ARTICLE TYPE

Competitive binding of HIF-1α and CITED2 to the TAZ1 domain of CBP from molecular simulations

Irene Ruiz-Ortiz[a] & David De Sancho,[a,b]

Many intrinsically disordered proteins (IDPs) are involved in complex signalling networks inside the cell. Their particular binding modes elicit different types of responses that can be subtly regulated. Here we study the binding of two disordered transactivation domains from proteins HIF-1α and CITED2, whose binding to the TAZ1 domain of CBP is critical for the hypoxic response. Experiments have shown that both IDPs compete for their shared partner, and that this competition is mediated by the formation of a ternary intermediate state. Here we use computer simulations with a coarse-grained model to provide a detailed molecular description of this intermediate. We find that the conserved LP(Q/E)L motif may have a critical role in the displacement of HIF-1α by CITED2 and show a possible mechanism for the transition from the intermediate to the bound state. We also explore the role of TAZ1 dynamics in the binding. The results of our simulations are consistent with many of the experimental observations and provide a detailed view of the emergent properties in the complex binding of these IDPs.

1 Introduction

In the last two decades, the study of intrinsically disordered proteins (IDPs) has changed the established paradigms of biomolecular transitions, both for folding and binding[1]. From extensive theoretical and experimental work, we have learnt that many proteins do not fold into regular, three dimensional, native structures. Instead, due to their unique sequence characteristics, IDPs may remain disordered, or fold only upon binding to their partners[2]. For this type of systems, there is an expanding repertoire of binding modes[3] that may be important to enable their key regulatory roles[4]. Much of the work on IDP binding described the mechanism in terms of either conformational selection or induced fit[5,6], and, more recently, the "dock and coalesce" model[7]. Also, from a kinetic standpoint, binding in an expanded, disordered state was proposed to be advantageous, as explained by the "fly-casting" model[8]. But more recently, we have learnt that IDPs can be more wildly dissimilar from folded proteins, and undergo multivalent binding through phosphorylated sites[9] or form complexes in their unfolded state with ultra-high affinity[10].

As important as characterizing the biophysics of coupled folding and binding processes, is understanding how these new modes of binding operate in the biological context[11]. The regulation of cell signalling in the case of IDPs is usually facilitated by promiscuous binding to different targets through alternative regions[12,13]. In other cases, multiple IDPs may bind to the same partner[14]. The peculiarities in the binding of IDPs may result in interesting effects, like allosteric modulation of one partner by another[13,15]. Here we focus in an interesting case involving two different IDPs, the C-TAD domain of the hypoxia inducible transcription factor HIF-1α and its negative feedback regulator, the CBP/p300-interacting transactivator 2 (CITED2), which bind to a shared partner protein, the TAZ1 domain of the CREB binding protein (CBP) or its homolog p300[16–25].

The HIF-1α-TAZ1-CITED2 system is involved in the regulation of cell survival in conditions of low oxygen levels (i.e. hypoxia), and for this reason it is highly relevant in the development of cancer therapies[26]. The intrinsically disordered domain of HIF-1α and CITED2 both bind to TAZ1 with very high affinity (K_D =10±2 nM) through largely overlapping regions of their shared partner[16,17,19,22] (see Figure 1a and Fig. S1 in Supplementary Information). A particularly important role in complex formation is played by the conserved LPXL motif (see Figure 1c). Surprisingly, in recent experiments on ternary mixtures involving both IDPs with TAZ1, increasing CITED2 concentrations were able to displace HIF-1α bound to TAZ1, while the reverse process was not observed[23]. The displacement was coupled to a 50-fold decrease of the apparent K_D for CITED2 (0.2±0.1 nM) and a 100-fold increase for HIF-1α (0.9±0.1 μM). To explain this counterintuitive result, Berlow et al invoked a transient intermediate state[23], although a detailed molecular description of this state is still lacking. The fine tuning of the competition between the IDPs for their shared partner may be critical for our understanding of the hy-

[a] *Donostia International Physics Center, 20018, Donostia-San Sebastián, Spain*
[b] *University of the Basque Country, Faculty of Chemistry, Paseo Manuel Lardizabal, 3, 20018 Donostia-San Sebastián, Spain. E-mail: david.desancho@ehu.eus*

† Electronic Supplementary Information (ESI) available: additional analysis of the simulations and four supplementary figures. See DOI: 10.1039/cXCP00000x/

