Mathematical Analysis of the NC Binding Model

Introduction

NC binds to small oligonucleotides following a complex binding pathway. The model shown in Fig. 3 is used for analysis of the SPR, FA, and TFQ assays. The mathematical details required for the analysis of each of these assays differ somewhat, although each analysis includes these core model statements. One implication of the model is that the composition of an NC:$(TG)_4$ mixture will depend upon the relative proportions of these two reactants: for example, when NC is in excess, as in the early phases of an experiment in which $(TG)_4$ is titrated into NC, then there will be a significant fraction of NON complexes present. This will also be true at the end of a reverse titration, when several equivalents of NC have been added to a $(TG)_4$ solution. Conversely, ONO will be a major constituent of mixtures containing an excess of $(TG)_4$ over NC. NO (i.e., 1:1) complexes will only be a significant component of the mixture when the concentrations of NC and $(TG)_4$ are approximately equal.

Surface Plasmon Resonance (SPR) Analysis

The SPR instrument's signals show the dynamics of a binding system (1). These signals contain three major components. First, the signals produced by the Biacore instrument include components due to the operation of the instrument (valve opening and closing, etc.), non-significant changes in buffer constituents, and other additive deterministic noise sources. These additive effects are removed in a signal grooming step described elsewhere (2). Second, as discussed below, the flow cell imposes physical limits on the SPR signal. Third, Fig. 3 of the paper is used as the basis for the SPR dynamic analysis. Here we discuss the second and third steps in the SPR analyses, taking into account the physics of the SPR instrument (3-6), and assume that the first step of data grooming has been completed successfully.

In some SPR experiments, the binding rate of analyte to the immobilized ligand on the chip is slow relative to the arrival rate of the analyte in the reaction volume (a nearly flat region along one side of the flow cell). In other systems, the binding of analyte and ligand is rapid relative to the arrival rate of analyte to the reaction region. When a binding reaction proceeds slowly relative to the arrival of analyte to the reaction region, unbound analyte builds up in the reaction region and soon is equal the concentration of the injected solution. In contrast, when a binding reaction proceeds quickly relative to the arrival rate of analyte, most or all of the analyte is bound up by the ligand as soon as the analyte arrives in the reaction region. In this case, the concentration of analyte in the reaction region cannot reach the concentration in the injected solution until the binding reaction slows as it nears a steady state. There are two indications that the binding of NC to $(TG)_4$ in our SPR experiments binds very quickly. First, we have assessed the rate at which NC interacts with $(TG)_4$ by pre-binding NC to immobilized $(TG)_4$ on an SPR chip, and then washing it off with a solution of $(TG)_4$. We found that the SPR signal due to the NC bound to the immobilized $(TG)_4$ drops essentially instantaneously under these conditions (data not shown). This observation demonstrates that NC can move from one TG4 molecule to another extremely rapidly, and suggests that the NC is essentially uniformly distributed in the reaction zone at relevant time scales.

Careful analysis of the kinetics of binding of NC to immobilized $(TG)_4$ in the SPR instrument also strongly supports the idea that transport of NC to the binding surface is the rate-limiting event in generation of the SPR signal. As discussed by Myszka et al. (5), if transport were fast relative to the binding reaction being analyzed, then the concentration of NC at the binding surface would be essentially constant over the duration of the reaction. Under these conditions, the rate of binding (i.e., the slope of the binding portion of the SPR curve) would decrease as the reaction proceeded and the $(TG)_4$ molecules became occupied with NC. In contrast, if transport is slow relative to binding, then the concentration of NC at the reaction surface will be virtually zero, and the rate of binding will be nearly constant throughout the binding reaction. We have analyzed a typical SPR curve for the binding of NC to $(TG)_4$ (shown in Fig. S1A) by constructing a transient phase diagram, plotting the slope of the curve (dRU/dt, black dots) against the signal itself. As shown in Fig. S1B, the slope is virtually constant over the entire range of the binding reaction, i.e., from ~ 40 to nearly 500 RU. In attempting to fit these results, we assumed that the change in NC concentration in the reaction area is always zero (hard transport). Hard transport assumes that all analyte (i.e., NC) molecules bind immediately upon arriving in the reaction zone until a steady state has been achieved.

