Cooperativity boosts affinity and specificity of proteins with multiple RNA-binding domains

Simon H. Stitzinger,¹ Salma Sohrabi-Jahromi,¹ and Johannes Söding^{1,2,*}

¹Quantitative and Computational Biology, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany.

²Campus-Institut Data Science (CIDAS), Göttingen, Germany.

* Correspondence: soeding@mpibpc.mpg.de

January 27, 2021

Abstract

Numerous cellular processes rely on the binding of proteins with high affinity to specific sets of RNAs. Yet most RNA binding domains display low specificity and affinity, to the extent that for most RNA-binding domains, the enrichment of the best binding motif measured by high-throughput RNA SELEX or RNA bind-n-seq is usually below 10-fold, dramatically lower than that of DNA-binding domains. Here, we develop a thermodynamic model to predict the binding affinity for proteins with any number of RNA-binding domains given the affinities of their isolated domains. For the four proteins in which affinities for individual domains have been measured the model predictions are in good agreement with experimental values. The model gives insight into how proteins with multiple RNA-binding domains can reach affinities and specificities orders of magnitude higher than their individual domains. Our results contribute towards resolving the conundrum of missing specificity and affinity of RNA binding proteins and underscore the need for bioinformatic methods that can learn models for multi-domain RNA binding proteins from high-throughput *in-vitro* and *in-vivo* experiments.

Introduction

RNA-binding proteins (RBPs) regulate various steps of mRNA biogenesis including RNA splicing, localization, translation, and degradation [1]. In order to ensure that these proteins bind the correct set of RNA molecules, the interactions have to be highly specific. This can be achieved through cooperative binding, in which two or more relatively weakly binding RNA-binding domains (RBDs) are combined, to form one high affinity complex [2, 3]. Around 45% of eukaryotic RBPs have at least two binding domains [4] (Figure 1). In addition, these domains often recognize very short and degenerate RNA motifs (~ 3 , rarely more than 5 nucleotides) [5, 6], while typical DNA-binding domains bind sequences of around 6 to 12 nucleotides [7, 8, 9]. Because of their larger binding sites, the dissociation constants (K_d) of most transcription factors are in the picomolar to nanomolar range, while individual RBDs often only bind with affinities in the micromolar to millimolar range.

RBPs use multi-domain RNA recognition to bind with high affinity and specificity, despite the low affinity of their individual domains. The affinity increase in cooperative binding stems from an increased local concentration of the RNA molecule at an unbound binding domain when another domain is already bound [10]. The neighboring RNA binding site is kept closer to the second RBD since the molecules are already linked. This adds to the background concentration, resulting in a much higher effective concentration c_{eff} . This high local concentration in turn increases the rate for the binding reaction. The concept of an effective concentration

has been introduced before [10, 11], and was used to thermodynamically model different systems with two binding sites, based on the interactions of the isolated sites [11, 12, 13, 14].

A general quantitative model of cooperative multi-domain interactions can be adapted to systems other than RNA binding. Cooperative binding is also utilized by antibodies [10, 13], DNA-binding proteins [12] and many components of phase-separated biological droplets [15, 16]. In all of these cases, the combination of multiple binding sites and their connection through flexible linkers increases binding affinity in a completely analogous way to multi-domain RNA binding.

For all these cooperative interactions it is important to understand the binding affinity of the complex. So far, existing models have only described cooperative binding between two domains, with flexible linkers between the domains of one binding partner. The concentration enhancing effects of flexible linkers were either modelled with a uniform distribution in a sphere with radius L around the first domain, where L is the length of the chain. Other approaches used the more detailed worm-like chain model.

Here, we develop an equilibrium thermodynamic model for cooperative RNA-protein interactions and use the worm-like chain model for the behavior of flexible polymer chains. We validate our analytical calculations with stochastic simulations using Gillespie’s algorithm [17, 18]. Using our model, we can show that the dissociation constant decreases exponentially with each added binding site, by which very high affinities and specificities can be achieved with low-affinity and

Figure 1: Many RBPs have more than one domain. Numbers of RNA binding domains per protein. The data in the graph was taken from the RNA-binding protein database (RBPDB) [4] and shows how many RBDs every protein in this database has. It includes proteins from human, mouse, Drosophila and Caenorhabditis elegans.

-specificity RNA binding domains. Our calculations can be extended to any number of binding sites and are able to also include situations, in which both binding partners contain flexible linkers between binding sites. We compare predictions of our model to four proteins, for which the affinities of individual domains have been measured and show that our estimated K_d s are in good agreement with the experimental values.

