Quantitative cooperative binding model for intrinsically disordered proteins interacting with nanomaterials

Da-Wei Li,[1*] Mouzhe Xie,[2] and Rafael Brüschweiler[1,2,3*]

[1]Campus Chemical Instrument Center, The Ohio State University, Columbus, Ohio 43210, USA

[2]Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, USA

[3]Department of Biological Chemistry and Pharmacology, The Ohio State University, Columbus, Ohio 43210, USA

[*]To whom correspondence should be addressed:

Da-Wei Li, Ph.D., E-mail: lidawei@gmail.com

Rafael Brüschweiler, Ph.D., E-mail: bruschweiler.1@osu.edu

ABSTRACT

Intrinsically disordered proteins (IDPs) can display a broad spectrum of binding modes and highly variable binding affinities when interacting with both biological and non-biological materials. A quantitative model of such behavior is important for the better understanding of the function of IDPs when encountering inorganic nanomaterials with the potential to control their behavior *in vivo* and *in vitro*. Depending on their amino-acid composition and chain length, binding properties can vary strongly between different IDPs. Moreover, due to differences in the physical chemical properties of clusters of amino-acid residues along the IDP primary sequence, individual residues can adopt a wide range of bound state populations. Quantitative experimental binding affinities with synthetic silica nanoparticles (SNPs) at residue-level resolution, which were obtained for a set of IDPs by solution NMR relaxation experiments, are explained here by a first-principle analytical statistical mechanical model termed SILC. SILC quantitatively predicts residue-specific binding affinities to nanoparticles and it expresses binding cooperativity as the cumulative result of pairwise residue effects. The model, which was parametrized for anionic SNPs and applied to experimental data of four IDP systems with distinctive binding behavior, successfully predicts differences in overall binding affinities, fine details of IDP-SNP affinity profiles, and site-directed mutagenesis effects with a spatial resolution at the individual residue level. The SILC model provides an analytical description of such types of fuzzy IDP-SNP complexes and may help advance understanding nanotoxicity and *in-vivo* targeting of IDPs by specifically designed nanomaterials.

INTRODUCTION

Intrinsically disordered proteins (IDPs) and protein regions, which account for more than thirty percent of the eukaryotic proteome,[1] are involved in a vast range of biologically important processes, such as regulation, catalysis, or biomineralization.[2-4] Their high structural flexibility enables IDPs to interact in unique ways with other biomolecules or inorganic materials.[5] Examples include the interaction of the disordered transactivation domain of the human tumor suppressor p53 (p53TAD) with the globular domain of MDM2 protein,[6] the interaction of the disordered prokaryotic ubiquitin-like protein (Pup) with mycobacterium proteasomal ATPase,[7] and the interaction of disordered silaffin peptides with inorganic silica surfaces.[8] In some cases, these interactions are known to be highly cooperative,[9-11] but the lack of quantitative information at atomic detail has prevented a deeper understanding of such interactions and the physical-chemical basis for the emergence of cooperativity has remained unaddressed. The hallmark of cooperativity in these systems is that individual residues specifically affect the binding free energy of other residues of the same polypeptide, i.e. the net free energy of binding is generally larger than the sum of the free energy contributions of individual residues. Atomic-detail, physics-based models are needed to understand and predict IDP interactions with other biomolecules, nanomaterials, and their role in the formation of membraneless organelles.[12-13]

Elucidating the cooperative binding modes of IDPs to nanomaterial surfaces is relevant as it may provide the means to selectively control IDP behavior in living cells using functionalized nanoparticles with specifically engineered surface modifications. Such work is of potential pharmaceutical value as it may lead to ways to prevent or combat IDP-related diseases, including neurodegeneration.[14-15] The accurate characterization of IDP-nanoparticle interactions can also help understand nanoparticle-induced toxicity and other health risks from exposure to nanomaterials.[16] Moreover, such knowledge sheds light on how IDPs participate in biomineralization processes, which in turn enable bioinspired and green synthesis of functional nanomaterials.[4, 17]

We recently developed an experimental approach, based on high-resolution NMR in solution, to quantitatively study the interactions of IDPs with silica nanoparticles (SNPs).[18-19] SNPs are widely used as matrix materials in nanomedicine for enhanced biomolecular imaging and targeted drug delivery. In addition, the chemically heterogeneous SNP surface, containing

negatively charged deprotonated silanol groups (-Si-O$^-$), hydrophilic protonated silanol groups (-Si-OH), and hydrophobic siloxane (-Si-O-Si-) patches,[20] permits the evaluation of both electrostatic and hydrophobic interactions that collectively contribute to IDP adsorption. In the current study, the SNPs have an average diameter of 20 nm and, hence, display much slower tumbling motion than free, i.e. unbound IDP residues. Upon interaction with SNPs, the tumbling motion of the interacting residues slows down considerably, which is manifested in an increase of transverse R_2 spin relaxation of backbone ^{15}N spins that can be accurately measured for each residue. The approach thereby provides quantitative residue-specific information about the propensity and population of an IDP when bound to the nanoparticle surface.

The residue-specific binding profiles of four IDPs to SNP surfaces, namely p53TAD, Pup, human α-synuclein (aSyn), and cytosolic loop region 2 (Cyto2) of the canine sodium-calcium exchanger NCX, have been experimentally determined providing direct evidence about binding cooperativity.[19] These data show that the average bound population of a polypeptide can vary widely for different IDPs. In addition, cooperativity effects were also revealed when comparing IDP binding profiles of wild-type vs. site-directed mutants.[19]

Since interaction properties of IDPs with SNP surfaces should only depend on their primary sequences, the binding affinities of the 20 free amino acids to anionic SNP surfaces were recently quantified.[18] They indicate the strongest affinities for Arg, Pro, Lys, Ile, His, Leu, Phe, and Val and minimal or negative affinities for Glu, Asp, Ser, Gly. Based on this information, the analytical "free residue interaction model" (FRIM) was developed by combining individual contributions from each with local neighboring effects in an additive manner.[19] FRIM is able to explain the segmental variations of the bound population for each protein, but does not adequately describe observed cooperativity effects, including the large differences in overall binding affinities between different IDPs. Specifically, FRIM requires an empirical global scaling factor for the prediction of ΔR_2, which can strongly vary from protein to protein and it cannot adequately reproduce $\Delta\Delta R_2$ behavior for wild-type vs. mutant proteins. This suggests that FRIM does not adequately account for the non-additive dependence of free energy contributions to binding by individual residues. These shortcomings point to the need for a physically more advanced binding model.

Besides FRIM, few mathematical binding models have been proposed that capture such types of situations.[21-25] The earliest model is the divalent binding model by Crothers and Metzer originally developed for antigen-antibody interactions.[23] It predicts the dependence of the binding constant of a 2nd binding event on the binding constant of an initial binding event. The model expresses this dependence as a function of the spatial relationship between the two binding sites and the acceptor concentration. Advanced versions of this model with more realistic linker properties have subsequently been proposed.[24-25] These models are, however, not directly applicable to the situation examined in this work, which is the interaction between IDPs and nanoparticles where each amino-acid residue acts as a ligand and simultaneously is also part of a linker.