When the binding reaction reaches steady state, the analyte concentration becomes identical throughout the flow cell, and has the value of the injected solution. As indicated by the red line in the Figure, these assumptions gave an excellent fit to the data. The results are, in contrast, entirely distinct from the data expected (blue line) if binding is slow relative to transport. The constant slope of the binding reaction seen here (black dots and red line) would appear to be a diagnostic test for the hard transport model, signaling that transport effects must be taken into account in analysis of the data.

We analyzed SPR results by first calculating a transport number (k_t) from the physical configuration of the flow cell and the flow rate. This parameter is equivalent to the "transport coefficient" described by Myszka et al. (5). We incorporated this term into the set of reactions shown in Fig. 3, and derived the stochastic vector differential equation shown in Equation [1]. This equation, which models the binding reaction seen in the reaction zone of the flow cell, is numerically integrated. The integrated vector signal is weighted by the molecular weights of the various species of complexes and summed to model the actual signal produced by the SPR instrument which should equal the sum of NO + 2NON + 2ONO. The model is compared with the actual SPR signal and the dynamic parameters are adjusted (using a Marquardt – Newton algorithm) to identify the fit signal that most closely matched the SPR signal in the least sum squares sense. Equation [1] may be used to fit a simple model to the SPR signals by setting some of the parameters to zero.

$$\begin{pmatrix} \frac{d}{dt}O \\ \frac{d}{dt}NO \\ \frac{d}{dt}NON \\ \frac{d}{dt}ONO \end{pmatrix} = \begin{pmatrix} -k_{no}\left[O\frac{k_tC + k_{no}K_{NO}NO + k_{non}K_{NON}NON}{k_t + k_{no}O + k_{non}NON} + K_{NO}NO\right] - k_{ono}\left(O \cdot NO - K_{ONO}ONO\right) \\ \left(k_{no}O - k_{non}NO\right)\left(\frac{k_tC + k_{no}K_{NO}NO + k_{non}K_{NON}NON}{k_t + k_{no}O + k_{non}NON}\right) - k_{no}K_{NO}NO + k_{non}K_{NON} - k_{ono}\left(O \cdot NO - K_{ONO}ONO\right) \\ k_{non}\left(NO\left(\frac{k_tC + k_{no}K_{NO}NO + k_{non}K_{NON}NON}{k_t + k_{no}O + k_{non}NON}\right) - K_{NON}NON\right) \\ k_{ono}\left(O \cdot NO - K_{ONO}ONO\right) \end{pmatrix} \quad [1]$$

where k_t = NC transport rate
k_{no} = on-rate for the NO complex
K_{NO} = equilibrium constant for the NO complex = K_D1 in Fig. 3
k_{non} = on-rate for the NON complex
K_{NON} = equilibrium constant for the NON complex = K_D2 in Fig. 3
k_{ono} = on-rate for the ONO complex
K_{ONO} = equilibrium constant for the ONO complex = K_D3 in Fig. 3
C = concentration of NC at the flow cell entrance.

These equations are integrated using a stiff integration algorithm and fit using a Marquardt – Newton method with the Matlab software. The amount of saturation signal for a putative 1:1 signal may be determined by multiplying the RU signal due to the immobilized nucleic acid (ligand) by the ratio of the molecular weights of the analyte and ligand. This value is displayed on several of the Figures included in the main paper.