Material and Methods

We will first describe our model and assumptions. We will then illustrate how the dissociation constant for a protein with two or three RNA binding domains can be derived, and generalize the derivation to any number of binding domains.

Simple cooperative binding model

The model describes the cooperative, multivalent binding of RNA-binding proteins possessing N RNA-binding domains to an RNA with N binding sites (Figure 2A). To be able to analytically calculate the effective dissociation constant for the proteins and its RNA substrate, we need to make two simplifying assumptions. First, we assume that each RNA-binding domain can only bind to a single, cognate binding site on the RNA, so domain 1 to RNA site 1, domain 2 to RNA site 2, and so on. We further assume that an RNA is at most bound by a single protein. This is good approximation as long as the local concentration of domains of the already bound protein at the RNA sites is much larger than the background protein concentration. In that regime, the first-bound protein will outcompete all other proteins from binding to its protein.

The transitions in this model from one binding con-

figuration to the next can be described by a coupled reaction system with 2^N different binding configurations, indexed by a binary string that indicates which sites are bound, e.g. (101) represents the configuration in which the first and third sites on the RNA are bound. Figure 2B shows the reaction network for $N = 3$ illustrated as a cube. The dissociation constants of the individual binding steps are called $K_{d,(x,y)}$, for a reaction between configuration x and configuration y . The effective concentration of the RNA-binding domain at another RNA site when one domain is already bound to its site is denoted as c_{d_i,L_i} , with d_i the distance between domains and L_i the chain length between RNA binding sites. For example if we consider the reaction $001 \rightleftharpoons 011$ we would write the equilibrium constant as

$$K_{d,(001,011)} = \frac{[001]}{[011]}, \quad (1)$$

and the concentration at the second binding site as c_{d_1,L_1} .

Effective concentrations using the worm-like chain model

To estimate the effective concentration of a protein domain at an RNA binding site if another domain of the protein already binds the RNA we model the RNA and the disordered peptide linker between RNA-binding domains as worm-like chains [19]. We use estimates of the persistence length l_p of RNA and disordered peptides, and we call L the chain length between RNA-binding sites. The probability density at an end-to-end distance of d , where d is the distance between binding domains on the protein, can be interpreted as the local concentration of the domain at its RNA binding site, as seen in figure 2A. Given a large enough chain length between the binding sites, $L \gg l_p$, the radial distribution function tends towards a Gaussian distribution,

$$c_{d_i,L_i} = \frac{1}{(2\pi\sigma^2)^{\frac{3}{2}}} \exp\left(-\frac{d_i^2}{2\sigma^2}\right), \quad (2)$$

with a variance of

$$\sigma^2 = \frac{2}{3}l_pL_i. \quad (3)$$

This assumes a rigid connection between protein domains. If the protein has flexible linkers between the binding domains, allowing them to move independently, we get (see Supplementary Methods for a more detailed derivation)

$$c_{R_1,R_2,L_{\text{RNA}},L_{\text{Protein}}} \approx \frac{1}{(2\pi\sigma^2)^{\frac{3}{2}}} \frac{\exp\left(-\frac{1}{2\sigma^2}(R_1 - R_2)^2\right) - \exp\left(-\frac{1}{2\sigma^2}(R_1 + R_2)^2\right)}{2R_1R_2/\sigma^2}, \quad (4)$$

with a variance of

$$\sigma^2 = \frac{2}{3}l_{p,\text{RNA}}L_{\text{RNA}} + \frac{2}{3}l_{p,\text{Protein}}L_{\text{Protein}}. \quad (5)$$

Figure 2: Thermodynamic model for cooperative RNA-protein interactions. (A) Illustration of an RNA with an RNA-binding protein (RBP) binding to it. All binding sites on the RNA are only bound by one domain of the RBD. Each of these interactions has its individual K_d . The worm like chain model is used to estimate the concentration $c_{d,L}$ of an RNA-binding domain at an RNA binding site when another domain at distance d in the protein is already bound to the RNA a linker length L away. **(B)** Representation of the reaction network at the example of three binding sites on the RNA. Every possible reaction step has a dissociation constant equal to the individual K_d for the domain-to-RNA-site interaction times the concentration of the domain at its cognate site (eq. 8).

