

Research Article

The Evidence of Cooperative Binding of a Ligand to G4 DNA

A. G. Kudrev

Saint Petersburg State University, 7/9 Universitetskaya Nab., St. Petersburg 199034, Russia

Correspondence should be addressed to A. G. Kudrev; kudrevandrei@mail.ru

Received 11 April 2017; Accepted 24 August 2017; Published 18 October 2017

Academic Editor: Chih-Ching Huang

Copyright © 2017 A. G. Kudrev. 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.

Intrinsic constants of the ligand binding with G4 DNA (guanine-rich DNA sequence) using quantitative standards can be convenient providing the assessment for elucidating the possibility of such structures participation in biochemical processes. In the present communication, the hard + soft modelling approach to calculate intrinsic constants of a ligand binding with short DNA molecule, particularly such as G4 DNA, has been proposed. The suggested approach has focused upon the quantitative evaluating of a mutual influence between sites and between bound ligands. The cross-validation between a new hard + soft modelling and conventional stepwise complex formation algorithm has been conducted. A number of simulated examples will illustrate the methodology. The experimental mole-ratio titration of TMPyP4 by G4 DNA $[(CG_3)_2CGC(AG_3)_2G]$ has been reexamined. The $[(CG_3)_2CGC(AG_3)_2G]$ that folds from a G-rich sequence found in the promoter region of c-kit oncogene can be considered as a molecule with two equivalent mutually influence binding sites.

1. Introduction

In the past decade, a great research effort around four-stranded guanine-rich DNA sequences known as G4-quadruplexes (G4) [1] and small molecule ligand interactions was motivated as both innovative basis for the design of novel inhibitors of telomerase activity [2] and selective fluorescent probes [3]. Recent bioinformatics studies have shown that in the human genome there are approximately 350,000 sequences that can potentially form G4 structures [4]. The current interest in G4 interactions is based on the concept that small molecules binding would stabilize these structures' formation, which causes an inhibition of telomerase activity, which plays a role in the immortal growth of cancer cells. Such small molecules-ligands are promising as novel anticancer drug candidates. Information in huge number of publications would be convenient to systematize using the stability of the complexes formed by the components of the systems in question or by the ability of ligand to bind G4 molecules present in the solution. However, in the present, such systematization is difficult to perform. The lack of reliable, quantitative characteristics for G4-quadruplexes binding properties complicates the development of bioinformatics tools, particularly for the analysis of involvement G4

in the inhibition of telomerase activities *in vivo* and for the discovery of new anticancer agents. To address this issue, robust binding models are needed to investigate G4-ligand interactions.

The TMPyP4 (5,10,15,20-tetrakis-(1-methyl-4-pyridyl)-21H,23H-porphyrin) is one ligand shown to be capable of binding to the G4-structures. The interaction between TMPyP4 and G4 has been extensively studied since it can inhibit the activity of telomerase upon binding to human telomeric sequences [5]. However, its binding mode is still controversial [6]. This cationic porphyrin contains a large π -planar structure and cationic functional groups providing strong π - π stacking and electrostatic interactions with G-quartets and phosphate groups on the G4 DNA. The variations in the UV-Vis absorption and fluorescence spectra of the binding TMPyP4 ligand are normally the result of a change in the molecular conformation of the ligand in addition to the change in the environmental polarity and strong transition moment dipole coupling between the base pairs of DNA and the ligand. The cationic porphyrin G4 interactions have been studied extensively using a wide variety of spectroscopic techniques and isothermal titration calorimetry (several examples reported previously are summarized in Table S1; see Table S1 in the Supplementary Material available online

at <https://doi.org/10.1155/2017/6780521>). The UV-visible optical spectroscopy (UV-VIS), fluorescence resonance energy transfer (FRET), and circular dichroism (CD) via the spectral changes provide crucial information about ligand G4 DNA interactions, such as stoichiometry, stabilization strength, and especially the selectivity for G4 in comparison to ds DNA [7]. However, the detail of the binding mode remains to be completely elucidated. Several models of the binding mode have been proposed to date, especially for intramolecular quadruplexes from human telomeric sequence [8].

The nature of the G4 DNA binding site is unmistakably crucial for understanding the binding stoichiometry which is one of the most important hard modelling parameters. The TMPyP4 planar molecule favours π -stacking with G-quartets, including end-stacking and intercalation, but also groove binding, and direct coordination to bases or the phosphate backbone is possible [7]. The cationic properties of TMPyP4 allow electrostatic interactions with the negative backbone of G4 DNA. In addition, the coexisting binding modes have been recently revealed [9]. As the porphyrin concentration increased, TMPyP4 started to stack along the DNA. The central flat part and peripheral N-methylpyridyl moiety of TMPyP4 have a common structure for self-stacking and intercalation; hence the patterns of shifts in UV-Vis and CD spectra were almost identical in the mentioned binding modes [10]. For either intercalation or end-stacking the stoichiometric ratio of ligand/G4 was found to be 2:1 in most cases [11]. However for an intramolecular antiparallel-parallel hybrid structure [d(AG3(T2AG3)3)], the weak anti-cooperativity of TMPyP4 accommodation onto two from three nonidentical sites was detected [12]. A complete picture is thus obtained, revealing the coexisting binding modes of intercalation and surface binding [12]. Subsequently, we can more precisely predict binding mode if in addition to the patterns of shifts in UV-Vis and CD spectra the stoichiometry and intrinsic binding constants are determined from experimental data.

