simulation which denotes the great changes in this region. On the contrary, there is less RMSF variation in the residues from 100 to 200 on the PCNA protein structure during the MD simulation. Therefore, there was fewer RMSF change during MD simulation in PCNA than Ki-67 and p27.

However, the migration analysis revealed that PCNA had great distance between CDK1 compared to Ki-67 and p27 during 5000 ps MD simulation survey. Although we found the largest distance between PCNA and CDK1 (Figure 13(a)), but their stable binding was shown in Figure 12 by RMSF analysis. In addition, the mean square displacement (MSD) of PCNA has smaller migration (Figure 13(b)). We further analyze the key residues of CDK1 for Ki-67 and p27 binding. The key binding residue includes Gly9, Ser10, Ile11, Leu12, Lys13, Lys14, and Val15 on Ki67-CDK1 protein structure dihedrals angle during simulation time of 5000 ps. All of the binding residues are stable during the whole MD simulation. However, there were only stable dihedrals angles in Gly9, Ser10, Leu12, and Val15 on p27-CDK1 protein complexes over

all MD simulation time. Finally, we found that all the key binding residues including Gly9, Ser10, Ile11, Leu12, Lys13, Lys14, and Val15 on PCNA-CDK1 protein structure had stable binding dihedrals angle during the whole MD simulation; no greater changes were found during the process of PCNA binding to CDK1 (Figure 14). The Cluster analyses were used to select the representative structure among all MD frames. For snapshot comparison assay, the represented structure selected from the last clustering groups for all MD frames of CDK1 complexes of Ki67, P27, and PCNA displaced at 4860 ps, 2740 ps, and 3880 ps, respectively, during simulation time of 5000 ps (Figure 15). Snapshot comparison study for CDK1 and target proteins for Ki-67, p27, and PCNA are showed in Figure 16. We found that the residues of CDK1 from 38 to 43 on CDK1 are getting close to Ki-67 from 0 ps to 4860 ps (Figure 16(a)). The result is correlated to RMSF analysis due to the high fluctuations on residues from 38 to 43 of CDK1 binds for Ki-67. For CDK1-p27 snapshot analysis, p27 is getting close to CDK1 from 0 ps to 2740 ps. The findings are also correlated to RMSF analysis because of high variations on residues from 25 to 40 of p27 binds for CDK1 (Figure 16(b)). We could only find the small change of one PCNA (blue) loop move away from CDK1 (green) at 3880 ps at residues from 100 to 200 on PCNA and CDK1 interaction (Figure 16(c)). This result is also correlated to RMSF analysis for small fluctuations on residues from 100 to 200 of PCNA binds for CDK1.

To sum up the results in our study and literature reviews, the molecular mechanism of Ki-67, p27, and PCNA interacting with CDK1 for cells proliferation and malignant transformation in patients with sinonasal IP is showed in Figure 17.