Calculating K_d values for two or three domains using equilibrium thermodynamics

As shown before, the total dissociation constant for a protein with two domains binding to a chain with two binding sites can be written in terms of association constants of its individual domains as

$$K_{a,\text{tot}}(2) = K_{a1} + K_{a2} + c_{d_1,L_1} K_{a1} K_{a2}. \quad (6)$$

To get some intuition for this equation, we consider the law of mass action for any binding step, when at least one domain is already bound, for example $01 \rightleftharpoons 11$:

$$K_{a1} = \frac{[11]}{[01]c_{d_1,L_1}} \quad (7)$$

and by rearranging we get

$$K_{a1}c_{d_1,L_1} = \frac{[11]}{[01]} = K_{a,(01,11)}. \quad (8)$$

Thus, equation 6 can also be written as

$$K_{a,\text{tot}}(2) = K_{a,(00,10)} + K_{a,(00,01)} + K_{a,(00,10)} K_{a,(10,11)}. \quad (9)$$

In the Supplementary Methods we derive the K_d for $N = 2$ for the alternative case where the two RNA-binding domains have the same specificities such that each of them can bind to any of the two binding motifs on the RNA.

To extend the previous treatment to three binding sites, we write the total dissociation constant K_d in terms of the individual concentrations of all binding configurations. Simple rearranging leads to the K_d in terms of

dissociation constants of the individual domains (Supplementary Methods),

$$K_{a,\text{tot}}(3) = K_{a1} + K_{a2} + K_{a3} + c_{d_1,L_1} K_{a1} K_{a2} + c_{|d_1+d_2|,L_1+L_2} K_{a1} K_{a3} + c_{d_2,L_2} K_{a2} K_{a3} + c_{d_1,L_1} c_{d_2,L_2} K_{a1} K_{a2} K_{a3}, \quad (10)$$

or in a more general notation

$$K_{a,\text{tot}}(3) = K_{a,(000,100)} + K_{a,(000,010)} + K_{a,(000,001)} + K_{a,(000,100)} K_{a,(100,110)} + K_{a,(000,100)} K_{a,(100,101)} + K_{a,(000,010)} K_{a,(010,011)} + K_{a,(000,100)} K_{a,(100,110)} K_{a,(110,111)}. \quad (11)$$

Calculating K_d values for N binding sites

To generalize this approach to N binding sites, we can similarly express the observed dissociation constant according to the law of mass action in terms of the equilibrium concentrations of all states:

$$K_{a,\text{tot}}(N) := \frac{\sum_{x \neq (0 \dots 0)} [x]}{[0 \dots 0][\text{RBP}]} = \sum_{x \neq (0 \dots 0)} \frac{[x]}{[0 \dots 0][\text{RBP}]}. \quad (12)$$

$[x]$ denotes here the concentration of the x 'th state, $[0 \dots 0]$ is the concentration of the unbound RNA and $[\text{RBP}]$ the concentration of the unbound protein. The individual terms can be interpreted as apparent association constants for a theoretical one-step reaction from the unbound state to state x , so that

$$K_{a,\text{tot}}(N) = \sum_{x \neq (0 \dots 0)} K_{a,(0 \dots 0,x)}. \quad (13)$$

The $K_{a,(0\dots 0,x)}$ can be written as a product

$$K_{a,(0\dots 0,x)} = \frac{1}{[\text{RBP}]} \frac{[x]}{[0\dots 0]} \quad (14)$$

$$= \frac{1}{[\text{RBP}]} \prod_{i=1}^{|x|} \frac{[p_x(i)]}{[p_x(i-1)]}, \quad (15)$$

where $|x|$ is the number of bound sites at state x and p_x describes a path from the unbound state $(0\dots 0)$ to state $x = p_x(|x|)$. We define $K_{a,p_x(i-1) \rightleftharpoons p_x(i)}$ as the association constant of the reaction between configurations $p_x(i-1)$ and $p_x(i)$ along the path p_x (Figure 2B). Because for the first factor we can write

$$\frac{[p_x(1)]}{[\text{RBP}][0\dots 0]} = K_{a,(0\dots 0) \rightleftharpoons p_x(1)}, \quad (16)$$

equation (15) can finally be written as

$$K_{a,(0\dots 0,x)} = \prod_{i=1}^{|x|} K_{a,p_x(i-1) \rightleftharpoons p_x(i)}. \quad (17)$$

It can readily be seen that

$$K_{a,\text{tot}}(N) = \sum_{x \neq (0\dots 0)} \prod_{i=1}^{|x|} K_{a,p_x(i-1) \rightleftharpoons p_x(i)} \quad (18)$$

is a generalization of equations (10) and (6).

