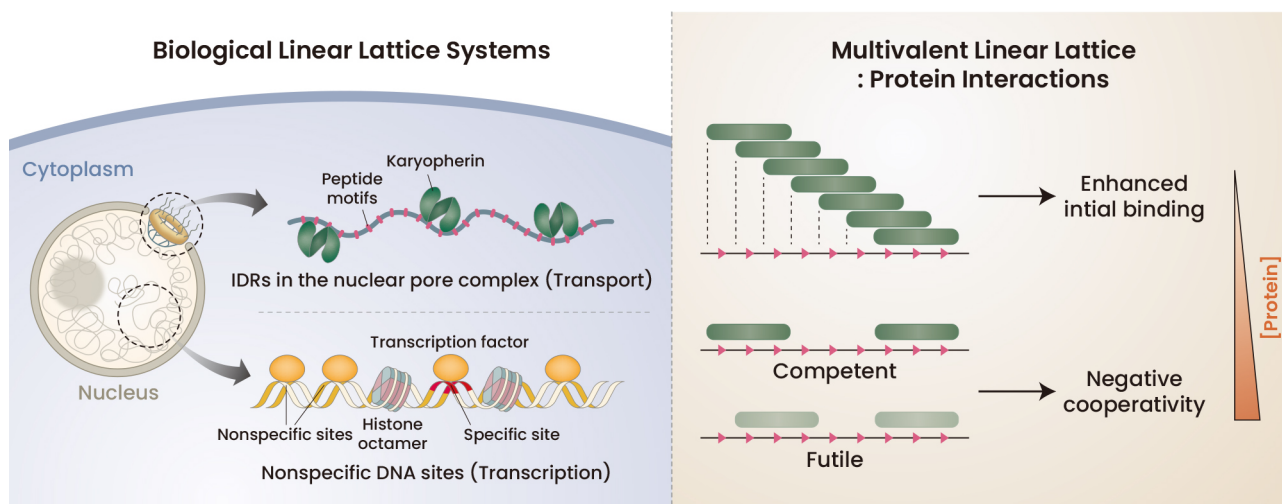


Minireview

Quantitative Frameworks for Multivalent Macromolecular Interactions in Biological Linear Lattice Systems

Jaejun Choi^{1,2}, Ryeonghyeon Kim^{1,2}, and Junseock Koh^{1,*}

¹School of Biological Sciences, Seoul National University, Seoul 08826, Korea, ²These authors contributed equally to this work.
 *Correspondence: junseockkoh@snu.ac.kr
<https://doi.org/10.14348/molcells.2022.0035>
www.molcells.org



Multivalent macromolecular interactions underlie dynamic regulation of diverse biological processes in ever-changing cellular states. These interactions often involve binding of multiple proteins to a linear lattice including intrinsically disordered proteins and the chromosomal DNA with many repeating recognition motifs. Quantitative understanding of such multivalent interactions on a linear lattice is crucial for exploring their unique regulatory potentials in the

cellular processes. In this review, the distinctive molecular features of the linear lattice system are first discussed with a particular focus on the overlapping nature of potential protein binding sites within a lattice. Then, we introduce two general quantitative frameworks, combinatorial and conditional probability models, dealing with the overlap problem and relating the binding parameters to the experimentally measurable properties of the linear lattice-

Received 6 March, 2022; revised 27 March, 2022; accepted 28 March, 2022; published online 27 June, 2022

eISSN: 0219-1032

©The Korean Society for Molecular and Cellular Biology.

©This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>.

protein interactions. To this end, we present two specific examples where the quantitative models have been applied and further extended to provide biological insights into specific cellular processes. In the first case, the conditional probability model was extended to highlight the significant impact of nonspecific binding of transcription factors to the chromosomal DNA on gene-specific transcriptional activities. The second case presents the recently developed combinatorial models to unravel the complex organization of target protein binding sites within an intrinsically disordered region (IDR) of a nucleoporin. In particular, these models have suggested a unique function of IDRs as a molecular switch coupling distinct cellular processes. The quantitative models reviewed here are envisioned to further advance for dissection and functional studies of more complex systems including phase-separated biomolecular condensates.

Keywords: biological linear lattice, combinatorial model, conditional probability model, multivalent binding, overlapping binding site

INTRODUCTION

Recent advances in cutting-edge biotechnologies have provided opportunities to observe unprecedented molecular details of various biological processes (Ha et al., 2022; Mahamid et al., 2016; Oikonomou and Jensen, 2017; Sigal et al., 2018). Interpretation of such observations requires quantitative models dissecting the underlying macromolecular interactions. In turn, the quantitative information allows further understanding and prediction of spatiotemporal regulation of specific cellular processes in dynamically changing environments. The complexity of macromolecular interactions ranges from simple 1:1 binding to formation of phase-separated condensates with multivalent binding among two or more components (Banani et al., 2017; Lyon et al., 2021; Shin and Brangwynne, 2017). In contrast to the 1:1 binding, multivalent interactions are difficult to describe with the simple mass action law but modeled with more sophisticated frameworks accounting for the presence of various molecular states (Bujalowski, 2006; Freire et al., 2009; Wyman and Gill, 1990). Furthermore, the quantitative models are often formulated with large numbers of parameters, and exemplary cases determining these parameters with suitable *in vitro* model systems and methods are exceedingly rare.

A linear or one-dimensional lattice is a relatively tractable multivalent system found in numerous cellular processes. Linear lattices present multiple binding motifs or domains to interact with diverse proteins or multiple copies of identical proteins (Fig. 1) (Cortese et al., 2008; Dunker et al., 2005; Fung et al., 2018). For instance, in many signaling pathways, scaffold proteins such as axin, BRCA1, and Ste5 recruit various target proteins via specific binding sites (Choi et al., 1994; Mark et al., 2005; Wodarz and Nusse, 1998). These scaffold-driven higher-order assemblies are predicted to colocalize and increase the local concentrations of the target proteins and thereby facilitate their interactions for efficient integration and propagation of diverse signals in the cell (Fig. 1A)

(Noutsou et al., 2011; Xue et al., 2013). Another example is the intrinsically disordered regions (IDRs) of some nucleoporins (Nups) present in the nuclear pore complex (NPC) (Fig. 1B) (Frey and Gorlich, 2007; Radu et al., 1995). The Nup IDRs mediate massive yet selective molecular transport between the nucleus and cytoplasm through specific interactions with karyopherin (Kap) proteins carrying macromolecular cargos (Koh and Blobel, 2015; Schoch et al., 2012). These interactions are achieved by multiple interspersed phenylalanine-glycine (FG) motifs on an IDR capturing several Kap molecules (Bayliss et al., 2000).

