Analysis of ligand binding to macromolecules using kinetic and polynomial approaches

Fatima Jamal [a], **Don Kulasiri** [a] **and Sandhya Samarasinghe** [a]

[a] Centre for Advanced Computational Solution (C-fCAS), Lincoln University, Christchurch, New Zealand
Email: Fatima.Jamal@lincolnuni.ac.nz

Abstract: One of the most interesting property exhibited by most of the proteins is its ligand binding and cooperative interaction ability. The cooperative interactions occur when ligand binding on one site alters the interactions at a distant ligand binding site of same protein. This cooperative and allosteric interactions are an important and interesting phenomenon, and they can be seen throughout biology. The first and probably the most studied cooperative molecule is Haemoglobin which was studied by Bohr, who showed that the binding curve of $Hb-O_2$ interaction is sigmoidal instead of hyperbolic. The process of cooperative binding and allosteric binding of ligands to the macromolecules or proteins adds complexity to the ligand binding process studies. To understand these functions we have to look beyond the scope of sequence of nucleotides encoding proteins required for a process and look into the network of interactive and interdependent regulatory networks as system which allows certain regulatory expressions to happen. The expression of these networks in the form of computer models that extracts the topology of events occurring in the systems representing it in a systematic way, makes the system easy to explain and predict. There are a number of ways in which a process or network can be modelled for easy explanations and understanding. In the current study we will be analysing ligand binding interactions using two different approaches (1) BP (Binding polynomial) approach; (2) Kinetic modelling approach.

The use of BP and kinetic modelling approach will be discussed in this study on two cooperative ligand binding proteins (1) TRAP (tryptophan RNA-binding Attenuation protein) and (2) CaM (Calmodulin). The BP approach has been used to analyse the cooperativity in the ligand binding sites in TRAP protein in *B. Subtilis* but it is new to be used on cooperative binding in CaM highlighting the possible use of this technique on the interaction analysis of cooperativity in Ca^{2+} – CaM system. The kinetic modelling approaches are more popular, and therefore have already been used on both TRAP and CaM analysis. In the current study we will explain and analyse these approaches in order to give an easier and detailed picture of the interaction to the readers, along with extracting useful insights from the models revealing better interaction properties and biological significances of different parameters and concentration values on the ligand binding process. The current study will not only help readers in getting detailed insight about the cooperative step-by-step binding on TRAP and CaM, but also help in getting an application based comparison of the two modelling approaches for analysing cooperativity.

Keywords: *Cooperativity, ligand binding, TRAP, CaM, computational biology, BP, ODEs*

1. INTRODUCTION

Protein and ligand interactions are one of the most important and common phenomenon occurring in living organisms (Dunn, Michael F, 2010). The ligand binding on proteins delivers a very important attribute to protein functioning with high specificity for a particular set of ligands and also delivers highly specific functions in return (Soderling, 1999). The process of ligand binding is a lot of times found to be cooperative as well as allosteric in nature. The ligand binding on proteins can either be positively or negatively cooperative in nature (Ricard and Cornish-Bowden, 1987). Proteins like TRAP (Gollnick 2001), haemoglobin (Pauling, 1935), myoglobin (Fenderson 1985), CAM (Faas et al., 2011), etc. are common examples of positive cooperativity in ligand binding. The Avidin (Zhao, 1993) and glyceraldehyde-3-phosphate and the enzyme glyceraldehyde-3-phosphate dehydrogenase (Koshland, 1968) are examples of negative cooperativity.