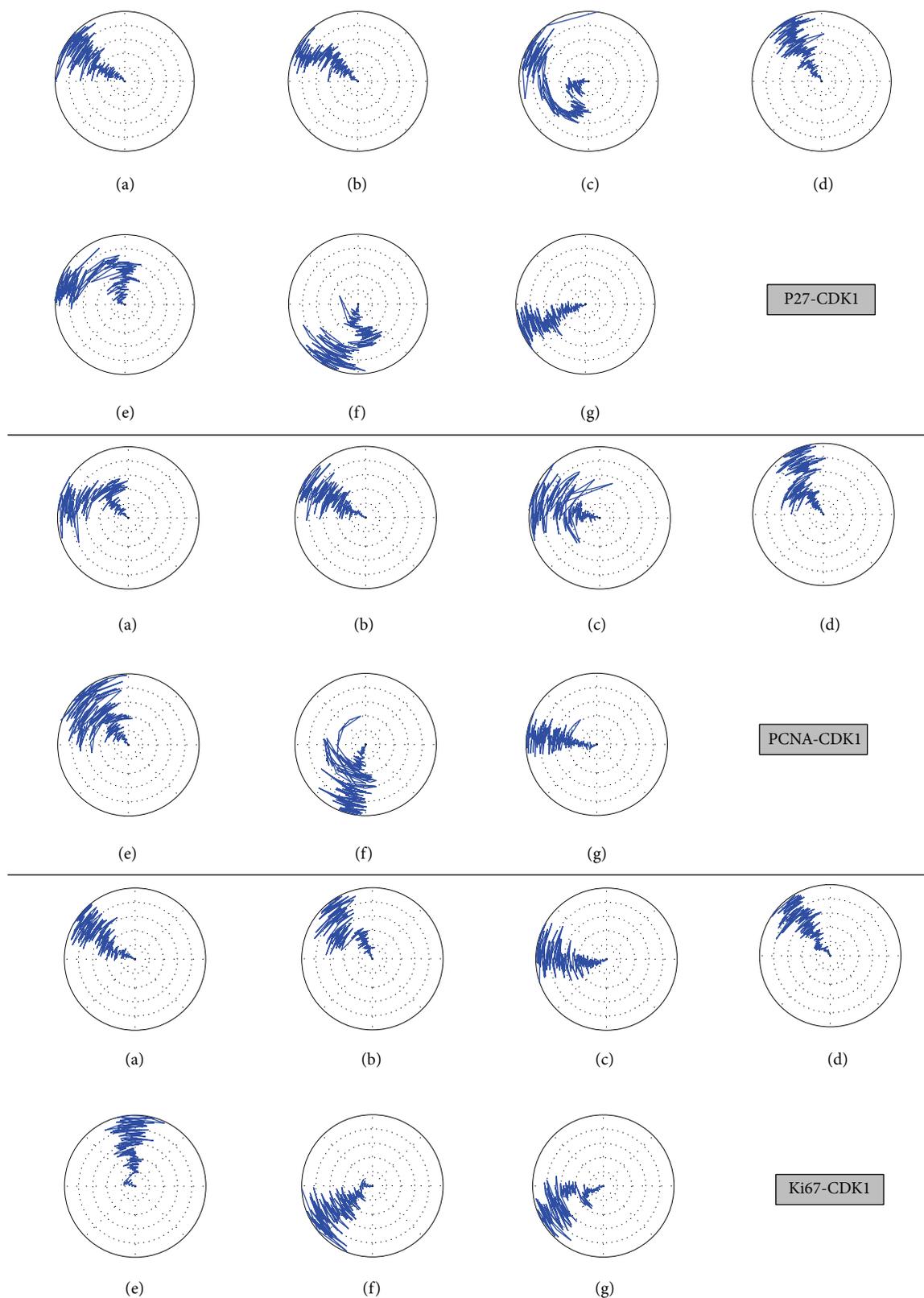


FIGURE 14: The dihedral angle of key binding residues: (a) Gly9, (b) Ser10, (c) Ile11, (d) Leu12, (e) Lys13, (f) Lys14, and (g) Val15 on CDK1 protein structure during simulation time of 5000 ps. The dihedral angles were calculated for CDK1 complexes with target binding proteins: Ki-67, p27, and PCNA over all MD simulation time.