Finally, nucleic acids are the most prominent linear lattice systems in the cell. In particular, the chromosomal DNA presents the enormous amount of repeating phosphate groups along its backbone, creating electrostatic potentials for nonspecific protein-DNA interactions (Fig. 1C) (Berg et al., 1981; Stracy et al., 2021). Such polyelectrolyte effect is a major driving force (Lohman et al., 1980; Record et al., 1976), particularly at low salt concentrations, for formation of nucleosomes (Shrader and Crothers, 1989; Widom, 1999) as well as for binding of chromatin architectural proteins such as HMG (high mobility group)-box proteins with little specificities for DNA base sequences (Dragan et al., 2004). Even specific DNA binding proteins typically engage their cationic amino acid side chains to neutralize DNA phosphate charges (Jen-Jacobson et al., 2000; Privalov et al., 2011). Thus, these proteins are expected to interact with nonspecific sites that are present in overwhelming excess over specific site in the chromosomal context. In addition, as the copy numbers of many transcription factors (TFs) are considered greater than

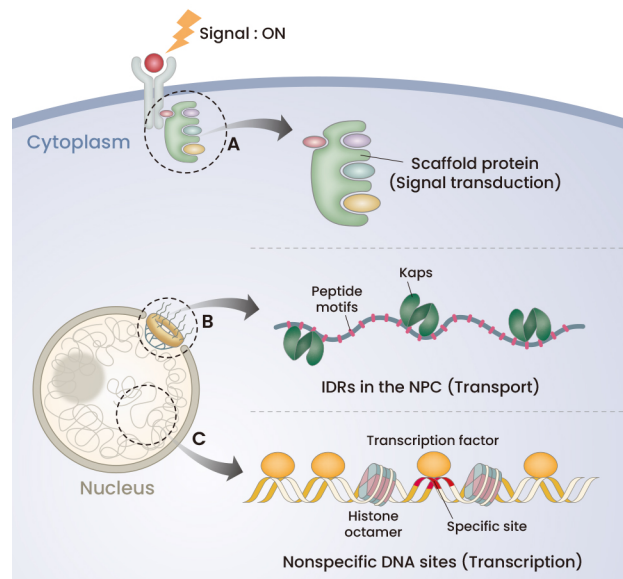


Fig. 1. Schematic illustration of the representative linear lattice systems in cellular processes. (A) Scaffold proteins recruiting diverse binding partners in signal transduction. (B) IDRs in the NPC binding multiple Kaps in nucleocytoplasmic transport. (C) Nonspecific sites on the chromosomal DNA for transcription factor binding.

those of their corresponding specific binding sites on DNA, the majority of these factors may exist *in vivo* as nonspecifically bound states (Bintu et al., 2005; Kao-Huang et al., 1977). The physiological impact of the nonspecific protein-DNA interaction is substantial as demonstrated in the classical study by the von Hippel group (von Hippel et al., 1974) as well as in the recent seminal work by the Phillips group (Brewster et al., 2014). Both groups used *Escherichia coli lac* repressor as a model system to investigate the interplay among the copy numbers of TFs and their binding sites on DNA, the specificity ratio, and the inducer binding affinity in bacterial gene expression. The quantitative models proposed in these studies accurately described and predicted the expression profiles of the genes under the repressor regulation by incorporating nonspecific protein-DNA interactions as a “sink” for RNA polymerase and *lac* repressor.

Taken together, numerous protein-protein and protein-nucleic acid interactions can be perceived as multivalent interactions mediated by linear lattices. Thus, quantitative models for linear lattice systems are indispensable in understanding a broad range of biological processes and may be further extended to dissect more complex systems including phase-separated biomolecular condensates. In this review, we go over two general mathematical frameworks, combinatorial and conditional probability models, for quantitative description of linear lattices. Prior to the detailed derivation of these models, the molecular features of multivalent interactions on a linear lattice will be qualitatively discussed in light of how they are fundamentally different from 1:1 binding or discrete-site systems. The derivation will be supplemented in [Supplementary Information](#) with some detailed mathematical procedures omitted but not immediately evident in the original articles. In the end, a couple of practical examples will be discussed where the models have been further extended and applied to highlight their physiological significance. The alternative methods of sequence generating functions and transfer matrix may be referred to the original and case studies for handling multiple binding modes, heterogeneous lattices, and lattice conformational changes (Bujalowski et al., 1989; Lifson, 1964; Schellman, 1974; Teif, 2007).

MOLECULAR FEATURES OF MULTIVALENT INTERACTIONS ON LINEAR LATTICES

It is straightforward to derive the quantitative models for the linear lattices that utilize discrete regions or domains to bind multiple distinct target proteins with the interaction stoichiometry of 1:1 for each target. In the absence of cooperativity among bound targets, the binding of each target can be handled, independent of binding of other targets, by the simple mass action law yielding a quadratic equation as a function of total concentrations of the lattice and the corresponding target. An advanced model has been derived by constructing a partition function for a linear lattice with cooperativities among bound targets (Cho et al., 2021).

Complexity arises when a target protein occupies two or more binding motifs on a linear lattice. We consider a linear lattice with a total of M motifs and a target protein occluding n consecutive motifs (Fig. 2) (Epstein, 1978; McGhee and

von Hippel, 1974). The binding motif can be any repeating unit including a base-pair or phosphate on DNA and a short peptide motif or a PTM (post-translational modification) moiety on an intrinsically disordered protein (IDP). As DNA or proteins have particular directions in denoting their motifs (5' to 3' end or N to C-terminus), target proteins are assumed to be polar as well in recognizing the motifs. It is further assumed that there is no partial binding where a target protein

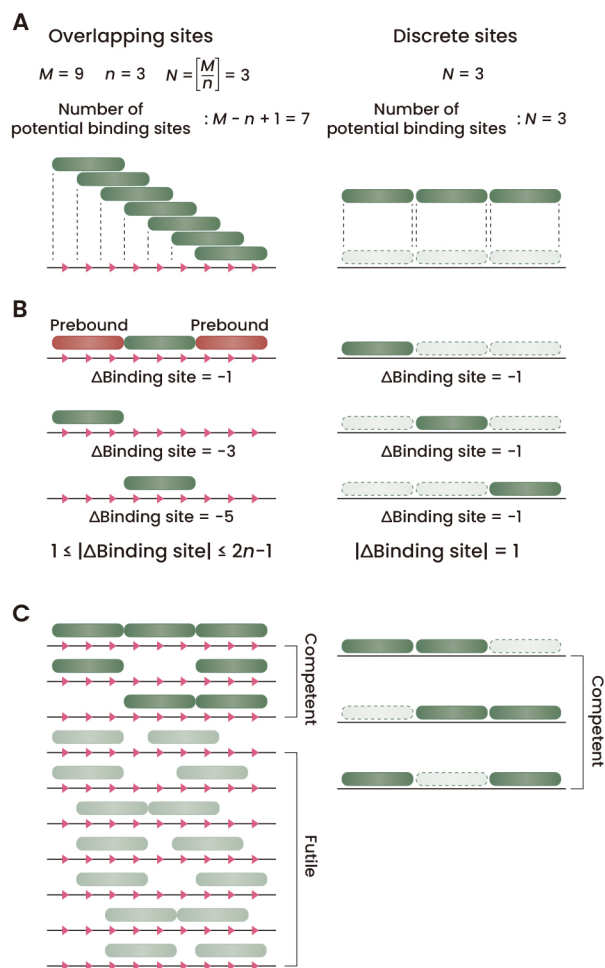


Fig. 2. Molecular features of multivalent interactions on a linear lattice ($M = 9$) where a protein occupies any $n (= 3)$ consecutive motifs. (A) The number of potential overlapping binding sites on a naked lattice (left panel) is greater as compared to a discrete-site system (right panel) with the same stoichiometry (N). (B) The number of potential binding sites eliminated upon binding of a protein depends on where the protein occupies on the lattice (left panel). In contrast, binding of a protein to the discrete-site system invariably eliminates only one potential binding site (right panel). (C) Possible configurations of the linear lattice with two proteins bound (left panel). Many configurations are futile for the last protein binding, resulting in apparent negative cooperativity among bound proteins. In contrast, all corresponding configurations in the discrete-site system are competent for binding (right panel).

occludes less than n motifs. Then, the target binding stoichiometry (N) is the greatest integer less than or equal to M/n ($N = \lfloor M/n \rfloor$). A fundamental nature of the linear lattice system becomes evident when a target protein binds to a naked lattice (left panel in Fig. 2A). Because the target protein occupies n consecutive motifs, any motifs except the rightmost $n-1$ positions can be starting points for target binding. Thus, potential target binding sites overlap and the number of such overlapping sites equals $M - n + 1$, obviously greater than the stoichiometry $\lfloor M/n \rfloor$. In contrast, for a conventional system in which a target protein binds discrete and isolated sites (right panel in Fig. 2A), the number of binding sites is simply equal to the stoichiometry $N = \lfloor M/n \rfloor$.

