

Research Article

Recognition of Errors in the Refinement and Validation of Three-Dimensional Structures of AC1 Proteins of Begomovirus Strains by Using ProSA-Web

Rajneesh Prajapat, Avinash Marwal, and R. K. Gaur

Department of Science, Faculty of Arts, Science and Commerce, Mody Institute of Technology and Science, Lakshmangarh, Sikar, Rajasthan 332311, India

Correspondence should be addressed to R. K. Gaur; gaurrajarshi@hotmail.com

Received 22 June 2013; Revised 2 October 2013; Accepted 2 October 2013; Published 2 January 2014

Academic Editor: Sílvia Bofill-Mas

Copyright © 2014 Rajneesh Prajapat 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 structural model of begomovirus AC1 protein is useful for understanding biological function at molecular level and docking study. For this study we have used the ProSA program (Protein Structure Analysis) tool to establish the structure prediction and modeling of protein. This tool was used for refinement and validation of experimental protein structures. Potential problems of protein structures based on energy plots are easily seen by ProSA and are displayed in a three-dimensional manner. In the present study we have selected different AC1 proteins of begomovirus strains (YP_003288785, YP_002004579, and YP_003288773) for structural analysis and display of energy plots that highlight potential problems spotted in protein structures. The 3D models of Rep proteins with recognized errors can be effectively used for *in silico* docking study for development of potential ligand molecules against begomovirus infection.

1. Introduction

Geminiviruses were recognized in 1978 by the International Committee on the Taxonomy of viruses on the basis of their unique virion morphology and possession of ssDNA as their genomic material [1, 2]. *Geminiviridae* is one of the largest plant virus family; its members have a circular, single-stranded DNA (ssDNA) genome of approximately 2.7–5.2 kb encapsulated within twinned (geminate) icosahedral virions. The protein coat of geminiviridae consists of one type protein molecule of about 28 kD molecular weight. Based on their genome arrangement and biological properties, geminiviruses are classified into one of four genera: Mastrevirus, Curtovirus, Topocovirus, and Begomovirus [3].

Begomoviruses, currently hold 200 species [4] and contain dicotyledonous infecting whitefly transmitted viruses in the family Geminiviridae, have either bipartite genomes (DNA-A and DNA-B) or monopartite genomes resembling DNA-A. DNA-A typically has six open reading frames (ORFs): AV1/V1 (coat protein, CP) and AV2/V2 (AV2/V2

protein) on the virion-sense strand and AC1/C1 (replication initiation protein, Rep), AC2/C2 (transcriptional activator, TrAP), AC3/C3 (replication enhancer, REn), and AC4/C4 (AC4/C4 protein) on the complementary-sense strand. DNA-B has two ORFs, encoding movement proteins: BV1 (nuclear shuttle protein, NSP) on the virus-sense strand and BC1 (movement protein, MP) on the complementary-sense strand [5].

Computational methods can be applied for the prediction of unknown structures of experimental and theoretical models of virus proteins [6, 7], but the problem in structural biology is the recognition of errors in experimental and theoretical models of protein structures. The ProSA tool (<https://prosa.services.came.sbg.ac.at/>) verifies the three-dimensional experimental and the theoretical models of protein structures that have prospective errors.

The application of computational methods [8, 9] and server (e.g., NAR web server) for the prediction of unknown structures adds a plethora of structural models [10, 11] to the study. The analysis of protein structures is generally

TABLE 1: Results summary of the Ramachandran plot.

