A STATISTICAL ANALYSIS OF RECEPTOR CLUSTERING USING RANDOM GRAPHS

H. Koeppl[1]*, L. J. Schumacher*[2] *and V. Danos*[3]

[1]School of Computer and Communication Sciences, EPFL, Switzerland
[2]St Johns College, University of Cambridge, United Kingdom
[3]School of Informatics, University of Edinburgh, United Kingdom

ABSTRACT

We address the problem of phase transitions in random site graphs. Such graphs were recently proposed as more useful abstractions of proteins and their network of interactions than standard graphs. A set of binding reactions between proteins give rise to the stochastic assembly of protein complexes. We use a rule-based language to represent and simulate such systems. Based on methods from statistical physics we investigate the dependency of the asymptotic size distribution of such complexes with respect to the site configuration and binding affinities of and between proteins, respectively. In particular we discuss a liquidity index of such abstract protein networks and apply it to the problem of chemoreceptor clustering in *E. coli*.

1. INTRODUCTION

High-throughput methods for measuring the binding of one protein to another, such as the yeast two-hybrid system have lead to a characterization of a large number of protein-protein interactions (PPI) in various model organisms. The availability of such interaction networks triggered a series of purely graph-theoretical analyses of such networks [1]. Recently it became clear however [2] that the incorporation of protein domain information into such analysis may affect the conclusions drawn in those previous studies. With respect to simple graphs a more appropriate abstraction for PPI networks with domain information are site graphs [3, 4]. Nodes of site graphs comprise domains or sites at which the graph's edges emerge or terminate.

Here we are interested in a description of the observed clustering of the chemoreceptors in *E. coli*. See [5] for the experimental evidence using cryo-electron tomography of such cluster or patch formation in chemotactic *E. coli* cells. Those tomography experiments shows patches that are generally ellipsoidal shapes of varying size with an average diameter of around 250 nm. The corresponding area of around 50 nm^2 for such a patch could host around 6500 receptors [5].

The involved chemoreceptors are transmembrane receptors and belong to the superfamily of methyl-accepting chemotaxis proteins (MCP) [6]. The residuals in the cytoplasmic part of the receptors that can be methylated were shown to be responsible for the adaptation of the signaling system and modulate the ligand affinity of the receptors. They are also supposed to be related to structural, conformational changes of receptors, i.e., whether their long cytoplasmic part is more stiff or more flexible. The two predominant receptor types are those for aspartate (Tar) and serine (Tsr), with several thousands of molecules per cell. Less abundant are receptors for dipetides (Tap), ribose and galatose (Trg) and redox potential (Aer) with a few hundreds per cell. Receptors of different type were shown to form hetero-trimers [6].

The function of clustering for chemotactic signal transduction is not entirely clear yet. However the extraordinary 50-fold amplification of the signaling cascade of ligand-concentration change to the flagella motor-bias change is partially attributed to some mechanism related to the clustering of the receptors [6]. One hyphothesis is that conformational change upon ligand binding of a receptor can spread to neighboring non-ligand bound receptors [7]. In a series of papers (see [7] and the references therein) Bray and co-workers devised an Ising-like lattice model for this conformational spread. Under this hypothesis the chemotactic system would have to maintain a delicate balance between sensitivity and dynamic range.

2. RANDOM SITE-GRAPHS

Inspired by the framework of [3] for *colored degree-driven random graphs*, we define an ensemble of site graphs that comply with an abstract notion of proteins and their interaction over binding domains [4]. We define a static random graph with sites $\mathcal{G}$ as the tuple $\mathcal{G} \equiv (N, \mathcal{D}, p(\mathbf{X}), \mathbf{T})$, with N the number of vertices, $\mathcal{D}$ the set of domain types, $\mathbf{T}$ the symmetric $|\mathcal{D}| \times |\mathcal{D}|$ domain preference matrix. The domain configuration of a particular vertex is represented by the vector $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{N}_0^{|\mathcal{D}|}$. Accordingly we denote $p : \mathcal{X} \to [0, 1]$ as the colored degree distribution $p(\mathbf{x})$ of $\mathcal{G}$. Fig. 1 shows an example of a site graph. The graph $\mathcal{G}$ defines an ensemble of static graphs and does not account for the growth of such graphs. To connect this notion with the dynamic assembly of proteins into aggregates we next consider the asymptotic state that such an stochastic assembly process reaches. The asymptotic state associates a probability over the number of bindings for any domain pair. It thus induces a distribution over an ensemble of randomly assembled graphs.