If the individual K_a s are large, the last term of the sum dominates

$$K_{a,\text{tot}}(N) \approx K_{a,1} \prod_{i=2}^N (K_{a,k} c_{d_{i-1}, L_{i-1}}). \quad (19)$$

This shows how each added binding site approximately multiplies the total K_a by a factor $K_{a,k} c_{d_{i-1}, L_{i-1}}$.

Simulation of cooperative binding with Gillespie algorithm

We validated our analytical calculations, described above using the Gillespie algorithm [17, 18] implemented in the Python library Gillespy2 [20]. We performed simulations of our model by defining all binding configurations as molecular entities in the simulation and determining the K_d based on trajectories of the systems (Supplementary methods).

Determining the model parameters

K_d values of individual binding domains are taken from experimental measurements like electrophoretic mobility shift assays (EMSA) or isothermal titration calorimetries (ITC). Distances between binding sites on the protein are Euclidian distances calculated based on available PDB structures. The contour lengths of ssRNA linkers between binding sites and the length of flexible linkers between protein domains are estimated as the number of nucleotides or amino acids multiplied with a length per base of 5.5 Å (mean of

Figure 3: Measured K_d s are in good agreement with model predictions. We found four RBPs composed of two RBDs for which dissociation constants of the full-length protein had been measured together with those of individual RBDs [27, 28, 29, 30, 31]. We used the simple thermodynamic model to estimate the effective K_d 's of the full-length RBPs from those of their individual domains and from linker lengths and protein binding site distances d and found agreement within a factor of ~ 5 . No free fitting parameters were used (See Supplementary Methods for details).

5 measurements) [21, 22, 23, 24, 25] or a length per aminoacid of 3.8 Å [26] respectively. The persistence length l_p of ssRNA and disordered proteins is estimated as 2.7 nm (mean of 5 measurements) [21, 22, 23, 24, 25] and 3.04 Å [26] respectively.

Results

The thermodynamic model correctly estimates measured dissociation constants

To validate our model, we estimated the dissociation constants for several RBPs by performing simulations and using our analytical approach (Figure 3). For these proteins the K_d values for individual domains and the whole protein have been measured experimentally. The proteins used were the zipcode binding protein 1 (ZBP1) [27], the heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) [28], the two terminal domains of the polypyrimidine tract binding protein (PTB) [29, 30] and the first four domains of the insulin-like growth factor 2 mRNA-binding protein 3 (IMP3 or IGF2BP3) [31]. The first three of these proteins consist of two rigidly linked domains. In contrast, IMP3 consists of three domain pairs with flexible linkers between the pairs. In our model the first two of the three IMP3 domain pairs were represented as two binding sites, connected by a flexible linker. All predictions were at least within a factor ~ 5 of the experimental value (see Supplementary Information), demonstrating the applicability of the model to multivalent, cooperative binding of RBDs to their RNA substrates.

K_d decreases exponentially with the number of binding sites

We then asked how the effective dissociation constants for RBPs depends on the number N of their RBDs (Figure 4A). We chose K_d values for RBDs and linker lengths in the ranges of typical RBPs. We observed an exponential increase in affinity with the number of binding sites with an increase by a factor $K_{a,k} c_{d,L}$ for each added domain (equation 19). The local concentration of the RBDs, $c_{d,L}$, depends on the linker length L between consecutive binding sites and the distance d between the consecutive RBDs as shown in equation (2). While the factor in real RBPs will depend on individual K_d s and distances between binding sites, the analysis shows that the global K_d can drop by orders of magnitude per domain added.

Contributions of individual domains to the total K_d becomes negligible after a threshold in the individual K_d

To further investigate the effect of domain K_d s to the total affinity, we calculated the dissociation constants for artificial RBPs with 3 domains, kept the K_d of the first domain constant and varied $K_{d,2}$ and $K_{d,3}$ (Figure 4B). As expected, the total K_d increases when the K_d of one individual domain is increased. According to equation (18), when $K_{a,k} c_{d_{i-1}, L_{i-1}} \leq 1$, or, equivalently, $K_{d,k} \geq c_{d_{i-1}, L_{i-1}}$, the contribution of domain k to the total K_d quickly saturates (horizontal line in Fig. 4B).