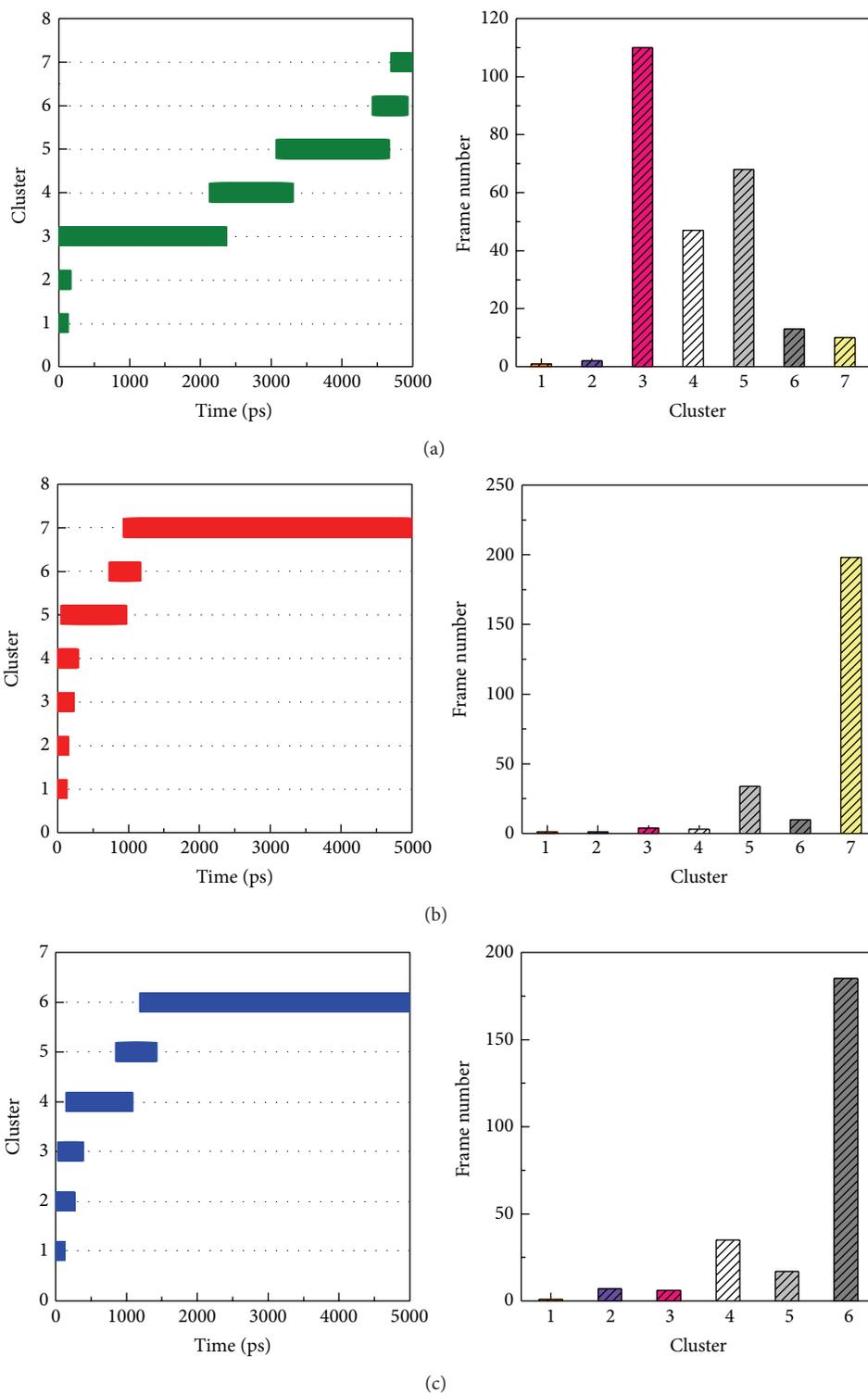


FIGURE 15: Cluster analyses of all CDK1 and the binding proteins: (a) Ki-67, (b) p27, and (c) PCNA during simulation time of 5000 ps; the represented structures were selected from the last clustering groups for all MD frames of CDK1 complexes. The represented structure of Ki-67, p27, and PCNA complex displaced at 4860 ps, 2740 ps, and 3880 ps, respectively.