Fluorescence Anisotropy

Fluorescence anisotropy signals show a binding signal that is based on fractional binding. Thus it is important to be able to estimate the stoichiometry in these assays. We estimate this parameter by relating the molar ratio of the reactants to the anisotropy. A simple describing function then is used to estimate the molar ratio at which the fractional binding becomes saturated.

$$anisotropy = G\frac{\left(C + V + MR\right) - \sqrt{\left(C + V + MR\right)^2 - 4V\,MR}}{2V} + offset \quad [2]$$

where: G = Gain
C = constant
MR = Molar Ratio (NC/$(TG)_4$)
V = "valency" = maximum $(TG)_4$ molecules bound per NC
offset = signal offset

Fitting the FA data with this equation gave a value of 'V' of 0.5, in complete agreement with the stoichiometry of ~ 2 that is evident in the inset to Fig. 5.

For the purposes of our analysis, we assumed that NO, ONO, and NON complexes all have the same anisotropy due to the flexible nature of the fluorophore used (7). Our treatment is based on Fig.3, and also takes into account the changes in the

fluorescence during the titration induced by binding of NC. This correction is performed using the R factor (8) as shown in Equation [3].

$$f_B = \frac{r - r_F}{(r - r_F) + R(r_B - r)} \quad [3]$$

where: f_B = fractional binding
r = observed anisotropy
r_F = anisotropy of the uncomplexed molecule
r_B = anisotropy of the completely complexed molecule
R = ratio of the complexed molecule's fluorescence intensity to the molecule's uncomplexed fluorescence intensity.

The R parameter is determined by fitting a logistic equation through the fluorescence intensity data, and estimating the ratio of intensity change as the titration progresses. Fractional binding, f_B, is numerically determined by analyzing Fig. 3 from the standpoint of conservation laws as shown in Equation [4] and normalizing to the concentration of oligonucleotide.

$$\begin{pmatrix} N_t \\ \\ O_t \end{pmatrix} = \begin{pmatrix} N + \frac{N \cdot O}{K_{NO}} + 2\frac{N^2 \cdot O}{K_{NO}K_{NON}} + \frac{N \cdot O^2}{K_{NO}K_{ONO}} \\ O + \frac{N \cdot O}{K_{NO}} + \frac{N^2 \cdot O}{K_{NO}K_{NON}} + 2\frac{N \cdot O^2}{K_{NO}K_{ONO}} \end{pmatrix} \quad [4]$$

Equation [3] and [4] are used in a Marquardt – Newton method (using the Matlab software) to estimate the equilibrium constants (K_x).

Tryptophan Quenching

The first step in analyzing the trytophan quenching is to estimate the stoichiometry of the reaction. We found that the fluorescence of NC was almost completely quenched at a molar ratio of 2 $(TG)_4$:1 NC. Thus, the binding reaction appears to be completed at this stoichiometry.

Following Vuilleumier et al., we model these binding reactions as using the following equation (9):

$$\nu = \frac{I_0 - I}{I_0 - I_t}\frac{O_t}{N_t} \quad [5]$$

where: ν = moles of oligonucleotide bound per mole of NC
I = observed fluorescence
I_0 = maximum fluorescence
I_t = minimum fluorescence
O_t = total oligonucleotide concentration
N_t = total NC concentration.

The function, ν, is numerically determined using Equation [5], and the K_x parameters are estimated by a Marquardt- Newton method using the Matlab software. The simple 1:1 binding reaction was modeled by eliminating the two right-hand columns from Equation [4]. In analogy with the anisotropy analysis, we assume that the fluorescence signal is

quenched to the same extent in NO, ONO, and NON complexes. Fits using only K_d1 and K_d2 or K_d3 are shown in Fig. S2.