Multidomain RNA binding proteins can achieve high specificity by modularity and a varying binding site density

It is known that the combination of multiple RNA binding domains is important for providing the specificity needed to bind to the correct target RNAs [32]. The density of binding sites on the RNA molecule is also an important indicator of binding affinity and specificity [33]. To investigate this effect, we calculated the K_d and the relative occupancy of RBPs as a function of the binding site density on the RNA (Figure 4C and D). To simplify, we assumed that the domains always bind sequential binding sites on the RNA. Intuitively, the occupancy increases with increasing binding site density for all numbers of protein domains. In proteins with multiple domains, however, the transition in the binding curve is much more rapid than for proteins with one domain, allowing high levels of occupancy even at low binding site densities. To quantify the cooperativity, we fitted a sigmoidal function $1/(1 + (D/D_0)^h)$ with binding site density on the RNA of D and Hill coefficient h . We observe that the Hill coefficient, which is a common measure of cooperativity, is close to the number of RBDs, although growing somewhat faster ($n_1 = 0.96$, $n_2 = 2.26$, $n_3 = 3.88$ and $n_4 = 5.65$ for one, two, three and four domains, respectively).

Discussion

Simplified assumptions can limit model accuracy

Our estimated K_d deviates to some extent from the experimental measurement (Figure 3). Some simplifications in the model can potentially explain these differences. Most notably a lot of linkers between RNA binding sites are very short. To estimate c_{eff} we use a Gaussian distribution, which is valid if the chain length is much larger than the persistence length of the RNA, $L \gg l_p$. If the chain length is shorter, the end-to-end distribution will not be radially symmetric anymore, but depend on the initial tangent orientation of the bound end [34]. It has been shown that only for $\frac{L}{2l_p} > 5.6$ the distribution has a single maximum near the origin in direction of the initial orientation and approaches a Gaussian for larger values [34]. The chain lengths in the examples given earlier correspond to rather stiff chains. Depending on the orientation of the next binding site in relation to the first, the effective concentration and consequently the final K_d can be overestimated or underestimated. In order to increase the accuracy of estimates for c_{eff} one would have to take into account other geometric properties of the protein next to the distance between binding sites. However, for short-range systems, the analytical solution to the worm-like chain becomes highly complex and with alternatives like molecular dynamics simulations the wanted simplicity and intuition of the model is likely to be lost.

In addition to the short RNA linkers, RNA secondary structure and unspecific binding might also play a role in decreasing accuracy of the predictions. Furthermore, the sequence of the RNA also influences its flexibility. A lot of measurements of the persistence length of ssRNA are being done with very repetitive sequences. This effect will most likely average out for long chains, but can still be present in short ranges considered here.

Thermodynamic model extends previous models of cooperative binding

The described model allows to estimate the thermodynamic behavior of cooperative interactions for any number of binding sites. There are existing quantitative descriptions of cooperative binding. Crothers and Metzger developed a model to determine the overall K_d of the two binding sites of an antibody, estimating c_{eff} with the particle-in-a-sphere model, assuming, that the RNA binding site is uniformly distributed inside a sphere with a radius of L around the first already bound binding site [10]. This model has been extended several times, taking into account different properties like chain length of the flexible linker between binding sites/domains and also transferring it into the context of RNA binding [11, 12, 13, 14]. All of these studies, however, derive dissociation constants for two domains. The results for $N = 2$ match our model and have now been extended to an arbitrary number of binding sites.

Figure 4: Dependence of the effective dissociation constants of RBPs on the number and K_d 's of RBDs, the effective concentration $c_{d,L}$, and the binding site density on the RNA. (A) Dissociation constants decrease exponentially with the number of binding domains. All RBD K_d 's were set to 10 μ M and distances d between rigidly linked binding domains to 2 nm. The relationship between the number of binding sites and the $K_{d,tot}$ is shown for RNA binding sites spaced at 10 nt, 20 nt, or 40 nt. (B) Individual RBDs contribute to the total K_d up to a threshold $c_{d,L}$ in the individual K_d . In the calculations we used a K_d of 10 μ M for the first domain. The K_d s for the second and third domain are $K_{d,2}$ and $K_{d,3}$, as indicated in the plot. This was calculated for equal distances between rigidly linked binding domains of 2 nm and an RNA linker length of 20 nt. (C) Dissociation constants decrease with the binding site density on the RNA. As a simplification we assumed that the domains bind sequentially the binding sites on the RNA. K_d s of individual domains were 100 μ M and the total RNA length was 200 nt. The horizontal line indicates a concentration of 0.1 μ M used for the calculations in Figure D. (D) Relative occupancy of the RNA as measured by $[RNA]/(K_d + [RNA])$ as a function of binding site density on the RNA at an RNA concentration of 0.1 μ M (horizontal line in Figure C). Curves show fits with sigmoidal Hill-functions, with Hill coefficients of $n_1 = 0.96$, $n_2 = 2.26$, $n_3 = 3.88$ and $n_4 = 5.65$ for one to four domains, respectively.