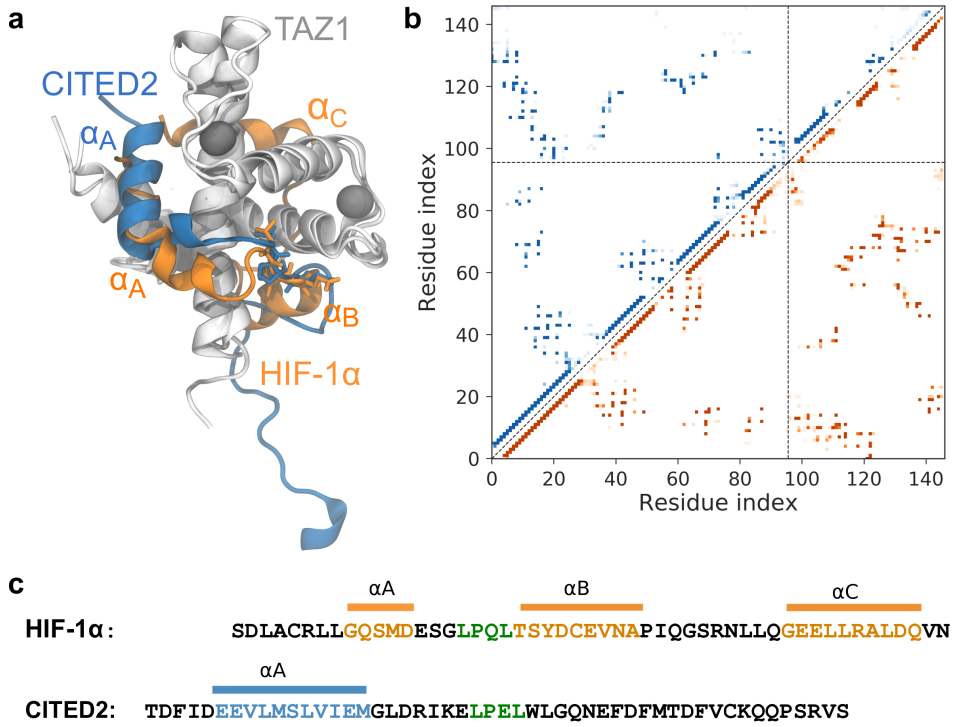

Fig. 1 (a) Cartoon representation of the complexes formed by HIF-1α (PDB id: 1l8c, orange) and CITED2 (1r8u, blue) with the TAZ1 domain of CBP (white). (b) Corresponding contact maps from the 20 experimental NMR models. The upper and lower triangles correspond to the CITED2:TAZ1 and HIF-1α:TAZ1 complexes, respectively. The color scale indicates the frequency of pairwise contacts within the NMR ensembles. The dashed vertical and horizontal lines mark the ends of TAZ1. (c) Amino acid sequences of HIF-1α and CITED2. We highlight the conserved LP(Q/E)L motif in green. Helical regions are marked with horizontal bars.

poxic response, as CITED2 acts as a negative feedback loop to control the response induced by HIF-1α[23].

Here we use molecular simulations to investigate the binding of HIF-1α and CITED2 to the TAZ1 domain of CBP. In the last years there have been remarkable improvements in force fields for explicit solvent, atomistic molecular dynamics (MD), which have been recalibrated to reproduce the properties of IDPs[27–30]. However, simulating full protein-protein binding with all atom MD simulations is extremely demanding even for binary systems[31,32]. Instead, we resort to simple structure-based, coarse grained models[33]. These models are based on the assumption that interactions in the 3D structure (i.e. native contacts) are the primary determinants of the folding/binding process. As a result they are consistent with the prescription for a funneled energy landscape in the theory of protein folding[34]. Coarse grained, structure-based models have consistently made successful, semi-quantitative predictions on protein folding, including the calculation of protein folding rates[35] and kinetic Φ-values[36]. Also for complex formation they have been successful in the description of the binding processes both for folded proteins[37] and for IDPs[38–40]. In fact one of us already applied this type of modelling to the binding of HIF-1α to TAZ1[40], including how additional contributions in the model, like electrostatics and non-native interactions, may change the results.

In this paper we expand that work to include the CITED2:TAZ1 complex and the emerging properties for the ternary mixture, using a simulation model that uses information from the topology of the binary complexes. Our main focus is on resolving a molecular picture of the ternary intermediate proposed by Berlow et al[23]. The paper is organized as follows: First, we describe the methodology, which involves producing a model that is consistent with the experimental structures of the TAZ1-bound states of both HIF-1α and CITED2. Next, we report the predictions from the consensus simulation model on the binding of each of the IDPs independently to their shared partner. We pay special attention to the influence of the binding on the dynamics of TAZ1, which has been recently investigated using NMR[24,25]. Finally, we show the emergent properties that we predict for the ternary complex and discuss the implications of our findings.

2 Materials and Methods

2.1 Coarse-grained simulation model

We start from the structure-based Karanicolas and Brooks model[41], which is coarse grained to the level of the C$^\alpha$ atoms. The potential energy function is defined as a sum of terms

$$V = V_{\text{bonds}} + V_{\text{angles}} + V_{\text{torsions}} + V_{\text{nonbonded}} \quad (1)$$

The first two energy terms in this equation are native-centric harmonic potentials for bonds and angles between beads in the model, with the equilibrium values being those in the reference (experimental) structure. The term for torsions is derived from PDB statistics. Finally, favourable non-bonded interactions are defined among amino acid pairs that are in contact in the native

conformation[41], and are described by the following expression,

$$V_{\text{nonbonded}} = \sum_{ij \in \text{contacts}} \varepsilon_{ij} \left[13 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 18 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{10} + \right. \\ \left. + 4 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \quad (2)$$

where ε_{ij} is the strength of the interaction between residues i and j, proportional to the Mijazawa and Jernigan contact energies[42], r_{ij} is the pairwise distance between beads in a instantaneous configuration and σ_{ij} is the distance in the reference conformation. For every pair of residues not forming contacts, the nonbonded potential is a repulsive term described as

$$V_{\text{nonbonded}} = \sum_{ij \notin \text{contacts}} \varepsilon_{\text{rep}} \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} \quad (3)$$

where ε_{rep} is a generic repulsive energy term related to the energy scale in the model and σ_{ij} is the repulsive radius[41].

2.2 Calibration of the model

Using this description, we have built our computational models based on the most representative conformers of the HIF-1α:TAZ1 and CITED2:TAZ1 experimental structures (with PDB ids 1l8c[16] and 1r8u[22], respectively, see Figure 1a). We note that this is a design decision, as coarse-grained simulations using as reference different NMR models in an ensemble may result in different behaviour[43,44]. For simplicity, we use as our reference the structure which is most central to the 20 different models in the ensembles reported for each of the complexes in the PDB.

The conformation of TAZ1 is slightly different in both experimental structures, with a C^{α}-*RMSD* of 2.7 Å. For this reason, we derived a consensus model for TAZ1, although we only calibrate the non-bonded interactions. Bonds and angles are very similar between both TAZ1 structures and in this work we use the values of the 1l8c structure for simplicity. The dihedral terms in the Karanicolas and Brooks model are statistical[41], and hence they do not change with the reference experimental structure. For the nonbonded interactions we produce a unified contact model including contacts from both experimental structures[45]. We follow a simple rule of thumb to produce the consensus model: whenever a contact is present in both reference structures for TAZ1, we include the contact with ε_{ij} and σ_{ij} values that are the mean for both experimental structures ; when a contact is only present in one structure, we keep its value of σ_{ij} but scale the interaction energy by $1/2$; finally, when a pair of residues is not in contact in any of the structures we choose the shortest repulsive core for the consensus model. We show representative examples of these three possibilities in Figure 2.