Competition Analysis

The interactions of NC with small oligonucleotides are so complex, as shown in the present paper, that a detailed analysis, using Equation [1], of the competition experiments in Figs 6 and 7 was not feasible. We chose to describe the results of competition assays using a describing function. The equation that provides some insight into this reaction is the logistic equation (10). The 4-parameter logistic equation we use here is:

$$y = IS_{\min} + \frac{A - IS_{\min}}{1 + \left(\frac{Conc}{IC_{50}}\right)^{b}} \qquad [6]$$

where y = normalized initial slope
IS_{min} = minimum normalized initial slope
A = normalized initial slope in absence of competitor
$Conc$ = Concentration of oligonucleotide
IC_{50} = Concentration of oligo reducing initial slope 50% of (1-IS_{min})
b = curvature of the competition curve.

As the concentration of competitor increases, the binding signal decreases. The concentration of competitor that reduces the binding signal to 50% of its final reduction is described with the IC_{50} parameter. At very large concentrations of competitor, the binding signal stabilizes to a final value. This remaining binding signal is described by the IS_{min} parameter (see Figure 7 of the main paper). The A and b parameters describe, together, the amplitude and curvature of the signal. These parameters may be fit through a least squares regression using SAS, Mathcad, or Sigma Plot.

Akaike's Information Criterion

The additional complexity and number of parameters in the complex model may not actually produce a statistically improved fit. One method to assess this question is through Akaike's information criterion (ΔAIC) (11). This statistic allows comparison between various models for a given data set, even when the models are not strictly subsuming models (12,13). Here we use the difference between two models (the complex model of Equation [1], and a simple Langmuir/Scatchard model) using the following odds ratio calculation:

$$\Delta AIC = -2\ln(likelihood\ ratio) + 2\big(p(A) - p(B)\big) \qquad [7]$$

where A = number of points
B = number of parameters.

Here the likelihood for a given fit using a given model is given by:

$$AIC = A\ln\big(root\ sum\ squares\ residuals\big) + 2B$$

The *ΔAIC* statistic imposes a penalty for each additional parameter. A positive value then indicates that the complex model better describes the signal despite the increased number of parameters.

The contribution of K_d2 and K_d3 to the analysis can also be demonstrated by attempting to fit the data without them. For example, Fig. 4 of the paper, in which $(TG)_4$ is titrated into 400 nM NC, shows the superior fit for the TFQ data obtained using K_d1, K_d2, and K_d3, compared to K_d1 alone. Fig. S2A presents the fit obtained with K_d1 and K_d2; it is evident that this model (dashed line) fits the data well up to ~ 350 nM $(TG)_4$; however, above this concentration, the data deviate from this model. This systematic deviation is presumably because ONO complexes become significant when the $(TG)_4$:NC ratio is near or greater than one. Conversely, a model containing only K_d1 and K_d3 (Fig. S2B) gives a very poor fit at $(TG)_4$ concentrations below ~ 550 nM. This systematic deviation demonstrates the necessity of postulating the formation of NON complexes for a satisfactory description of the data when NC is in excess. The curve for the simple fit (dashed line in Fig. 4 of the paper) contains both of these systematic deviations.

Isothermal Titration Calorimetry Analysis

One other approach with which we characterized the interactions between NC and $(TG)_4$ was ITC. The methods used were as follows: the calorimetric titrations were performed on a VP-ITC titration calorimeter from Microcal, Inc. (Northampton, MA). Titrations were carried out at 35 °C in 1x HBS, 100 μM TCEP, 1 μM $ZnCl_2$, 5 mM β-mercaptoethanol, and 0.05% polyethylene glycol 20000. In a typical "forward" experiment in which a solution of $(TG)_4$ was injected into a solution of NC protein, the concentrations of the NC protein and $(TG)_4$ were 18 μM and 210 μM, respectively. 10-μl aliquots of a $(TG)_4$ solution were injected from a 250-μl syringe into a rapidly mixing (400 rpm) solution of NC (cell volume = 1.3472 ml). Control experiments involved injecting identical amounts of $(TG)_4$ solution into buffer without NC. In the reverse experiment, the concentration of NC protein and $(TG)_4$ were 119 μM and 10 μM, respectively. For the N-term mutant NC experiment, the concentrations of the mutant NC protein and $(TG)_4$ oligo were 187 μM and 20 μM, respectively. All other experimental conditions were identical and as described above.