As the linear lattice subsequently binds more target proteins, its overlapping nature generates additional features further deviating from the discrete-site system. The number of potential binding sites eliminated upon binding of a protein depends on where the protein occupies on the lattice. When a protein binds to a gap exactly n motifs long between prebound proteins on the lattice, only one potential site is removed. Instead, if a gap is longer than $3n - 2$ motifs, binding of a protein to this region can eliminate as many as $2n - 1$ sites. For instance, binding of a protein with the site size of $n = 3$ to the three leftmost motifs on a linear lattice with a total of nine motifs eliminates three potential binding sites (the second figure in the left panel of Fig. 2B). Alternatively, if the protein occupies the three motifs at the center of the lattice, five potential binding sites are eliminated (the third figure in the left panel of Fig. 2B). However, in the discrete-site system, protein binding invariably eliminates only one potential binding site (right panel in Fig. 2B). Finally, it is difficult to completely saturate the linear lattice since the overlapping protein binding increasingly accumulates gaps with less than n motifs that are futile for binding. This point is explicitly illustrated in Fig. 2C (left panel) listing all possible configurations of the linear lattice with $\lfloor M/n \rfloor - 1$ proteins bound. Among them, many are futile configurations with the n free (unoccupied) motifs scattered over the lattice and must rearrange the bound proteins to create a site with n consecutive motifs for the last protein binding. Such a rearrangement or reduction in number of lattice configurations corresponds to a loss of mixing entropy, culminating in apparent negative cooperativity among bound proteins. In contrast, the number of available binding sites is independent of the configuration of bound proteins in the discrete-site system (right panel in Fig. 2C). In summary, because of the overlapping nature of multivalent linear lattice-target interactions, a linear lattice initially presents binding sites greater than the stoichiometry and thereby enhances protein binding as compared to a discrete-site system. However, with density of bound proteins increased, the effect of the overlapping binding is reversed, attenuating saturation of the linear lattice.

The following sections review the quantitative models penetrating the overlap problem of the linear lattice to yield the mathematical formulations relating the binding parameters to experimentally measurable properties of the lattice-target interactions. A core element of each model is the computation of the number of possible configurations for a given density of bound proteins on a lattice.

QUANTITATIVE FRAMEWORK FOR LINEAR LATTICE-PROTEIN INTERACTIONS: COMBINATORIAL MODEL

A complete set of parameters for description of linear lattice-protein interactions consists of the binding stoichiometry (N), binding constant (K), and cooperativity (ω) among bound proteins. As discussed above, the binding stoichiometry (N) is determined by the numbers of all motifs on a lattice (M) and those occupied by a target protein (n , termed site size) ($N = \lfloor M/n \rfloor$). The binding constant (K) corresponds to the affinity between a protein and a site n motifs long. Cooperativity can arise from pairwise interactions between any two proteins bound to a linear lattice. Although there are in principle iC_2 pairs on a lattice with i (≥ 2) proteins bound, the models discussed in this review formulate cooperativity only for the interaction between nearest neighbors (i.e., a pair of contiguously bound proteins without any intervening free motifs). Thus, the cooperativity parameter (ω) is equivalent to an equilibrium constant for formation of a direct “contact

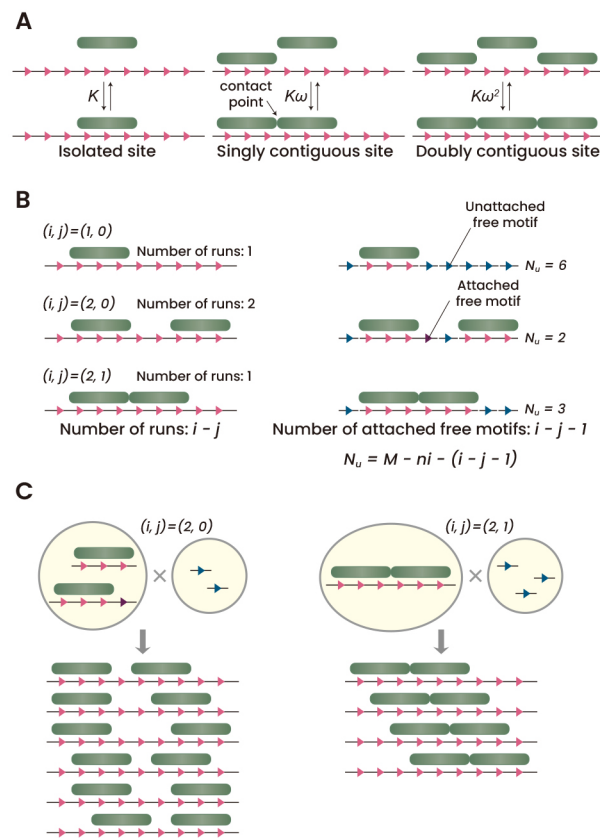


Fig. 3. Calculation of the number of distinct lattice configurations with i proteins bound and j contact points. (A) Three distinct types of protein binding sites on a linear lattice and the definitions of K and ω . (B) Dissection of a linear lattice into two distinct physical elements, runs and unattached free motifs. The $i - j - 1$ leftmost runs are attached at their right-hand end with a free motif (termed attached free motif). (C) Creation of the distinct lattice configurations by combining the two elements.

point^o between a pair of bound proteins. Then, under these definitions, a linear lattice presents three distinct types of protein binding sites (Fig. 3A): 1) an isolated site with the binding constant K ; 2) a singly contiguous site with the binding constant $K\omega$; 3) a doubly contiguous site with the binding constant $K\omega^2$. If $\omega > 1$ (or $0 < \omega < 1$), the nearest neighbor interaction is favorable (or unfavorable) and the protein binding is positively (or negatively) cooperative. For $\omega = 1$, bound proteins are independent of each other and the binding is noncooperative.

A fundamental relationship between the binding parameters and experimental variables can be derived by constructing a partition function for a linear lattice (Freire et al., 2009; Wyman and Gill, 1990). The partition function is a sum of relative probabilities or statistical weights of all possible protein-bound states of a linear lattice with a free lattice assigned as a reference state of unit relative probability (i.e., statistical weight = 1). Then, the statistical weight of a lattice with i proteins bound and j contact points among them is given by $(K[P])^i \omega^j$ where $[P]$ is the free protein concentration. However, in order to account for the presence of multiple configurations for a given set of (i, j) , the statistical weight must be multiplied by the degeneracy term $P_M(i, j)$, the number of distinct ways to distribute i proteins on a lattice with M motifs and j contact points. Then, the partition function (Z) is given by the following equations:

$$Z = \sum_{i=0}^N \sum_{j=0}^{i-1} P_M(i, j) (K[P])^i \omega^j \quad (1)$$

The average number of proteins bound per lattice (or binding density, ν), which is a principal quantity to be measured in all binding experiments, can be formulated from the partition function:

$$\nu = \frac{\partial \ln Z}{\partial \ln [P]} = \left\{ \sum_{i=0}^N \sum_{j=0}^{i-1} i P_M(i, j) (K[P])^i \omega^j \right\} \times \left\{ \sum_{i=0}^N \sum_{j=0}^{i-1} P_M(i, j) (K[P])^i \omega^j \right\}^{-1} \quad (2)$$