In order to approximately recapitulate the experimental dissociation constants in the nM range, we scaled the values of ε_{ij} in the HIF-1α:TAZ1 and CITED2:TAZ1 complexes. This involved decreasing the interaction strength of native contacts by 2% for CITED2:TAZ1 and increasing it by 4% for HIF-1α:TAZ1. In the simulations of the ternary complex, nonbonded interactions between HIF-1α and CITED2 are defined simply using the repulsive term from Equation 3.

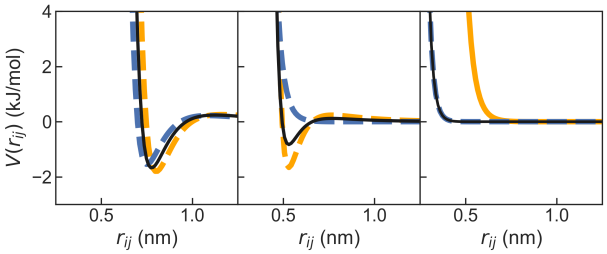

Fig. 2 Intra-TAZ1 nonbonded potentials for the HIF-1α:TAZ1 (orange), CITED2:TAZ1 (blue) and consensus (black) models. We show three representative cases, corresponding to a contact shared between both experimental structures (left), a contact present in one and absent in the other structure (center), and the repulsive cores being defined at different distances (right).

Finally, in the experimental structures of the TAZ1 complexes with HIF-1α and CITED2, there are three Zn^{2+} cations (see Figure 1a), each of them coordinated to a His and three Cys residues. These metal cations are essential for TAZ1 to retain a stable fold[16,21]. In our coarse grained representation, we mimic their effect by introducing soft bonds with equilibrium distances calculated as the mean from all the NMR models from both structures and a spring constant of 1000 kJ/mol/nm.

2.3 Molecular simulations

Using the model described above, we have run Langevin dynamics simulations of free TAZ1, the binary mixtures of TAZ1 and both IDPs (HIF-1α:TAZ1 and CITED2:TAZ1), and the ternary mixture (HIF-1α:CITED2:TAZ1) using a the Gromacs package (version 4.0.5[46]). To calculate the potentials of mean force for the binary complex, we have run umbrella sampling simulations using harmonic potentials on values of the intermolecular fraction of native contacts, Q, ranging between 0 and 1. All umbrella simulations were run at 300 K. For the ternary complex, we run temperature replica exchange for 22 different temperatures, ranging between 290 and 395 K, evenly spaced by 5 K. Replica swaps were attempted every 5000 steps. In all cases, a time step of 10 fs was used to propagate the dynamics with a leap-frog stochastic dynamics integrator, using a friction coefficient of 0.2 ps^{-1}.

2.4 Analysis of the simulations

We monitor folding and binding using the fraction of native contacts, Q, as the average degree of contact formation

$$Q = 1/N_{ij} \sum_{i,j \in \text{contacts}} \left(1 + \exp[\beta(d_{ij} - \gamma d_{ij,0})]\right)^{-1}, \quad (4)$$

In this equation, the sum runs over the N_{ij} pairs of residues (i,j) forming native contacts, d_{ij} and $d_{ij,0}$ are the distances between a pair of beads in the instantaneous and reference configurations, respectively, and β and γ are adjustable parameters that take the values of 50 nm^{-1} and 1.2, respectively[47]. Because the unbound state collapses onto a single bin when histogramming on Q, we

also calculate the d_{RMS}, defined as[48]

$$d_{\mathrm{RMS}} = \left(1/N_{ij} \sum_{i,j \in \mathrm{contacts}} (d_{ij} - d_{ij,0})^2\right)^{1/2}, \quad (5)$$

Results for different runs were combined to estimate potentials of mean force using the weighted histogram analysis method (WHAM)[49]. Errors in the free energy surfaces were determined using block averaging. We estimate the value of the dissociation constant K_{D} from the populations of the bound and unbound states (p_{b} and p_{u}) and the protein concentration in our simulation box as

$$K_{\mathrm{D}} = \frac{p_{\mathrm{u}}^2 [\mathrm{Protein}]}{p_{\mathrm{b}}} \quad (6)$$

as before[40]. In this expression, the bound population p_b is calculated integrating the Boltzmann weights along Q as $p_{\mathrm{b}} = \int_{Q_{\mathrm{TS}}}^{1} \exp(-\beta F(Q)) dQ$, where $\beta = 1/k_{\mathrm{B}}T$ is the inverse thermal energy $F(Q)$ is the free energy and Q_{TS} is the value corresponding to the barrier top in the potential of mean force. Then, the unbound population is simply $p_{\mathrm{u}} = 1 - p_{\mathrm{b}}$.

3 Results and discussion

3.1 Binding free energy landscapes for the binary complex models

Using the consensus models for both the HIF-1α:TAZ1 and the CITED2:TAZ1 complexes, calibrated to recover K_{D} values in the experimental (i.e. nM) range, we have run umbrella sampling simulations at room temperature. In Figure 3a we show the potentials of mean force for the projections on the umbrella coordinate, i.e. the fraction of intermolecular native contacts, Q. The potentials of mean force exhibit the characteristic sharp minimum at $Q = 0$, where all the unbound state collapses, and a broad basin for the bound state, as found generally for simulations of IDPs[38]. The sharp barrier at low values of Q is qualitatively consistent with the results from experimental Φ-value analysis[50], which indicate that native hydrophobic contacts are not present in the transition state.