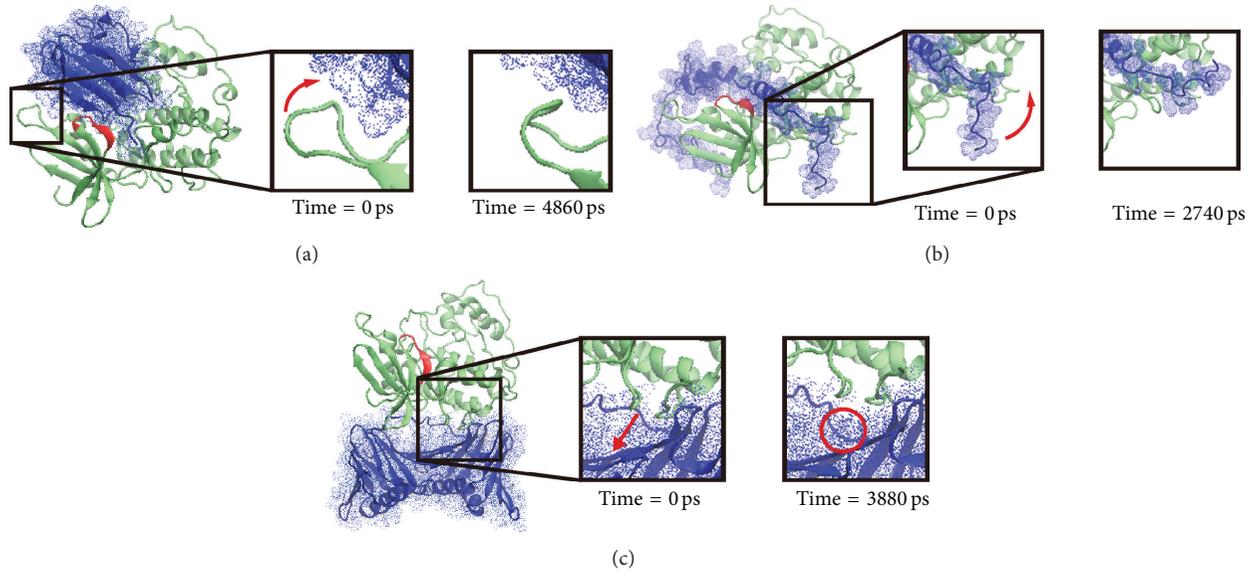


FIGURE 16: Comparison between the initial snapshot (0 ps) and represented conformation for all CDK1 complexes. (a) One of CDK1 (green) loop moved approach to Ki-67 (blue) at 4860 ps. (b) The protein structure of p27 (blue) bound to CDK1 (green) more tightly at 2740 ps. (c) One of PCNA (blue) loop moved away from CDK1 (green) at 3880 ps.

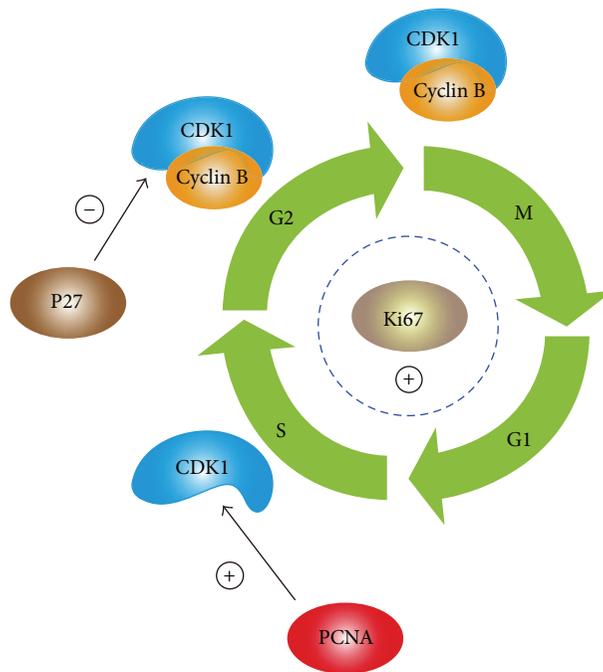


FIGURE 17: The molecular mechanism of Ki-67, p27, and PCNA in cell cycle.

#### 4. Discussion

Although the inverted papillomas rarely occurred in the sinonasal tract, but the easy recurrent nature and malignant transformation often troubles patients and physicians. Life-long regular follow-up is needed for early detection of recur-

rence or malignant change and could lead to better disease control for IPs patients [9, 20].