We first measured the heat released as the oligonucleotide was titrated into a solution of 18 μM NC. As shown in Fig. S3A, there was a monotonic decrease in the exothermic heat of binding with each successive injection of $(TG)_4$ until saturation was achieved at a molar ratio of approximately 1.0. It is interesting to note that the titration curve is centered around a molar ratio of 0.5, in agreement with the idea, discussed above, that a single oligonucleotide molecule can bind two NC molecules in a NON-type complex (14). The total heat released over the course of the titration was ~18 kcal/mol.

We also performed a "reverse" ITC experiment in which NC was titrated into a solution of 10 μM $(TG)_4$. As seen in Fig. S3B (red curve), the results here were strongly suggestive of the complexity of the reaction. No clear saturation was attained in this titration; the curve exhibited a small peak at a molar ratio of 0.5 (when one NC molecule

had been added per two $(TG)_4$ molecules) and was steepest at a molar ratio of 1.0. A total of ~ 12 kcal/mol was released over the course of the titration. Taken together, the ITC data did not point to a precise stoichiometry for the interactions between NC and $(TG)_4$, but this is perhaps not surprising in light of the reactants' apparent ability to form NO, ONO, NON, and other, higher-order complexes (see above). The dependence of the total heat released upon the direction of the titration also underscores the complexity of the interactions between NC and $(TG)_4$.

In addition, we varied the conditions of the titration in an effort to probe the role of electrostatic forces in the binding of NC to $(TG)_4$. In one such experiment, NC was added to $(TG)_4$ as in Fig. S3B, but the NaCl concentration was 250 mM rather than 150 mM. As shown in the blue curve in Fig. S3B, the profile was very similar to that in 0.15M NaCl, but the total heat released was ~8 rather than 12 kcal/mol. We also tested the binding of the "N-term" mutant protein, in which the five lysine and arginine residues between positions 3 and 14 of NC had been replaced with alanine. The profile observed with mutant protein in 150 mM NaCl was also similar to the wild-type, but only 4-5 kcal/mol were released (data not shown). The dramatic reduction in the total heat released in the presence of additional counterions, or when the charge on NC had been decreased, suggests that coulombic attractions play a major role in the interactions between NC and $(TG)_4$.