Here, we propose a new, statistical-mechanics based polyvalent binding model, termed SILC for SImultaneous Linker-ligand Cooperative binding, to quantitatively explain cooperative IDP-SNP binding with a resolution at the individual amino-acid residue level. The model is shown to represent in good approximation the partition function of the protein of the bound state by a product of pairwise residue contributions. Parametrization of the model using NMR-derived residue-specific binding affinities of the four IDPs shows that the model has quantitative predictive power being able to reproduce experimentally observed cooperative binding properties. In the following sections, the new model is first formally described, then parametrized using experimental data, and finally applied to explain cooperative binding behavior manifested in several different IDP systems.

Theory

In the SILC binding model introduced here, each amino acid interacts with the SNP surface through either hydrophobic or electrostatic interactions. The direct interaction between residues, which is restricted to sequential residues due to chemical bonding, is modeled as a constraint with energy $R_{i,i+1}$ keeping the distance between sequential residues essentially fixed. We first describe the model for a linear polypeptide chain that contains only 3 amino-acid residues before it is generalized to a polypeptide chain of arbitrary length. The partition function Z of the 3-residue system is

$$Z = \int\int\int d\mathbf{r}_1\, d\mathbf{r}_2\, d\mathbf{r}_3 \exp(-\beta(E_1 + E_2 + E_3 + R_{12} + R_{23})) \quad (1)$$

where E_1, E_2, and E_3 are interaction energies between residues 1, 2, and 3 with the SNP surface, respectively, where E_i only depends on the type of residue i and $\beta = 1/(k_B T)$. Because the distances between neighboring residues 1, 2 and 2, 3 are essentially fixed, this is accounted for by the pairwise distance-dependent energy terms R_{12} and R_{23}, which can be enforced by a steep harmonic inter-residue distance-dependent energy potential that is centered around the equilibrium distance between neighboring residues. Eq. (1) can be rearranged with the goal to approximate Z as a product of individual residue contributions. First, Eq. (1) can be expressed as a nested integral

$$Z = \int d\mathbf{r}_1 \{\exp(-\beta E_1) \int d\mathbf{r}_2 [\exp(-\beta(E_2 + R_{12}) \int d\mathbf{r}_3 \exp(-\beta(E_3 + R_{23})))]\} \quad (2)$$

Residue 2 affects residue 1 directly via R_{12}, whereas residue 3 affects residue 1 only indirectly via the intermedidate residue 2. Next, the energy term R_{23} is replaced by a new energy R_{13}, which is less restrictive than R_{23} by allowing low energy arrangements over a larger distance range than R_{23} and at the same time eliminating the explicit dependence on residue 2:

$$Z = \int d\mathbf{r}_1 \{\exp(-\beta E_1) \int d\mathbf{r}_2 [\exp(-\beta(E_2 + R_{12}) \int d\mathbf{r}_3 \exp(-\beta(E_3 + R_{13})))]\} \quad (3)$$

The functional form of R_{13} will be empirically parametrized below. Eq. (3) can be rewritten as

$$Z = \int d\mathbf{r}_1 \{\exp(-\beta E_1) \int d\mathbf{r}_2 \exp(-\beta(E_2 + R_{12})) \int d\mathbf{r}_3 \exp(-\beta(E_3 + R_{13}))\} \quad (4)$$

providing an expression for Z that only includes pairwise interactions with residue 1. It permits direct computation of the population p_1 of conformers with residue 1 bound to the nanoparticle surface (see below).

Eq. (4) can be generalized to any residue i in the polypeptide chain with N residues

$$Z = \int d\mathbf{r}_i \{\exp(-\beta E_i) \prod_{n \neq i} \int d\mathbf{r}_n \exp(-\beta(E_n + R_{i,n}))\} \quad (5)$$

where the energy terms $R_{i,n}$ have the same general form as R_{13} of Eq. (4) independent of residue type. The bound population p_i of residue i can then be expressed as the fraction of the partition function Z with residue i bound *vs.* the total partition function:

$$p_i = \frac{\int_{\mathbf{r}_i \in bound} d\mathbf{r}_i \{\exp(-\beta E_i) \prod_{n \neq i} \int d\mathbf{r}_n \exp(-\beta(E_n + R_{i,n}))\}}{\int_{\mathbf{r}_i \in bound} d\mathbf{r}_i \{\exp(-\beta E_i) \prod_{n \neq i} \int d\mathbf{r}_n \exp(-\beta(E_n + R_{i,n}))\} + \int_{\mathbf{r}_i \in unbound} d\mathbf{r}_i \{\exp(-\beta E_i) \prod_{n \neq i} \int d\mathbf{r}_n \exp(-\beta(E_n + R_{i,n}))\}} \quad (6)$$

In the case that population p_i is small, which applies for the experimental conditions used in this work, the denominator of Eq. (6) can be simplified:

$$p_i = \frac{\int_{\mathbf{r}_i \in bound} d\mathbf{r}_i \{\exp(-\beta E_i) \prod_{n \neq i} \int d\mathbf{r}_n \exp(-\beta(E_n + R_{i,n}))\}}{\int_{\mathbf{r}_i \in unbound} d\mathbf{r}_i \{\exp(-\beta E_i) \prod_{n \neq i} \int d\mathbf{r}_n \exp(-\beta(E_n + R_{i,n}))\}} \quad (7)$$

Eq. (7) can be further simplified by recognizing that the integrals over $\mathbf{r}_n$ have a trivial dependence on $\mathbf{r}_i$ for most of the allowed range of $\mathbf{r}_i$. If residue *i* is bound to the surface, the range of positions $\mathbf{r}_i$ is fully restricted and if residue *i* is in an unbound state far away from the nanoparticle surface, the precise location $\mathbf{r}_i$ has little effect on the integrals over $\mathbf{r}_n$. Only in the intermediate regime where residue *i* is located in the vicinity of the nanoparticle surface, the effect on $\mathbf{r}_n$ becomes more complicated. It is then possible to factorize the integrals in Eq. (7) in both the numerator and denominator by averaging out the $\mathbf{r}_i$-dependence of $\int d\mathbf{r}_n \exp(-\beta(E_n + R_{i,n}))$:

$$p_i = \frac{\int_{\mathbf{r}_i \in bound} d\mathbf{r}_i \exp(-\beta E_i)}{\int_{\mathbf{r}_i \in unbound} d\mathbf{r}_i \exp(-\beta E_i)} \prod_{n \neq i} \frac{\int_{i \in bound} d\mathbf{r}_n \exp(-\beta(E_n + R_{i,n}))}{\int_{i \in unbound} d\mathbf{r}_n \exp(-\beta(E_n + R_{i,n}))} \quad (8)$$

The ratio of the integrals over $\mathbf{r}_i$ in Eq. (8) is the intrinsic bound population of free amino acid *i* at the nanoparticle concentration [SNP]. Because $\int d\mathbf{r}_n \exp(-\beta(E_n + R_{i,n}))$ only depends on the intrinsic binding property of amino acid *n* and distance restraint between residue *i* and *n*, the bound population can be approximated as

$$p_i \approx \frac{[SNP]}{[SNP]_0} p_i^{intrinsic} \prod_{n \neq i} q_{n,|n-i|} \quad (9)$$

where $p_i^{intrinsic}$ is the bound population of the free amino acid *i* to the SNPs. $[SNP]_0$ is a reference nanoparticle concentration and

$$q_{n,|n-i|} = \frac{\int_{i \in bound} d\mathbf{r}_n \exp(-\beta(E_n + R_{i,n}))}{\int_{i \in unbound} d\mathbf{r}_n \exp(-\beta(E_n + R_{i,n}))} \quad (10)$$

are termed “binding contribution factors” (BCF), which only depend on the amino-acid type of residue n and the sequential distance between residues *n* and *i*.