The evaluation of intrinsic binding constants which belong to the G4 DNA-ligand-binding site seems to be the first important step toward understanding the ligand exertion of influence on G4 DNA. These interactions are difficult to monitor because of complex structures and topologies of G4-quadruplexes, different sites and potential modes of binding, and the presence of other secondary nucleic acid structures. The problem is related to the high and very dynamic topological plasticity of quadruplexes, and it is very difficult to study the binding of high-affinity ligands without evoking the great topological consequences of the ligands binding that in turn strongly influence the second binding event. Quantitative description of interaction G4 DNA with TMPyP4 has been performed using either the direct fit method [13] or the independent binding sites model [14] or the conventional stepwise complex formation model [15]. All these methods do not take into account the mutual influence (MI) in the course of successive ligand binding. Reasonably in addition to the different classes of binding sites, the cooperativity of interactions should be taken into account [16].

The thermodynamic effect of cooperativity may be due to changes in the macromolecule's conformation or redistribution of the electron density at the site due to the influence through oligomers bonds or by the MI between bound ligands via space [17]. A variety of methods have been developed for computational explanation of cooperative interactions of a ligand with set of binding sites [18]. If these sites are equivalent and independent, relevant experimental data may be analysed toward the thermodynamic binding constant and the number of binding sites in terms of a linear Scatchard plot [19]. Since this presumes a most simple binding model, such an approach is susceptible to substantial misinterpretations in practice [20].

The McGhee and von Hippel model which takes into account the MI between nearest neighbouring bound ligands has been the most widely used [21] in combination with the Scatchard equation for description of small molecules with homogeneous polymers. Sometimes those equations are incorrectly used to describe interactions with G4 and other DNA nonhomogeneous oligomers even though the authors of the model have concluded that the binding of ligands known to have heterogeneous binding sites on the DNA cannot be adequately treated based on the conditional probability approach [21]. Although Scatchard analysis is very much attractive owing to pictorial rendition, it has serious implications of unreliability regarding the shape of a Scatchard plot and it is evident that this method is not appropriate for quantitative evaluation of the binding of TMPyP4 with G4. The Job plots analysis of the ligand binding with the G-quadruplex sites accounting for their MI has been described in [12].

To compare the interpretation of spectrometrically measured data using the canonical stepwise model with that of mutually influencing binding sites, we considered the combination of hard model-based fitting and soft model-free Singular Value Decomposition (SVD) [22]. The Rank Annihilation Factor Analysis (RAFA) has also aided in unambiguous revealing of the cooperative interactions. This solution should improve our understanding of the mechanism of the G4 folding and binding at a molecular level, which in turn will help in highly efficient and selective telomerase inhibitors discovery.

2. Computational Details

2.1. Description of the Model. To simulate G4-ligand interactions the model based upon a multiple set of binding sites has been proposed [12]. When binding with G4 DNA is encountered, then each monomeric ligand is assumed to be bound to one of the sites on the DNA oligomer. The cooperativity is the enhancing of binding affinity in comparison with the sole binding, but anticooperativity is the reduction of binding affinity. The secondary interaction is an important source of ligand binding cooperativity. In DNA-ligand complexes, the arrangement of the ligands decides the total stability of the complex and elucidates that secondary interactions between neighbouring ligands can contribute additional stabilization [23]. To find a sign of cooperative interactions just the influencing of ligand on another ligand

bound in the close vicinity has been taken into account. Consider intrinsic constants which govern the binding of ligand to a set of sites. In general case, the binding of N ligands among N vacancies gives 2^N possible variants of coordination. The one-dimension matrix of site-specific equilibrium constants \mathbf{K}_{1-N} (\mathbf{K}_{1-N} are constants for a ligand binding to each isolated site $1 \cdots N$) and matrix $\mathbf{\Omega}$ of the MI corrections (cooperativity parameters $[\omega_1 \ \omega_2 \ \cdots \ \omega_N]$) are constructed in accordance with hypothesis about structure of adducts formed in DNA oligomer saturation by a ligand. The values of \mathbf{K}_{1-N} and ω_i could be found simultaneously as was described previously [24].

We will start here with looking at the simplest of situations for ligand binding, namely, the interaction with bidentate oligomer. The matrix of configurations \mathbf{M} has been introduced in order to describe the detailed set of configurations adopted by the macromolecule in microstates. Each line of the matrix \mathbf{M} displays a microstates configuration through a sequence of zeroes in those positions where the sites of multidentate macromolecule are not occupied by a ligand, and with a sequence of ones in those positions where the sites are bound to the ligand. In the course of the macromolecule sites saturation, previously bound ligands can affect the binding of the following ligands. For the oligomer containing two nonidentical binding sites, the matrix \mathbf{M}^* can be written as

$$\mathbf{M}^* = \begin{bmatrix} 0 & 0 \\ K_1 & 0 \\ 0 & K_2 \\ K_1\omega & K_2\omega \end{bmatrix}. \quad (1)$$

The product of the nonzero elements in \mathbf{M}^* provides the column of stability constants of microscopic species

$$\mathbf{B} = \prod_{j=1}^N (\mathbf{M}^* - \mathbf{M} + \mathbf{E}); \quad (2)$$

here the matrix \mathbf{E} is all ones and the same size as \mathbf{M} . The estimates of the true nonlinear variables $\mathbf{K}_{1,2}$ and ω allow calculating the matrix \mathbf{C}_{DNAL} of DNA oligomer species of concentration for the i th total concentration of ligand and DNA, correspondingly C_L and C_D .

$$\mathbf{C}_{\text{DNAL}} = \begin{bmatrix} [\text{DNA}]_1 & [\text{DNAL}']_1 & [\text{DNAL}'']_1 & [\text{DNAL}_2]_1 \\ \cdots & \cdots & \cdots & \cdots \\ [\text{DNA}]_{\text{Np}} & [\text{DNAL}']_{\text{Np}} & [\text{DNAL}'']_{\text{Np}} & [\text{DNAL}_2]_{\text{Np}} \end{bmatrix}. \quad (3)$$