S. no.	Protein accession no.	Description	Virus strain	Residues in favoured region	Residues in allowed region	Residues in outlier region
1	YP_003288785	Rep	<i>Sweet potato leaf curl Lanzarote virus</i>	87.3%	11.4%	1.3%
2	YP_002004579	Rep	<i>Sweet potato leaf curl Spain virus</i>	79.7%	12.7%	7.6%
3	YP_003288773	Rep	<i>Sweet potato leaf curl Canary virus</i>	85.5%	12.8%	1.7%

a difficult and cumbersome exercise. The new service presented here is a straightforward and easy to use extension of the classic ProSA program, which exploits the advantages of interactive web-based applications for the display of scores and energy plots that highlight potential problems spotted in protein structures. To check 3D models of protein structures for potential errors, ProSA [12] is a widely used tool. Its range of application includes error recognition in experimentally determined structures [13–15], theoretical models [16–19], and protein engineering [20, 21]. For *in silico* ligand designing to be an effective inhibitor, Rep protein of selected begomovirus strains (YP_003288785, YP_002004579, and YP_003288773), which is responsible for replication, was used. This is the highlight of this study.

2. Materials and Methods

For the present study different bioinformatics tools and databases were used for molecular modeling of Rep protein of begomovirus strains (YP_003288785, YP_002004579, and YP_003288773), for example, GenBank-NCBI, PDB (Protein Data Bank), UCLA-DOE, RAMPAGE server, and so forth. Rep proteins sequence of begomovirus strain (YP_003288785, YP_002004579, and YP_003288773) was retrieved in FASTA format from NCBI database for homology modeling. Homology modeling procedure was performed in four basic sequential steps: template selection, target template alignment, model construction, and model assessment, and ProSA tool was used for potential errors detection [22]. ProSA-web requires the atomic coordinates of the model to be evaluated. The *z*-score indicates overall model quality and measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations. *Z*-scores outside a range characteristic for native proteins indicate erroneous structures. In order to facilitate interpretation of the *z*-score of the specified protein, its particular value is displayed in a plot that contains the *z*-scores of all experimentally determined protein chains.

Sequences retrieved from NCBI:

```
>gi|262530246|ref|YP_003288785.1| Rep [Sweet
potato leaf curl Lanzarote virus]MPRAGRNF
IKAKNYFLTYPCSLTKEEALDQLLHLNTPTNKKFIKICR
ELHENGEPHLVLLQFEGNYQCTNRFFDLVSPSRSSHFH
PNIQRAKSSSDVKSYVDKDGDTIEWGEFQVDGRSARGGQ
TANDAAAEALNSGSKEAALQIIREKLPEKFIFQYHNLG
LDRIFFSPPPSVYSSPFSSSFNAVPDIISDWAENVMDS
ARPDRPISIVIEGPRIGKTVMWARS LGPHNYLCGHLDSL
PVYSNSAWYNVIDDVNPQYLKHFKEFMGAQKDWQSNCKY
```

```
KPVQIKGGIPTIFLCNPGEGSSFKLWLDKPEQGALKNWAT
ANAIFCDVQSPFVWQEEVSHSGATAHRGEGQEES
>gi|194271409|ref|YP_002004579.1| Rep [Sweet
potato leaf curl Spain virus]
MPRAGRNFNINAKNYFLTYPCSSISKEEALAQILNIPTAVN
KKFIKICRELHEDGQPHLVLLQFEGKFQCTNQRLFDLVS
QTRSAHFHPNIQRAKSSSDVKSYVDKDGDTLEWGEFQVDG
RSARGGQQTANDAAAEALNAGSKDAALQIIREKLPEKFIF
QYHNLVSNLDRIFSPPPSVYSSPFSSSFNNVPDIISDWA
AENVMDSAAARPERPISIVIEGPRSMGKTVMWARS LGPHNYL
CGHLDSLSPKVYSNSAWYNVIDDVNPQYLKHFKEFMGAQKD
WQSNCKYGKPVQIKGGIPTIFLCNPGEGSSFKLWLDKPEQ
EALKNWAVKNAVFCVDSPFWIQEEVSHSGTNRGGQEEP
EENS
>gi|262530241|ref|YP_003288773.1| REP [Sweet
potato leaf curl Canary virus]
MPRKQGFRVQAKNIFLTYPKCSLSKEQALEQLRATHCPSD
KLFIRVSQEKHQDQSLHLHLVLIQFKGAEFKNPRHFDLHH
PHNSSQFHPNFQAAKSSSDVKSYIEKDGDYLDWGEFQIDG
RSARGGQQTANDAAAEALNAGSKEAALQIIREKLPEKYIF
QYHNLVSNLDRIFSPPPAVYCSPFSSSFNNVPDIISDWA
AENVMDSAAARPDRIPIVIEGPRIGKTVMWARS LGPHNYL
CGHLDSLSPKVYSNSAWYNVIDDVNPQYLKHFKEFMGAQKD
WQSNCKYGKPVQIKGGIPTIFLCNPGEGSSFKLWLDKPE
QEALKNWALKNAIFCDVQSPFVWQEEVSGAGAITRSSEE
GQEES
```