The BCFs $q_{n,|n-i|}$ must fulfill certain general properties: for small values |*n*–*i*|, the restraining energy $R_{i,n}$ requires that residue *n* must be close to (far away from) the binding surface if residue *i* is in the bound (unbound) state. This implies that $q_{n,0} \propto \int_{n \in bound} d\mathbf{r}_n \exp(-\beta E_n)$ and, hence, $q_{n,0}$ is proportional to $p_n^{intrinsic}$, i.e., the effect of a residue *n* on the binding of residue *i* is directly proportional to the intrinsic binding affinity of the free amino acid form of residue *n*. On the other hand, if |*n*–*i*| becomes very large, $q_{n,|n-i|}$will approach 1, since the conformational space sampled by residue *n* does not depend on whether residue *i* is bound or unbound. These conclusions are independent of the details of the functional form of E_n, except that E_n should be more favorable (i.e. lower) near the surface of the nanoparticle for positively charged and hydrophobic residues and less favorable (i.e. higher) for negatively charged residues. For attractive interactions, $q_{n,|n-i|} > 1$ and for repulsive interactions, $q_{n,|n-i|} < 1$ (approaching 1 for $|n-i| \to \infty$).

The explicit mathematical form of $q_{n,|n-i|}$ can be specified based on an empirical parametrization using high-quality experimental data. In the following we use the following inverse quadratic functional form:

$$q_{n,|n-i|} = 1 + \frac{q_{n,0} - 1}{1 + a_n |n-i| + b_n |n-i|^2} \tag{11}$$

where |*n*–*i*| is the sequential distance between the two residues *n* and *i* and a_n, b_n are empirical parameters that depend on the amino-acid type of residue *n*. The specific values of the model parameters are determined by non-linear least squares fitting to our experimental NMR data (see Table 1). We term this model SILC.

MATERIALS AND EXPERIMENTAL METHODS

Sample preparation. NMR samples of four intrinsically disordered proteins were prepared: the transactivation domain of human tumor suppressor p53 (p53TAD) and several of its mutants, human α-synuclein (aSyn) and several of its mutants, prokaryotic ubiquitin-like protein (Pup), and the cytoplasmic segment 2 at the C-terminus of the canine Na^+/Ca^{2+} exchanger (Cyto2) (all protein sequences are listed in Table S1). Site-directed mutagenesis, expression and purification of ^{15}N or $^{15}N,^{13}C$ isotopically-labeled samples followed identical procedures as described previously and in the Supporting Information. All final NMR samples consisted of 170 μM IDPs in 20 mM sodium phosphate buffer (pH 7.0) that also contained 100 mM NaCl and 10% D_2O.

SNP preparation. Colloidal SNPs (AkzoNobel Bindzil 2040) were dialyzed in the same buffer using 10 kDa molecular-weight-cut-off semi-membrane for at least 16 h before mixing with IDPs. These near-spherical SNPs have a relatively narrow size distribution (19.5 ± 5.3 nm in diameter by transmission electron microscopy and 25.3 ± 8.9 nm by dynamic light scattering) and are negatively charged at the surface (ζ-potential –23.0 ± 7.4 mV) at neutral pH as characterized previously.[26] A range of different SNP concentrations was optimized for each IDP so that the maximum effective R_2 rates measured by spin relaxation experiment (see below) fall between 5 and 25 s^{-1}. This range was found to be a viable compromise between a large dynamic range for ΔR_2 and sufficient signal-to-noise, which together determine the measurement accuracy. Table S2 summarizes the concentrations and conditions of all samples used.

NMR relaxation experiments. ^{15}N-NMR spin relaxation parameters, R_1 and $R_{1\rho}$, of each protein were measured in the presence and absence of SNPs. Pseudo-3D experiments based on 2D ^{15}N-^1H HSQC experiments with residue-specific peak intensities were recorded at six different relaxation delays. The residue-specific R_1 and $R_{1\rho}$ rates were obtained by fitting the intensity vs. delay curves by a single exponential, and the R_2 rates were subsequently extracted based on their R_1 and $R_{1\rho}$ rates using ^{15}N spin-lock field strength and resonance offset parameters of each cross-peak. All NMR experiments were conducted on a Bruker Avance III HD spectrometer operated at 850 MHz ^1H frequency equipped with a TCI cryogenic probe at 298 K.

Extraction of bound-state populations. Experimental transverse NMR relaxation data reflect bound-state populations as follows. For a two-site exchange process in the fast exchange regime

where each amino-acid residue *i* exists either in a free or an SNP-bound state with populations p_i^f and p_i^b, the bound populations can be directly determined from

$$p_i^b = \frac{\Delta R_{2,i}}{R_2^b} \tag{12}$$

where $\Delta R_{2,i} = R_{2,i}^{SNP} - R_{2,i}^f$ is the R_2 difference in the presence and absence of SNPs. R_2^b is the (estimated) R_2 value for a residue that is bound to the SNP for 100% of the time (see Eq. (S4) of Ref. [18]). Eq. (12) is further supported by data shown in Figure S3.