Here $[\text{DNA}] = C_D(1 + [L]^S \mathbf{B})^{-1}$ is an equilibrium concentration of free DNA oligomer and $[\text{DNAL}'_s] = [\text{DNA}][L]^S \mathbf{B}^T$ are G4 DNA species equilibrium concentrations; $[L]$ is an equilibrium concentration of free ligand; $S = 1, 2$ is a stoichiometric coefficient; Np is the number of experimental points. In our case the sum of microforms equilibrium concentrations $[\text{DNAL}']$ and $[\text{DNAL}'']$ gives the concentration

of oligomer species corresponding to single bound ligand. This species in conventional stepwise complex formation model is regarded as the 1 : 1 complex. Once we know \mathbf{C}_{DNAL} , we can calculate concentrations of free sites and these bound with a ligand.

$$\mathbf{C}_{\text{MonL}} = [\mathbf{C}_{\text{DNAL}} [\mathbf{E} - \mathbf{M}] \ \mathbf{C}_{\text{DNAL}} \mathbf{M}]. \quad (4)$$

2.2. The Validation of the Proposed Model. According to Lambert–Beer’s law, a spectrometrically measured data matrix \mathbf{A} is the product of a concentration matrix \mathbf{C} and a matrix \mathbf{S} of molar absorptivities of species that contribute to the absorption of light at any wavelength (species pure spectra). The corresponding general matrix equation is given as

$$\mathbf{A} = \mathbf{C}\mathbf{S} + \boldsymbol{\tau}; \quad (5)$$

here $\boldsymbol{\tau}$ is a matrix of residuals. The number of rows in matrix \mathbf{A} is equal to the number of studied solutions with different components concentrations. The number of columns in matrix \mathbf{A} is equal to the number of wavelengths. To simulate the data matrix \mathbf{A} a concentration matrix $\mathbf{C} = \mathbf{C}_{\text{MonL}}$ has been computed in accordance with equilibrium model of two-step ligand binding. The total light absorbance is due to the absorption of the occupied Mon_1L , Mon_2L and unoccupied sites Mon_1 , Mon_2 of oligomer molecules and absorption of free and coordinated ligand L . The data set was designed to simulate an interaction ligand with oligomer (TMPyP4 with G-quadruplex DNA), which has two binding sites. Since the rank R of data matrix \mathbf{A} is equal to the rank of \mathbf{C} or \mathbf{S} , whichever is smaller, and since the rank of either \mathbf{C} or \mathbf{S} can be no larger than the total number of spectral species, the rank of \mathbf{A} can be no larger than the total number of spectral species, provided the number of wavelengths and the number of studied solutions are both greater than the number of spectral species. For simplicity in the current simulations residuals $\boldsymbol{\tau}$ were put to zero. Thus the rank R is the number of singular values of $\mathbf{A}\mathbf{A}^T$ that are bigger than machine precision (10^{-10} – 10^{-13}).

2.3. The Least Square (LS) Fitting Procedure. The fitting procedure should deconvolve the data matrix \mathbf{A} into matrices of concentrations \mathbf{C} and molar absorptivities of a matrix \mathbf{S} . Multiplication of both sides of (5) on a pseudoinverse [22] matrix $\mathbf{C}^+ = (\mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T$, computed with an initial estimate of the required parameters, gives a current estimation \mathbf{S}^* of the true matrix of molar parameters \mathbf{S} . The best set of linear parameters \mathbf{S}^* for system in question can be calculated explicitly in a linear least-squares calculation by

$$\mathbf{S}^* = \mathbf{C}^+ \mathbf{A}. \quad (6)$$

In order to evaluate linear as well as nonlinear parameters for a chemical model such as that described above, a decomposition of the experimental data matrix \mathbf{A} is executed under a nonnegativity constraint. The product of \mathbf{S}^* and \mathbf{C}^+ gives \mathbf{A}_{calc} at the current values for nonlinear parameters.

Consequently, the iteration formula for refinement of the required nonlinear parameters can be written as

$$\mathbf{A}_{\text{calc}} = [\mathbf{C}^+ \mathbf{A}]_{>0} \mathbf{C}^* + \boldsymbol{\tau}, \quad (7)$$

where $[\dots]_{>0}$ is the operator that substituted negative values to zero values (nonnegativity constraint only for spectral profiles was applied); $\boldsymbol{\tau}$ is the residual matrix with unexplained data variance (here τ is machine precision). Relevant experimental data should be analysed toward the intrinsic association binding constants \mathbf{K}_{1-2} and intrinsic parameter of cooperativity ω to obtain a best fit to the data. When optimal nonlinear parameters were found, the pure species spectra have been calculated in accordance with (6).

The Levenberg–Marquardt nonlinear LS algorithm is used for refinement of parameters [25]. The iterative procedure described above is used to find the minimum of the sum of squares of deviations of all absorbance data matrix elements which were calculated using theoretical model, from the corresponding matrix considered on the current iterative step, calculated in turn by using (7). Refinement was stopped when the relative difference between consecutive iterations fell below a threshold value. The Hamilton factor was used in the present work to compare matrices by applying the following equation:

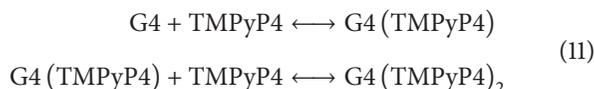
$$\text{PE}(\mathbf{F}_{\text{exp}}, \mathbf{F}_{\text{calc}}) = \sqrt{\frac{\text{Trace}[(\mathbf{F}_{\text{exp}} - \mathbf{F}_{\text{calc}})(\mathbf{F}_{\text{exp}} - \mathbf{F}_{\text{calc}})^T]}{\text{Trace}[\mathbf{F}_{\text{exp}} \mathbf{F}_{\text{exp}}^T]}} \times 100\%, \quad (8)$$

where Trace is the sum of the diagonal elements. Equation (8) was used to calculate lack of fit (lof). In this case, $\mathbf{F}_{\text{exp}} = \mathbf{A}_{\text{exp}}$; $\mathbf{F}_{\text{calc}} = \mathbf{A}_{\text{calc}}$. Second, (8) was used to compare the distribution diagrams calculated by application of conventional stepwise chemical model from the ones simulated by

$$\mathbf{C}_{\text{MonL}} = [[\text{Mon}_1] \quad [\text{Mon}_2] \quad [\text{Mon}_1L] + [\text{Mon}_1L2] \quad [\text{Mon}_2L] + [\text{Mon}_2L2]]. \quad (10)$$

SVD applied for \mathbf{A}_{1a} yields $R = 3$ indicating the three optically active chemical species in the full spectrum and local rang $R_L = 2$ indicating the species in spectral window of TMPyP4.

To fit the data, in accordance with proposed approach, the conventional noncooperative model of stepwise complex formation first has been tested.



Here stepwise equilibrium constants $x(1) = \lg \mathbf{K}_1$ and $x(2) = \lg \mathbf{K}_2$ are independent variables and $\omega = \text{constant} = 1$.

the cooperative model. In this case, a concentration matrix $\mathbf{F}_{\text{exp}} = \mathbf{C}_{\text{DNAL}}$ was calculated using a model, and a matrix $\mathbf{F}_{\text{calc}} = \mathbf{C}_{\text{calc}}$ was calculated in turn by using (3) on the last LS iterative step. In all cases, Matlab-2014® software was used in the calculations.

3. Results and Discussion

3.1. Cooperative Binding (Data Matrix \mathbf{A}_{1a}). The UV-Vis absorption spectra changes of an aqueous solution in binding reactions of TMPyP4 and G4 in all known cases are typically similar. To simulate data set one have to construct bathochromic red shifts 15–20 nm and hypochromicities at λ_{max} (60–70%) of the Soret band were observed during coordination TMPyP4 to G4. Spectra \mathbf{S}_L and \mathbf{S}_{Mon} used to build the data sets \mathbf{A}_{exp} are shown in Supplementary Material (Figure S1). In the simplest case, we assume that spectrum \mathbf{S}_{MonL} is distorted superposition of \mathbf{S}_L and \mathbf{S}_{Mon} . This situation is a binding without spectral interference. More general case of spectral interference (SI) is the mutual perturbation of the ligand spectra bound in the vicinity.

Let us consider the model defined by equilibrium constants $\lg \mathbf{K}_{1-2} = [6.5 \ 6.5]$ and $\omega = 2.0$, governing the formation of a MonL species (two equal sites plus a positive binding cooperativity between sites). Simulated data matrix (see Supplementary Material, Figure S2) \mathbf{A}_{exp} should be written as

$$\mathbf{A}_{\text{exp}} = [L] \mathbf{S}_L + \mathbf{C}_{\text{MonL}} \mathbf{S}_M + \boldsymbol{\tau}. \quad (9)$$

Here elements of the matrix $\boldsymbol{\tau}$ are put to zero; \mathbf{S}_L is molar absorptivities of free ligand. Given that two sites are equivalent and have no SI the matrix of molar absorptivities of free and occupied sites I-II is $\mathbf{S}_M = [\mathbf{S}_{\text{Mon}_1} \ \mathbf{S}_{\text{Mon}_2} \ \mathbf{S}_{\text{Mon}_1L} \ \mathbf{S}_{\text{Mon}_2L}]^T$. The concentration matrix used to calculate the absorbance data matrix is

However the LS fitting gives a poor fit to the target matrix \mathbf{A}_{exp} . The reproduced \mathbf{A}_{1a} matrix is computed as

$$\mathbf{A}_{\text{calc}} = [L] \mathbf{S}_L + \mathbf{C}_{\text{DNAL}} \mathbf{S}_F + \boldsymbol{\tau}. \quad (12)$$

Here $\mathbf{S}_F = [\mathbf{S}_{\text{DNA}} \ \mathbf{S}_{\text{DNAL}} \ \mathbf{S}_{\text{DNAL}_2}]^T$; τ is lof. Once we analysed simulated data matrix \mathbf{A}_{1a} using a model of independent binding sites, taking into account the existence of 2 sites of 2 types, the best fit was model with two equivalent binding sites $\lg \mathbf{K}_{1-2} = [7.04 \ 7.04]$, and a lof is 0.2%.

Figure 1(a) shows distribution diagrams of hypothetical DNA oligomer between species DNA, DNAL, and DNAL₂. As can be seen from Figure 1(a), the resolved distribution of the species significantly differs from the theoretical one (PE 18.4%). Figure 1(b) shows simulated and resolved pure species