Procheck [23] outcomes are displayed in the form of profile search and Ramachandran plots. The models were checked with Verified-3D server [24] and Ramachandran plot at RAMPAGE [25] server. PDB files of Rep protein were used for evaluation through ProSA-web (<https://prosa.services.came.sbg.ac.at/prosa.php/>) that requires the atomic coordinates' file of protein.

3. Results and Discussion

A particular intention of the ProSA-web application is to encourage structure depositors to validate their structures before they are submitted to PDB and to use the tool in early stages of structure determination and refinement. Rep proteins 3D models with recognized errors were used for development of potential ligand molecules against begomovirus infection through docking process. A good quality Ramachandran plot has over 90% in the most favored regions [26] but the Ramachandran plot of YP_003288785.pdb has only 87.3% of residues in the most favoured regions. Therefore it is a near to good quality model (Table 1, Figure 1(a)). Similarly, the Ramachandran plot of YP_002004579 (Figure 1(b))

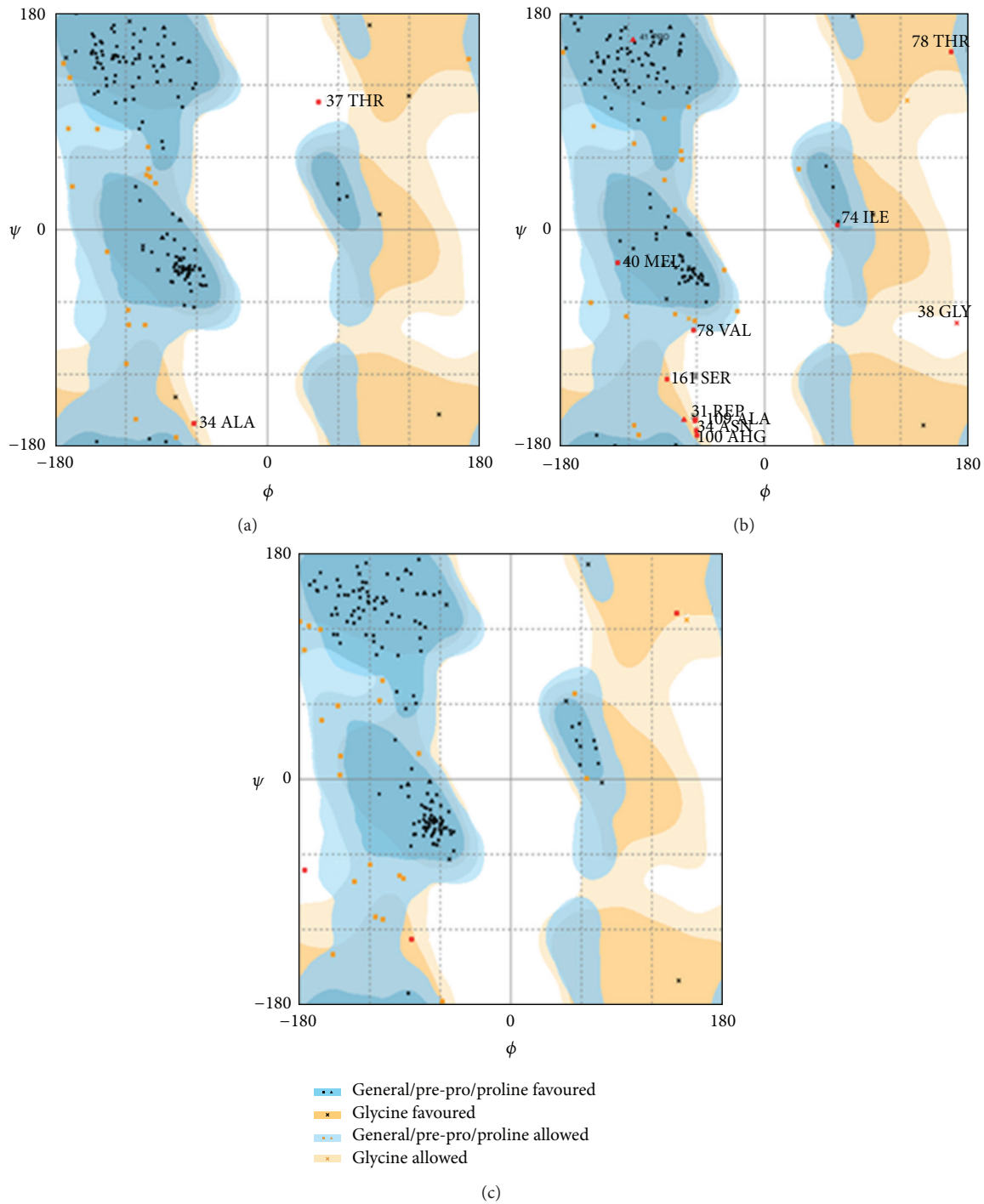


FIGURE 1: Ramachandran plot of Rep proteins of (a) *Sweet potato leaf curl Lanzarote virus* [YP_003288785], (b) *Sweet potato leaf curl Spain virus* (YP_002004579), and (c) *Sweet potato leaf curl Canary virus* (YP_003288773).

and YP_003288773 (Figure 1(c)), respectively, has 79.7% and 85.5% residues in the most favored regions. Figure 2 shows the results for a monomer of Rep proteins of *Sweet potato leaf curl Lanzarote virus*, *Sweet potato leaf curl Spain virus*, and *Sweet potato leaf curl Canary virus* [27].

The ProSA-web results indicate that Rep proteins have features characteristic for native structures. Figure 2(a) depicts the ProSA-web z-scores of all protein chains in PDB

(Table 2) determined by X-ray crystallography (light blue) or NMR spectroscopy (dark blue) with respect to their length [22]. The plot shows only chains with less than 1,000 residues and a z-score of 10. The z-scores of Rep proteins are highlighted as large dots.

Figure 2(b) shows the energy plot of Rep proteins. The energy plot shows the local model quality by plotting energies as a function of amino acid sequence position i . Residue