In addition, previous models have only described flexible linkers between binding sites on one binding partner. Since both the RNA is a semiflexible polymer and there are many examples of proteins with flexible linkers between the domains, this has been extended in our model to allow flexible linkers on both binding partners.

Other mechanisms for increasing affinity and specificity

Our model presented here assumes that RBPs bind to well defined binding sites. In many cases, however, the short and degenerate motifs on the RNA can be repeated, creating the possibility for multiple binding registers [2, 3, 35]. This can be seen for example in the HuR C-terminal RRM binding to AU-rich RNA regions [36] and also in PTB, one of our examples, which binds to polypyrimidine tracts [37]. When the repeat regions are long enough, the proteins can bind in more

than one arrangement. The effects on the affinity by encompassing n binding registers in the RNA motif can be estimated through a simple statistical consideration by a linear factor of n .

Moreover, the concept of "fuzziness" has been qualitatively described by Olsen et al. [38]. It describes the more general situation when every RNA binding site can at least to some degree bind to every protein domain. We calculate this effect in our model for two binding sites (Supplementary Methods). Including fuzzy binding in the calculations increases the number of possible bound configurations and thus the complexity of the combinatorics. However, it does not qualitatively change the results that we present here.

Multidomain RNA binding can promote phase separation

It is well known that phase-separated biological droplets, which function to concentrate and organize molecules inside the cell, form multivalent networks. These multivalent interactions can arise from weak interactions between intrinsically disordered regions of the proteins and/or by multivalency through multiple connected domains [16]. We are, however, still lacking a good quantitative description of the underlying mechanisms behind phase separation in biology. Many stages of RNA metabolism also involve phase separation [16, 39, 40]. These condensates that form around RNA molecules have been called scaffolded condensates [41, 42]. Understanding cooperative multidomain binding in the simple case described here, could also give some insight into the formation of phase-separated RNA-protein complexes.

Conclusion

We have shown here a general method for evaluating the affinity of multivalent cooperative binding interactions based on the affinity of the individual non-cooperative interactions. This model underlines the large importance of cooperativity in multivalent binding. It demonstrates how highly specific and affine interactions can be made possible even at very low concentrations. Besides RNA-protein binding this model can be applied to many different systems such as antibodies or the binding of proteins to intrinsically disordered regions in proteins.

Code and data availability

The code used for the simulations and calculations is available at https://github.com/soedinglab/cooperative_rbp. The protein structures used in the validation of the model are available under the PDB accession codes 2n8l, 6dcl, 2adc, 6fq1 and 6gqe.

Contributions

SHS implemented the algorithms and conducted all the analysis. JS conceptualized the idea. SSJ and JS supervised research. SHS, SSJ, and JS wrote the manuscript.

Funding

We acknowledge support by the focus program SPP2191 of the Deutsche Forschungsgemeinschaft.