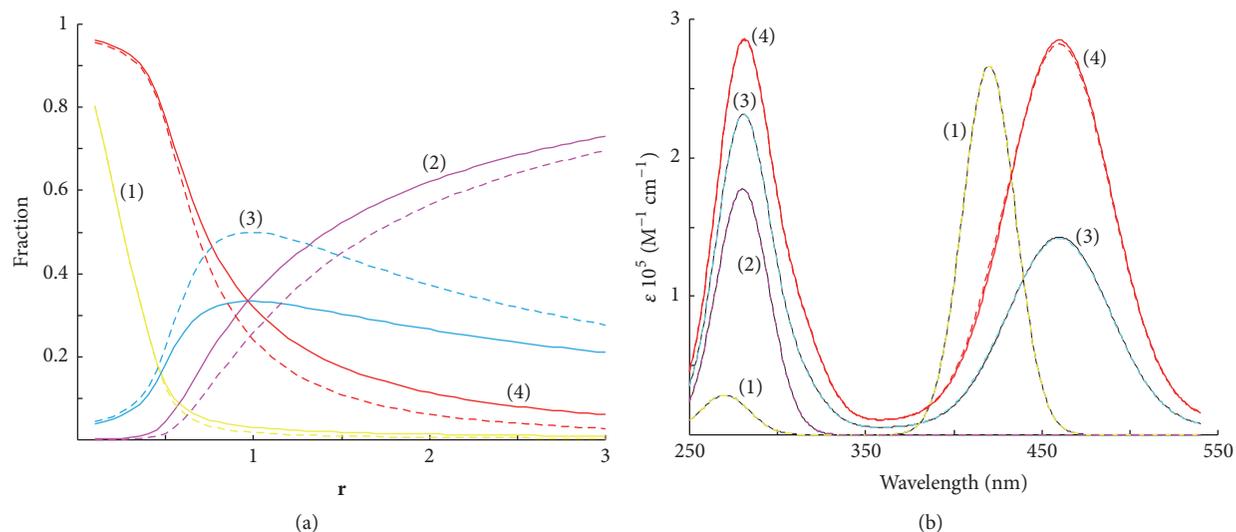


FIGURE 1: Distribution diagram (a) and pure species spectra (b) resolved by deconvolution of \mathbf{A}_{1a} . Species fractions calculated as a function of the DNA/ligand molar ratio r ($r = C_D/C_L$; $C_L = 5 \mu\text{mol/L}$) in accordance with the model: $\lg \mathbf{K}_{1-2} = [7.04 \ 7.04]$, $\omega = 1 = \text{const}$, shown as dotted lines. Profiles simulated in accordance with a binding model: $\lg \mathbf{K}_{1-2} = [6.5 \ 6.5]$, $\omega = 2.0$, shown as solid lines. (1) L , (2) DNA, (3) DNAL, and (4) DNAL_2 .

spectra. The shape of spectral profiles is due to calculated monomers spectra. The spectrum of species DNAL_2 is doubled spectrum of DNAL, $2\mathbf{S}_{\text{DNAL}} = \mathbf{S}_{\text{DNAL}_2}$.

RAFA of the Data Matrix \mathbf{A}_{1a} . Given that spectra \mathbf{S}_{DNA} and \mathbf{S}_L are easily measured experimentally, we can calculate the matrix \mathbf{A}' .

$$\mathbf{A}' = \mathbf{A}_{\text{exp}} - [L] \mathbf{S}_L - [\text{DNA}] \mathbf{S}_{\text{DNA}} + \tau. \quad (13)$$

With the result of (13), \mathbf{A}_{exp} can be divided into two parts, where the known species L and DNA only contribute to the one part. The other part \mathbf{A}' contains linear combination of the unknown spectra. The SVD of matrix \mathbf{A}'_{1a} indicated no rank annihilation of \mathbf{A}_{1a} including concentration window of the ligand. This result conforms that spectral species are inaccurately recognized.

The difference between theoretical and calculated fractions of free sites and these bound with a ligand, that is, Mon_1 , Mon_2 , Mon_1L , Mon_2L , $\text{Mon}L_2$, also can be seen. However, the concentration of free monomer is calculated correctly. The matrix \mathbf{A}'' may be written in the form:

$$\begin{aligned} \mathbf{A}'' = & \mathbf{A}_{\text{exp}} - [L] \mathbf{S}_L \\ & - [[\text{Mon}_1] \ [\text{Mon}_2]] [\mathbf{S}_{\text{Mon}_1} \ \mathbf{S}_{\text{Mon}_2}]^T + \tau. \end{aligned} \quad (14)$$

The reduced data matrix \mathbf{A}''_{1a} has been obtained by subtracting the contribution of the proposed model species from the complete data set. The rank of the matrix \mathbf{A}''_{1a} is less than the rank of the full \mathbf{A}_{1a} matrix $R'' = R - 1$. So in the full spectral range one species is correctly recognized.

The rank annihilation completes the conclusion that sites are indistinguishable monomers.

3.2. Anticooperative Binding without SI (Data Matrix \mathbf{A}_{1b}). To simulate a data matrix \mathbf{A}_{1b} obeying (9), the same spectra as in the previous example were used (see Figure S1). Concentrations are calculated from the model $\lg \mathbf{K}_{1-2} = [6.5 \ 6.5]$, $\omega = 0.3$. Fitting the absorbance data matrix using conventional stepwise model without cooperativity gives the $\text{lof}(\mathbf{A}_{1b}) \approx 0.0\%$. Here the best fit gives model considering 2 sites of 2 different types $\lg \mathbf{K}_{1-2} = [6.79 \ 5.16]$. The stepwise model provides a good fit to simulated data. Consequently, we have a model ambiguity as a result of the rotational ambiguity. All species fractions quantified very well (see Table 1).

RAFA of the Data Matrix \mathbf{A}_{1b} . As can be seen from Table 1, SVD is applied on \mathbf{A}_{1b} , \mathbf{A}'_{1b} , and \mathbf{A}''_{1b} giving the different result as for the above case of positive cooperativity. One can see the rank annihilation in full spectral range and in a spectral window of ligand. This result suggests that subtracted spectral species are correctly recognized. The rank of the matrix \mathbf{A}''_{1b} is less than the rank of the full \mathbf{A}'_{1b} matrix $R'' = R' - 1$. So in the full spectral range extra one species is correctly recognized.

3.3. Cooperative Binding with SI (Data Matrix \mathbf{A}_{1c}). In the previous examples, we have used an assumption that all G4-ligand complexes formed possess the same absorbance spectrum, irrespective of the binding mode. However this assumption is violated in most actual systems, probably in each system with multiple binding sites. We will use SI when binding in each site leading to a change in the spectrum of ligand in the vicinity. Given that bound ligands have SI, an experimental matrix \mathbf{A}_{1c} should obey (12) where the matrix of molar absorptivities is

TABLE 1: Intrinsic binding constants calculated for simulated ligand G4 DNA equilibrium systems.