The Ki-67 promoted the initiation of G1 phase from G0 in IPs cells, and it persisted expression for cell cycle in proliferation phase. Ki-67 is a protein, which affects cell cycle in the proliferation phase from G1-S-G2-M except the

G0 phase [8, 11, 14]. In the literature reporting the p53, p63, p21, and p27 mutations induced sinonasal IP with SCC transformation, there were still debates [6, 11, 14]. Ki-67 is recently reported to the occurrence of sinonasal neoplasm [21]; the aggressive tumor behavior correlated to higher Ki-67 index and caused nasal epithelium to severe dysplasia and even squamous cell carcinoma [22]. The elevated Ki-67 could even be found in the IPs synchronously contained squamous cell carcinomas. Ki-67 will also affect p21, p27, and CDKs [8, 11, 12, 14]. However, we could not conclude whether the decreased p27 is caused by elevated expression of Ki-67 in sinonasal IPs tissues. Further study is warrant to elucidate the relationship between Ki-67 and P27.

The tumor suppression function of p53 is related to many cancers reported by Katori et al. and Gujrathi et al. [22, 23]. They suggested that testing for p53 may help to screen out papilloma lesions with a potential for dysplasia or carcinoma; however, we did not find it significant in multivariate analysis. This is because of the limited patient numbers in our study. However, not only p53 but also p63 is elevated expressed in IPs with malignant transformation [11].

Increased proliferative activity with elevated Ki-67 expression of tumor cells were reported as an important prognostic marker in many human tumors; it is also important in IPs recurrence and cancerization. Especially, Ki-67 suspect directly affects the cell cycle during proliferation phase and could interact with p53, p21, and p27 tumor suppressor genes and modulates cell cycle by affecting the checkpoint of G1 phase [8, 24, 25]. The Ki-67 recently found to have interaction between CDK1 in nature structure and molecular biology [26] and CDK1 partake the main role in cell proliferation and even malignant transformation [10]. We suspect that the Ki-67 not only initiated the IPs cells entrance into G1 phase of cell cycle but also caused malignant transformation in cell cores by affecting the CDK1.

The clinical roles of p21 and p27 to the head and neck SCC cancerization are still debates, but we found that the lower level of p27 also revealed in sinonasal IPs with malignant transformation in our study. Although there were few studies of p21 and p27 reporting the correlation to human IPs with cancerization, the debates still remained. Some were with Oncel et al. [11, 27] and some were against [25, 28] observed.

Besides Ki-67, we also found the PCNA, a predictor for IPs malignant transformation in the collaboration with CDK1. In our IPs with malignant transformation, both PCNA and Ki-67 were elevated by IHC stains.

As elevated PCNA expression showed in IP with malignant transformation from our survey, we suspected the PCNA to be important factor to induce cancerization for patients with sinonasal IP. Recently, it is also found the elevated PCNA in IP compared to the sinonasal polyps by Mumbuc et al. [3]. The PCNA could interact with CDK1 and promote the cell entry to the cell cycle to the consequent proliferation and cancerization.

Finally, we survey the PLUNC to the sinonasal IPs with cancerization since it is frequently reported to cause the nasopharyngeal cancer formation. In our prior study, the PLUNC expression was decreased in pseudomonas sinusitis in chronic disease status [29]. The PLUNC was also correlated

to the chronic rhinosinusitis with multiple bacteria colonization [17]. And, besides, the PLUNC was also supposed to have anti-infection and antibiofilm function [30, 31]. Because it was supposed to have anticancerization of nasopharyngeal carcinoma [32], but we found no correlation of sinonasal IPs with SCC transformation. On the contrary, we could only find the elevated PLUNC expression in patients with sinonasal IPs with multiple recurrences and revision sinus surgery. The further mechanism and reasons for elevated PLUNC expression and multiple recurrences should be further surveyed in the future.