The results in Figure 3 recapitulate those in previous work with coarse grained models for these proteins, despite the changes introduced in the simulation model. In particular, for HIF-1α:TAZ1 we find an intermediate state, which we term HI, emerging at values of $Q \simeq 0.3$, just like in previous simulation studies[40,51]. In this intermediate HIF-1α binds only through one of its helices to TAZ1[40], primarily the C-terminal α_{C}-helix, which forms more contacts with TAZ1 than α_{A} or α_{B} (see Figure 1b). The prediction of an intermediate state and a highly dynamic α_{A} helix is consistent with experimental NMR measurements on this system[23], in particular with the low $\{^1\mathrm{H}\} - ^{15}\mathrm{N}$ values reported for the region encompassing the α_{A} helix and the LPQL motif, which imply high flexibility[23]. The consistency with previous simulation results for HIF-1α:TAZ1 binding[40,51] suggests that the adjustments made to produce the consensus model for the intra-TAZ1 contacts do not compromise the predictions from the original Karanicolas and Brooks prescription[41]. We note that for the similar coarse-grained SMOG model[52,53], the binding of HIF-1α to TAZ1 does not involve any intermediate states[54]. In the case of the CITED2:TAZ1 complex, both the Karanicolas and Brooks model and the SMOG model result in two-state binding[54].

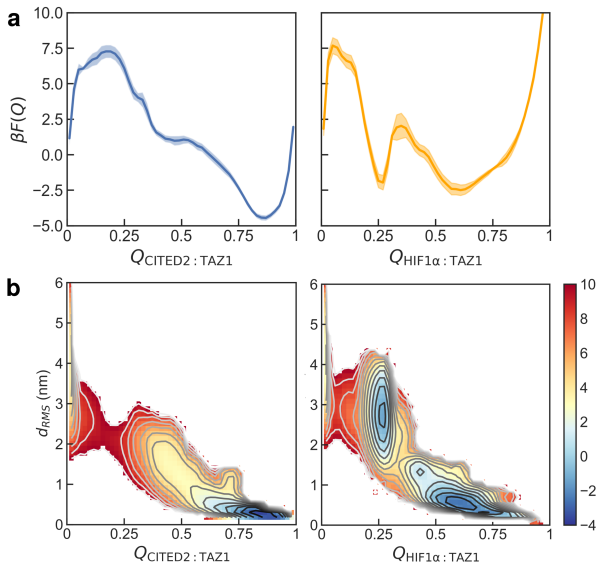

Fig. 3 Free energy landscapes for binding of HIF-1α (orange) and CITED2 (blue) to TAZ1. (a) Free energy landscapes for the projection on the fraction of native contacts (Q). Errors are shown as bands. (b) Two dimensional free energy landscapes for the projection on both Q and d_{RMS}. Free energies are shown in units of $k_{\mathrm{B}}T$.

As we have mentioned, in the projection on Q all the unfolded state collapses onto a single value ($Q = 0$). To resolve the heterogeneity of the unfolded state, we also show the free energy landscapes for the projection on the d_{RMS}, defined as the mean squared pairwise distance for native contacts (see Methods and Figure 3b). This projection highlights the differences in binding scenarios between HIF-1α and CITED2 to their shared partner TAZ1. While CITED2 undergoes cooperative binding overcoming a single free energy barrier, the binding of the α-helices of HIF-1α are decoupled. Differences in the binding between HIF-1α:TAZ1 and CITED2:TAZ1 are unsurprising given that the topologies of these complexes are also distinct (see contact maps in Figure 1b). While HIF-1α folds forming three helices upon binding to TAZ1, CITED2 only forms one helix and keeps a 30-residues long stretch of its sequence unfolded in the bound state.

3.2 Binding mechanism of the IDPs to TAZ1

The umbrella sampling simulations employed for the estimation of free energy surfaces are useful to determine relative stabilities of states and barriers between them, but they do not contain information about the conformational dynamics. In order to observe binding transitions for both IDPs, we have run a set of equilibrium simulations starting from conformations within the unbound state. In order to do this, we randomly picked distances between TAZ1 and each of the IDPs where both proteins are separated and let them progress. This results in multiple binding events, that let us characterize the binding mechanisms.

In Figure 4 we show time series data for the projection on the intermolecular Q from two representative binding trajectories. In

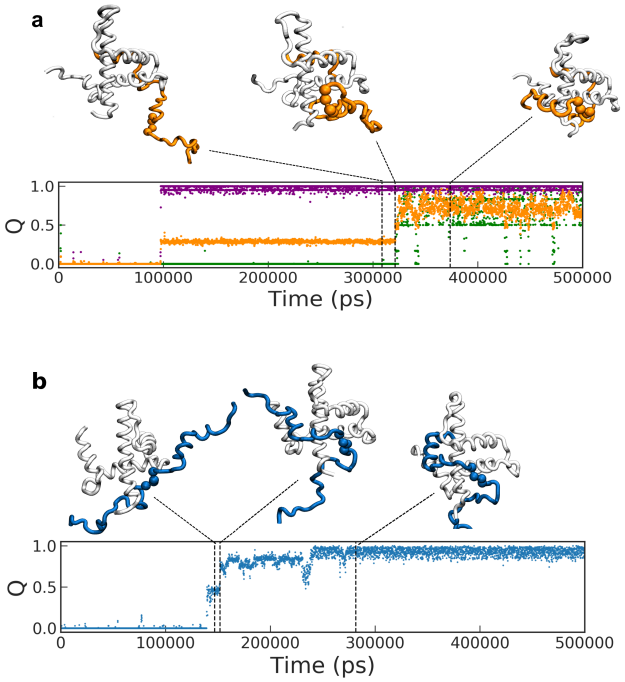

Fig. 4 Time series data for the fraction of intermolecular native contacts (Q) and corresponding snapshots for two representative binding trajectories for binary complexes HIF-1α:TAZ1 (a) and CITED2:TAZ1 (b). In (a) we show Q for all the contacts (orange), together with that for helices α_A (green) and α_C (violet). Vertical dashed lines indicate time stamps for selected conformations.