RESULTS

Evidence for global cooperative binding

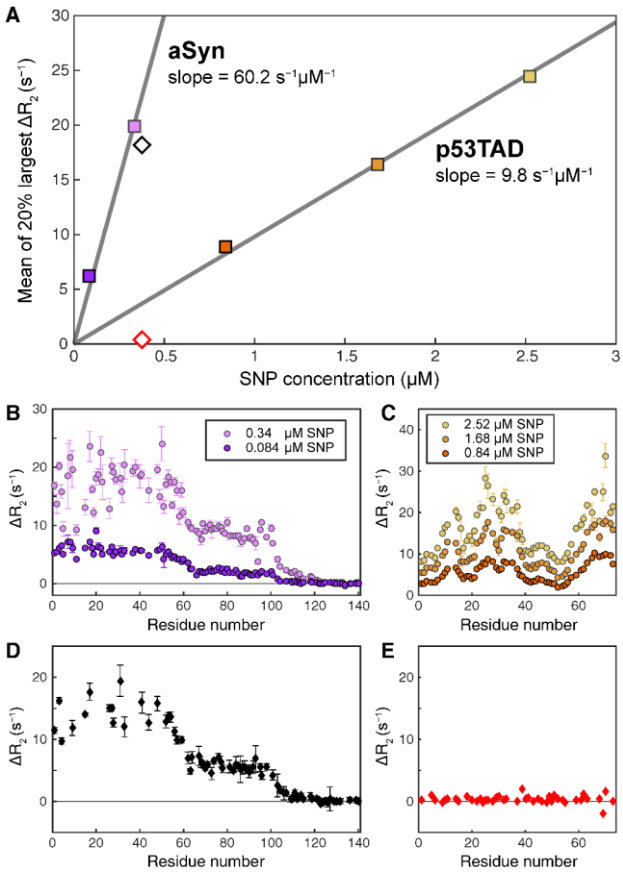


Figure 1. ^{15}N-ΔR_2 values of IDPs as a function of amino-acid residue sequence and SNP concentration. (A) Average ΔR_2 (of largest 20% residues) of aSyn and p53TAD display a linear dependence on SNP concentration, although with significantly different slopes (gray lines). Experimental interaction profiles of (B) aSyn and (C) p53TAD at different SNP concentrations measured for individually prepared samples. (D,E) Experimental interaction profiles of mixture of (D) 170 μM aSyn and (E) 170 μM p53TAD present in same NMR sample with 0.38 μM SNPs. The expected binding strengths of the IDPs of (D,E) are indicated as an open black diamond for aSyn and an open red diamond for p53TAD in Panel (A).

Experimental binding affinity profiles of the wild-type forms of the four IDPs, namely p53TAD, aSyn, Pup, and Cyto2, to silica nanoparticles were measured by determining the residue-specific ^{15}N ΔR_2 values, which is the difference between ^{15}N R_2 in the presence and absence of SNPs. The binding affinities grow linearly with the SNP concentration as is demonstrated for aSyn and

p53TAD in Figure 1A,B,C. Importantly, each IDP displays its individual slope of ΔR_2 vs. SNP-concentration, which is defined by the average of the largest 20% ΔR_2 values of each protein divided by the SNP concentration. The slope for aSyn is about 6.1 times larger than that for p53TAD, which shows that at identical protein and SNP concentration α-synuclein has a much stronger binding affinity than p53TAD. This finding is corroborated by a competitive binding assay of the two proteins: when the two proteins at a concentration of 170 μM each are mixed with 0.38 μM SNPs, aSyn-SNP binding completely dominates that of p53TAD (Figure 1D,E). This behavior cannot be explained by the FRIM model, which was developed previously for the characterization of IDP-SNP interactions. Since the FRIM model only applies in a regime of weak cooperativity, Figure 1 suggests that a new model is needed that can accommodate stronger cooperativity effects. The experimental data of Figure 1 therefore represent an important benchmark to test the new binding model.

Table 1. Fitted model parameters that enter Eq. (11)[a] of the SILC model

Residue type[b]	$q_{n,0}$	a_n	b_n	Residue type[b]	$q_{n,0}$	a_n	b_n
N-term	1.96	0.0982	0.00305	His (H)	1.130	0.521	0
Lys (K)	1.96	0.0982	0.00305	Ile (I)	1.132	0.521	0
Arg (R)	2.44	0.0982	0.00305	Leu (L)	1.126	0.521	0
Asp (D)	0.668	0.0982	0.00305	Met (M)	1.086	0.521	0
Glu (E)	0.668	0.0982	0.00305	Asn (N)	1.052	0.521	0
C-term	0.668	0.0982	0.00305	Gln (Q)	1.053	0.521	0
Pro (P)	1.167	0.521	0	Thr (T)	1.051	0.521	0
Ser (S)	1.050	0.521	0	Val (V)	1.095	0.521	0
Ala (A)	1.054	0.521	0	Trp (W)	1.078	0.521	0
Cys (C)	1.058	0.521	0	Tyr (Y)	1.058	0.521	0
Phe (F)	1.114	0.521	0	Gly (G)	1.051	0.521	0

[a] For standard SNP concentration

[b] Amino acids together with one-letter amino-acid abbreviations

Fitting of model parameters

The parameters of the SILC binding model were determined based on Eq. (9) using a nonlinear least-squares fit to the bound populations of the wild-type forms of all four IDPs. Experimental binding population data of IDP mutants were subsequently used for validation. Optimal results were obtained with 9 fitting parameters. Five of these parameters are the $q_{n,0}$ values of (i) positively charged lysine and the N-terminus, (ii) positively charged arginine, (iii) negatively charged glutamate, aspartate and the C-terminus, (iv) proline, which is the strongest interacting non-charged residue, and (v) serine, which is the weakest interacting neutral residue. The remaining fitting parameters were a_n, b_n for all charged residues, and a_n (with $b_n = 0$) for all other residues, and the global scaling factor $c = p_{\text{intrinsic}} / q_{n,0}$ for all residue types. The $q_{n,0}$ of all other non-charged residues were linearly interpolated between $q_{n,0}$ of serine and proline according to the experimental binding strengths of the free amino acids. Experimental profiles were all linearly scaled to a standard SNP concentration of 0.17 μM. The model parameters were determined by a non-linear least squares fit that minimizes deviation between experimental and predicted binding populations using a trusted region reflective algorithm as implemented in the software package Ceres-Solver was employed.[27] The minimization was run 5000 times using randomly selected initial parameters and the best fitted set is reported as final result (Table 1). At our standard SNP concentration of 0.17 μM, the fitted global scaling factor c is 0.702. Although identification of the global minimum in a high-dimensional space is not always guaranteed, identical best solutions were obtained over 100 times providing confidence that the optimal solution indeed corresponds to the global optimum.

The optimized binding contribution factors (BCF) $q_{n,|n-i|}$ (Eq. (11)) of a residue n to another residue i are depicted in Figure 2. These factors depend on the type of residue n as well as the separation $|n–i|$ along the primary sequence. The positively charged Arg and Lys residues have the strongest effect on their neighboring residues due their attractive interactions with the negatively charged SNP surface consisting of negatively charged (deprotonated) silanol groups - Si-O$^-$. The effect decreases with increasing residue separation (i.e. BCFs approach 1) but it remains noticeable even >25 residues away. The negatively charged Asp and Glu residue show the opposite effect by diminishing the interactions of neighboring residues with the SNP surface. Neutral residues promote attraction, although on a significantly smaller scale than Arg and Lys.