	Data matrix	$\lg \mathbf{K}_{1,2}, \omega$ theoretical model	$\lg \mathbf{K}_{1,2}, \omega$ LS calc. parameters	lof %	PE _F %	PE _S %	R (R _L)	R' (R' _L)	R'' (R'' _L)
<i>Identical sites</i>									
Cooperative binding without SI	A _{1a}	[6.5 6.5], 2.0	[7.04 7.04], 1.0*	0.2	18.4	1.0	3 (2)	3 (2)	2 (2)
Anticooperative binding without SI	A _{1b}	[6.5 6.5], 0.3	[6.79 5.16], 1.0*	≈0**	≈0**	≈0**	3 (2)	2 (1)	1 (1)
Cooperative binding with SI	A _{1c}	[6.5 6.5], 2.0	[6.75 6.75], 1.0*	0.8	19.4	2.3	4 (3)	4 (3)	3 (3)
Anticooperative binding with SI	A _{1d}	[6.5 6.5], 0.3	[6.79 5.16], 1.0*	≈0**	≈0**	≈0**	4 (3)	2 (2)	2 (2)
<i>Different sites</i>									
Binding without SI	A _{2a}	[7 6], 1	[6.74 6.74], 0.58	≈0**	≈0**	≈0**	4 (3)	2 (2)	2 (2)

R is the rank of full data matrix **A**; R_L is the rank of data matrix part covering spectral window of a ligand; SI is spectral interference; * parameter is fixed; **1, lof < 10⁻⁴ ≈ 0; C_L + C_P = 5 μM.

$S_M = [S_{Mon_1} \ S_{Mon_2} \ S_{Mon_1L} \ S_{Mon_2L} \ S_{Mon_1L2} \ S_{Mon_2L2}]^T$ (see Supplementary Materials, Figure S1b) and

$$C_{MonL} = [[Mon_1] \ [Mon_2] \ [Mon_1L] \ [Mon_2L] \ [Mon_1L2] \ [Mon_2L2]] \quad (15)$$

is concentration matrix calculated from the before used binding model $\lg K_{1-2} = [6.5 \ 6.5]$, $\omega = 2.0$. The distribution plot of resolved oligomer species is very similar to these shown on Figure 1. SVD is applied on A_{1c} yielding four main factors defining absorbance variance ($R = 4$), and the intrinsic absorption of the ligand is determined by three factors ($R_L = 3$). With the application of (13), the rank annihilation was not achieved. This suggests that all forms are calculated with poor accuracy. Application of (14) for total spectral range gives $R'' = 3$ which means the correct calculation of only the monomer contribution.

3.4. Anticooperative Binding with SI (Data Matrix A_{1d}).

Consider the binding model $\lg K_{1-2} = [6.5 \ 6.5]$, $\omega = 0.3$. The model is used to simulate anticooperative binding with SI. In this example, the model ambiguity is the result of LS fitting. The presence of SI does not affect the shape of oligomer distribution diagram shown on Figure 2. However, the result of RAFA is not the same. The identification of sites identity is ambiguous in this case. As can be seen from Table 1, the result of SVD application on matrices A_{1d} , A'_{1d} , and A''_{1d} is the same as for these for two types of sites with different spectral characteristics (see below).

3.5. The Binding to G4 DNA with Different Sites (Data Matrix A_{2a}).

Consider lastly an oligomer which has different binding sites. Consequently each site should have different affinities for the ligand and may have a different absorbance signal. The model with a strong binding site and weaker binding site $\lg K_{1-2} = [7 \ 6]$ to simulate the data matrix was used. Simulated data can be decomposed using the model of two equivalent binding sites and anticooperativity $\lg K_{1-2} = [6.74 \ 6.74]$, $\omega = 0.58$ with excellent precision (see Table 1). In this case to simulate data use (12), but reproduced data matrix is calculated in accordance with (9). Resolved distribution of G4 oligomer species well reproduce those used for simulation. The rank of a data matrix constructed in accordance with (12) is equal to the rank of C ($R = 4$). The rank of both A_{2a} and A_{2b} in spectral window ($R_L = 3$) is less than the number of spectral species in full spectral range because the light absorption by monomer species is zero in this interval. Nevertheless, the rank of A' along with A'' two units less compared with A can be seen due to rotational ambiguity. Moreover we can see the same sequences of R (R_L), R' (R'_L), and R'' (R''_L) regardless whether or not there is spectral MI. Thus, given example of binding shows inapplicability of RAFA to indicate the fact that sites are different or the same sites bind in anticooperative fashion.

3.6. The TMPyP4 Binding with Nucleotide Sequence 5'-(CG3)2CGC(AG3)2G-3' Experimental Data (Matrix A_{3e}).

To illustrate the applicability of the suggested approach to model porphyrin binding with a macromolecule, a porphyrin TMPyP4 binding to 5'-(CG3)2CGC(AG3)2G-3' was selected as experimental example. UV-absorbance experimental data published previously [15] has been used for further numerical treatment. In the original work [15] mole-ratio experiments to study the G4-ligand interaction were performed. When solution of G4 is added to solution of TMPyP4 in buffer pH 7.1, at 25°C the induced bathochromic red shifts and hypochromicities for the Soret band clearly indicate the DNA-ligand complexation. A stepwise complex formation model has been applied in [15] to interpret the experimental data (matrix A_{3e}). Conventional stepwise complex formation algorithm best fits the data matrix which included one strong binding site and one weak binding site ($\lg K_{1-2} = [7.07 \ 5.79]$), $\text{Iof}(A_{3e}) \approx 0.25\%$. However, the same fit gives a model with two equivalent anticooperative binding sites. The calculated intrinsic binding constants K_{1-2} and the cooperativity parameter ω are summarized in Table 2. The calculated distribution diagrams for this mole-ratio spectrometric titration of G4 by TMPyP4 for both models are compared in Figure 2. For this G4 DNA example of binding interaction, the LS fitting discloses a model ambiguity. The values of R'' (R''_L) we have given in Table 2 equal the values of all these quantities which are given in Table 1 for anticooperative binding with identical sites without SI. Based on this observation, one can suggest a model of anticooperative assembling of TMPyP4 and G4 DNA with identical binding sites without SI. In addition, the calculated pure spectra for species involved in the equilibrium are shown in Figure 2. The spectra are calculated according to the following equation:

$$\begin{aligned} S &= A''_{3e} [[Mon_1L] + [Mon_2L] \ [Mon_1L2]]^+ \\ &= A'_{3e} [[Mon_1L] + [Mon_1L2] \ [Mon_2L] + [Mon_2L2]]^+ . \end{aligned} \quad (16)$$

The spectra associated with the ligand are compared in Figure 2(b). Some difference between Mon_1L and Mon_2L spectra can be seen. But this very little difference is probably due to insufficiency of free sites and free ligand fixed spectra measurements. The calculation of pure absorption monomers spectra is impossible because of rank deficiency after subtracting contribution of free ligand absorbance from full data matrix A_{3e} (see Table S2 in the Supplementary Material).

In summary, the methodology of the absorbance data matrix numerical treatment for G4 DNA/ligand equilibrium system has been successfully developed using a hard modelling matrix method and Rank Annihilation Factor Analysis. Accurate identification of binding sites identity

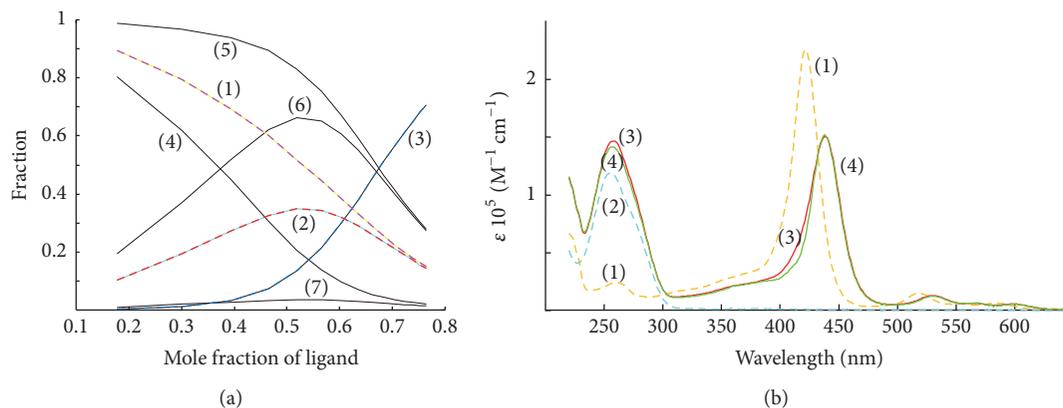


FIGURE 2: Distribution diagram (a) and pure species spectra (b) resolved by deconvolution of A_{3e} . Dashed lines are calculated in accordance with anticooperative model: $\lg K_{1-2} = [6.79 \ 6.79]$, $\omega = 0.43$; solid lines are calculated in accordance with stepwise model $\lg K_{1-2} = [7.07 \ 5.79]$, $\omega = 1.0$; (a) (1) $[Mon_1]/C_D = [Mon_2]/C_D$, (2) $[Mon_1L]/C_D = [Mon_2L]/C_D$, (3) $[Mon_1L2]/C_D = [Mon_2L2]/C_D$, (4) $[Mon_1]/C_D$, (5) $[Mon_2]/C_D$, (6) $[Mon_1L]/C_D$, (7) $[Mon_2L]/C_D$; (b) (1) TMPyP4, S_L ; (2) $Mon_1 = Mon_2$, S_{Mon} ; (3) Mon_1L , S_{Mon_1L} ; (4) Mon_2L , S_{Mon_2L} ; pH 7.1, $T = 25^\circ C$.

TABLE 2: Intrinsic binding constants for TMPyP4 binding to G4 oligomer calculated from experimental UV-Vis titrations (pH 7.1, $T = 25^\circ C$).

A model	Data matrix	$\lg K_{1-2}, \omega$ LS calculated parameters	lof %	PE_F %	rank A $R(R_L)$	rank A' $R'(R'_L)$	rank A'' $R''(R''_L)$
Stepwise complex formation	A_{3e}	$[7.07 \ 5.79]$, 1.0*	0.25	$\approx 0^{**}$	3 (2)	2 (1)	—
Cooperative binding	A_{3e}	$[6.79 \ 6.79]$, 0.43	0.25	$\approx 0^{**}$	—	—	1 (1)

R is the rank of full data matrix A ; R_L is the rank of data matrix part covering spectral window of a ligand 340–540 nm.

and the mutual influence between bound ligands, which is essential to G4 DNA *in vivo* function annotation, can be done for sites involved in cooperative ligand binding. Unfortunately the discovery of the sites identity is hindered when sites are involved in anticooperative binding. It is shown that the spectrometrically measured titration curves can be described ambiguously using the model of two different and independent binding sites as well as that of two mutually influencing identical sites in the case of anticooperative binding.

Conflicts of Interest

The author declares that they have no conflicts of interest.