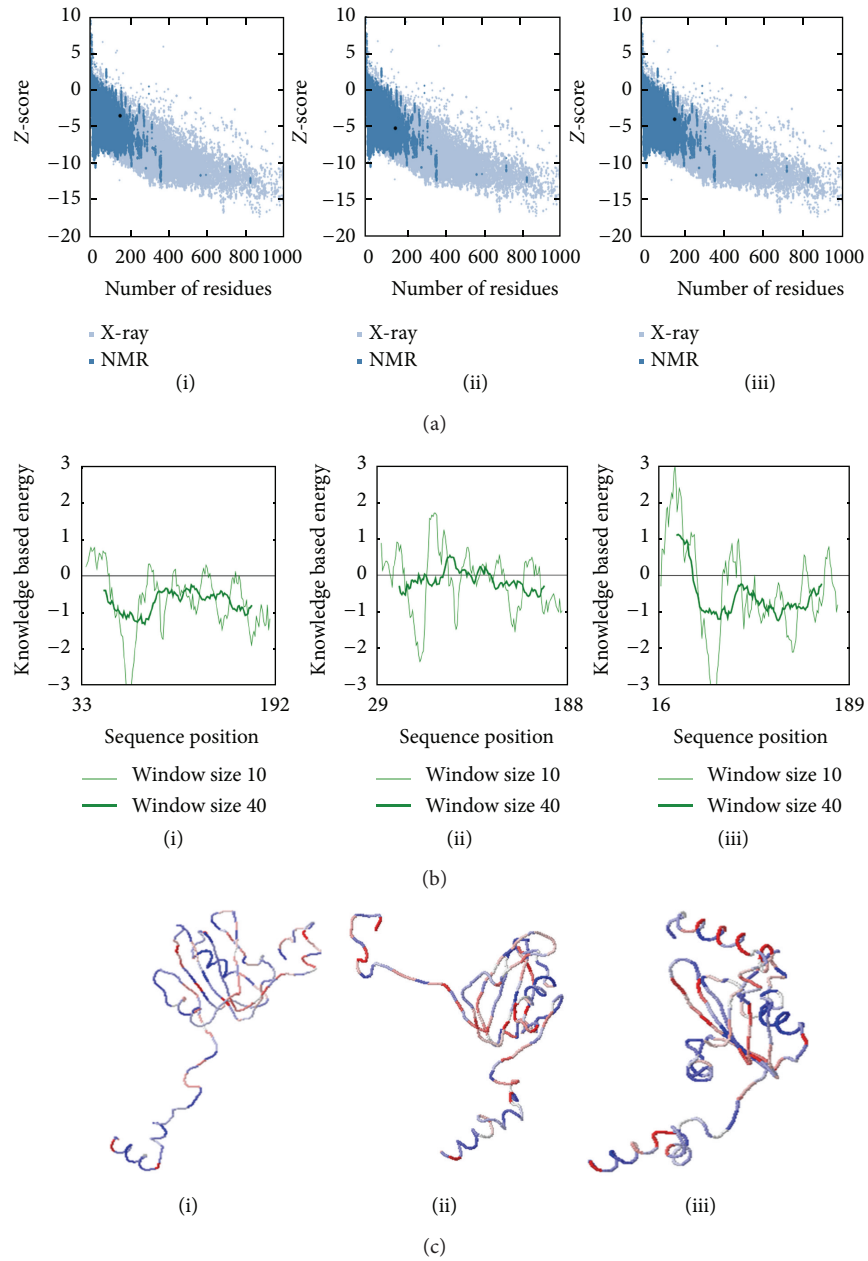


FIGURE 2: (a) ProSA-web z-scores of all protein chains: investigation of three Rep proteins structures of (i) *Sweet potato leaf curl Lanzarote virus*, (ii) *Sweet potato leaf curl Spain virus*, and (iii) *Sweet potato leaf curl Canary virus* using the ProSA-web service (YP_003288785, YP_002004579, and YP_003288773). (b) Energy plot of all three Rep proteins: investigation of three Rep proteins structures of (i) *Sweet potato leaf curl Lanzarote virus*, (ii) *Sweet potato leaf curl Spain virus*, and (iii) *Sweet potato leaf curl Canary virus* using the ProSA-web service (YP_003288785, YP_002004579, and YP_003288773). (c) Jmol Ca trace of Rep proteins: investigation of three Rep proteins structures of (i) *Sweet potato leaf curl Lanzarote virus*, (ii) *Sweet potato leaf curl Spain virus*, and (iii) *Sweet potato leaf curl Canary virus* using the ProSA-web service (YP_003288785, YP_002004579, and YP_003288773).

TABLE 2: Z-score of selected Rep proteins.

S. no.	Accession no.	Protein	Virus strain	Z-score
1	YP_003288785	Rep	<i>Sweet potato leaf curl Lanzarote virus</i>	-5.09
2	YP_002004579	Rep	<i>Sweet potato leaf curl Spain virus</i>	-3.33
3	YP_003288773	Rep	<i>Sweet potato leaf curl Canary virus</i>	-3.86

energies averaged over a sliding window are plotted as a function of the central residue in the window. A window size of 80 is used due to the large size of the Rep protein chain (default: 40). In general, positive values correspond to problematic or erroneous parts of a model. A plot of single residue energies usually contains large fluctuations and is of limited value for model evaluation. Hence the plot is smoothed by calculating the average energy over each 40-residue fragment si , $i + 39$, which is then assigned to the “central” residue of the fragment at position $i + 19$.