These attractive interactions are likely to be of a hydrophobic nature, similar to those observed in molecular dynamics (MD) simulations of biomolecule-SNP interactions, caused by the hydrophobic nature of surface siloxane groups -Si-O-Si-.[28]

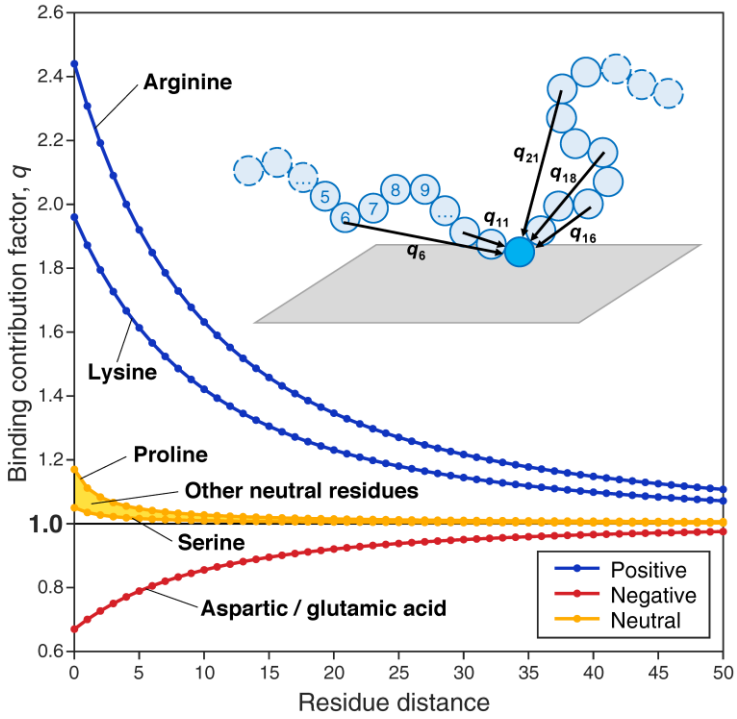


Figure 2. Optimized parameters of SILC model. The contribution of each residue to nanoparticle binding of a residue of interest is modeled by a residue-specific hyperbolic multiplicative binding contribution factor q, which depends on the inter-residue distance. Positive (negative) q values have a longer range and indicate attractive (repulsive) electrostatic interactions, which dominate q values belonging to neutral residues (hydrophobic interactions).

The bound population of Eq. (8) corresponds to conditions where the protein concentration is highly diluted and the bound population is then independent of the precise protein concentration. When the protein concentration is increased and becomes comparable to the effective concentration of protein binding sites on the surfaces of the nanoparticles, as is the case in our experiments, the bound population will decrease with increasing protein concentration due to saturation. For example, under standard SNP concentration, the experimental binding population of aSyn is 6 times larger than p53TAD (using the mean of the strongest 20% residues of each protein to assess binding). However, when aSyn and p53TAD are simultaneously present, their binding ratio is not necessarily 6-fold as a result of the possible existence of multiple layers of bound protein or SNP aggregation. Relative binding strengths of

any two proteins can be experimentally determined using direct competitive binding experiments as a function of SNP and protein concentrations. Because SILC is relatively insensitive to the binding strength ratio, we assume here that the binding strength of each protein is proportional to its bound population under standard SNP concentration. For example, our model (Eqs. (9), (11)) can be fitted almost equally well when the binding strength ratio of aSyn to p53TAD is changed from 6 to 20.

The parametrized form of the SILC model has been implemented as a webserver, which accepts the IDP amino-acid sequence as input and returns the residue-specific bound population profile for anionic silica nanoparticles at http://spin.ccic.osu.edu/index.php/frimsi/.

Application of SILC model to 4 IDPs

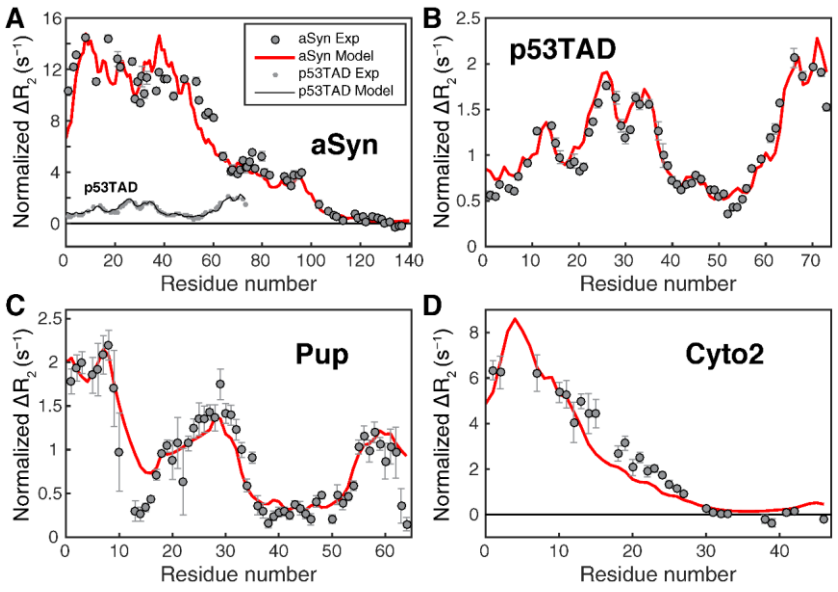


Figure 3. Experimental interaction profiles with SNPs (solid circles) of four representative IDPs (170 μM) normalized to a standard SNP concentration (0.17 μM) in comparison with SILC model prediction (thick red lines): (A) aSyn, (B) p53TAD, (C) Pup, (D) Cyto2. In Panel (A), the p53TAD profile is to scale plotted as aSyn, which shows an approximately 6-fold difference in the absolute binding strength between aSyn and p53TAD at same SNP concentration.

Application of the SILC model to the wild-type forms of the 4 IDPs is shown in Figure 3. The four IDPs display distinctly different affinity profiles reflecting their different amino-acid compositions and lengths. aSyn and Cyto2 display a similar negative gradient of their affinities from the N- to the C-terminus. However, in terms of absolute affinities, aSyn interacts with SNPs

about twice as strongly as Cyto2. By contrast, p53TAD and Pup have a markedly more varied profile with regions with elevated affinity followed by regions with lower affinity and *vice versa*. The absolute binding affinities of these two proteins are substantially lower than those of aSyn and Cyto2. The prediction by the SILC model (red line) reproduces the experimental profiles