The Computer-Aided Drug design (CADD) can used to further investigate clinical or illness research, include illness research [33], risk factor studies [34], case reports, and molecular mechanism. In our docking and molecular dynamic results of PCNA, Ki-67, and p27 to CDK1, we found that all three of them had stable docking to the CDK1. And the p27 were supposed to have more stable interaction to the CDK1 to work as an inhibitor for sinonasal IPs with cancerization. The PCNA and Ki67 were considered as promoters and promote cell proliferation and cancerization in sinonasal IPs.

In conclusion, this is a first study showing that the Ki-67, PCNA, and p27 are all important in IPs recurrences and cancerization via CDK1. And we also are the first ones to find elevated PLUNC expression in multiple recurrence sinonasal IPs. However, further mechanism survey is still warranted in the future. The current computational studies are all compatible with our wet lab studies which make our study more trustable and could apply to the future treatment toward patients with such disease. Therefore, patients with sinonasal IP and expressed elevated Ki-67, PCNA, and decreased p27 should be considered to have higher possibilities of malignant transformation and should be followed more closely in clinical practice [35].

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

The research was supported by grants from the National Science Council of Taiwan (NSC101-2314-B-039-013-MY3, NSC102-2325-B039-001, NSC102-2221-E-468-027), from China Medical University (CMU102-BC-3), from Asia University (ASIA100-CMU-2, ASIA101-CMU-2, and 102-ASIA-07), and from China Medical University Hospital (DMR-102-001, DMR-103-025, DMR-103-058, DMR-103-001, and DMR-103-096). This study is also supported in part by Taiwan Department of Health Clinical Trial and Research Center of Excellence (DOH102-TD-B-111-004) and Taiwan Department of Health Cancer Research Center of Excellence (MOHW103-TD-B-111-03), and CMU under the aim for Top University Plan of the Ministry of Education, Taiwan.