In order to further narrow down those regions in the model that contribute to a bad overall score, ProSA-web visualizes the 3D structure of the protein using the molecule viewer, Jmol. Figure 2(c) illustrates the Jmol Ca trace of Rep proteins. Residues are colored from blue to red in the order of increasing residue energy.

4. Conclusion

PDB files sometimes contain errors and generally remain unknown until the corresponding revisions are made available to the structural community. Hence, ProSA is a diagnostic tool that is based on the statistical analysis of all available protein structures. By using subsequent independent X-ray analysis, we studied Rep proteins of *Sweet potato leaf curl Lanzarote virus*, *Sweet potato leaf curl Spain virus*, and *Sweet potato leaf curl Canary virus* that are known to be incorrect, yielding a completely different conformation. The 3D models of Rep proteins with recognized errors can be effectively used for *in silico* docking study for development of potential ligand molecules against begomovirus infection.

Conflict of Interests

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

Authors' Contribution

Rajneesh Prajapat and Avinash Marwal contributed equally to the work.

Acknowledgments

The authors would like to acknowledge a vote of thanks to the Department of Biotechnology (DBT Project no. BT/PR13129/GBD/27/197/2009) and the Department of Science and Technology (DST Project no. SR/FT/LS-042/2009), India, for their financial support.

References

- [1] R. E. Matthews, “Classification and nomenclature of viruses,” *Intervirology*, vol. 12, no. 3–5, pp. 129–296, 1979.
- [2] R. M. Goodman, “Geminiviruses,” *Journal of General Virology*, vol. 54, pp. 9–21, 1981.
- [3] J. Stanley, D. M. Bisaro, R. W. Briddon et al., “Family Geminiviridae,” in *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*, pp. 301–326, Academic Press, 2005.
- [4] C. M. Fauquet, R. W. Briddon, J. K. Brown et al., “Geminivirus strain demarcation and nomenclature,” *Archives of Virology*, vol. 153, no. 4, pp. 783–821, 2008.
- [5] M. R. Rojas, C. Hagen, W. J. Lucas, and R. L. Gilbertson, “Exploiting chinks in the plant’s armor: evolution and emergence of geminiviruses,” *Annual Review of Phytopathology*, vol. 43, pp. 361–394, 2005.
- [6] A. Marwal, A. Sahu, R. Prajapat, D. K. Choudhary, and R. K. Gaur, “Molecular and recombinational characterization of begomovirus infecting an ornamental plant *Alternanthera sessilis*: a new host of *Tomato Leaf Curl Kerala Virus* reported in India,” *Science International*, vol. 1, pp. 51–56, 2013.
- [7] R. Prajapat, A. Marwal, A. Sahu, and R. K. Gaur, “Molecular in silico structure and recombination analysis of betasatellite in *Calotropis procera* associated with begomovirus,” *Archives of Phytopathology and Plant Protection*, vol. 45, pp. 1980–1990, 2012.
- [8] A. Marwal, A. Sahu, R. Prajapat, D. K. Choudhary, and R. K. Gaur, “First report of association of begomovirus with the leaf curl disease of a common weed *Datura innoxia*,” *Indian Journal of Virology*, vol. 23, pp. 83–84, 2012.
- [9] R. Prajapat, A. Marwal, A. Sahu, and R. K. Gaur, “First report of *Begomovirus* infecting *Sonchus asper* in India,” *Science International*, vol. 1, pp. 108–110, 2013.
- [10] J. A. Fox, S. McMillan, and B. F. F. Ouellette, “A compilation of molecular biology web servers: 2006 update on the Bioinformatics links directory,” *Nucleic Acids Research*, vol. 34, pp. W3–W5, 2006.
- [11] H. M. Berman, S. K. Burley, W. Chiu et al., “Outcome of a workshop on archiving structural models of biological macromolecules,” *Structure*, vol. 14, no. 8, pp. 1211–1217, 2006.
- [12] M. J. Sippl, “Recognition of errors in three-dimensional structures of proteins,” *Proteins*, vol. 17, no. 4, pp. 355–362, 1993.
- [13] L. Banci, I. Bertini, F. Cantini et al., “Solution structure and intermolecular interactions of the third metal-binding domain of ATP7A, the Menkes disease protein,” *The Journal of Biological Chemistry*, vol. 281, no. 39, pp. 29141–29147, 2006.
- [14] O. Llorca, M. Betti, J. M. González, A. Valencia, A. J. Márquez, and J. M. Valpuesta, “The three-dimensional structure of an eukaryotic glutamine synthetase: functional implications of its oligomeric structure,” *Journal of Structural Biology*, vol. 156, no. 3, pp. 469–479, 2006.
- [15] K. Teilum, J. C. Hoch, V. Goffin, S. Kinet, J. A. Martial, and B. B. Kragelund, “Solution structure of human prolactin,” *Journal of Molecular Biology*, vol. 351, no. 4, pp. 810–823, 2005.
- [16] D. Petrey and B. Honig, “Protein structure prediction: inroads to biology,” *Molecular Cell*, vol. 20, no. 6, pp. 811–819, 2005.
- [17] K. Ginalski, “Comparative modeling for protein structure prediction,” *Current Opinion in Structural Biology*, vol. 16, no. 2, pp. 172–177, 2006.
- [18] R. Panteri, A. Paiardini, and F. Keller, “A 3D model of Reelin subrepeat regions predicts Reelin binding to carbohydrates,” *Brain Research*, vol. 1116, no. 1, pp. 222–230, 2006.
- [19] J. Mansfeld, S. Gebauer, K. Dathe, and R. Ulbrich-Hofmann, “Secretory phospholipase A2 from *Arabidopsis thaliana*: insights into the three-dimensional structure and the amino acids involved in catalysis,” *Biochemistry*, vol. 45, no. 18, pp. 5687–5694, 2006.