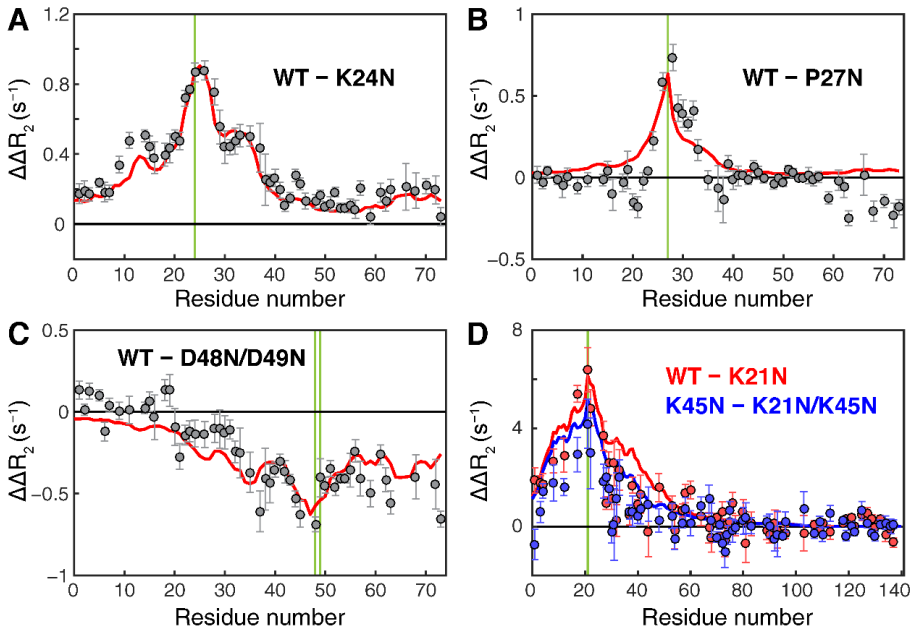


Figure 4. Experimental $\Delta\Delta R_2$ values (solid circles) vs. SILC model predictions (thick lines) for (A) p53TAD[WT – K24N], (B) p53TAD[WT – P27N], (C) p53TAD[WT – D48N/D49N], (D) aSyn[WT – K21N] and aSyn[K45N – K21N/K45N], highlighting the effects of mutations. The slightly, yet systematically elevated $\Delta\Delta R_2$ values of aSyn[WT – K21N] over aSyn[K45N – K21N/K45N] provide direct experimental evidence of the positive cooperativity of K21 and K45 residues, since the absence of K45 allosterically reduces the overall contribution of K21 to the binding of other residues (see Supporting Information). The mutation sites are indicated by vertical green lines.

remarkably well both in terms of the local variations and the absolute magnitude of binding. This demonstrates that the SILC model is capable to realistically reproduce the global binding properties of this diverse set of IDPs.

Cooperativity effects are not only expected to affect the global binding properties, but also impact local behavior. For this purpose, we compared the behavior of mutants of aSyn and p53TAD with their wild-type forms. Specifically, for p53TAD the single-site mutants K24N and P27N and the double mutant D48N/D49N were produced and subjected to residue-specific ^{15}N ΔR_2 analysis with the results shown in Figure 4A,B,C. The depicted $\Delta\Delta R_2$ profiles correspond to the ΔR_2 difference profiles p53TAD[WT – K24N], p53TAD[WT – P27N], and p53TAD[WT –

D48N/D49N]. As expected, all three profiles show the strongest $\Delta\Delta R_2$ effect at the position of the mutations with large effects also visible for neighboring residues, which gradually diminish for larger separation along the amino-acid sequence. The p53TAD[WT – D48N/D49N] has a dominantly negative profile reflecting that the binding affinity is increased when the negatively charged aspartate residues D48 and D49 are replaced by neutral asparagine. The effects of K24N and D48N/D49N have a significantly longer range than the effect of P27N, which is accurately captured by the model of Eq. (10) where charged and neutral residues have different a_n, b_n parameters (Table 1). The same characteristic long-range effect exerted by charged residues on neighboring residues is observed for aSyn for the single mutant aSyn[WT – K21N] and the double mutant aSyn[K45N – K21N/K45N] (Figure 4D).

The experimental ΔR_2 profile of p53TAD[WT – K24N] has in addition to the main maximum at residue 24 clearly visible additional maxima on either side, namely around residues E11, L14 and S33, L35, which could not be explained by FRIM. By contrast, the SILC binding model is capable of reproducing such finer features of the affinity profile (Figure 4A). They are a direct reflection of local cooperativity between K24 and the neighboring residue clusters. Since both neighboring clusters have an elevated binding propensity on their own (Figure 3B), the enthalpic binding forces are able to more successfully counteract the conformational entropy effects for these clusters as compared to other parts of the polypeptide chain. A further increase in the bound population of the central cluster by substituting the neutral N24 by the more favorably interacting K24 will then translate into an increase of the bound state of all neighboring residues. Due to the product nature of SILC, it disproportionately affects neighboring residue clusters with an intrinsically higher bound population resulting in the appearance of these additional maxima in the $\Delta\Delta R_2$ profile observed both in the experiment and reproduced by SILC.

For aSyn, the experimental $\Delta\Delta R_2$ profile of aSyn[WT – K21N] has a clear tendency to exceed the one of aSyn[K45N – K21N/K45N] (red vs. blue circles in Figure 4D). Again, this is a reflection of entropic binding cooperativity akin to the one of p53TAD[WT – K24N]. Nanoparticle binding of aSyn K21N benefits from the presence of K45, which binds strongly to the SNP surface and, through reduction of the conformational entropy of neighboring residues, facilitates N21 binding, resulting in a relatively large $\Delta\Delta R_2$ effect. By contrast, aSyn

K21N/K45N benefits to a lesser extent from an entropy reduction, since N45 has a substantially weaker binding affinity to SNPs than K45. As a consequence, in aSyn K21N/K45N, SNP binding of N21 has to overcome the larger entropy cost, which results in a weaker $\Delta\Delta R_2$ effect. The observed trend is clearly observable, although relatively noisy, because each $\Delta\Delta R_2$ profile requires the collection of 4 ^{15}N-^{1}H HSQC-based R_2 experiments, each based on a R_1 and $R_{1\rho}$ experiment, applied to WT or mutant 1 vs. mutant 2 in the presence and absence of SNPs. Hence, the observed trend between aSyn[WT – K21N] and aSyn[K45N – K21N/K45N] is based on the cumulative differential effects of no fewer than 16 R_1 or $R_{1\rho}$ relaxation experiments, each with a finite signal-to-noise ratio.

DISCUSSION

The interaction of an IDP with a nanoparticle is conceptually related to the binding of multiple ligands to a receptor where the ligands are sequentially connected to each other by flexible linkers. Such systems can achieve distinct binding behavior through cooperativity. Previously, analytical mathematical models for divalent and multivalent binding have been proposed where each fragment can only adopt either a bound or an unbound state. In this case, the total apparent association constant can be expressed as product of the individual association constants and effective protein concentration, which depend on the linker properties, such as their lengths and internal flexibilities.[23] Subsequent work used more realistic linker models to improve agreement with experiment.[24-25] These divalent/multivalent binding models work best in the case of strong positive cooperativity, where only free states (i.e. none of the fragments is bound) and fully bound states have significant populations thereby greatly reducing the complexity of the model.

The divalent/multivalent binding model is inadequate for the treatment of IDP-SNP interactions, because the IDP residues simultaneously serve as ligands and linkers. This requires the inclusion of partially bound states, which are absent in divalent/multivalent binding models where a ligand can only be either bound or unbound. By contrast, our model incorporates a continuous spatial distribution of each IDP residue instead of just two discrete states “bound” and “unbound”. Positively charged residues tend to have a higher population near the SNP surface, even when they are still in an unbound state, i.e., while retaining their translational and

rotational motional degrees of freedom. As a consequence, positively charged residues in a near-bound state reduce the entropic cost of binding of neighboring residues and thereby facilitate their binding to the SNP surface in a cooperative manner. The "continuum" representation of the unbound and bound states in our model enables the realistic representation of the difference between charged and hydrophobic residues, where charged residues have a significantly longer range effect on other residues along the primary sequence than neutral and hydrophobic residues. Such behavior cannot be captured by the above-mentioned two-state models. The cooperativity or anti-cooperativity effect of each residue for each other residue is determined by the binding contribution factor q, which has a smooth dependence on residue separation for all residues independent of their charge. The model includes the cumulative effect of such two-body interactions and shows accurate performance for the IDPs studied here with their different interaction propensities. The model could be accurately validated based on high-resolution solution NMR data. This requires that the bound IDP populations are relatively small; otherwise strong line broadening could cause the disappearance of signals. For IDPs with binding affinities that are similar to the four proteins studied, our model is expected to predict accurate affinity profiles. For systems with significantly stronger IDP-nanoparticle interactions even relatively small inaccuracies in model parameters can have a sizeable effect on affinity predictions as a direct consequence of binding cooperativity, which exponentiates the effect of variations in model parameters. However, there are currently no experimental methods available to quantitatively probe strong IDP-nanoparticle interactions at residue-level resolution. Our model can be adopted to describe IDP binding with nanoparticles and nanomaterials other than anionic SNPs.