## References

- [1] W. Lawson and Z. M. Patel, "The evolution of management for inverted papilloma: an analysis of 200 cases," *Otolaryngology—Head and Neck Surgery*, vol. 140, no. 3, pp. 330–335, 2009.
- [2] K. C. Chen, S. S. Chang, H. J. Huang, T. Lin, Y. Wu, and C. Y. Chen, "Three-in-one agonists for PPAR- $\alpha$ , PPAR- $\gamma$ , and PPAR- $\delta$  from traditional Chinese medicine," *Journal of Biomolecular Structure and Dynamics*, vol. 30, no. 6, pp. 662–683, 2012.
- [3] S. Mumbuc, M. Karakok, T. Baglam, E. Karatas, C. Durucu, and Y. Kibar, "Immunohistochemical analysis of PCNA, Ki67 and p53 in nasal polyposis and sinonasal inverted papillomas," *Journal of International Medical Research*, vol. 35, no. 2, pp. 237–241, 2007.
- [4] P. Strojjan, A. Ferlito, V. J. Lund et al., "Sinonasal inverted papilloma associated with malignancy: the role of human papillomavirus infection and its implications for radiotherapy," *Oral Oncology*, vol. 48, no. 3, pp. 216–218, 2012.
- [5] E. Giotakis, I. P. Gomatos, L. Alevizos et al., "Apoptotic and proliferative status in HPV (+) and HPV (-) inverted papilloma patients. Correlation with local recurrence and clinicopathological variables," *Pathology Research and Practice*, vol. 208, no. 6, pp. 338–343, 2012.
- [6] M. Saegusa, H. Nitta, M. Hashimura, and I. Okayasu, "Down-regulation of p27(Kip1) expression is correlated with increased cell proliferation but not expression of p21(waf1) and p53, and human papillomavirus infection in benign and malignant tumours of sinonasal regions," *Histopathology*, vol. 35, no. 1, pp. 55–64, 1999.
- [7] M. J. Schwerer, A. Sailer, K. Kraft, K. Baczako, and H. Maier, "Patterns of p21waf1/cip1 expression in non-papillomatous nasal mucosa, endophytic sinonasal papillomas, and associated carcinomas," *Journal of Clinical Pathology*, vol. 54, no. 11, pp. 871–876, 2001.
- [8] G. Altavilla, A. Staffieri, G. Busatto, A. Canesso, L. Giacomelli, and G. Marioni, "Expression of p53, p16INK4A, pRb, p21WAF1/CIP1, p27KIP1, cyclin D1, Ki-67 and HPV DNA in sinonasal endophytic Schneiderian (inverted) papilloma," *Acta Oto-Laryngologica*, vol. 129, no. 11, pp. 1242–1249, 2009.
- [9] G. Buiet, X. Montbarbon, B. Fleury et al., "Inverted papilloma with associated carcinoma of the nasal cavity and paranasal sinuses: treatment outcomes," *Acta Oto-Laryngologica*, vol. 132, no. 1, pp. 80–85, 2012.
- [10] D. Santamaría, C. Barrière, A. Cerqueira et al., "Cdk1 is sufficient to drive the mammalian cell cycle," *Nature*, vol. 448, no. 7155, pp. 811–815, 2007.
- [11] S. Oncel, T. Cosgul, A. Calli, C. Calli, and E. Pinar, "Evaluation of P53, P63, P21, P27, Ki-67 in Paranasal Sinus Squamous Cell Carcinoma and Inverted Papilloma," *Indian Journal of Otolaryngology and Head and Neck Surgery*, vol. 63, no. 2, pp. 172–177, 2011.
- [12] T. Masubuchi, Y. Tada, S. I. Maruya et al., "Clinicopathological significance of androgen receptor, HER2, Ki-67 and EGFR expressions in salivary duct carcinoma," *International Journal of Clinical Oncology*, 2014.
- [13] M. Ekholm, S. Beglerbegovic, D. Grabau et al., "Immunohistochemical assessment of Ki67 with antibodies SP6 and MIB1 in primary breast cancer: a comparison of prognostic value and reproducibility," *Histopathology*, 2014.
- [14] K. Hellman, D. Lindquist, C. Ranhem, E. Wilander, and S. Andersson, "Human papillomavirus, p16, and Ki-67 in relation to clinicopathological variables and survival in primary carcinoma of the vagina," *The British Journal of Cancer*, vol. 110, no. 6, pp. 1561–1570, 2014.
- [15] I. Y. Shin, N. Y. Sung, Y. S. Lee et al., "The expression of multiple proteins as prognostic factors in colorectal cancer: cathepsin D, p53, COX-2, epidermal growth factor receptor, C-erbB-2, and Ki-67," *Gut Liver*, vol. 8, no. 1, pp. 13–23, 2014.
- [16] Y. He, G. Zhou, Y. Zhai et al., "Association of PLUNC gene polymorphisms with susceptibility to nasopharyngeal carcinoma in a Chinese population," *Journal of Medical Genetics*, vol. 42, no. 2, pp. 172–176, 2005.
- [17] Y. A. Tsou, M. T. Peng, Y. F. Wu et al., "Decreased PLUNC expression in nasal polyps is associated with multibacterial colonization in chronic rhinosinusitis patients," *European Archives of Oto-Rhino-Laryngology*, vol. 271, no. 2, pp. 299–304, 2014.
- [18] Accelrys, *Discovery Studio Client v2.5*, Accelrys, Inc., San Diego, Calif, USA, 2009.
- [19] S. Pronk, S. Páll, R. Schulz et al., "GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit," *Bioinformatics*, vol. 29, no. 7, pp. 845–854, 2013.
- [20] V. Sciarretta, I. J. Fernandez, P. Farneti, and E. Pasquini, "Endoscopic and combined external-transnasal endoscopic approach for the treatment of inverted papilloma: analysis of 110 cases," *European Archives of Oto-Rhino-Laryngology*, 2013.
- [21] M. J. Schwerer, A. Sailer, K. Kraft, and H. Maier, "Cell proliferation and p27Kip1 expression in endophytic schneiderian papillomas," *Laryngoscope*, vol. 112, no. 5, pp. 852–857, 2002.
- [22] H. Katori, A. Nozawa, and M. Tsukuda, "Relationship between p21 and p53 expression, human papilloma virus infection and malignant transformation in sinonasal-inverted papilloma," *Clinical Oncology*, vol. 18, no. 4, pp. 300–305, 2006.
- [23] C. Gujrathi, I. Pathak, J. Freeman, and S. Asa, "Expression of p53 in inverted papilloma and malignancy associated with inverted papilloma," *Journal of Otolaryngology*, vol. 32, no. 1, pp. 48–50, 2003.
- [24] S. Gunia, F. Fritzsche, M. May, D. Liebe, and S. Koch, "Do different disorders in cell cycle regulation account for different biological behavior of urothelial and sinonasal inverted papillomas?" *Pathobiology*, vol. 75, no. 1, pp. 34–41, 2008.
- [25] A. Affolter, S. Helmbrecht, S. Finger, K. Hörmann, and K. Götte, "Altered expression of cell cycle regulators p21, p27, and p53 in tumors of salivary glands and paranasal sinuses," *Oncology Reports*, vol. 13, no. 6, pp. 1089–1094, 2005.
- [26] I. J. Byeon, H. Li, H. Song, A. M. Gronenborn, and M. D. Tsai, "Sequential phosphorylation and multisite interactions characterize specific target recognition by the FHA domain of Ki67," *Nature Structural and Molecular Biology*, vol. 12, no. 11, pp. 987–993, 2005.
- [27] H. R. Choi, S. A. Tucker, Z. Huang et al., "Differential expressions of cyclin-dependent kinase inhibitors (p27 and p21) and their relation to p53 and Ki-67 in oral squamous tumorigenesis," *International Journal of Oncology*, vol. 22, no. 2, pp. 409–414, 2003.
- [28] N. Keleş, B. Erdamar, A. Kaur, and K. Değer, "p21, p53, and p27 Kip1 alterations in benign and malignant tumors of sinonasal epithelium," *Otolaryngology—Head and Neck Surgery*, vol. 129, no. 1, pp. 77–84, 2003.
- [29] Y. A. Tsou, C. M. Chen, T. C. Lin et al., "Decreased SPLUNC1 expression is associated with Pseudomonas infection in surgically treated chronic rhinosinusitis patients who may require repeated sinus surgery," *Laryngoscope*, vol. 123, no. 4, pp. 845–851, 2013.