The expression of genes encoding for proteins required for synthesis of Tryptophan in *B.Subtilis* is regulated by a protein TRAP made up of 11 identical subunits forming a ring shaped complex. The expression is regulated under conditions when free tryptophan levels are high within cell or when the amino acylation of tRNA-tryptophan is low. Under low levels of free tryptophan within cell, the TRAP protein complex is not activated and the transcription of protein coding region is continued. When free tryptophan levels are high, the TRAP complex gets activated (by binding 11 L-tryptophan ligand molecules) and binds on the mRNA transcript. The binding of trp on TRAP is found to be cooperative in nature, the reason and source of this cooperativity is a subject of research. Another ligand binding protein is CaM which is a regulatory protein that changes the activity and adds specificity to different signalling molecules (e.g., adenylyl cyclase, protein kinases and more than 100 others) related to Ca^{2+} signalling in neuronal cells. The Ca^{2+} binding on CaM is found to allosterically change the CaM protein structure which helps in binding of other molecules and helping a diverse range of interactions to take place (Xia, 2005). CaM has 2 lobes, the N and C lobes with each having 2 Ca^{2+} binding sites (Faas et al., 2011). The 2 lobes and four binding sites of CaM are found to be working at different binding affinities for Ca^{2+} as well as other proteins bringing about cooperativity. These homotropic and heterotrophic cooperativities seen in Cam-Ca^{2+} interactions is very important and interesting to study (Cheung, 1980).

Although a lot of studies have already been done on ligand binding on CaM and TRAP proteins, the kinetics of the interaction are known to be adding complexity to study this protein (Gollnick 2001), (Faas et al., 2011). In this paper we explore and compare two computational modelling methods on the CaM and TRAP ligand binding, in order to get answers related to their conformational changes and applicability of these approaches on heterotrophic and homotropic cooperative proteins. The two modelling methods, BP approach and ODE (Ordinary Differential Equation) approach, are used to analyse cooperativity in this paper and the ODE method is more adaptable and user-friendly as compared to the BP approach.

2. ANALYSIS USING ODE APPROACH

The TRAP-trp and CaM-Ca^{2+} interaction using the ordinary differential approach is explained in this section of the paper. The ordinary differential equation approach is useful to provide a model for a known system where the rate of change of a component can be expressed in terms of levels of other components in the system. Since the protein interactions with its ligands are based on their concentration change with time, the ODE type modelling approach is a good way of expressing and analysing the reaction process (Lang, 1997 and Molenaar, 2000). In the current section, the ODEs and the assumptions driving these ODEs from their ligand binding models are described.

2.1 Network Model and ODEs for TRAP

One of the example of ODE modelling used for TRAP is the research paper by Kleckner et al, 2013. The research is based on the assumption that if cooperativity is present, then the ligand binding process can be divided into few steps explained by the changes in reaction rates (a number of steps were studied with this model but the model producing the best fit to the binding data was selected in the paper, which is a two-step bind-bind model). The model considers binding at initial 2 binding sites highlighting the significant reaction rate change occurring after the initial binding has happened. The two-step model (Figure 1a) is given below, W is the ligand (trp), a pair of unbounded binding sites are represented as T_T, the species with one ligand bound can either be TW_T or T_TW and a fully bounded binding sites as TW_TW. The ODEs and fitted parameters are described in the paper (Kleckner et al, 2013).

The values for k1f (1st binding), k1b (dissociation of 1st ligand), k2f (2nd binding) and k2b (dissociation of 2nd ligand) are obtained by fitting the equations on the fluorescence binding data obtained via experimental studies.

Therefore, the present modelling approach utilises the binding data for generation of ODE models which can then be utilised to obtain fitted parameters that can give insights about the binding process.