We define the dynamics of assembling a member of $\mathcal{G}$ through a set of rules [8]. Rules can be thought of reactions with the important difference that rules can act on partially defined chemical species. A single rule can thus account for many reactions. In this work we focus on the unconditional binding rules, i.e., the probability of a bond

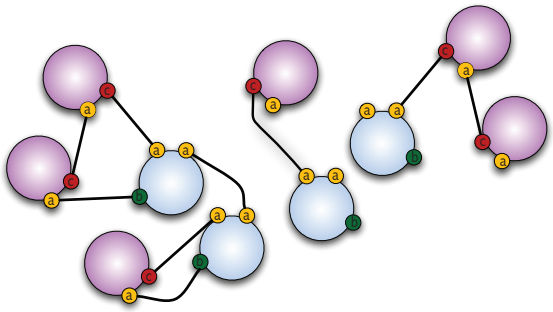

Figure 1. Example of site graph with domain set $\mathcal{D} = \{a, b, c\}$.

between two domain does not depend on conditions other than that the two domains must be unbound. Classical chemical kinetics for such an assembly process is inappropriate as it leads to a combinatorial number of chemical species. Instead we describe the process through the abundance of the $|\mathcal{B}|$ distinct bond types, with $\mathcal{B} \subset \mathcal{D}^2$ the set of unique domain pairs, between which a bond can be formed (this can immediately be read off from the contact map of a rule-set, e.g. Fig. 2). Let us declare the random variable $Z_{ij}(t)$ for the number of bonds between domain $(i,j) \in \mathcal{B}$ at time t. Let us further denote d_j, $j \in \mathcal{D}$ as the total number of domains of type j in the assembly. The bond number $Z_{ij}(t)$ is then bounded from above by $M_{ij} = \min(d_i, d_j)$. We assume in the following that d_j is an invariant of the ensemble of random graphs for every type $j \in \mathcal{D}$. Detailed balance of probabilities gives

$$\begin{aligned} &P[\mathbf{Z}(t+\tau) = \mathbf{z}] = \\ &\{1 - P[\mathbf{Z}(t+\tau) \in \mathcal{E}(\mathbf{z}) \,|\, \mathbf{Z}(t) = \mathbf{z}]\}\, P[\mathbf{Z}(t) = \mathbf{z}] \\ &+ \sum_{\mathbf{y} \in \mathcal{I}(\mathbf{z})} P[\mathbf{Z}(t+\tau) = \mathbf{z} \,|\, \mathbf{Z}(t) = \mathbf{y}]\, P[\mathbf{Z}(t) = \mathbf{y}], \end{aligned}$$

where the first summands accounts for not leaving, within the interval $[t, t+\tau)$, the state $\mathbf{Z}(t) = \mathbf{z}$ to the exit states $\mathcal{E}(\mathbf{z})$ of the transition graph underlying the discrete-state continuous-time Markov chain. In turn, the second term gives the probability to reach the state $\mathbf{Z}(t) = \mathbf{z}$ from input states $\mathcal{I}(\mathbf{z})$ within the interval $[t, t+\tau)$. Here we assume all reactions to be reversible and thus $\mathcal{E}(\mathbf{z}) = \mathcal{I}(\mathbf{z}) \equiv \mathcal{U}(\mathbf{z})$ for all $\mathbf{z}$ with

$$\begin{aligned} \mathcal{U}(\mathbf{z}) = &\left\{\mathbf{y} \mid \forall (i,j) \in \mathcal{B}, \mathbf{y} = \mathbf{z} + \mathbf{e}_i \mathbf{e}_j^T \text{ if } z_{ij} < M_{ij}\right\} \\ &\cup \left\{\mathbf{y} \mid \forall (i,j) \in \mathcal{B}, \mathbf{y} = \mathbf{z} - \mathbf{e}_i \mathbf{e}_j^T \text{ if } z_{ij} > 0\right\}, \end{aligned}$$

where we introduced $\mathbf{e}_i$ as the i-th unit vector of a $|\mathcal{D}|$-dimensional basis. Relying on mass-action kinetics the master equation reads