Our previously developed FRIM for IDP-SNP binding focused on weak cooperativity by including the effects of neighboring residues as additive contributions.[18] Although FRIM successfully reproduces the experimentally observed binding affinity trends along the polypeptide chain, it does not accurately predict the absolute binding affinities of the different IDPs. FRIM also does not reproduce some of the finer details of the affinity profiles, which are revealed by site-directed mutagenesis derived $\Delta\Delta R_2$ profiles (Figure 4). SILC, by contrast, produces accurate absolute affinity profiles and at the same time explains unique features found in $\Delta\Delta R_2$ profiles. At the level of binding free energies, the SILC model represents cooperativity among all residues by a sum of all pairwise cooperativities, thereby rendering the problem

mathematically tractable in analytical form. The good agreement with experiment provides strong evidence that the model is valid for moderately strong binding proteins, including aSyn. However, for systems that display very strong positive cooperativity, such as IDPs with many more positively charged residues, SILC is not valid any longer and the divalent/multivalent binding models[21-25] may be better suited.

Both the SILC model and the NMR spin relaxation data primarily reflect the binding properties of individual residues. Sometimes, however, it can be informative to view binding from a "protein-centric" perspective, where a protein molecule is defined to be in the bound state if *at least* one residue is bound to the SNP. The bound protein population P_{protein} can then be empirically expressed as

$$P_{protein} = 1 - \prod_{i=1}^{N} (1 - p_i)^{\lambda} \qquad (13)$$

where p_i is the bound population of residue i (i = 1,…,N) predicted by the SILC model (Eq. (9)), $0 < \lambda \leq 1$ is an empirical scaling factor to account for the cooperative binding properties of amino acid residues that are close in sequence, and $\lambda = 1$ for independent binding. A higher degree of cooperativity corresponds to generally larger p_i and smaller λ values. The p53TAD binding experiment at 170 μM protein and 2.52 μM SNP concentration, the protein centric binding population $P_{\text{protein}} \cong 0.14$ for $\lambda = 1$ and $P_{\text{protein}} \cong 0.08$ for $\lambda = 0.5$.

One of the hallmarks of IDPs in biological systems is their ability to form fuzzy complexes with receptors.[29-30] The IDP-SNP systems should serve as paradigms to study such effects with residue-level resolution by solution NMR spectroscopy. Our model incorporates the heterogeneous nature of IDP conformers interacting with the nanoparticle surface. For each IDP, it is a large number of different IDP binding modes that cumulatively produce the binding profile. This is analogous to the co-existence of multiple binding modes between an IDP that interacts with the surface of a folded protein receptor.[31] The bound conformers that are dominantly populated will sensitively depend on the number and strengths of specific and non-specific interactions between the IDP and the folded protein, which can encompass broad distributions of the kind observed here or, alternatively, much narrower distributions in cases where a unique network of complementary interactions between the IDP and the receptor is preferentially formed.

CONCLUSION

We have introduced SILC as a statistical mechanical model to quantitatively explain cooperative binding between four different IDPs and anionic silica nanoparticles at the level of individual residues. SILC represents the partition function of the bound state with contributions from all residue pairs. Most numerical values for the model parameters directly stem from ΔR_2-relaxation experiments of free amino acids interacting with nanoparticles and $\Delta\Delta R_2$ differences of IDP-binding profiles due to site-specific mutations. It is found that cooperativity plays a critical role for the understanding of the IDP-SNP interaction data both in terms of the fine structure of the binding profiles and prediction of the differences between the absolute binding affinities of the different IDPs, which includes the correct prediction of competitive binding between aSyn and p53TAD.

Certain nanoparticles are known to transfect cells and may cause toxicity. The detailed and quantitative understanding of nanoparticle interactions with an array of intracellular components, including IDPs, is likely to advance our understanding of nanotoxicity. Conversely, it may be possible to utilize existing and engineer new nanoparticles with specific surface modifications that embody optimized physical-chemical properties to target certain IDPs *in vitro* or *in vivo*.[32] Such targeting may help prevent or mitigate certain types of IDP-related diseases.[14, 16, 33] Such efforts will benefit from the SILC model to make accurate and rapid predictions of quantitative binding properties to identify specific IDP regions or entire IDPs.

SUPPORTING INFORMATION

The Supporting Information is available free of charge at …

- Additional information about sample preparation, NMR experiments, and data analysis
- Tables with amino-acid sequences of IDPs and SNP concentrations
- Figures of ^{15}N-^{1}H HSQC spectra of all 4 proteins, experimental reproducibility of R_2 profiles, comparison between R_2 and η_{xy}.

- Plots of BCFs as cooperativity free energy contributions.

We have introduced SILC as a statistical mechanical model to quantitatively explain cooperative binding Acknowledgements

This work was supported by the U.S. National Science Foundation (grant MCB-1715505 to R.B.). All NMR experiments were performed at the CCIC NMR facility at The Ohio State University.

Conflict of interest

The authors declare no conflict of interest.

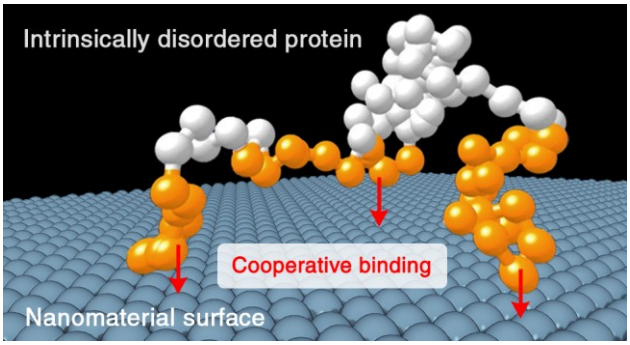
Keywords: Intrinsically disordered proteins • nanoparticles • cooperativity • statistical mechanical cooperative binding model • NMR spectroscopy • silica nanoparticles