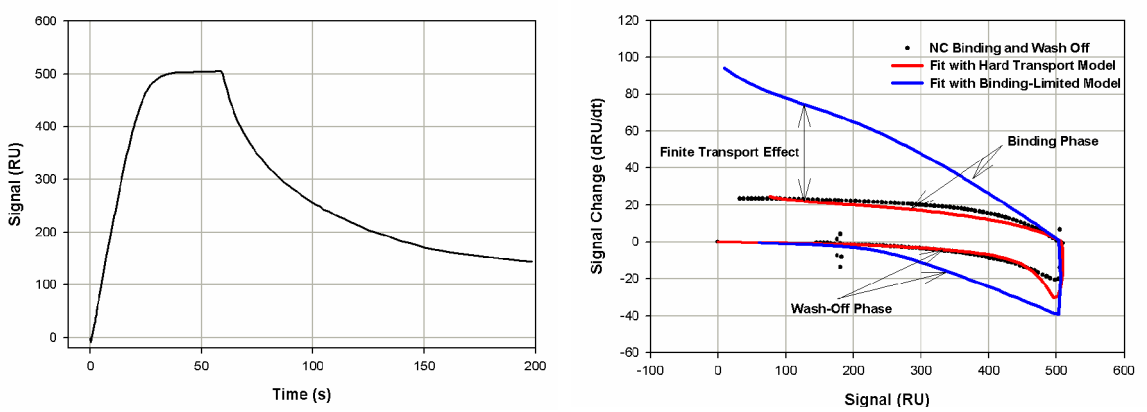

Figure S1. Characteristics of Transport-limited Binding. (A) SPR plot for binding of NC to $(TG)_4$. A solution of 200 nM NC was passed across an SPR chip containing 157 RU of $(TG)_4$, under the conditions described in the paper. (B) Transient phase diagram in which the data in Fig. 1A (black dots) is analyzed by plotting the slope of the SPR signal (dRU/dt) *vs.* the signal. In the initial or "binding" phase of the experiment, the slope is positive with a value ~ 22 RU/s as the amount of complex (measured in RUs) increases from ~0 to 500 from left to right. In the wash-off phase, the slope is at first negative, at RU = 500, and then gradually increases to zero at the end of the experiment as the signal declines from 500 to 0 (right to left in graph). The blue curve is a fit of the data in which binding is assumed to be slow relative to transport; it can be seen here that the slope would decrease continually as the binding proceeded. The red curve is a fit of the data in which binding is assumed to be rapid relative to transport; here the slope is nearly constant throughout the binding phase of the experiment.

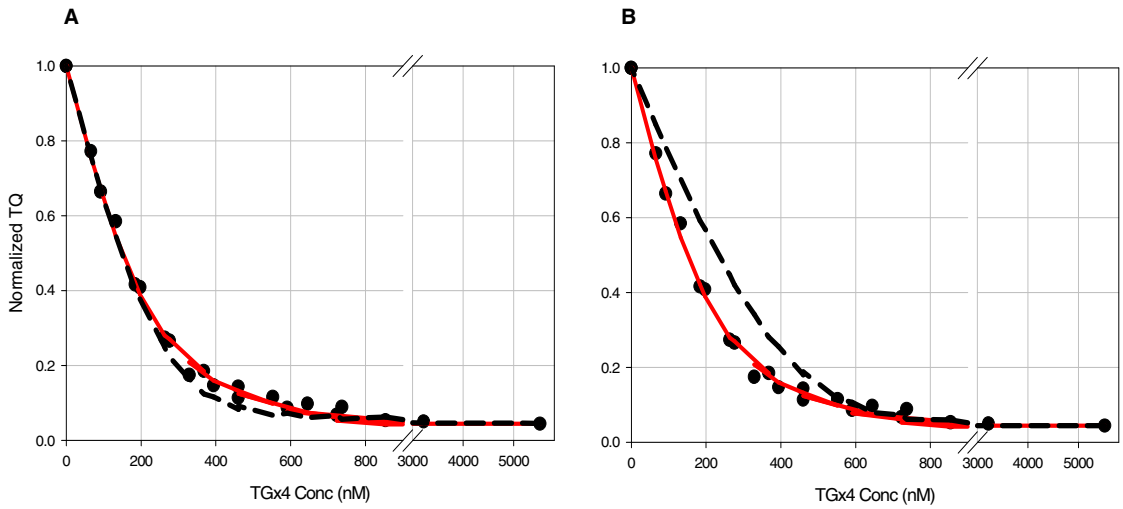

Figure S2. Requirement for both K_d2 and K_d3 for satisfactory modeling of TFQ data. (A) Fit with only K_d1 and K_d2. The dashed line shows the attempt to fit the data (black dots) using only K_d1 and K_d2. (B) Fit with only K_d1 and K_d3. The dashed line shows the attempt to fit the data (black dots) using only K_d1 and K_d3. The fit obtained using all three K_d's is shown by the solid red line in both panels.