$$\begin{aligned} \frac{\mathrm{d}P(\mathbf{Z}(t) = \mathbf{z})}{\mathrm{d}t} = &\sum_{(i,j) \in \mathcal{B}} \gamma_{ij}^{-}(z_{ij}+1) P[\mathbf{Z}(t) = \mathbf{z} + \mathbf{e}_i \mathbf{e}_j^T] \\ &+ \sum_{(i,j) \in \mathcal{B}} \gamma_{ij}^{+}(\nu_i + 1)(\nu_j + 1) P[\mathbf{Z}(t) = \mathbf{z} - \mathbf{e}_i \mathbf{e}_j^T] \\ &- \sum_{(i,j) \in \mathcal{B}} (\gamma_{ij}^{-} z_{ij} + \gamma_{ij}^{+} \nu_i \nu_j) P[\mathbf{Z}(t) = \mathbf{z}], \end{aligned} \tag{1}$$

with the number of free type i sites $\nu_i = d_i - \sum_{j=1}^{|\mathcal{D}|} z_{ij}$ and the kinetic on and off-rates γ_{ij}^{+} and γ_{ij}^{-}, respectively. Equation (1) is of a differential-difference type and not solvable analytically, in general. It can be solved however, for the particular case of *non-competitive* binding in the asymptotic case [9]. That is, every domain-type can only be bound to exactly one other domain-type. We focus on this subsequently because the presented receptor clustering model exhibits this feature. After appropriate permutation, the corresponding matrix $\mathbf{T}$ obeys the block-diagonal form with 2×2 anti-diagonal blocks (we exclude self-binding). This pattern is shared with the matrices of on and off kinetic rates. The probability of an admissible unbinding or binding of an edge between two particular domain types is depending on the number of established edges of this type and on the number of free domains of proper type, respectively. The number of free domains is now however not depending on edges of other type. In other words the joint density of bond counts becomes $P(\mathbf{Z}(t) = \mathbf{z}) = \prod_{(i,j) \in \mathcal{D}^2} P(Z_{ij}(t) = z_{ij})$. Thus the problems decouples into $|\mathcal{B}|$ independent subproblems. In this case it is convienient to work with the number of a free domains of a particular type $\nu_i = d_i - z_{ij}$, where domain j is now the unique partner for domain i. For the asymptotic $\bar{G}(\omega_i) \equiv \lim_{t \to \infty} G(\omega_i, t)$ of the moment generating function $G(\omega_i, t) = \sum_{\nu_i \in \mathbb{N}_0} P(\nu_i, t) \omega_i^{\nu_i}$, $|\omega_i| < 1$ we obtain from (1) the scalar second-order ordinary differential equation

$$\begin{aligned} 0 = &\omega_i \frac{\mathrm{d}^2}{\mathrm{d}\omega_i^2} \bar{G}(\omega_i) + (d_i - d_j + 1 + \Gamma_{ij}\omega_i) \frac{\mathrm{d}}{\mathrm{d}\omega_i} \bar{G}(\omega_i) \\ &- \Gamma_{ij} d_i \bar{G}(\omega_i) \end{aligned} \tag{2}$$

with the equilibrium dissociation constant $\Gamma_{ij} \equiv \gamma_{ij}^{-}/\gamma_{ij}^{+}$. The solution of this equation is the confluent hypergeometric function $\bar{G}(\omega_i) = \alpha\ {}_1F_1(-d_j, d_i - d_j + 1; -\Gamma_{ij}\omega_i)$, for $d_i \geq d_j$, where α is a scaling factor that can be determined through the requirement $\bar{G}(1) = 1$. A series representation for this function reads

$${}_1F_1(a, b; \omega) = \sum_{n=0}^{\infty} \frac{(a)_n \omega^n}{(b)_n n!} = \frac{(-a)!(b-1)!}{(b-a-1)!} L_{-a}^{b-1}(\omega)$$

where $(a)_n$ the denotes the rising factorial $(a)_n \equiv a(a+1)(a+2)\cdots(a+n-1)$. The function can also be expressed by Laguerre functions $L_{-a}^{b-1}(\omega)$ with $a < 0$ and $b \geq 1$. The asymptotic expected number of free sites of type i is then obtained by differentiating $\bar{G}(\omega_i)$ and subsequent evaluation at $\omega_i = 1$

$$\langle \nu_i \rangle = \Gamma_{ij} \frac{d_j}{d_i - d_j + 1} \frac{{}_1F_1(-d_j + 1, d_i - d_j + 2; -\Gamma_{ij})}{{}_1F_1(-d_j, d_i - d_j + 1; -\Gamma_{ij})} \tag{3}$$

where we incorporated the appropriate normalization. As we assumed the total number of domains of a particular type to be an invariant of our random ensemble, the expected number of corresponding bonds computes to $\langle z_{ij} \rangle = d_i - \langle \nu_i \rangle$.