Likewise, the average number of contact points per lattice can be calculated from a partial derivative of the partition function:

$$\bar{j} = \frac{\partial \ln Z}{\partial \ln \omega} = \left\{ \sum_{i=0}^N \sum_{j=0}^{i-1} j P_M(i, j) (K[P])^i \omega^j \right\} \times \left\{ \sum_{i=0}^N \sum_{j=0}^{i-1} P_M(i, j) (K[P])^i \omega^j \right\}^{-1} \quad (3)$$

The final task in constructing the partition function is to derive the expression for $P_M(i, j)$. Here we follow the original combinatorial derivation of $P_M(i, j)$ (Epstein, 1978), highlighting the concept behind the mathematical procedures. A linear lattice with i proteins bound and j contact points may be dissected into two physical elements. The first element is a "run" defined as a distinct cluster of contiguously bound proteins, and the number of runs can be calculated as $i - j$ (Fig. 3B). Because there is at least one free motif between runs, each of the $i - j - 1$ leftmost runs must be attached with a free motif on the right side. The second element is the remaining free motifs and there are $M - ni - (i - j - 1)$ unattached free motifs ($\equiv N_u$). Then, the number ($\equiv N_c$) of ways of mixing these two elements to create the distinct lattice

configurations equals the number of distributing $i - j$ runs (accompanied with the $i - j - 1$ attached free motifs) and N_u unattached free motifs into $N_u + i - j$ slots (Fig. 3C):

$$N_c = \frac{(N_u + i - j)!}{N_u! (i - j)!} \quad (4)$$

In this expression, all runs have been treated as identical elements, regardless of the actual number of bound proteins in each run. Therefore, in order to complete the derivation of $P_M(i, j)$, the function N_c must be multiplied by the number ($\equiv N_p$) of distinct ways to distribute i proteins into $i - j$ runs:

$$N_p = \frac{(i - 1)!}{j! (i - j - 1)!} \quad (5)$$

The equation N_p is mathematically equivalent to the number of partitions of the integer i into $i - j$ positive integers. Finally, $P_M(i, j)$ is derived as the following equation:

$$P_M(i, j) = N_c N_p = \frac{(M - ni + 1)! (i - 1)!}{(M - ni - i + j + 1)! (i - j)! j! (i - j - 1)!} \quad (6)$$

For noncooperative binding ($\omega = 1$), the number of contact points j becomes irrelevant and $P_M(i, j)$ reduces to $P_M(i)$, the number of ways of mixing i proteins and $M - ni$ free motifs to build distinct lattice configurations:

$$P_M(i) = \frac{(M - ni + i)!}{(M - ni)! i!} \quad (7)$$

Then, the partition function for noncooperative binding can be written in a simplified form:

$$Z = \sum_{i=0}^N P_M(i) (K[P])^i \quad (8)$$

In practice, the total lattice and protein concentrations ($[L]_{tot}$ and $[P]_{tot}$), rather than the free protein concentration ($[P]$), are known experimental variables. The total concentrations are related to each other and other binding parameters through a simple mass balance equation:

$$[P]_{tot} = [P] + \nu [L]_{tot} \quad (9)$$

For a given set of binding parameters and reactant concentrations, this mass balance equation can be solved for $[P]$ by the numerical procedures such as the Newton-Raphson and the bisection method (Hamming, 1986). In turn, this solution allows calculation of the relative probabilities of all lattice configurations and the ensemble-averaged quantities including Eqs. 2 and 3. Thus, the combinatorial method is straightforward and intuitive in constructing a partition function which illustrates distribution among various protein-bound states of a linear lattice as a function of lattice and protein concentrations. However, this method is difficult to apply to a very long linear lattice (i.e., $M \gg n$) because the number of possible lattice configurations may be too large and potentially cause an overflow problem in computation.

QUANTITATIVE FRAMEWORK FOR LINEAR LATTICE-PROTEIN INTERACTIONS: CONDITIONAL PROBABILITY MODEL

Several quantitative frameworks have been proposed to treat an “infinitely” long linear lattice ($M \gg n$), particularly relevant for proteins nonspecifically binding the chromosomal DNA. Among these frameworks, we review the conditional probability model originally presented in the seminal work by McGhee and von Hippel (1974). In this model, the conditional probabilities have been formulated for the particular states (free or bound) of two consecutive motifs on a linear lattice. For instance, the conditional probability ff (or fb_1) is defined as, given a randomly chosen free motif, the probability of the subsequent righthand side motif being free (or bound by the left end of a protein) (Supplementary Fig. S1). In addition, the conditional probability $b_n f$ (or $b_n b_n$) is defined as, given a motif bound by the right end of a protein, the probability of the subsequent motif being free (or bound by the left end of a protein) (Supplementary Fig. S1). The conditional probabilities were then used to derive an expression for the average number of free binding sites per lattice at a given binding density. This elegant approach yielded a modified form of the Scatchard equation:

$$\frac{v}{[P]} = K \cdot \{ \text{Average number of free binding sites per lattice} \} \quad (10a)$$

$$\frac{\theta}{[P]} = K(1 - n\theta) \left[\frac{(2\omega - 1)(1 - n\theta) + \theta - R}{2(\omega - 1)(1 - n\theta)} \right]^{n-1} \cdot \left[\frac{1 - (n+1)\theta + R}{2(1 - n\theta)} \right]^2 \quad (10b)$$

$$R = \sqrt{[1 - (n+1)\theta]^2 + 4\omega\theta(1 - n\theta)} \quad (10c)$$

where θ corresponds to the average number of proteins bound per motif (i.e., $\theta = v/M$).

Referring to Supplementary Information for the detailed mathematical procedures of the derivation, we focus on a few intuitive limiting cases leading to the interpretations of this equation consistent with the molecular features of the linear lattice system (McGhee and von Hippel, 1974).

1) In the case of $\omega = 1$ (noncooperative binding), by using L'Hospital's rule, the equation can be reduced to the following (see Supplementary Information for detailed mathematical procedures):

$$\lim_{\omega \rightarrow 1} \left(K(1 - n\theta) \left[\frac{(2\omega - 1)(1 - n\theta) + \theta - R}{2(\omega - 1)(1 - n\theta)} \right]^{n-1} \cdot \left[\frac{1 - (n+1)\theta + R}{2(1 - n\theta)} \right]^2 \right) = K(1 - n\theta) \left[\frac{1 - n\theta}{1 - (n-1)\theta} \right]^{n-1} \quad (11)$$

Note that, for $n = 1$ (no overlap between bound proteins), the equation further reduces to the original Scatchard equation $\theta/[L] = K(1 - \theta)$ in which the term $(1 - \theta)$ simply represents the fraction of free motifs. Because the squared bracket term in Eq. 11 is always less than unity for $n \geq 2$, the fraction of free motifs competent for binding is smaller than the total fraction of free motifs $1 - n\theta$. Therefore, this result quantitatively supports that, even without genuine interactions among bound proteins (i.e., $\omega = 1$), apparent negative cooperativity arises from the overlap among potential binding sites and consequent futile gaps shorter than n motifs.

2) In the case of $\omega = 0$ (infinite negative-cooperativity), Eq. 10b reduces to the following expression:

$$\frac{\theta}{[P]} = K \cdot \{1 - (n+1)\theta\} \left[\frac{1 - (n+1)\theta}{(1 - n\theta)} \right]^n \quad (12)$$

This reduced form simply corresponds to Eq. 11 with $n = n + 1$. The increased binding site size demonstrates that, if the interaction between bound proteins is extremely unfavorable, there is apparently no contact point between any adjacently bound proteins. Instead, they are separated by a persistent free motif. This result clearly demonstrates the fundamental relationship between binding site size and cooperativity.