2.2 Network Model and ODEs for CaM

To study the kinetics of Ca^{2+} binding to CaM a two-step model is proposed by Faas et al. This model for CaM similar to TRAP model, describes the cooperative changes occurring in the ligand binding sites after the initial binding has occurred. The model diagram is given in figure 1b where X can be either N or C depending upon the lobes being studied. The tensed or unbounded state of CaM is represented by T, which changes to a relaxed R state after initial ligand binding. The k_{on} and k_{off} describe the binding and dissociation rates of Ca^{2+} and their values are different for C and N lobes. The parameter values and binding rates were obtained by fitting the two-step model to the fluorescence data for ligand binding. The ODEs and parameter values can be obtained from the original paper (Faas et al., 2011).
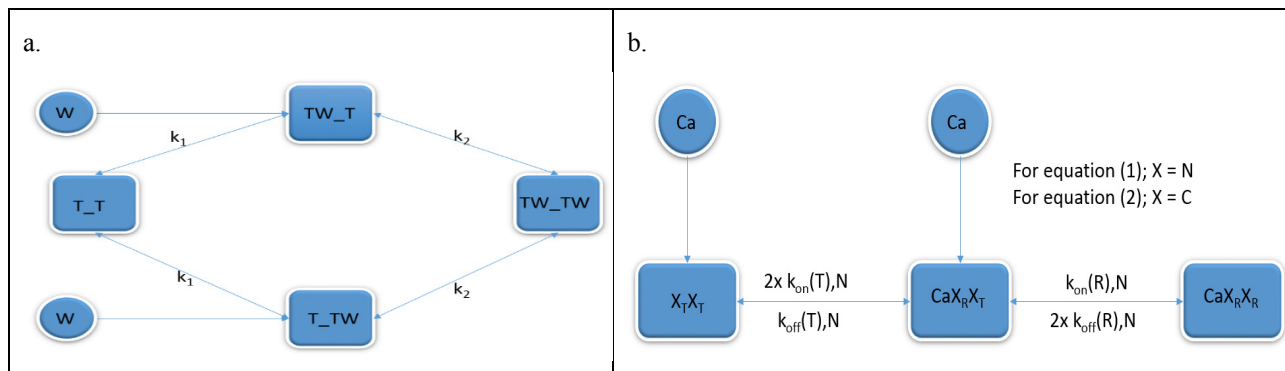
Figure 1. a) The network diagram of CaM-Ca^{2+} interaction model where (Ca) shows states of CaM requiring Ca^{2+} and b)The two states interaction model of trp binding on TRAP.

3. ANALYSIS USING BP APPROACH (TRAP and CaM)

The BP (*I*) based approach can be used with different methods of derivations and calculations. All of the methods calculate the values of concentration of bounded and unbounded form of proteins and concentration of free ligand along with parameters like binding and interaction constants, these values are useful as they describe the ligand binding process in each protein. BP contains all the information about a system, as *I* is the sum of interactions in all possible configurations of ligand binding sites in a protein. In contrast to the parameters obtained for a particular model, the parameters for binding polynomial have a general validity for a system under study. The value of binding polynomial is dependent upon the structure of the protein (number of binding sites and shape) and the concentration of free ligand molecules (c_t). This approach is fast and a systematic way of analysing the ligand binding process in any macro-molecule, if its structure is known (Saroff, 1997). This method of modelling has already been used and described by Saroff et al., but this is a new method to be applied for a heterotopic ligand binding protein CaM.

The generation of a BP was formally based on matrix combinatorial method, which is complex and time consuming in case of large number of binding sites. Therefore, Saroff and Kiefer developed a fast combinatorial algorithm for deriving BP for any number of sites, either large or small, overlapping or linear or circular protein. In the original model Saroff and Keifer found that with this method the generation of BP was 1000 times faster than Mathematica based matrix formulations for up to 50 number of binding sites. The original algorithm is written in FORTRAN for 11 identical binding sites on TRAP and is defined by 3 loops stated below, and the BP is generated by summing the values of bounded ligands generated by the model.

1) Number of bound ligands on CaM (Lb)
2) Number of occupied-occupied interactions (Ipp)
3) Number of occupied sites at the end (Nen).

The model for TRAP is explained in Saroff et al., and the model described for TRAP was adopted and changed to CaM and Ca^{2+} interaction. The polynomial thus obtained is discussed in the section below.