3. SIMPLE CLUSTERING MODEL

There is experimental evidence that the basic functional unit of the receptors is a trimer of receptor dimer and that

the tips of the cytoplasmic end of those three receptors dimers can bind, independent of the presence of CheW or CheA.

In [7] Bray and co-workers proposed an interesting model for receptor clustering. First, some facts about the involved agents. The adaptor CheW is a small monomeric protein and is supposed to bind the receptor tip as well as to CheA. CheA forms dimers, whereas a single CheA monomer has five domains, i.e., P1 phoshorylation domain, P2 the CheY and CheB binding site, P3 the dimerization domain, P4 catalytic domain, P5 regulatory domain, that allows to couple CheA to the receptor and to CheW. Analyzing the 3D protein structure Shimizu et al. [7] argue that CheW fits well into the groove between two receptor homodimers. Their proposal is that 3 CheW bind the three grooves in a trimer of dimer receptor-bundle and that at each CheW one part of the CheA dimer is bound on its P5 domain.

The above observations allow us to construct a simple rule-based model [8] for the cluster formation. In particular, we want to determine whether a set of dissociation rates corresponding to the rule set can give rise to a lattice the size of which scales with the size of the protein ensemble. We assume that the basic receptor unit of a trimer of receptor homo-dimers has formed and we spend three binding sites for such a trimer to bind three adaptors CheW, i.e., trimeric receptor agent $\mathtt{Tar}(\mathtt{a},\mathtt{a},\mathtt{a})$ (where we allow the interface of an agent to be a multi-set of domains). We assume for the monomeric adaptor protein CheW two binding sites, i.e., agent $\mathtt{CheW}(\mathtt{b},\mathtt{c})$. Following the experimental evidence we spend two binding sites (the P5 domains) for the dimer CheA, i.e., agent $\mathtt{CheA}(\mathtt{d},\mathtt{d})$. We start with the reversible recruitment of the adaptor proteins to the receptor.

$$\mathtt{Tar}(\mathtt{a}),\mathtt{CheW}(\mathtt{b}) \underset{\gamma^-_{ab}}{\overset{\gamma^+_{ab}}{\rightleftharpoons}} \mathtt{Tar}(\mathtt{a}^1),\mathtt{CheW}(\mathtt{b}^1) \qquad (4)$$

According to the syntax of the deployed language [8], the rules encode that binding of CheW to the receptor trimer is independent of the binding status of neighboring sites of the receptor. Next, we represent the binding of the kinase CheA to the adaptor CheW, where we do not require that the adaptor has already bound the receptor.

$$\mathtt{CheW}(\mathtt{c}),\mathtt{CheA}(\mathtt{d}) \underset{\gamma^-_{cd}}{\overset{\gamma^+_{cd}}{\rightleftharpoons}} \mathtt{CheW}(\mathtt{c}^1),\mathtt{CheA}(\mathtt{d}^1) \qquad (5)$$

For a regular hexagonal structure the two sites of a CheA agent should connect to two different receptor trimers. This is not enforced in this rule-set and CheA can loop back and bind at another CheW on the same receptor trimer. However it is questionable whether in nature such back looping is prohibited. It may well be that the resulting array is not a perfect hexagonal lattice. Furthermore, in simulation studies the frequency of such loops was very low. Here we chose the population size according to the assumed stochiometry and therewith allow for total pairing of the agents, i.e., in the ratio $2\times\mathtt{Tar}(\mathtt{a},\mathtt{a},\mathtt{a})$, $6\times\mathtt{CheW}(\mathtt{b},\mathtt{c})$ and $3\times\mathtt{CheA}(\mathtt{d},\mathtt{d})$. The contact map [8] corresponding to this rule-set is depicted in Fig. 2

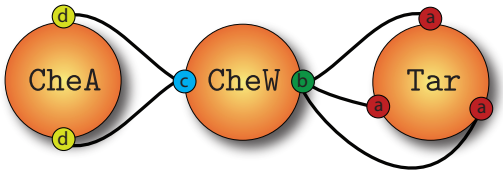

Figure 2. Contact map for the simple receptor clustering model.