References

- Dreyfuss, G., Kim, V. N., and N., K. (2002) Messenger-RNA-binding proteins and the messages they carry. *Nat. Rev. Mol. Cell. Biol.*, **3**, 195–205.
- Mackereth, C. D. and Sattler, M. (2012) Dynamics in multi-domain protein recognition of RNA. *Curr. Opin. Struct. Biol.*, **22**, 287–296.
- Helder, S., Blythe, A. J., Bond, C. S., and Mackay, J. P. (2012) Determinants of affinity and specificity in RNA-binding proteins. *Curr. Opin. Struct. Biol.*, **38**, 83–91.
- Cook, K. B., Kazan, H., Zuberi, K., Morris, Q., and Hughes, T. R. (2010) RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res.*, **39**, D301–D308.
- Dominguez, D., Freese, P., Alexis, M. S., Su, A., Hochman, M., Palden, T., Bazile, C., Lambert, N. J., Van Nostrand, E. L., Pratt, G. A., Yeo, G. W., Graveley, B. R., and Burge, C. B. (2018) Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Mol. Cell*, **70**, 854–867.
- Ray, D., Kazan, H., Cook, K. B., Weirauch, M. T., Najafabadi, H. S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., Na, H., Irimia, M., Matzat, L. H., Dale, R. K., Smith, S. A., Yarosh, C. A., Kelly, S. M., Nabet, B., Mecnas, D., Li, W., Laishram, R. S., Qiao, M., Lipshitz, H. D., Piano, F., Corbett, A. H., Carstens, R. P., Frey, B. J., Anderson, R. A., Lynch, K. W., Penalva, L. O. F., Lei, E. P., Fraser, A. G., Blencowe, B. J., Morris, Q. D., and Hughes, T. R. (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, **499**, 172–177.
- Jolma, A., Yan, J., Whittington, T., Toivonen, J., Ritta, K. R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., Palin, K., Vaquerizas, J. M., Vincetelli, R., Luscombe, N. M., Hughes, T. R., Lemaire, P., Ukkonen, E., Kivioja, T., and Taipale, J. (2013) DNA-Binding Specificities of Human Transcription Factors. *Cell*, **152**, 327–339.
- Franco-Zorilla, J. M., López-Vidriero, I., Carrasco, J. L., Godoy, M., Vera, P., and Solano, R. (2014) DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc. Natl. Acad. Sci. USA*, **111**, 2367–2372.
- Zhu, C., Byers, K. J., McCord, R. P., Shi, Z., Berger, M. F., Newburger, D. E., Saulrieta, K., Smith, Z., Shah, M. V., Radhakrishnan, M., Philippakis, A. A., Hu, Y., De Masi, F., Pacek, M., Rolfs, A., Murthy, T. ad LaBear, J., and Bulky, M. L. (2009) High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome Res.*, **19**, 556–566.
- Crothers, D. M. and Metzger, H. (1972) The influence of polyvalency on the binding properties of antibodies. *Immunochemistry*, **9**, 341–357.
- Shamoo, Y., Abdul-Manan, N., and Williams, K. R. (1995) Multiple RNA binding domains (RBDs) just don't add up. *Nucleic Acids Res.*, **23**, 725–728.
- Zhou, H.-X. (2001) The Affinity-Enhancing Roles of Flexible Linkers in Two-Domain DNA-Binding Proteins. *Biochemistry*, **40**, 15069–15073.
- Zhou, H.-X. (2003) Quantative Account of the Enhanced Affinity of Two Linked scFvs Specific for Different Epitopes on the Same Antigen. *J. Mol. Biol.*, **329**, 1–8.
- Bobrovnik, S. A. (2007) The influence of rigid or flexible linkage between two ligands on the effective affinity and avidity for reversible interactions with bivalent receptors. *J. Mol. Recognit.*, **20**, 253–262.
- 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.
- Sabari, B. S., Dall'Agnese, A., and Young, R. A. (2020) Biomolecular Condensates in the Nucleus. *Trends Biochem. Sci.*, **45**, 961–977.
- Gillespie, D. T. (1976) A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. *J. Comput. Phys.*, **22**, 403–434.
- Gillespie, D. T. (1977) Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, **81**, 2340–2361.