4. RESULTS AND DISCUSSION

4.1 ODE Analysis

The models discussed in section 2 of this paper are simulated in order to get results or inferences that give some insightful knowledge. The binding kinetics of both the proteins being studied (TRAP and CaM) were analysed in their respected research papers. The experimental binding data was used to simulate both the models at different experimental conditions and model fitting was used to determine the parameter values.

In case of TRAP kinetic analysis by Kleckner et al., once the fitting is done (using DYNAFIT) and the parameter values are known, the model is simulated using the initial conditions and other inputs based on the experimental conditions and the results are compared. The figure 2 shows the results of experiment and model data overlapping which confirms the suitability of the model. The ODE modelling helped Kleckner et al. to infer some useful insights based on the results obtained, they suggested that ligand binding on TRAP occurs in two steps (initial binding that cooperatively promotes the further binding on TRAP). The binding and dissociation rates of the model shows the affinities of first and second binding and it is found that the second binding is moderately faster (1.4-2.5 fold), confirming a positive cooperativity in ligand binding on TRAP.
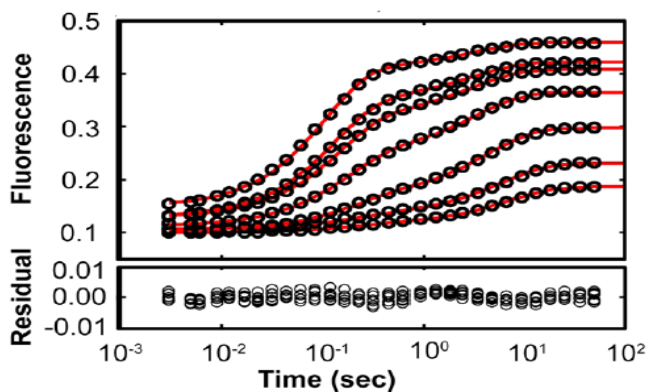
Figure 2. The fluorescence response of the fully-bounded pair of binding site at different initial TRAP concentrations. The red lines are the fit to the experimental data as obtained from the 2-step BB model. The increase in concentration dependant fluorescence with time can be seen, which is divided into 2 kinetics steps. The initial slow response is seen till time scale 10^{-1} sec and then a fast increase in the fluorescence is clearly visible (Kleckner et al, 2013).

In CaM binding analysis by Faas et al. the binding dynamics were also measured using fluorescence analysis, and the data was then analysed using the two step cooperative binding model given in section 2 above. The accuracy of fits were determined by analysing model over a wide range of experimental data based on fitting methods (Faas et al. 2007). To analyse the difference in binding at each lobe, the experiments were done using CaM mutants which only support binding either at N or C lobe at a time. The analysis using 2 step model confirmed that Ca^{2+} binds with higher affinity to C lobe than N lobe and that both the lobes show cooperativity but at different levels and by different mechanisms (Faas et al., 2011).

4.2 Results of the BP Analysis

The BP generated for 11 binding sites on TRAP is given in (Saroff, 1997) and a similar binding polynomial generated for four number of binding sites as in CaM is given in equation below.

$$I = 1 + 2\alpha_1 c_t k_t + 2\alpha_1{}^2 c_t k_t + \alpha_1{}^2 c_t{}^2 k_t{}^2 + 2\alpha_1{}^3 c_t{}^2 k_t{}^2 + 2\alpha_1 \alpha_2 c_t{}^2 k_t{}^2$$
$$+ \alpha_1{}^2 \alpha_2 c_t{}^2 k_t{}^2 + 2\alpha_1{}^2 \alpha_2 c_t{}^3 k_t{}^3 + 2\alpha_1 \alpha_2{}^2 c_t{}^3 k_t{}^3 + \alpha_2{}^3 c_t{}^4 k_t{}^4 \qquad \dots(1)$$

The *I* is the sum of BP values and α1 and α2 in equation represent unoccupied-occupied and occupied-occupied interactions respectively for both TRAP and CaM. The parameters k_t = intrinsic constant for ligand binding to protein and c_t= concentration of free protein. This polynomial is the sum of all the possible interactions happening among the neighbouring sites after or before ligand binding. The BP shows how the interaction between adjacent subunits will differ if both are bound or unbound or if one is bound while other is not.