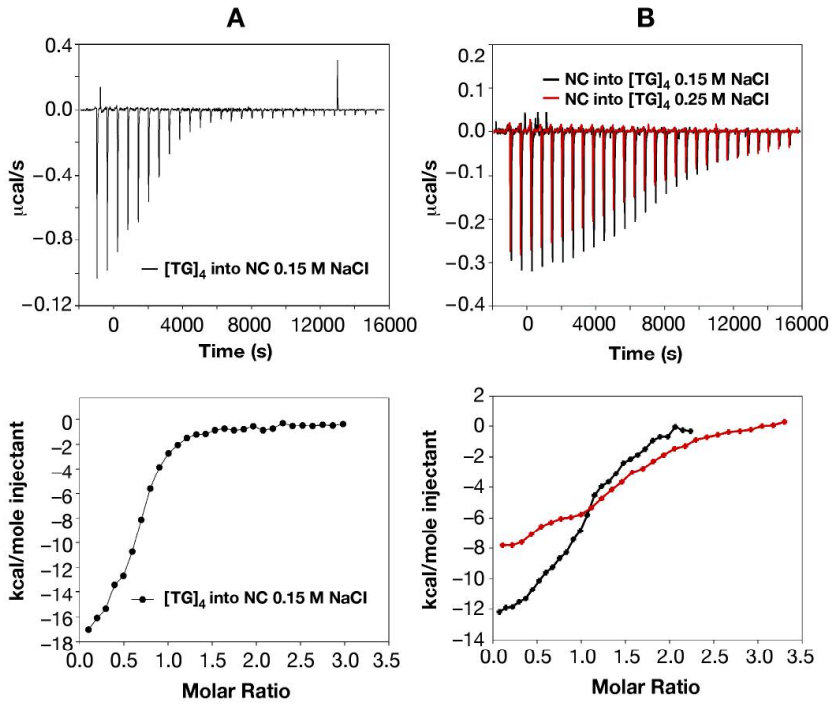

Figure S3. ITC analysis of the interaction of NC with $(TG)_4$. (A) $(TG)_4$ was injected into a solution of 18 μM NC. (B) NC was injected into a solution of 10 μM $(TG)_4$ (black traces). In another experiment, NC was injected into a solution of 20 μM $(TG)_4$ in a buffer containing 250 mM NaCl (red traces). Upper panels: direct measurement of heat released in response to injections. Lower panels: plot of kcal/mole injectant versus molar ratio.

References

1. Karlsson, R. (1999) *J Mol Recognit*, **12,** 285-292.
2. Myszka, D.G. (1999) *J. Mol. Recognit*, **12,** 279–284
3. Doyle, M.L., Myszka, D.G. and Chaiken, I.M. (1996) *J Mol Recognit*, **9,** 65-74.
4. Myszka, D.G., Morton, T.A., Doyle, M.L. and Chaiken, I.M. (1997) *Biophys Chem*, **64,** 127-137.
5. Myszka, D.G., He, X., Dembo, M., Morton, T.A. and Goldstein, B. (1998) *Biophys J*, **75,** 583-594.
6. Edwards, D.A. (2004) *J Math Biol*, **49,** 272-292.
7. Gokulrangan, G., Unruh, J.R., Holub, D.F., Ingram, B., Johnson, C.K. and Wilson, G.S. (2005) *Anal Chem*, **77,** 1963-1970.
8. Lakowicz, J.R. (1999) *Principles of Fluorescence Spectroscopy*. Second Edition ed. Kluwer Academis / Plenum Publishers, New Yrok.
9. Vuilleumier, C., Bombarda, E., Morellet, N., Gerard, D., Roques, B.P. and Mely, Y. (1999) *Biochemistry*, **38,** 16816-16825.
10. Kuzmic, P., Hill, C., Kirtley, M.P. and Janc, J.W. (2003) *Anal Biochem*, **319,** 272-279.
11. Forster, M.R. (2000) *J Math Psychol*, **44,** 205-231.
12. Wicherts, J.M. and Dolan, C.V. (2004) *Structural Equation Modeling*, **11,** 45-50.
13. Zucchini, W. (2000) *J Math Psychol*, **44,** 41-61.
14. Wiseman, T., Williston, S., Brandts, J.F. and Lin, L.N. (1989) *Anal Biochem*, **179,** 131-137.