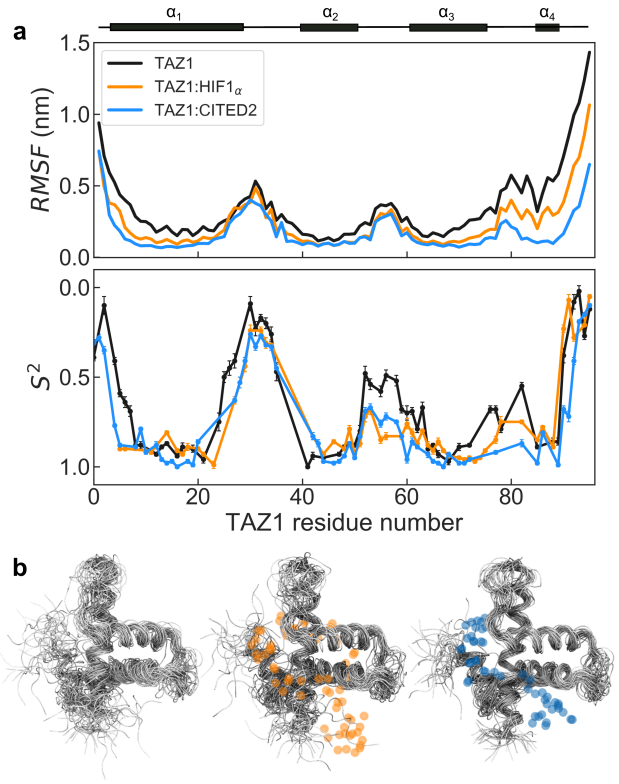

Fig. 5 (a) Comparison of the fluctuations in equilibrium simulations of TAZ1 (top) against the experimental S^2 parameters by Berlow et al[25] (bottom). In both panels we show values for free TAZ1 (black), and for its complexes with HIF-1α (orange) and CITED2 (blue). At the top, we show schematically the DSSP assignment of secondary structure, with bars corresponding to α-helices. (b) Overlay of representative snapshots from free TAZ1 (left) and TAZ1 bound to HIF-1α (center) and CITED2 (right). In the bound conformations, coloured circles correspond to C^α positions of residues that form native contacts with TAZ1.

the case of HIF-1α, the initial binding may occur through the interaction of TAZ1 with either helix α_A or α_C. However, we find the latter to be the dominant pathway (Figure 4a). In a subsequent step, the rest of the protein, including the LPQL motif, wraps around TAZ1. In the case of CITED2, we find that the initial approach typically involves the LPEL residues (Figure 4b), but this does not result in a distinct conformational state. Right after the initial encounter, the remainder of the contacts involving the α_A helix gradually form upon binding to TAZ1. The C-terminal tail of the protein, which does not involve intermolecular native contacts, remains unstructured in the bound state.

There are hence two predictions of the structure-based model on the binary binding processes. First, our simulations show that CITED2 binds to TAZ1 with much higher cooperativity than HIF-1α, as originally proposed from CD experiments on both complexes[19] and later confirmed by NMR[23]. Second, the conserved LPXL from CITED2 and HIF-1α, which interact with an overlapping region of TAZ1, have very different roles in the binding process. While CITED2 may form its first few contacts through its LPEL motif, the LPQL of HIF-1α binds only in a second step to TAZ1.

3.3 Native fluctuations in the consensus TAZ1 model

While much of the focus in the study of IDP binding has been on the disordered part of the system, recent studies have also stressed the importance of the interplay with the dynamics of the binding partner[55]. Proteins that act as binding partners of IDPs do not behave like mere scaffolds that disordered ligands bind to. Instead, they can adjust their flexibility in relation to their interactions with ligands. This conclusion has been drawn for TAZ1 from structural studies in the free state[21] and bound to a collection of its intrinsically disordered partners, including HIF-1α, CITED2, RelA-TA2 or TAD-STAT2[16,22,25,56–58].

We check these effects in a our simulations for free TAZ1 and in its complexes with both HIF-1α and CITED2. In Figure 5a we report the root mean square fluctuations (*RMSF*) in the different stable states of TAZ1 from molecular simulations with the consensus models. Our results can be compared, albeit at a qualitative level, with Berlow et al's NMR order parameters[25], which measure the amplitude of backbone motions (see Figure 5a, bottom). Unfortunately, given the resolution of our simulation model, we cannot calculate the values of the S^2 order parameters directly, as in the coarse grained description we lack information about the N-H bond vector whose dynamics are probed with these measurements[59]. Still, the agreement of the trends in the S^2 order parameters and the calculated *RMSF* values is remarkable, and serves as validation of our coarse description of the stabilizing effects of Zn^{2+} cations in TAZ1 (see Methods).

First, we focus on the differences in fluctuations along the pro-

tein sequence for both bound and unbound TAZ1. In all cases, we find higher values of the *RMSF* in the hinge regions of TAZ1 and at the termini, while helical segments fluctuate less (Figure 5a). This is in agreement with the trends in the S^2 values, which indicate a more restricted mobility for the four amphipathic helices and large amplitude fluctuations primarily in loops. Second, we compare the bound states against the free form of TAZ1. In the bound states with either HIF-1α or CITED2, the calculated *RMSF* for TAZ1 decreases considerably with respect to the unbound TAZ1. This effect is also captured in the shifts of the probability distributions of both the *RMSD* and the intramolecular Q for TAZ1 contacts (see Figure S3 in the Supplementary Information and in selected snapshots from the different states in Figure 5b). These rigidification effects are consistent with the higher S^2 order parameters in the bound forms of TAZ1 from NMR experiments. Experimentally, an exception to this general trend is the loop between α_1 and α_2 for HIF-1α-bound TAZ1[25,60]. This exception is also recovered in the simulations, which indicate a similarly fluctuating loop regardless of the state of TAZ1. Finally, we also capture the rank order of rigidification between the two complexes. The differences in the S^2 order parameters between both complexes indicate that CITED2:TAZ1 forms the least fluctuating complex[25], as we also see in the simulations (Figure 5a). The main differences between the bound states are found in helices α_1 and, most importantly, α_4. The cleft formed between these two helices is the overlapping region where both ligands bind to TAZ1. Therefore, larger fluctuations in this area for the HIF-1α:TAZ1 complex may help CITED2 recognize and displace its partner[25].