As used by Saroff and Keifer, the fitting of the binding data on equation 2 can be used to derive the parameter values. In equation 2, v_t is average number of ligand molecules bound on the protein, *I* is the binding polynomial and c_t is the concentration of the free ligand molecules. The average bound density and concentration of free ligands are the values important to obtain a binding isotherm curve. The shape of this binding isotherm is a

decent way to visualise any cooperativity (should be steep sigmoidal, if present) in ligand binding process on a protein.

$$v_t = \frac{\partial \ln I}{\partial \ln c_t} \qquad \qquad \dots(2)$$

For CaM-Ca^{2+} interaction, the experimental fluorescence data (Crouch, 1980) was used to obtain the parameter values for the binding polynomial by fitting the data with equation 2. The fitting was done with MCMC toolbox (http://helios.fmi.fi/~lainema/mcmc/) using mean absolute error function (MAE) and the best parameter values were selected based on the lowest MAE value for 1000 realizations. In case of TRAP (Saroff, 1997) the experimental data fitting was done using Nelder and Mead method (Nelder, 1965) with best fit determined using least sum of square error. Results in the form of binding isotherm plots with the experimental data are given in the figure below. For CaM, the results of fitting and experimental data show a close fit proving the validity of the model. The binding isotherm for TRAP is obtained from Saroff, 1997. The sigmoidal curve of the isotherm shows the presence of cooperativity in ligand binding in CaM as well as TRAP.

a) CaM: $\alpha_1 = 0.57$; $\alpha_2 = 0.51$; log $k_t = 5.2$ b) TRAP: $\alpha_1 = 0.50$; $\alpha_2 = 0.60$; log $k_t = 5.2$

(Saroff, 1997)
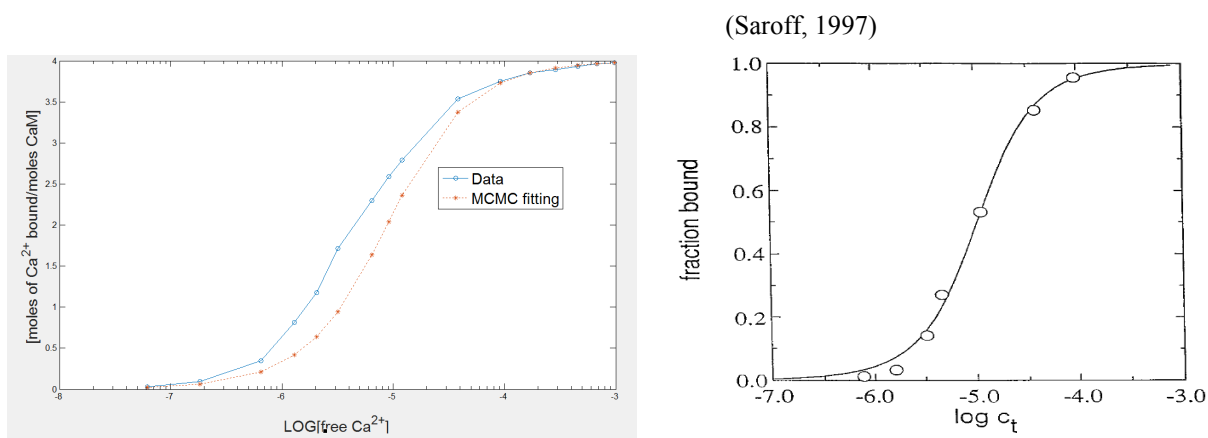
Figure 3. a) The binding isotherm of CaM-Ca^{2+} (blue) from the experimental data and (red) obtained from MCMC fitting for parameter values. b) The binding isotherm for TRAP-trp interaction, the sigmoidal curve is the experimental data and the dots represent the fit from model. The values of parameters obtained after fitting for CaM and from the Saroff, 1997 for TRAP are also given above.