3) Further insight can be provided at the molecular level from the partial derivatives of Eqs. 10b and 11 with respect to θ at the limiting condition of $\theta \rightarrow 0$ (see Supplementary Information for detailed mathematical procedures):

$$\frac{\partial(\theta/[P])}{\partial\theta} \Big|_{\theta=0} = \frac{\partial(v/[P])}{\partial v} \Big|_{v=0} = K(2\omega - 2n - 1) \quad (13)$$

Based on Eq. 10a, the partial derivative can be interpreted as a net change in the average numbers of all three types (Fig. 3A) of binding sites, weighted by their corresponding binding constants, upon binding of one protein to a naked ($v = 0$) lattice. As illustrated in Fig. 2B, the binding of a protein to a sufficiently long region eliminates a total of $2n - 1$ potential binding sites. In addition, the binding converts the two adjacent isolated binding sites into two singly contiguous binding sites ($2K\omega$). Hence, a total of $(2n - 1) + 2$ isolated binding sites has been eliminated ($-(2n + 1) \cdot K$). Likewise, the partial derivative of Eq. 11 at $\theta \rightarrow 0$ is given by:

$$\frac{\partial(\theta/[P])}{\partial\theta} \Big|_{\theta=0} = -K(2n - 1) \quad (14)$$

Therefore, in the noncooperative case, the binding of one ligand to a naked lattice simply eliminates $2n - 1$ potential binding sites.

Taken together, although the conditional probability method is based on the different conceptual framework as compared to the combinatorial approach, the final formulation provides intuitive interpretations fully consistent with the molecular features of the linear lattice systems. In practice, Eq. 10b is rearranged and incorporated into a mass balance equation relating the binding parameters to the total concentrations of lattice motif and protein ($[M]_{\text{tot}}$ and $[P]_{\text{tot}}$):

$$[P] = \frac{\theta}{K(1 - n\theta)ff^{n-1}C^2} \quad (15a)$$

$$ff = \left[\frac{(2\omega - 1)(1 - n\theta) + \theta - R}{2(\omega - 1)(1 - n\theta)} \right] \quad (15b)$$

$$C = \left[\frac{1 - (n+1)\theta + R}{2(1 - n\theta)} \right] \quad (15c)$$

$$[P]_{\text{tot}} = [P] + \theta[M]_{\text{tot}} \quad (15d)$$

$$[P]_{tot} = \frac{\theta}{K(1-n\theta)ff^{n-1}C^2} + \theta[M]_{tot} \quad (15e)$$

Eq. 15e can be numerically solved for θ at given values of $[M]_{tot}$ and $[P]_{tot}$. When interactions of proteins with short linear lattices (e.g., DNA oligomers) are analyzed, the equation can be partially corrected for the assumption of infinite lattice length by applying an “end effect” constant, $(M - n + 1) / M$, to the term ff^{n-1} (Tsodikov et al., 2001).

APPLICATION AND EXTENSION OF THE QUANTITATIVE MODELS

Competition among multiple binding modes in protein-nucleic acid interactions

Spatiotemporal regulation of transcription is achieved by interactions between TFs and their specific binding sites on DNA. Because of the enormous number of nonspecific sites on the chromosomal DNA, binding of TFs to these regions must be taken into account to accurately predict the occupancy of the specific sites and thereby the transcription profiles of the corresponding genes (Brewster et al., 2014; von Hippel et al., 1974). In order to recapitulate the essential features of the competition between specific and nonspecific DNA binding, the conditional probability model was extended and applied to a hypothetical two-component (TF and infinitely long DNA with a few embedded specific sites) system. While the 1:1 interaction between TF and a specific site is fully described by the binding constant K_{sp} , the nonspecific binding is characterized by the binding site size n (in base-pairs), the binding constant K_{ns} , and the cooperativity parameter ω . Then, combining Eq. 10b with the mass-action law for the 1:1 specific binding, the TF concentrations of free, specifically, and nonspecifically bound forms ($[TF]$, $[TF]_{sp,b}$, $[TF]_{ns,b}$) can be derived as the following equations:

$$[TF] = \frac{\theta}{K_{ns}(1-n\theta)ff^{n-1}C^2} \quad (16a)$$

$$[TF]_{sp,b} = K_{sp} \frac{[D_{sp}]_{tot}[TF]}{1 + K_{sp}[TF]} \quad (16b)$$

$$[TF]_{ns,b} = \theta[M]_{tot} \quad (16c)$$

where $[D_{sp}]_{tot}$ and $[M]_{tot}$ are the total concentrations of the specific site and the nonspecific binding motif (base-pair), respectively. Substituting Eq. 16a for $[TF]$ in Eq. 16b, the mass balance equation for the total TF concentration ($[TF]_{tot} = [TF] + [TF]_{sp,b} + [TF]_{ns,b}$) can be numerically solved for θ . The final outcome of the calculation is the fractional occupancy of the specific site ($Y_{sp} = [TF]_{sp,b}/[D_{sp}]_{tot}$) as a function of total concentration ratio between TF and the specific site ($[TF]_{tot}/[D_{sp}]_{tot}$) ranging from 0 to 10 (upper panels in Figs. 4A and 4B). In the calculation, the ratio K_{sp}/K_{ns} (termed specificity ratio) (Fig. 4A) or the total nonspecific motif concentrations (Fig. 4B) was varied over orders of magnitude while the nonspecific binding site size and cooperativity were fixed at the constant values for simplicity ($n = 10$, $\omega = 1$).

At a given specificity ratio and a total motif concentration,

as the concentration ratio $[TF]_{tot}/[D_{sp}]_{tot}$ is increased, the fractional occupancy of the specific site by TF monotonically increases with an apparent hyperbolic feature (upper panels in Figs. 4A and 4B). However, the underlying distribution of TF exhibits a dynamic shift from specifically to nonspecifically bound states (bottom panels in Figs. 4A and 4B). For higher specificity ratio or lower nonspecific motif concentrations, the specific complex is predominant in the regime $[TF]_{tot}/[D_{sp}]_{tot} < 1$, leading to a steep rise in occupancy of the specific site. Consequently, the transition to the nonspecifically bound state is achieved at higher concentration ratio. Therefore, under these conditions, a relatively small amount of TF is required to saturate the specific site and thereby fully activate transcription. Conversely, for lower specificity ratio or higher nonspecific motif concentrations, the nonspecific binding significantly competes with the specific binding even at low $[TF]_{tot}/[D_{sp}]_{tot}$ (bottom panels in Figs. 4A and 4B), attenuating saturation of the specific site (upper panels in Figs. 4A and 4B). These simulations suggest that, since protein-DNA interactions are generally sensitive to many cellular conditions such as salt concentration and osmotic stress, changes in these variables potentially fine-tune the specificity ratio of TFs and thereby the corresponding transcription levels. Furthermore, a change in chromosome packing may indirectly affect the TF-specific site interaction by altering the nonspecific site concentrations. Taken together, nonspecific protein-DNA interactions, via change in either specificity ratio or abundance of nonspecific sites, can modulate the occupancies of specific TF binding sites and consequently reprogram the gene-specific transcriptional activities.

Competitions between specific and nonspecific binding or among multiple nonspecific binding modes have been observed in numerous *in vitro* protein-DNA interactions as well (Bujalowski et al., 1988; Rajendran et al., 1998). Even studies using short oligonucleotides have shown similar competitions due to significantly low specificity ratios (Holbrook et al., 2001; Koh et al., 2008). In order to accurately determine a specific binding constant, the linear lattice models must be applied or further advanced to tease apart the contributions from multiple binding modes to the observed binding signal (Tsodikov et al., 2001).