The picture emerging from both experiments and simulations for TAZ1 is one of a highly dynamic partner for IDP binding. TAZ1 explores a continuum of possible forms with varying flexibility depending on the ligand, as opposed to different conformational states. There is a clear hierarchy in terms of rigidity, that ranges from the most dynamic free form to the most static CITED2-bound state. Although there may be additional implications for the regulation of gene expression, the primary functional role of the conformational plasticity of TAZ1 is to fine tune the binding affinities through modulation of entropic contributions[24].

3.4 Free energy landscape for competitive binding

We now focus on the emergent behaviour of the consensus models when the three proteins are included in the simulation box. We recall that the only additional tweak of the interaction parameters with respect to the binary simulation models, is the inclusion of a repulsive core that avoids overlaps between the two IDPs (see Methods). In this case, we have run extensive replica exchange simulations at temperatures ranging between 290 and 395 K, in order to increase the efficiency in the sampling of conformational space. The results from these simulations are robust, as the average properties that we calculate were converged after half the total duration of the simulations (5 microseconds) that we have run at each temperature (see Figure S3 in Supplementary Information).

The inclusion of the three molecules results in a more complex

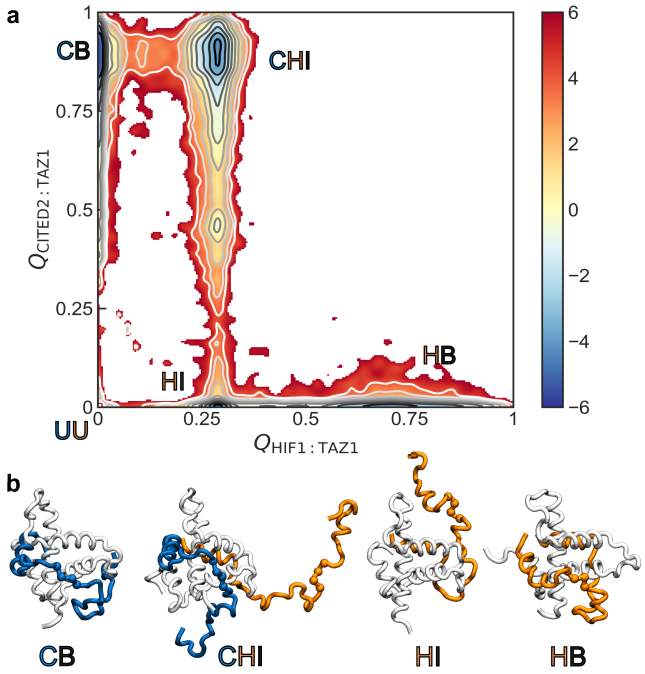

Fig. 6 (a) Two dimensional free energy landscape for the projection on the fraction of intermolecular native contacts, Q, from replica exchange simulations of the ternary complex. Free energies are shown in units of k_BT. Names of relevant bound states are overlaid on the figure. (b) Representative snaphots of the stable bound states: CITED2 bound (CB), HIF-1α bound (HB), HIF-1α intermediate (HI) and ternary intermediate (CHI).

landscape where both binary and ternary complexes may readily form. In Figure 6a we show the potential of mean force for the projection on the inter-molecular fraction of native contacts at 300 K. As expected, the presence of the alternative binding partner does not influence the dominant bimolecular free energy basins. For HIF-1α:TAZ1, we find the intermediate HI state and the fully bound HB state at $Q_{\mathrm{CITED2:TAZ1}} = 0$. Conversely, for the CITED2:TAZ1 complex, the bound state (CB) appears in the potential of mean force at $Q_{\mathrm{HIF1}\alpha\mathrm{:TAZ1}} = 0$. Ternary simulations with both IDPs present not only recapitulate observations from binary complex formation, but also result on a ternary intermediate state, that we term CHI (see Figure 6a and b). The intermediate appears at $Q_{\mathrm{HIF1}\alpha\mathrm{:TAZ1}} \simeq 0.3$ and values of $Q_{\mathrm{CITED2:TAZ1}}$ corresponding to the bound state. We note that there is an additional, more shallow basin at $Q_{\mathrm{HIF1}\alpha\mathrm{:TAZ1}} \simeq 0.3$ and $Q_{\mathrm{CITED2:TAZ1}} \simeq 0.4$, which may have an important role as an "encounter complex" in the displacement of HIF-1α by CITED2 (see below).

In Figure 6b we show representative snapshots of the stable states (excluding the unbound state, UU). As the binary states do not differ from those in the simulations with only one IDP, we focus on the ternary CHI intermediate, which is highly dynamic as originally proposed[23]. In this state, the C-terminal helix (α_C) of HIF-1α remains bound to TAZ1, while CITED2 binds through its initial alpha helix α_A and the LPEL motif. Hence, as anticipated by the simulations of the binary complexes, we observe very different behaviours for the LPEL motif from CITED2 and the homologous LPQL motif from HIF-1α, with the latter being within

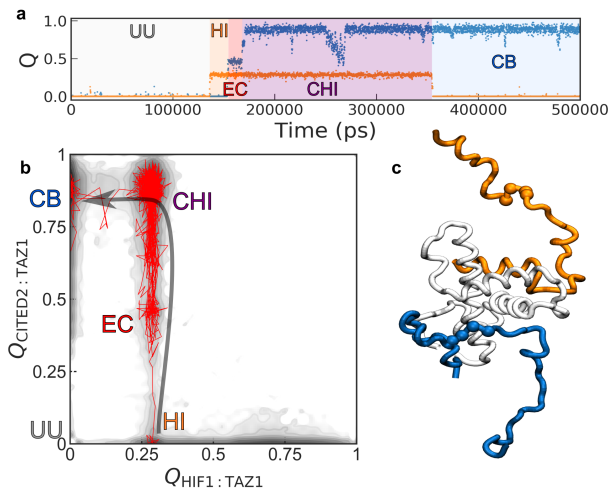

Fig. 7 Binding mechanism for the ternary complex. (a) Time series data for the fraction of native contacts for a trajectory where CITED2 displaces HIF-1α. Coloured bands indicate segments in the trajectory corresponding to different conformational states (labelled). Note the "encounter complex" (EC) in red. (b) Overlay of the values of Q in the same trajectory (red lines) on the potential of mean force. The arrow marks the proposed mechanism to progress from the HI intermediate to the CB state. (c) Representative structure of the ternary "encounter complex".

unbound segment of the protein. This is important because these residues are critical for the stability of the complexes and competition between both IDPs has been proposed to be mediated by the LPXL region.