The values of parameter can be used as a tool to infer some insights in the source of cooperativity visible in the binding isotherm. It is suggested by Saroff and Keifer that in case of TRAP, the value of α_2 is greater than α_1, showing an increase in interaction energy after the initial ligand binding therefore, a destabilization energy gives rise to positive cooperativity. Based on the above suggestions and parameter values, it can be suggested that the source of cooperativity in case of CaM is from the stabilizing interactions as the interaction between already bounded sites enhances the process of further binding.

5. COMPARATIVE ANALYSIS

Both the approaches have their own merits and demerits and their differences cannot be explained without taking into consideration certain factors that affect the modelling process. The availability of experimental data, consideration of structural information and the type of protein the approach is used for plays a big role in describing the preference of certain approach over the other. As the type of protein and the ligands binding to it are same for both the approaches compared here, it makes it easier to compare the feasibility of these approaches for Ca^{2+} binding to CaM. But the applicability of these approaches to different ligand binding proteins is the main purpose of comparing these approaches, so that they can be applied to different ligand binding molecules later.

The generation of BP in the first approach is quite fast as the polynomial can be generated using the algorithm given by Saroff and Keifer, whereas the ODE type approach is based on the generation and selection of right set of ODEs in order to achieve a good model for a specific protein. The generation, testing and simulation of

each set of ODE for a protein can be a tedious job lengthening the model generation process, whereas BP generation can be straight forward and quick approach for model generation.

The BP approach requires a set of experimental data in order to get the parameter values for the polynomial, without which the extraction of inferences is not possible. The ODE type approach on the other hand require experimental in order to propose set of ODEs for a particular reaction as well as for fitting and simulating the results. Although the requirement of data should not be considered as a demerit any of these approaches because even though it makes the modelling process longer to proceed but the binding data can be generated virtually based on the literature and other binding information available. The BP approach requires the structural knowledge and information for its polynomial generation as well as fitted parameter selection. The lack of structural information, in case of a new protein or proteins whose structure is not or cannot be studied well can be a demerit questioning the universal applicability of this type of approach.

Obtaining and explaining insights are the main purpose of these modelling approaches as this is what elevates the modelling data explanations over the experimental data. The values obtained from set of ODEs are straight forward and easy to understand, as the input and output both work on the changes in concentration (for these models) although, having the right set of ODEs is the main task. All the different ODE model suggestions are required to be tested with the experimental data in order to obtain the closest model as done for both CaM and TRAP. Although there could be any number of interactions for big or small proteins, the good thing about ODE models is that it can be summed up into the small number of ODEs and still represent the interaction well. As in case of TRAP, it has 11 number of binding sites and 11 TRAP-trp interactions but still it is well represented by a two-step model by Kleckner et al. In case of heterotrophic protein CaM, where different lobes show different levels of cooperativity the ODE based models can have similar or different equations and interactions can be analysed individually. The BP approach on the other hand did not take into consideration the difference in ligand binding interactions on N and C lobe in CaM, but takes the protein's overall cooperative interactions into account. The use of BP generation algorithm is preferred for proteins with large number of binding sites like TRAP, as the generation of polynomial for such large numbers is a lengthy process. But the polynomial generation for small proteins like TRAP is not as lengthy, minimising the preferable use of this approach for smaller proteins.

The inferences obtained through the results in both the approaches hold valuable insights and reasonable explanations which can be used for visualising and explaining ligand binding process in any protein. But for the current proteins it is found that the generation and analysis of ODEs was a simpler approach and the inferences obtained hold better value and broader explanations.