Competition among distinct target proteins for binding to an intrinsically disordered protein

IDPs often utilize short peptide motifs to recruit multiple distinct targets or multiple copies of an identical target (Cumberworth et al., 2013; Hong et al., 2020; Wright and Dyson, 2015). These IDPs are collectively termed hubs and involved in signal transduction and macromolecular transport. A representative example is Nup153, a subunit of the NPC, that contains a long C-terminal IDR (~600 amino acids in length) (Krull et al., 2004). The IDR presents multiple FG-motifs to interact with Kaps carrying macromolecular cargos into and out of the nucleus. Multiple hydrophobic pockets on the Kap surface are the primary binding sites for the FG-motifs (Bayliss et al., 2000).

A recent thermodynamic study has developed an advanced combinatorial model to demonstrate that the Nup153 IDR comprises a high-affinity 1:1 binding site and a series of

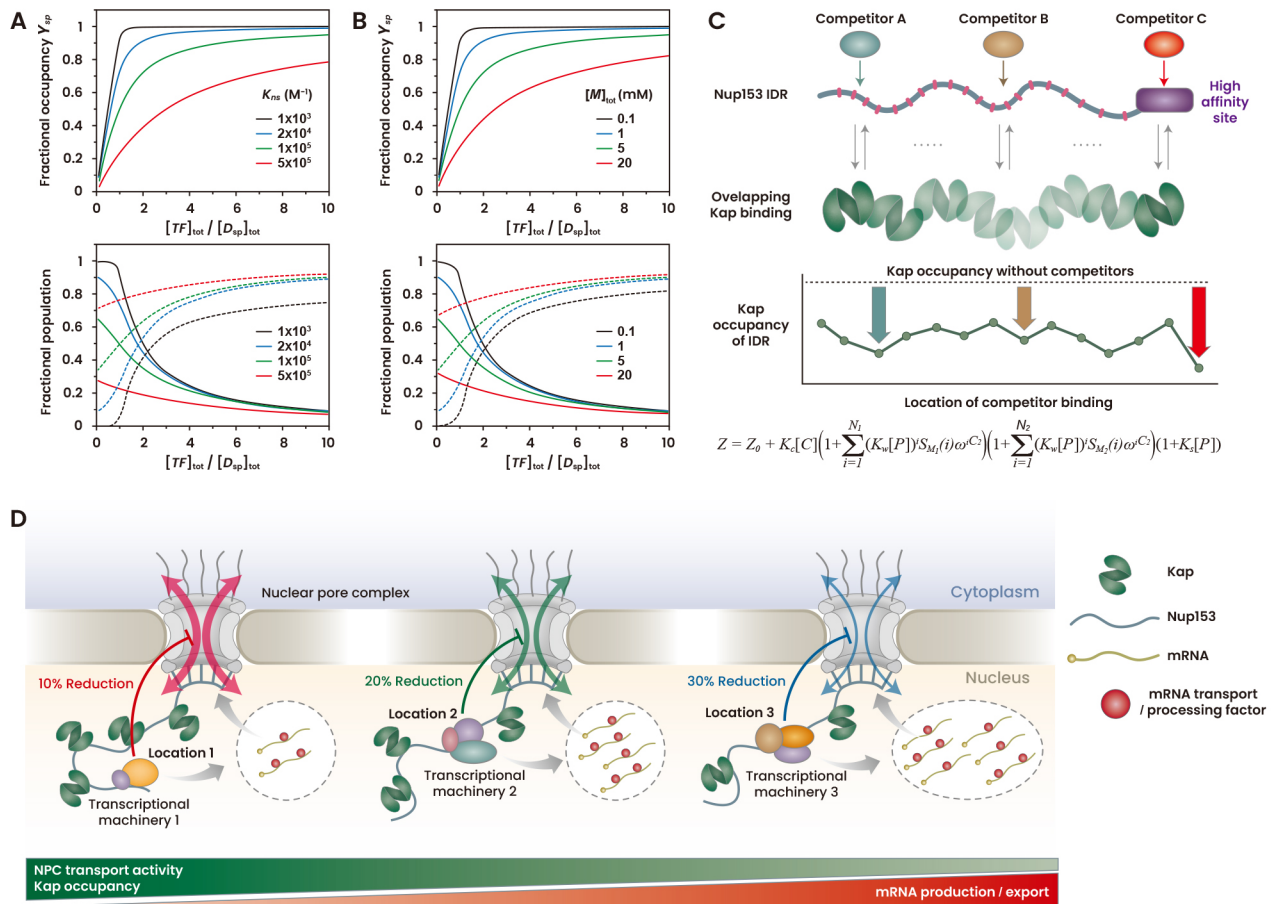


Fig. 4. Application and extension of the quantitative models for linear lattice systems. (A and B) Effects of nonspecific protein-DNA interactions on transcription. Upper panels: Using an extended conditional probability model (Eq. 16), the fractional occupancy of specific DNA sites ($Y_{sp} = [TF]_{sp,b} / [D_{sp}]_{tot}$) for binding of a hypothetical TF was calculated as a function of molar ratio $[TF]_{tot} / [D_{sp}]_{tot}$ for various sets of interaction parameters. Bottom panels: The corresponding fractional distribution of TF between specifically (solid curves) and nonspecifically (dashed curves) bound states were calculated. In these calculations, the value of K_{ns} (A) or the concentration of nonspecific motifs ($[M]_{tot}$) (B) was varied with the fixed values of $K_{sp} = 1 \times 10^{12} M^{-1}$, $n = 10$ bp, and $\omega = 1$ ($[M]_{tot} = 5$ mM in (A); $K_{ns} = 1 \times 10^5 M^{-1}$ in (B)). (C) Quantitative model for assembly of the Nup153 IDR hub with multiple interaction partners and competitors (adapted from Cho et al., 2021). The Nup153 IDR presents a high-affinity 1:1 Kap binding site (purple) and a series of low-affinity sites for overlapping binding of multiple Kaps. Kap occupies multiple dipeptide (FG) motifs (pink vertical bars). Using advanced combinatorial models, fine-tuning of the Kap occupancy of Nup153 IDR was predicted as a function of location of the competitor binding site. In the partition function Z , Z_0 corresponds to the partition function of the Nup153 IDR in the absence of competition; $K_c[C]$ represents the competitor binding; The terms in the brackets are the partition functions for two subregions of the low-affinity sites separated by the competitor binding; $(1 + K_s[P])$ represents the 1:1 interaction of Kap with the high-affinity site. (D) On the basis of the multivalent, overlapping IDR-Kap interaction, the Nup153 IDR is proposed to function as a molecular switch to couple nucleocytoplasmic transport to transcription.

low-affinity sites for binding of multiple Kaps (Fig. 4C) (Cho et al., 2021). Calorimetric data of various protein concentrations and IDR lengths were scrutinized to further show that the overlapping binding of Kaps to the low-affinity sites results in apparent negative cooperativity. Because the Nup153 IDR potentially interacts with nuclear proteins involved in transcription and chromatin organization (Kadota et al., 2020; Kasper et al., 1999), this study has constructed composite combinatorial models to test how the multivalent Kap binding would be affected by competitive binding of nuclear proteins (Fig. 4C). Remarkably, the simulation has revealed that the Kap

occupancy of the low-affinity region can be fine-tuned by changing the location of the competitor binding site (Fig. 4C). This delicate modulation arises from the molecular feature of the overlapping binding: The number of potential Kap binding sites eliminated by the competition is determined by the position of the competitor binding site (Fig. 2B). Therefore, assuming that the Kap occupancy is a proxy for the transport activity of the NPC, it is conceivable that the Nup153 IDR functions as a molecular switch coupling specific nuclear processes to distinct transport states. For instance, a strong promoter may be coupled to the NPC activity in such a way

that specific TFs or co-activators associated with the strong promoter target a location in the Nup153 IDR that considerably reduces the Kap occupancy (Fig. 4D). As a consequence of the reduced general transport activity mediated by Kaps, a large amount of mRNA transcribed from the strong promoter may be efficiently exported through the NPC (Fig. 4D). Although awaiting experimental validation, the coupling mechanism built upon multivalent, overlapping IDP-target interactions may contribute to the functional versatility of the IDP hubs in dynamic cellular processes. This exemplary study demonstrates that the original combinatorial model can be readily expanded by simple mathematical operations to account for additional complexities in linear lattice-protein interactions including heterogeneous binding sites.