- 19 Kratky, O. and Porod, G. (1949) Röntgenuntersuchungen gelöster Fadenmoleküle. *Recl. Trav. Chim. Pays-Bas*, **68**, 1106–1122.
- 20 Abel, J. H., Drawert, B., Hellander, A., and Petzold, L. R. (2016) GillesPy: A Python Package for Stochastic Model Building and Simulation. *IEEE Life Sci. Lett*, **2**, 35–38.
- 21 Chi, Q., Wang, G., and Jiang, J. (2013) The persistence length and length per base of single-stranded DNA obtained from fluorescence correlation spectroscopy measurements using mean field theory. *Physica A*, **393**, 1072–1079.
- 22 Chen, H., Meisburger, S. P., Pabit, S. A., Sutton, J. L., Webb, W. W., and Pollack, L. (2012) Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proc. Natl. Acad. Sci. USA*, **109**, 799–804.
- 23 Laurence, T. A., Kong, X., Jäger, M., and Weiss, S. (2005) Probing structural heterogeneities and fluctuations of nucleic acids and denaturated proteins. *Proc. Natl. Acad. Sci. USA*, **48**, 17348–17353.
- 24 Murphy, M. C., Rasnik, I., Cheng, W., Lohman, T. M., and Ha, T. (2004) Probing Single-Stranded DNA Conformational Flexibility Using Fluorescence Spectroscopy. *Biophys. J.*, **86**, 2530–2537.
- 25 Mills, J., Vacano, E., and Hagerman, P. J. (1999) Flexibility of single-stranded DNA: use of gapped duplex helices to determine the persistence lengths of Poly(dT) and Poly(dA). *J. Mol. Biol.*, **285**, 245–257.
- 26 Zhou, H.-X. (2001) Loops in Proteins Can Be Modeled as Worm-Like Chain. *J. Phys. Chem. B*, **105**, 6763–6766.
- 27 Nicastro, G., Candel, A. M., Uhl, M., Oregioni, A., Hollingworth, D., Martin, S. R., and Ramos, A. (2017) Mechanism of β -actin mRNA Recognition by ZBP1. *Cell Rep.*, **31**, 1187–1199.
- 28 Kooshapur, H., Choudhury, N. R., Simon, B., Mühlbauer, M., Jussopow, A., Fernandez, N., Jones, A. N., Dallmann, A., Gabel, F., Camilloni, C., Michlewski, G., Caceres, J. F., and Sattler, M. (2018) Structural basis for terminal loop recognition and stimulation of pri-miRNA-18a processing by hnRNP A1. *Nat. Commun.*, **9**, 2479.
- 29 Lamichhane, R., Daubner, G. M., Thomas-Crusells, J., Auweter, S. D., Manatschal, C., Austin, K. S., Valniuk, O., Allain, F. H.-T., and Rueda, D. (2010) RNA looping by PTB: Evidence using FRET and NMR spectroscopy for a role in splicing repression. *Proc. Natl. Acad. Sci. USA*, **107**, 4105–4110.
- 30 Oberstrass, F. C., Auweter, S. D., Erat, M., Y., H., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D. L., and Allain, F. H.-T. (2005) Structure of PTB bound to RNA: Specific Binding and Implications for Splicing Regulation. *Science*, **309**, 2054–2057.
- 31 Schneider, T., Hung, L.-H., Aziz, M., Wilmen, A., Thaum, S., Wagner, J., Janowski, R., Müller, S., Schreiner, S., Friedhoff, P., Hüttemaier, S., Niessing, D., Sattler, M., Schlundt, A., and Bindereif, A. (2019) Combinatorial recognition of clustered RNA elements by the multidomain RNA-binding protein IMP3. *Nat. Commun.*, **10**, 2266.
- 32 Lunde, B. M., Moore, C., and Varani, G. (2007) Modular design for efficient function. *Nat. Rev. Mol. Cell Biol.*, **8**, 470–490.
- 33 Danilenko, M., Dalgliesh, C., Pagliarini, V., Naro, C., Ehrmann, I., Feracci, M., Kheirollahi-Chadegani, M., Tyson-Capper, A., Clowry, G. J., Fort, P., Dominguez, C., Sette, C., and Elliott, D. J. (2017) Binding site density enables paralog-specific activity of SLM2 and Sam68 proteins in Neurexin2AS4splicing control. *Nucleic Acids Res.*, **45**, 4120–4130.
- 34 Spakowitz, A. J. and Wang, Z.-G. (2005) End-to-End distance vector distribution with fixed end orientations for the wormlike chain model. *Phys. Rev. E*, **72**, 041802.
- 35 Sohrabi-Jahromi, S. and Söding, J. (2021) Thermodynamic modeling reveals widespread multivalent binding by RNA-binding proteins. *unpublished*.
- 36 Ripin, N., Boudet, J., Duszczek, M. M., Hinniger, A., Faller, M., Krepl, M., Gadi, A., Schneider, R. J., Šponer, J., Meisner-Kober, N. C., and Allain, F. H.-T. (2019) Molecular basis for AU-rich element recognition and dimerization by the HuR C-terminal RRM. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 2935–2944.
- 37 Auweter, S. D., Oberstrass, F. D., and Allain, F. H.-T. (2006) Solving the Structure of PTB in Complex with Pyrimidine Tracts: An NMR Study of Protein-RNA Complexes of Weak Affinities. *J. Mol. Biol.*, **367**, 174–186.
- 38 Olsen, J. G., Teilum, K., and Kragelund, B. B. (2017) Behaviour of intrinsically disordered proteins in protein-protein complexes with an emphasis on fuzziness. *Cell. Mol. Life Sci.*, **74**, 3175–3183.
- 39 Fay, M. M. and Anderson, P. J. (2018) The Role of RNA in Biological Phase Separations. *J. Mol. Biol.*, **430**, 4685–4701.
- 40 Lin, Y., Protter, D. S. W., Rosen, M. K., and Parker, R. (2015) Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol. Cell*, **60**, 208–219.
- 41 Rhine, K., Vidaurre, V., and Myong, S. (2020) RNA droplets. *Annu. Rev. Biophys.*, **49**, 247–265.
- 42 Söding, J., Zwicker, D., Sohrabi-Jahromi, S., Boehning, M., and Kirschbaum, J. (2020) Mechanism for Active Regulation of Biomolecular Condensates. *Trends Cell Biol.*, **30**, 4–14.