REFERENCES

Dunn, Michael F. (Apr 2010). Protein–Ligand Interactions. *John Wiley & Sons Ltd, Chichester*. http://www.els.net [doi: 10.1002/9780470015902.a0001340.pub2]

Soderling, T. R. (1999). The Ca $^{2+}$–calmodulin-dependent protein kinase cascade. *Trends in biochemical sciences*, *24*(6), 232-236.

Ricard, J., & Cornish-Bowden, A. (1987). Co-operative and allosteric enzymes: 20 years on. *The FEBS Journal*, *166*(2), 255-272.

Elliott, M. B., Gottlieb, P. A., & Gollnick, P. (2001). The mechanism of RNA binding to TRAP: initiation and cooperative interactions. *Rna*, *7*(1), 85-93.

Pauling, L. (1935). The oxygen equilibrium of hemoglobin and its structural interpretation. *Proceedings of the National Academy of Sciences*, *21*(4), 186-191.

Royer, W. E., Love, W. E., & Fenderson, F. F. (1985). Cooperative dimeric and tetrameric clam haemoglobins are novel assemblages of myoglobin folds. *Nature*, *316*(6025), 277-280.

Faas, G. C., Raghavachari, S., Lisman, J. E., & Mody, I. (2011). Calmodulin as a direct detector of Ca^{2+} signals. *Nature neuroscience*.

Zhao, S., Walker, D. S., & Reichert, W. M. (1993). Cooperativity in the binding of avidin to biotin-lipid-doped Langmuir-Blodgett films. *Langmuir*, *9*(11), 3166-3173.

Conway, A., & Koshland, D. E. (1968). Negative cooperativity in enzyme action. Binding of diphosphopyridine nucleotide to glyceraldehyde-3-phosphate dehydrogenase. *Biochemistry*, *7*(11), 4011-4023.

Xia, Z., & Storm, D. R. (2005). The role of calmodulin as a signal integrator for synaptic plasticity. *Nature reviews. Neuroscience*, *6*(4), 267.

Cheung, W. Y. (1980). Calmodulin plays a pivotal role in cellular regulation. *Science*, *207*(4426), 19-27.

Lang, S. (1997). Ordinary differential equations. In *Undergraduate Analysis* (pp. 538-562). Springer New York.

Mattheij, R., & Molenaar, J. (2002). *Ordinary differential equations in theory and practice*. Society for Industrial and Applied Mathematics.

Kleckner, I. R., McElroy, C. A., Kuzmic, P., Gollnick, P., & Foster, M. P. (2013). Homotropic cooperativity from the activation pathway of the allosteric ligand-responsive regulatory trp RNA-binding attenuation protein. *Biochemistry*, *52*(49), 8855-8865.

Saroff, H. A., & Kiefer, J. E. (1997). Analysis of the Binding of Ligands to Large Numbers of Sites: The Binding of Tryptophan to the 11 Sites of thetrpRNA-Binding Attenuation Protein. *Analytical biochemistry*, *247*(1), 138-142.

Helton, J. C., Johnson, J. D., Sallaberry, C. J., & Storlie, C. B. (2006). Survey of sampling-based methods for uncertainty and sensitivity analysis. *Reliability Engineering & System Safety*, *91*(10), 1175-1209.

Ling, H., Kulasiri, D., & Samarasinghe, S. (2010). Robustness of G1/S checkpoint pathways in cell cycle regulation based on probability of DNA-damaged cells passing as healthy cells. *Biosystems*, *101*(3), 213-221.

Marino, S., Hogue, I. B., Ray, C. J., & Kirschner, D. E. (2008). A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of theoretical biology*, *254*(1), 178-196.

Crouch, T. H., & Klee, C. B. (1980). Positive cooperative binding of calcium to bovine brain calmodulin. *Biochemistry*, *19*(16), 3692-3698.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The computer journal*, *7*(4), 308-313.

Faas, G. C., Schwaller, B., Vergara, J. L., & Mody, I. (2007). Resolving the fast kinetics of cooperative binding: Ca2+ buffering by calretinin. *PLoS biology*, *5*(11), e311.