CONCLUSION

Linear lattice systems and their multivalent interactions with target proteins often regulate dynamic cellular processes. Because of the overlapping target binding sites on a linear lattice, quantitative understanding of such interactions requires a fundamentally different framework as compared to simple 1:1 binding or discrete-site systems. In this review, we discussed the two prevalent approaches in unraveling the linear lattice systems, namely combinatorial and conditional probability models. Constructing the lattice partition functions from the combinatorial approach is straightforward and readily expandable in data analysis and predictions as illustrated in the Nup153 IDR–Kap interaction. On the other hand, the conditional probability model provides invaluable physical insights consistent with the molecular features of the multivalent linear lattice–target interactions. Furthermore, this method is suitable in simulating *in vivo* nucleic acid systems of apparent infinite lattice length. These frameworks may serve as a cornerstone to develop sophisticated models to analyze more complex cellular processes including competition among multiple DNA binding proteins on nucleosomal DNA (Segal and Widom, 2009) as well as formation of phase-separated condensates involving multiple components (Lyon et al., 2021).

Note: Supplementary information is available on the *Molecules and Cells* website (www.molcells.org).

ACKNOWLEDGMENTS

This work was supported by Samsung Science & Technology Foundation and Research (SSTF-BA1802-09) and the National Research Foundation (2019R1C1C1011640).

AUTHOR CONTRIBUTIONS

J.C. and J.K. analyzed the data. J.C., R.K., and J.K. wrote the manuscript.

CONFLICT OF INTEREST

The authors have no potential conflicts of interest to disclose.

ORCID

Jaejun Choi <https://orcid.org/0000-0002-9546-6692>
Ryeonghyeon Kim <https://orcid.org/0000-0003-4542-5339>

Junseock Koh <https://orcid.org/0000-0002-2920-2656>

REFERENCES

- Banani, S.F., Lee, H.O., Hyman, A.A., and Rosen, M.K. (2017). Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* *18*, 285–298.
- Bayliss, R., Littlewood, T., and Stewart, M. (2000). Structural basis for the interaction between FxFG nucleoporin repeats and importin-beta in nuclear trafficking. *Cell* *102*, 99–108.
- Berg, O.G., Winter, R.B., and von Hippel, P.H. (1981). Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* *20*, 6929–6948.
- Bintu, L., Buchler, N.E., Garcia, H.G., Gerland, U., Hwa, T., Kondev, J., Kuhlman, T., and Phillips, R. (2005). Transcriptional regulation by the numbers: applications. *Curr. Opin. Genet. Dev.* *15*, 125–135.
- Brewster, R.C., Weinert, F.M., Garcia, H.G., Song, D., Rydenfelt, M., and Phillips, R. (2014). The transcription factor titration effect dictates level of gene expression. *Cell* *156*, 1312–1323.
- Bujalowski, W. (2006). Thermodynamic and kinetic methods of analyses of protein-nucleic acid interactions. From simpler to more complex systems. *Chem. Rev.* *106*, 556–606.
- Bujalowski, W., Lohman, T.M., and Anderson, C.F. (1989). On the cooperative binding of large ligands to a one-dimensional homogeneous lattice: the generalized three-state lattice model. *Biopolymers* *28*, 1637–1643.
- Bujalowski, W., Overman, L.B., and Lohman, T.M. (1988). Binding mode transitions of *Escherichia coli* single strand binding protein-single-stranded DNA complexes. Cation, anion, pH, and binding density effects. *J. Biol. Chem.* *263*, 4629–4640.
- Cho, B., Choi, J., Kim, R., Yun, J.N., Choi, Y., Lee, H.H., and Koh, J. (2021). Thermodynamic models for assembly of intrinsically disordered protein hubs with multiple interaction partners. *J. Am. Chem. Soc.* *143*, 12509–12523.
- Choi, K.Y., Satterberg, B., Lyons, D.M., and Elion, E.A. (1994). Ste5 tethers multiple protein kinases in the MAP kinase cascade required for mating in *S. cerevisiae*. *Cell* *78*, 499–512.
- Cortese, M.S., Uversky, V.N., and Dunker, A.K. (2008). Intrinsic disorder in scaffold proteins: getting more from less. *Prog. Biophys. Mol. Biol.* *98*, 85–106.
- Cumberworth, A., Lamour, G., Babu, M.M., and Gsponer, J. (2013). Promiscuity as a functional trait: intrinsically disordered regions as central players of interactomes. *Biochem. J.* *454*, 361–369.
- Dragan, A.I., Read, C.M., Makeyeva, E.N., Milgotina, E.I., Churchill, M.E., Crane-Robinson, C., and Privalov, P.L. (2004). DNA binding and bending by HMG boxes: energetic determinants of specificity. *J. Mol. Biol.* *343*, 371–393.
- Dunker, A.K., Cortese, M.S., Romero, P., Iakoucheva, L.M., and Uversky, V.N. (2005). Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* *272*, 5129–5148.
- Epstein, I.R. (1978). Cooperative and non-cooperative binding of large ligands to a finite one-dimensional lattice. A model for ligand-oligonucleotide interactions. *Biophys. Chem.* *8*, 327–339.
- Freire, E., Schon, A., and Velazquez-Campoy, A. (2009). Isothermal titration calorimetry: general formalism using binding polynomials. *Methods Enzymol.* *455*, 127–155.
- Frey, S. and Gorlich, D. (2007). A saturated FG-repeat hydrogel can reproduce the permeability properties of nuclear pore complexes. *Cell* *130*, 512–523.
- Fung, H.Y.J., Birol, M., and Rhoades, E. (2018). IDPs in macromolecular complexes: the roles of multivalent interactions in diverse assemblies.

Curr. Opin. Struct. Biol. 49, 36-43.

Ha, T., Kaiser, C., Myong, S., Wu, B., and Xiao, J. (2022). Next generation single-molecule techniques: imaging, labeling, and manipulation in vitro and in cellulo. *Mol. Cell* 82, 304-314.

Hamming, R.W. (1986). *Numerical Methods for Scientists and Engineers* (2nd Edition) (New York, NY: Dover).

Holbrook, J.A., Tsodikov, O.V., Saecker, R.M., and Record, M.T., Jr. (2001). Specific and non-specific interactions of integration host factor with DNA: thermodynamic evidence for disruption of multiple IHF surface salt-bridges coupled to DNA binding. *J. Mol. Biol.* 310, 379-401.

Hong, S., Choi, S., Kim, R., and Koh, J. (2020). Mechanisms of macromolecular interactions mediated by protein intrinsic disorder. *Mol. Cells* 43, 899-908.