4. RESULTS

With the model definition in Section 3 and the expression for the expected number of bond for every $(i,j)\in\mathcal{B}$ at equilibrium (3) we have the ingredients in place to determine the constituents of $\mathcal{G}=(N,\mathcal{D},p(X),\mathbf{T})$. Based on this, we will here soley focus on the percolation threshold or liquidity index for such a random site graph. Namely we know from [4, 3] that this threshold corresponds to the condition that the largest eigenvalue (in absolute value) of the matrix $\mathbf{TE}$ approaches one. The matrix $\mathbf{E}$ constitutes the matrix of second-order moments of the colored degree distribution $p(\mathbf{X})$, i.e., $E_{ij}=\langle x_i x_j\rangle-\delta_{ij}\langle x_i\rangle$. A graph above this threshold (supercritical) features a single connected component that comprises the major part of the vertices, whereas the remaining components account for $\log(N)$-order of vertices. The size of that giant component increases with increasing graph size N. Graphs below that threshold are said to be subcritical.

The degree distribution function $p(\mathbf{X})$ for an ensemble of proteins is predetermined through their respective copy numbers and their domain configurations. For our model we have four domain-types $\mathcal{D}=\{a,b,c,d\}$ and the only domain configuration with non-zero probability are $\mathbf{x}_T=(3,0,0,0)$, $\mathbf{x}_W=(0,1,1,0)$ and $\mathbf{x}_A=(0,0,0,2)$ for the receptor trimer, the adaptor CheW and the autokinase CheA, respectively (refer to Fig. 2). The distribution $p(\mathbf{X}=\mathbf{x})$ denotes the probability that a randomly picked vertex has domain configuration $\mathbf{x}$. If we assume that the copy numbers of our emsemble obeys the stoichiometric ratio, we obtain $p(\mathbf{X}=\mathbf{x}_T)=\frac{2}{11}$, $p(\mathbf{X}=\mathbf{x}_W)=\frac{6}{11}$ and $p(\mathbf{X}=\mathbf{x}_A)=\frac{3}{11}$. This ratio allows for total pairing of domains, thus $\langle x_i\rangle=\frac{6}{11}$ for all $i\in\mathcal{D}$. The second-order moments then read

$$\mathbf{E}=\frac{6}{11}\begin{pmatrix}2&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&1\end{pmatrix}. \qquad (6)$$

As we deal with non-competitive binding the domain-preference matrix exhibits the discussed block-diagonal structure

$$\mathbf{T}=\begin{pmatrix}0&T_{ab}&0&0\\T_{ba}&0&0&0\\0&0&0&T_{cd}\\0&0&T_{dc}&0\end{pmatrix}. \qquad (7)$$

The relation between $\mathbf{T}$ and expected number of bonds at equilibrium $\langle z_{ij}\rangle$ is as follows. The probability that

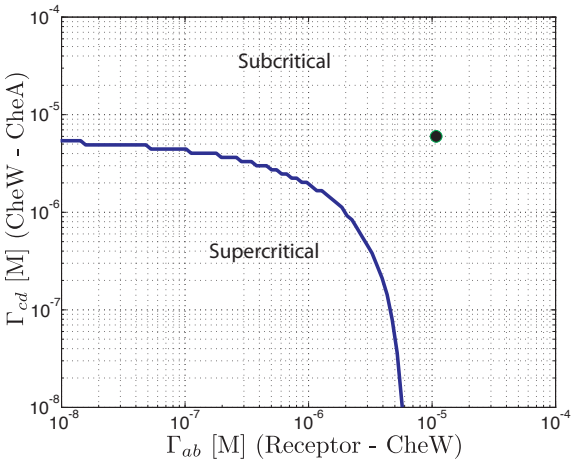

Figure 3. Phase diagram showing the partitioning into a subcritical and supercritical region; reported pair of dissociation constants [10] indicated as solid dot.