3.5 A putative mechanism for the displacement of HIF-1α by CITED2

Clearly, the emergent free energy basins in the ternary landscape may have mechanistic importance for the transitions between the ternary intermediate and the binary complexes. We run short equilibrium simulations of the ternary system to resolve molecular transitions between the different states. In these runs, we find binary complex formation as the most likely event. These transitions do not differ from the results reported for the simulations of the binary mixtures (see Figure 4). As one might expect, the presence of the alternative binding partner in the simulation does not perturb the mechanism of the transition if only one IDP binds. This might be different in a model that explicitly included the effects of long-range charged interactions, possibly pointing to a limitation of our study. However, the charge densities of HIF-1α and CITED2 are very similar (positively and negatively charged residues are about 10 and 20%, respectively, for both proteins). Hence we think it is unlikely that this would play a significant part in influencing bimolecular binding.

Additionally to the binary binding processes, the CB and HI states are connected to the ternary intermediates in the potential of mean force in Figure 6a. Transitions involving these states readily appear in our equilibrium runs of the ternary system. We show a trajectory involving a transition between the binary intermediate (i.e. HI) state and the ternary CHI intermediate state of the complex in Figure 7a (90 other trajectories from different initial states, where we find the expected hops between free energy basins, are summarized in the Supplementary Information, Figure S4). Initially, the α_C helix of HIF-1α binds to form the HI intermediate. In the HI state, the cleft between the α_1 and α_4 helices of TAZ1 is still free to interact. From this state, the complex could commit to the fully bound form (HB). Instead, in this specific trajectory we find that the LPEL motif of CITED2 interacts with the corresponding groove in TAZ1 and binds to form the CHI intermediate. The structure of the "encounter complex" (EC) formed after the initial attack via the LPEL motif is still highly flexible and has relatively few specific interactions[61,62] (see Figure 7c). Next, CITED2 reaches its fully bound state, while HIF-1α remains partly bound (CHI state). Finally, HIF-1α detaches from the complex and the system reaches the CB state. We note that this multi-step mechanism (marked by an arrow in Figure 6b) closely resembles the one proposed from the experimental results[23].

4 Conclusions

In this study we have considered the case of a ternary system involving two IDPs, HIF-1α and CITED2, and their highly dynamic binding partner, TAZ1. Combined efforts from different groups, using an array of methods including ITC, NMR and fluorescence anysotropy experiments[16,17,19,21–23,25,60], have resulted in a detailed description of the behaviour of these molecules and their binding mechanisms. In isolation, TAZ1 is a highly dynamic protein that sacrifices part of its conformational entropy upon binding to its partners[25,60], more so in the case of CITED2. Experiments on the binary mixtures of TAZ1 with HIF-1α and CITED2 indicated similarly strong (i.e. nM) binding for both IDPs. More surprisingly, NMR and fluorescence experiments on ternary mixtures including TAZ1 and both IDPs suggested a kinetically driven displacement of HIF-1α by CITED2[23]. An intermediate state with the α_C helix of HIF-1α still bound to TAZ1 was invoked by the authors.

In this work we have studied this system using coarse grained molecular simulations with a potential energy function defined by the native contacts alone (i.e. with no explicit description of unspecific interactions like electrostatics or non-native contacts). One of our predictions for the binary complexes is that HIF-1α:TAZ1 binding occurs via an intermediate where only the α_C helix is bound to TAZ1, while CITED2:TAZ1 binding is two-state. Although one of us has shown that the weight of the intermediate in the ensemble will vary with details of the model[40], experimental evidence points to a special role for the α_C:TAZ1 interactions, in agreement with our predictions. On one hand ITC experiments have shown that this region alone is responsible for the formation of a 200 μM complex[63]. Also, the α_A and (to a lesser extent) the LPQL motif have decreased $\{^1H\}$-^{15}N NOE values in NMR experiments, indicating that these are the most dynamic regions of the complex[23]. Also, Freedman et al[19] and later Berlow and her coworkers[23] proposed that differences in the binding cooperativity of CITED2 and HIF-1α may be key to enable the regulatory role of this hypoxic switch. The results from our simulations confirm the strong coupling in the binding of CITED2 and weak coupling in the case of HIF-1α.

We also provide a detailed description of the competitive binding process in the presence of both IDPs, together with snapshots for the binding mechanism. In our binding trajectories, the LPEL motif of CITED2 has a critical role in the formation of an encounter complex with TAZ1 (EC in Figure 7), taking advantage of the flexibility of the HIF-1α complex in the LPQL region. From this encounter complex, the system progresses to the ternary binding intermediate (CHI) that closely recapitulates the description in the original work[23]. The conclusive details on the different metastable states *en-route* to the CITED2:TAZ1 complex may be resolved in the future using simulations with an atomistic description of protein-protein interactions and the influence of the solvent, and experiments with methods like relaxation-dispersion NMR for the detection of "invisible states"[64] or smFRET[65].

Even if we have stressed several points of agreement between experiment and simulation, there are other aspects that our simulations do not reproduce. First, from HSQC experiments of ^{15}N-labelled TAZ1 with both IDPs in a 1:1:1 molar ratio, Berlow et al concluded that only the CITED2:TAZ1 complex was present. On the contrary, in our simulations we find that the free energy basins for the CITED2:TAZ1 and HIF-1α:TAZ1 complexes remain present. This discrepancy may be due to the additional stable intermediate (CHI) in the ternary mixture, which very much resembles the CITED2:TAZ1 complex and may result in indistinguishable resonances in the NMR spectra. Second, competitive binding assays resulted in a decrease of K_D for CITED2:TAZ1 (0.2 nM) and an increase of K_D for HIF-1α (0.9 μM). While we cannot figure out a thermodynamic argument for the values of the binary equilibrium constants to change, we must note that these experimental values correspond to apparent K_D. Again, one possibility is that the decreased K_D for CITED2:TAZ1 accounts for the combined populations of the CB and CHI intermediate. Another possible explanation is that the simple description from topology alone is not sufficient to account for these subtle effects.