- [30] A. Balakrishnan, S. A. Marathe, M. Joglekar, and D. Chakravorty, "Bactericidal/permeability increasing protein: a multifaceted protein with functions beyond LPS neutralization," *Innate Immunity*, vol. 19, no. 4, pp. 339–347, 2013.
- [31] L. Gakhar, J. A. Bartlett, J. Penterman et al., "PLUNC is a novel airway surfactant protein with anti-biofilm activity," *PLoS ONE*, vol. 5, no. 2, Article ID e9098, 2010.
- [32] Y. Zhou, Z. Zeng, W. Zhang et al., "Identification of candidate molecular markers of nasopharyngeal carcinoma by microarray analysis of subtracted cDNA libraries constructed by suppression subtractive hybridization," *European Journal of Cancer Prevention*, vol. 17, no. 6, pp. 561–571, 2008.
- [33] S.-C. Hsu, J.-H. Lin, S.-W. Weng et al., "Crude extract of *Rheum palmatum* inhibits migration and invasion of U-2 OS human osteosarcoma cells by suppression of matrix metalloproteinase-2 and -9," *BioMedicine*, vol. 3, no. 3, pp. 120–129, 2013.
- [34] W.-Y. Lin, H.-P. Liu, J.-S. Chang et al., "Genetic variations within the PSORS1 region affect Kawasaki disease development and coronary artery aneurysm formation," *BioMedicine*, vol. 3, no. 2, pp. 73–81, 2013.
- [35] W.-L. Liao and F.-J. Tsai, "Personalized medicine: a paradigm shift in healthcare," *BioMedicine*, vol. 3, no. 2, pp. 66–72, 2013.