Jen-Jacobson, L., Engler, L.E., and Jacobson, L.A. (2000). Structural and thermodynamic strategies for site-specific DNA binding proteins. *Structure* 8, 1015-1023.

Kadota, S., Ou, J., Shi, Y., Lee, J.T., Sun, J., and Yildirim, E. (2020). Nucleoporin 153 links nuclear pore complex to chromatin architecture by mediating CTCF and cohesin binding. *Nat. Commun.* 11, 2606.

Kao-Huang, Y., Revzin, A., Butler, A.P., O'Conner, P., Noble, D.W., and von Hippel, P.H. (1977). Nonspecific DNA binding of genome-regulating proteins as a biological control mechanism: measurement of DNA-bound Escherichia coli lac repressor in vivo. *Proc. Natl. Acad. Sci. U. S. A.* 74, 4228-4232.

Kasper, L.H., Brindle, P.K., Schnabel, C.A., Pritchard, C.E., Cleary, M.L., and van Deursen, J.M. (1999). CREB binding protein interacts with nucleoporin-specific FG repeats that activate transcription and mediate NUP98-HOXA9 oncogenicity. *Mol. Cell Biol.* 19, 764-776.

Koh, J. and Blobel, G. (2015). Allosteric regulation in gating the central channel of the nuclear pore complex. *Cell* 161, 1361-1373.

Koh, J., Saecker, R.M., and Record, M.T., Jr. (2008). DNA binding mode transitions of Escherichia coli HU(alpha-beta): evidence for formation of a bent DNA-protein complex on intact, linear duplex DNA. *J. Mol. Biol.* 383, 324-346.

Krull, S., Thyberg, J., Bjorkroth, B., Rackwitz, H.R., and Cordes, V.C. (2004). Nucleoporins as components of the nuclear pore complex core structure and Tpr as the architectural element of the nuclear basket. *Mol. Biol. Cell* 15, 4261-4277.

Lifson, S. (1964). Partition functions of linear-chain molecules. *J. Chem. Phys.* 40, 3705-3710.

Lohman, T.M., deHaseth, P.L., and Record, M.T., Jr. (1980). Pentylsine-deoxyribonucleic acid interactions: a model for the general effects of ion concentrations on the interactions of proteins with nucleic acids. *Biochemistry* 19, 3522-3530.

Lyon, A.S., Peeples, W.B., and Rosen, M.K. (2021). A framework for understanding the functions of biomolecular condensates across scales. *Nat. Rev. Mol. Cell Biol.* 22, 215-235.

Mahamid, J., Pfeffer, S., Schaffer, M., Villa, E., Danev, R., Cuellar, L.K., Forster, F., Hyman, A.A., Plitzko, J.M., and Baumeister, W. (2016). Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science* 351, 969-972.

Mark, W.Y., Liao, J.C., Lu, Y., Ayed, A., Laister, R., Szymczynska, B., Chakrabarty, A., and Arrowsmith, C.H. (2005). Characterization of segments from the central region of BRCA1: an intrinsically disordered scaffold for multiple protein-protein and protein-DNA interactions? *J. Mol. Biol.* 345, 275-287.

McGhee, J.D. and von Hippel, P.H. (1974). Theoretical aspects of DNA-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. *J. Mol. Biol.* 86, 469-489.

Noutsou, M., Duarte, A.M., Anvarian, Z., Didenko, T., Minde, D.P., Kuper, I.,

de Ridder, I., Oikonomou, C., Friedler, A., Boelens, R., et al. (2011). Critical scaffolding regions of the tumor suppressor Axin1 are natively unfolded. *J. Mol. Biol.* 405, 773-786.

Oikonomou, C.M. and Jensen, G.J. (2017). Cellular electron cryotomography: toward structural biology in situ. *Annu. Rev. Biochem.* 86, 873-896.

Privalov, P.L., Dragan, A.I., and Crane-Robinson, C. (2011). Interpreting protein/DNA interactions: distinguishing specific from non-specific and electrostatic from non-electrostatic components. *Nucleic Acids Res.* 39, 2483-2491.

Radu, A., Moore, M.S., and Blobel, G. (1995). The peptide repeat domain of nucleoporin Nup98 functions as a docking site in transport across the nuclear pore complex. *Cell* 81, 215-222.

Rajendran, S., Jezewska, M.J., and Bujalowski, W. (1998). Human DNA polymerase beta recognizes single-stranded DNA using two different binding modes. *J. Biol. Chem.* 273, 31021-31031.

Record, M.T., Jr., Lohman, M.L., and De Haseth, P. (1976). Ion effects on ligand-nucleic acid interactions. *J. Mol. Biol.* 107, 145-158.

Schellman, J.A. (1974). Cooperative multisite binding to DNA. *Isr. J. Chem.* 12, 219-238.

Schoch, R.L., Kapinos, L.E., and Lim, R.Y. (2012). Nuclear transport receptor binding avidity triggers a self-healing collapse transition in FG-nucleoporin molecular brushes. *Proc. Natl. Acad. Sci. U. S. A.* 109, 16911-16916.

Segal, E. and Widom, J. (2009). From DNA sequence to transcriptional behaviour: a quantitative approach. *Nat. Rev. Genet.* 10, 443-456.

Shin, Y. and Brangwynne, C.P. (2017). Liquid phase condensation in cell physiology and disease. *Science* 357, eaaf4382.

Shrader, T.E. and Crothers, D.M. (1989). Artificial nucleosome positioning sequences. *Proc. Natl. Acad. Sci. U. S. A.* 86, 7418-7422.

Sigal, Y.M., Zhou, R., and Zhuang, X. (2018). Visualizing and discovering cellular structures with super-resolution microscopy. *Science* 361, 880-887.

Stracy, M., Schweizer, J., Sherratt, D.J., Kapanidis, A.N., Uphoff, S., and Lesterlin, C. (2021). Transient non-specific DNA binding dominates the target search of bacterial DNA-binding proteins. *Mol. Cell* 81, 1499-1514. e6.

Teif, V.B. (2007). General transfer matrix formalism to calculate DNA-protein-drug binding in gene regulation: application to OR operator of phage lambda. *Nucleic Acids Res.* 35, e80.

Tsodikov, O.V., Holbrook, J.A., Shkel, I.A., and Record, M.T., Jr. (2001). Analytic binding isotherms describing competitive interactions of a protein ligand with specific and nonspecific sites on the same DNA oligomer. *Biophys. J.* 81, 1960-1969.

von Hippel, P.H., Revzin, A., Gross, C.A., and Wang, A.C. (1974). Non-specific DNA binding of genome regulating proteins as a biological control mechanism: I. The lac operon: equilibrium aspects. *Proc. Natl. Acad. Sci. U. S. A.* 71, 4808-4812.

Widom, J. (1999). Equilibrium and dynamic nucleosome stability. *Methods Mol. Biol.* 119, 61-77.

Wodarz, A. and Nusse, R. (1998). Mechanisms of Wnt signaling in development. *Annu. Rev. Cell Dev. Biol.* 14, 59-88.

Wright, P.E. and Dyson, H.J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* 16, 18-29.

Wyman, J. and Gill, S.J. (1990). *Binding and Linkage: Functional Chemistry of Biological Macromolecules* (Mill Valley, CA: University Science Books).

Xue, B., Romero, P.R., Noutsou, M., Maurice, M.M., Rudiger, S.G., William, A.M., Jr., Mizianty, M.J., Kurgan, L., Uversky, V.N., and Dunker, A.K. (2013). Stochastic machines as a colocalization mechanism for scaffold protein function. *FEBS Lett.* 587, 1587-1591.