References

- [1] S. Neidle and S. Balasubramanian, *Quadruplex Nucleic Acids*, Royal Society of Chemistry, Cambridge, UK, 2006.
- [2] E. P. Calvo and M. Wasserman, "G-Quadruplex ligands: potent inhibitors of telomerase activity and cell proliferation in *Plasmodium falciparum*," *Molecular & Biochemical Parasitology*, vol. 207, article 33, 2016.
- [3] E. Largy, A. Granzhan, F. Hamon, D. Verga, and M.-P. Teulade-Fichou, "Visualizing the quadruplex: from fluorescent ligands to light-up probes," *Topics in Current Chemistry*, vol. 330, pp. 111–178, 2013.
- [4] J. L. Huppert, "Hunting G-quadruplexes," *Biochimie*, vol. 90, no. 8, pp. 1140–1148, 2008.
- [5] Q. Li, J.-F. Xiang, Q.-F. Yang, H.-X. Sun, A.-J. Guan, and Y.-L. Tang, "G4LDB: a database for discovering and studying G-quadruplex ligands," *Nucleic Acids Research*, vol. 41, no. 1, pp. D1115–D1123, 2013.
- [6] P. Murat, Y. Singh, and E. Defrancq, "Methods for investigating G-quadruplex DNA/ligand interactions," *Chemical Society Reviews*, vol. 40, no. 11, pp. 5293–5307, 2011.
- [7] Q. Cao, Y. Li, E. Freisinger, P. Z. Qin, R. K. Sigel, and Z. Mao, "G-quadruplex DNA targeted metal complexes acting as potential anticancer drugs," *Inorganic Chemistry Frontiers*, vol. 4, no. 1, pp. 10–32, 2017.
- [8] R. Gargallo, R. Eritja, and A. G. Kudrev, "Spectrometric study of the oligodeoxyribonucleotide protonation in aqueous solution," *Russian Journal of General Chemistry*, vol. 80, no. 3, pp. 485–492, 2010.
- [9] T. Qin, K. Liu, D. Song, C. Yang, and H. Su, "Porphyrin bound to i-motifs: intercalation versus external groove binding," *Chemistry—An Asian Journal*, vol. 12, no. 13, pp. 1578–1586, 2017.
- [10] T. Uno, K. Aoki, T. Shikimi, Y. Hiranuma, Y. Tomisugi, and Y. Ishikawa, "Copper insertion facilitates water-soluble porphyrin binding to rA·rU and rA·dT base pairs in duplex RNA and RNA·DNA hybrids," *Biochemistry*, vol. 41, no. 43, pp. 13059–13066, 2002.

- [11] D.-F. Shi, R. T. Wheelhouse, D. Sun, and L. H. Hurley, "Quadruplex-interactive agents as telomerase inhibitors: synthesis of porphyrins and structure-activity relationship for the inhibition of telomerase," *Journal of Medicinal Chemistry*, vol. 44, no. 26, pp. 4509–4523, 2001.
- [12] A. G. Kudrev, "Calculation of cooperativity and equilibrium constants of ligands binding to G-quadruplex DNA in solution," *Talanta*, vol. 116, pp. 541–547, 2013.
- [13] A. J. Bhattacharjee, K. Ahluwalia, S. Taylor et al., "Induction of G-quadruplex DNA structure by Zn(II) 5,10,15,20-tetrakis(N-methyl-4-pyridyl)porphyrin," *Biochimie*, vol. 93, no. 8, pp. 1297–1309, 2011.
- [14] J. B. Chaires, "Analysis and interpretation of ligand - DNA binding isotherms," *Methods in Enzymology*, vol. 340, pp. 3–22, 2001.
- [15] S. Manaye, R. Eritja, A. Aviñó, J. Jaumot, and R. Gargallo, "Porphyrin binding mechanism is altered by protonation at the loops in G-quadruplex DNA formed near the transcriptional activation site of the human c-kit gene," *Biochimica et Biophysica Acta—General Subjects*, vol. 1820, no. 12, pp. 1987–1996, 2012.
- [16] A. G. Kudrev, "Dependence of the formation function of a complex with an oligomer on the number of monomers in its Chain," *Russian Journal of General Chemistry*, vol. 72, pp. 1501–1507, 2002.
- [17] A. S. Mahadevi and G. N. Sastry, "Cooperativity in noncovalent interactions," *Chemical Reviews*, vol. 116, no. 5, pp. 2775–2825, 2016.
- [18] T. L. Hill, *Cooperativity Theory in Biochemistry: Steady State and Equilibrium Systems*, Springer, New York, NY, USA, 1985.
- [19] G. Scatchard, "The attractions of proteins for small molecules and ions," *Annals of the New York Academy of Sciences*, vol. 51, pp. 660–672, 1949.
- [20] I. M. Klotz, "Numbers of receptor sites from scatchard graphs: facts and fantasies," *Science*, vol. 217, no. 4566, pp. 1247–1249, 1982.
- [21] J. D. McGhee and P. H. von Hippel, "Theoretical aspects of DNA-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice," *Journal of Molecular Biology*, vol. 86, pp. 469–489, 1974.
- [22] G. H. Golub and C. F. van Loan, *Matrix Computations*, Johns Hopkins Series in the Mathematical Sciences, Johns Hopkins University Press, London, UK, 2nd edition, 1989.
- [23] A. G. Kudrev, "Cooperative binding of 2,2'-bipyridine into polynucleotide poly(A)-poly(U-) in an alkaline aqueous solution," *Biopolymers*, vol. 99, no. 9, pp. 621–627, 2013.
- [24] A. G. Kudrev, "Cooperativity during binding of a ligand to a multidentate oligomer," *Polymer Science Series A*, vol. 55, no. 10, pp. 586–594, 2013.
- [25] M. Maeder and A. D. Zuberbühler, "Nonlinear least-squares fitting of multivariate absorption data," *Analytical Chemistry*, vol. 62, no. 20, pp. 2220–2224, 1990.