an randomly picked domain-pair (i, j) is bond to each other computes asymptotically to the relative frequency of present (i, j) bonds $\langle z_{ij} \rangle$ with respect to all possible bonds, i.e.,

$$\frac{\langle z_{ij} \rangle}{N^2 \langle x_i \rangle \langle x_j \rangle} \equiv \frac{T_{ij}}{N},$$

that establishes the relation between the domain preference and the equilibrium number of expected bonds. In order to compute $\langle z_{ij} \rangle$ for the receptor cluster model we need to determine the dissociation constants Γ_{ij}. According to direct pull-down assays in [10] the receptor - CheW adaptor dissociation constant amounts to $\Gamma_{ab} = 10.8 \pm 0.7$ μM, while the CheW - CheA dissociation constant was measured to be $\Gamma_{cd} = 6.0 \pm 0.2$ μM. With a reported cell volume of 1.4×10^{-15} L for *E. coli* we obtain $\Gamma_{ab} = 9100$ and $\Gamma_{cd} = 5060$ molecules per cell. With an estimated total [11] of 17 ± 2 μM (14330 per cell) Tsr and Tar receptors and the assumption on the stochiometric ratio we can determine the two nonzero entries of $\mathbf{T}$. The largest eigenvalue in absolute terms of $\mathbf{TE}$ turns out to be $\lambda = 0.79$. Although close to the percolation threshold the model is subcritical for this set of dissociation constants. We next sample over these constants in order to determine the percolation boundary. In Fig. 3 we show the phase plane that is partitioned into a subcritical and supercritical phase.

5. CONCLUSION

We connected the framework of random site-graphs with a notion of abstract protein interaction networks by looking at the asymptotic probability distribution of such a network. For unconditional and non-competitive binding the mean occupancy level of sites is given analytically. With this level and the domain configuration of the involved proteins the percolation threshold or liquidity index for the network can be determined. Based on recent experimental evidence we devised a model for chemoreceptor clustering in *E. coli* and analyze its criticality. The model's subcriticality for constants in [10] gives rise to interesting questions. Is there another stoichiometry between receptor, CheW and CheA? May the *in vivo* dissociation constants be different from the ones in [10]?

References

[1] A.-L. Barabasi and Z. N. Oltvai, "Network biology: Understanding the cell's functional organization," *Nature Reviews*, vol. 5, pp. 101–113, 2004.

[2] P. M. Kim, L. J. Lu, Y. Xia, and M. B. Gerstein, "Relating three-dimensional structures to protein networks provides evolutionary insights," *Science*, vol. 314, pp. 1938–1941, 2006.

[3] B. Söderberg, "Random graph models with hidden color," *Acta Physica Polonica B*, vol. 34, no. 10, pp. 5085–5102, 2004.

[4] V. Danos and L. Schumacher, "How liquid is biological signalling?," *Theoretical Computer Science*, vol. 410, no. 11, pp. 1003–1012, 2009.

[5] P. Zhang, C. M. Khursigara, L. M. Hartnell, and S. Subramaniam, "Direct visualization of escherichia coli chemotaxis receptor arrays using cryo-electron microscopy," *Proc. Natl. Acad. Sci. USA*, vol. 104, no. 10, pp. 3777–3781, 2007.

[6] G. L. Hazelbauer, J. J. Falke, and J. S. Parkinson, "Bacterial chemoreceptors: high performance signaling in networked arrays," *Trends in Biochemical Sciences*, vol. 33, no. 1, pp. 9–19, 2008.

[7] T. S. Shimizu, N. Le Novere, M. D. Levin, A. J. Beavil, B. J. Sutton, and D. Bray, "Molecular model of a lattice of signalling proteins involved in bacterial chemotaxis," *Nature Cell Biology*, vol. 2, pp. 792–796, 2000.

[8] V. Danos, J. Feret, W. Fontana, R. Harmer, and J. Krivine, "Rule-based modelling of cellular signalling," in *Proceedings of the 18th International Conference on Concurrency Theory (CONCUR'07)*, Sep 2007, Lecture Notes in Computer Science.

[9] I. G. Darvey, B. W. Ninham, and P. J. Staff, "Stochastic models for second-order chemical reaction kinetics. the equilibrium state," *The Journal Of Chemical Physics*, vol. 45, no. 6, pp. 2145–2155, 1966.

[10] M. S. Boukhvalova, F. W. Dahlquist, and R. C. Stewart, "CheW binding interaction with CheA and Tar," *Journal of Biological Chemistry*, vol. 277, no. 25, pp. 22251–22259, 2002.

[11] M. Li and G. L. Hazelbauer, "Cellular stoichiometry of the components of the chemotaxis signaling complex," *Journal of Bacteriology*, vol. 186, no. 12, pp. 3687–3694, 2004.