REFERENCES

1. Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T., Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **2004,** *337*, 635.
2. Oldfield, C. J.; Dunker, A. K., Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* **2014,** *83*, 553.
3. Wright, P. E.; Dyson, H. J., Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **2015,** *16*, 18.
4. Boskey, A. L.; Villarreal-Ramirez, E., Intrinsically disordered proteins and biomineralization. *Matrix Biol.* **2016,** *52-54*, 43.
5. Uversky, V. N., Dancing Protein Clouds: The Strange Biology and Chaotic Physics of Intrinsically Disordered Proteins. *J. Biol. Chem.* **2016,** *291*, 6681.
6. Shan, B.; Li, D. W.; Bruschweiler-Li, L.; Brüschweiler, R., Competitive binding between dynamic p53 transactivation subdomains to human MDM2 protein: implications for regulating the p53.MDM2/MDMX interaction. *J. Biol. Chem.* **2012,** *287*, 30376.
7. Wang, T.; Darwin, K. H.; Li, H., Binding-induced folding of prokaryotic ubiquitin-like protein on the Mycobacterium proteasomal ATPase targets substrates for degradation. *Nat. Struct. Mol. Biol.* **2010,** *17*, 1352.
8. Zerfass, C.; Buchko, G. W.; Shaw, W. J.; Hobe, S.; Paulsen, H., Secondary structure and dynamics study of the intrinsically disordered silica-mineralizing peptide P5 S3 during silicic acid condensation and silica decondensation. *Proteins* **2017,** *85*, 2111.
9. Olsen, J. G.; Teilum, K.; Kragelund, B. B., Behaviour of intrinsically disordered proteins in protein-protein complexes with an emphasis on fuzziness. *Cell. Mol. Life Sci.* **2017,** *74*, 3175.
10. Shen, Q.; Shi, J.; Zeng, D.; Zhao, B.; Li, P.; Hwang, W.; Cho, J. H., Molecular Mechanisms of Tight Binding through Fuzzy Interactions. *Biophys. J.* **2018,** *114*, 1313.
11. Zhang, L.; Li, M.; Liu, Z., A comprehensive ensemble model for comparing the allosteric effect of ordered and disordered proteins. *PLoS Comput. Biol.* **2018,** *14*, e1006393.
12. Gomes, E.; Shorter, J., The molecular language of membraneless organelles. *J. Biol. Chem.* **2019,** *294*, 7115.
13. Chong, P. A.; Forman-Kay, J. D., Liquid-liquid phase separation in cellular signaling systems. *Curr. Opin. Struc. Biol.* **2016,** *41*, 180.
14. Yoo, S. I.; Yang, M.; Brender, J. R.; Subramanian, V.; Sun, K.; Joo, N. E.; Jeong, S. H.; Ramamoorthy, A.; Kotov, N. A., Inhibition of amyloid peptide fibrillation by inorganic nanoparticles: functional similarities with proteins. *Angew. Chem. Int. Ed. Engl.* **2011,** *50*, 5110.
15. Roshanfekrnahzomi, Z.; Badpa, P.; Esfandiari, B.; Taheri, S.; Nouri, M.; Akhtari, K.; Shahpasand, K.; Falahati, M., Silica nanoparticles induce conformational changes of tau protein and oxidative stress and apoptosis in neuroblastoma cell line. *Int. J. Biol. Macromol.* **2019,** *124*, 1312.
16. Cabaleiro-Lago, C.; Quinlan-Pluck, F.; Lynch, I.; Lindman, S.; Minogue, A. M.; Thulin, E.; Walsh, D. M.; Dawson, K. A.; Linse, S., Inhibition of amyloid β protein fibrillation by polymeric nanoparticles. *J. Am. Chem. Soc.* **2008,** *130*, 15437.
17. Sharma, V.; Srinivasan, A.; Roychoudhury, A.; Rani, K.; Tyagi, M.; Dev, K.; Nikolajeff, F.; Kumar, S., Characterization of protein extracts from different types of human teeth and insight in biomineralization. *Sci .Rep.* **2019,** *9*, 9314.
18. Xie, M.; Hansen, A. L.; Yuan, J.; Brüschweiler, R., Residue-Specific Interactions of an Intrinsically Disordered Protein with Silica Nanoparticles and their Quantitative Prediction. *J. Phys. Chem. C* **2016,** *120*, 24463.
19. Xie, M.; Li, D. W.; Yuan, J.; Hansen, A. L.; Brüschweiler, R., Quantitative Binding Behavior of Intrinsically Disordered Proteins to Nanoparticle Surfaces at Individual Residue Level. *Chemistry* **2018,** *24*, 16997.

20. Rimola, A.; Costa, D.; Sodupe, M.; Lambert, J. F.; Ugliengo, P., Silica surface features and their role in the adsorption of biomolecules: computational modeling and experiments. *Chem. Rev.* **2013,** *113*, 4216.
21. Jencks, W. P., On the Attribution and Additivity of Binding-Energies. *Proc. Natl. Acad. Sci. U.S.A.* **1981,** *78*, 4046.
22. Kitov, P. I.; Bundle, D. R., On the nature of the multivalency effect: A thermodynamic model. *J. Am. Chem. Soc.* **2003,** *125*, 16271.
23. Crothers, D. M.; Metzger, H., Influence of Polyvalency on Binding Properties of Antibodies. *Immunochemistry* **1972,** *9*, 341.
24. Zhou, H. X., Quantitative account of the enhanced affinity of two linked scFvs specific for different epitopes on the same antigen. *J. Mol. Biol.* **2003,** *329*, 1.
25. Bobrovnik, S. A., 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.* **2007,** *20*, 253.
26. Zhang, B.; Xie, M.; Bruschweiler-Li, L.; Bingol, K.; Brüschweiler, R., Use of Charged Nanoparticles in NMR-Based Metabolomics for Spectral Simplification and Improved Metabolite Identification. *Anal .Chem.* **2015,** *87*, 7211.
27. Agarwal, S.; Mierle, K. *ceres-solver*.
28. Shi, B.; Shin, Y. K.; Hassanali, A. A.; Singer, S. J., DNA Binding to the Silica Surface. *J. Phys. Chem. B* **2015,** *119*, 11030.
29. Fuxreiter, M.; Tompa, P., Fuzzy complexes: a more stochastic view of protein function. *Adv. Exp. Med. Biol.* **2012,** *725*, 1.
30. Arbesu, M.; Iruela, G.; Fuentes, H.; Teixeira, J. M. C.; Pons, M., Intramolecular Fuzzy Interactions Involving Intrinsically Disordered Domains. *Front. Mol. Biosci.* **2018,** *5*, 39.
31. Arbesu, M.; Maffei, M.; Cordeiro, T. N.; Teixeira, J. M.; Perez, Y.; Bernado, P.; Roche, S.; Pons, M., The Unique Domain Forms a Fuzzy Intramolecular Complex in Src Family Kinases. *Structure* **2017,** *25*, 630.
32. Roach, P.; Farrar, D.; Perry, C. C., Surface tailoring for controlled protein adsorption: effect of topography at the nanometer scale and chemistry. *J Am. Chem. Soc.* **2006,** *128*, 3939.
33. Vio, V.; Jose Marchant, M.; Araya, E.; J Kogan, M., Metal nanoparticles for the treatment and diagnosis of neurodegenerative brain diseases. *Curr. Pharm. Des.* **2017,** *23*, 1916.

TOC figure


Intrinsically disordered protein
Cooperative binding
Nanomaterial surface