- [20] M. K. Beissenhirtz, F. W. Scheller, M. S. Viezzoli, and F. Lisdat, "Engineered superoxide dismutase monomers for superoxide biosensor applications," *Analytical Chemistry*, vol. 78, no. 3, pp. 928–935, 2006.
- [21] M. Wiederstein and M. J. Sippl, "Protein sequence randomization: efficient estimation of protein stability using knowledge-based potentials," *Journal of Molecular Biology*, vol. 345, no. 5, pp. 1199–1212, 2005.
- [22] M. Wiederstein and M. J. Sippl, "ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins," *Nucleic Acids Research*, vol. 35, pp. W407–W410, 2007.
- [23] R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton, "PROCHECK: a program to check the stereochemical quality of protein structures," *Journal of Applied Crystallography*, vol. 26, pp. 283–291, 1993.
- [24] K. M. Goh, N. M. Mahadi, O. Hassan, R. N. Zaliha, R. A. Rahman, and R. M. Illias, "Molecular modeling of a predominant β -CGTase G1 and analysis of ionic interaction in CGTase," *Biotechnology*, vol. 7, no. 3, pp. 418–429, 2008.
- [25] S. C. Lovell, I. W. Davis, W. B. Arendall et al., "Structure validation by α geometry: ϕ, ψ and $C\beta$ deviation," *Proteins*, vol. 50, no. 3, pp. 437–450, 2003.
- [26] J. Xiao, Z. Li, M. Sun, Y. Zhang, and C. Sun, "Homology modeling and molecular dynamics study of GSK3/SHAGGY-like kinase," *Computational Biology and Chemistry*, vol. 28, no. 3, pp. 179–188, 2004.
- [27] G. Chang and C. B. Roth, "Structure of MsbA from *E. coli*: a homolog of the multidrug resistance ATP binding cassette (ABC) transporters," *Science*, vol. 293, no. 5536, pp. 1793–1800, 2001.