In the finishing stages of this work, we have become aware of two other studies with similar simulation models that include an explicit electrostatic term[66,67], which gives the opportunity to compare contributions from topology alone and additional energy terms. In one of these studies, Chu et al. use an alternative model by Clementi et al[52], which is also coarse-grained to the C$^\alpha$-level[66]. The details of the coarse-grained model[52] are slightly different: for example, in the Clementi model the dihedral terms in the potential energy function are structure-based, while in the Karanicolas and Brooks model they are based on PDB statistics[41,52]. In the other study, by Wang and Brooks, the same coarse-grained model as in this work was used, although adding electrostatics, and different simulation and analysis techniques were used. As inevitably happens in coarse grained modelling, there are design decisions for building the ternary model that diverge between all three works, like the description of the Zn^{2+} cations in TAZ1. Regardless of the differences, all three works agree in the prediction of a very similar ternary intermediate state, with very high $Q_{\text{CITED2:TAZ1}}$ and low $Q_{\text{HIF1}\alpha\text{:TAZ1}}$, and the Karanicolas and Brooks model either with or without electrostatics, results in an additional intermediate state (HI) in the binding of HIF-1α to TAZ1[67]. It is remarkable that, as we show here, all the information about the stable states in the competitive binding mechanism is encoded in topology alone, even if the relative weights of the different states will be sensitive to details in the parametrization. Future work will help to further understand the features of this ternary binding energy landscape and its role in the regulation of the hypoxic response and exploit them in anti-cancer therapies.

Acknowledgments

IR acknowledges a PhD studentship from DIPC. DDS acknowledges Grants PGC2018-099321-B-I0 and RYC-2016-19590 from the Spanish Ministry of Science, Research and Universities and an Ikerbasque Research Fellowship. We wish to thank Robert B. Best for generously sharing computational tools for running the simulations, Antonio Rey for carefully reading the manuscript and Gerhard Hummer for useful discussions.

References

1 A. K. Dunker, C. J. Brown, J. D. Lawson, L. M. Iakoucheva and Z. Obradovic, *Biochemistry*, 2002, **41**, 6573–6582.

2 V. N. Uversky, J. R. Gillespie and A. L. Fink, *Proteins*, 2000, **41**, 415–427.

3 H. Y. J. Fung, M. Birol and E. Rhoades, *Curr. Opin. Struct. Biol.*, 2018, **49**, 36 – 43.

4 J. Gsponer, M. E. Futschik, S. A. Teichmann and M. M. Babu, *Science*, 2008, **322**, 1365–1368.

5 P. Csermely, R. Palotai and R. Nussinov, *Trends Biochem. Sci.*, 2010, **35**, 539 – 546.

6 M. Knott and R. B. Best, *J. Chem. Phys.*, 2014, **140**, 175102.

7 H.-X. Zhou, X. Pang and C. Lu, *Phys. Chem. Chem. Phys.*, 2012, **14**, 10466–10476.

8 B. A. Shoemaker, J. J. Portman and P. G. Wolynes, *Proc. Natl. Acad. Sci. U.S.A.*, 2000, **97**, 8868–8873.

9 A. Bah, R. M. Vernon, Z. Siddiqui, M. Krzeminski, R. Muhandiram, C. Zhao, N. Sonenberg, L. E. Kay and J. D. Forman-Kay, *Nature*, 2015, **519**, 106.

10 A. Borgia, M. B. Borgia, K. Bugge, V. M. Kissling, P. O. Heidarsson, C. B. Fernandes, A. Sottini, A. Soranno, K. J. Buholzer, D. Nettels *et al.*, *Nature*, 2018, **555**, 61.

11 P. E. Wright and H. J. Dyson, *Nat. Rev. Mol. Cell Biol.*, 2015, **16**, 18.

12 A. Cumberworth, G. Lamour, M. Babu and J. Gsponer, *Biochem. J.*, 2013, **454**, 361–369.

13 A. C. M. Ferreon, J. C. Ferreon, P. E. Wright and A. A. Deniz, *Nature*, 2013, **498**, 390.

14 L. Dahal, T. O. Kwan, J. J. Hollins and J. Clarke, *J. Mol. Biol.*, 2018, **430**, 2468 – 2477.

15 H. N. Motlagh, J. O. Wrabl, J. Li and V. J. Hilser, *Nature*, 2014, **508**, 331–339.

16 S. A. Dames, M. Martinez-Yamout, R. N. De Guzman, H. J. Dyson and P. E. Wright, *Proc. Natl. Acad. Sci. U.S.A.*, 2002, **99**, 5271–5276.

17 S. J. Freedman, Z.-Y. J. Sun, F. Poy, A. L. Kung, D. M. Liv-

ingston, G. Wagner and M. J. Eck, *Proc. Natl. Acad. Sci. U.S.A.*, 2002, **99**, 5367–5372.
18 G. L. Semenza, *Proc. Natl. Acad. Sci. U.S.A.*, 2002, **99**, 11570–11572.
19 S. J. Freedman, Z.-Y. J. Sun, A. L. Kung, D. S. France, G. Wagner and M. J. Eck, *Nat. Struct. Mol. Biol.*, 2003, **10**, 504.
20 S. Bhattacharya and P. J. Ratcliffe, *Nat. Struct. Mol. Biol.*, 2003, **10**, 501–503.
21 R. N. De Guzman, J. M. Wojciak, M. A. Martinez-Yamout, H. J. Dyson and P. E. Wright, *Biochemistry*, 2005, **44**, 490–497.
22 R. N. De Guzman, M. A. Martinez-Yamout, H. J. Dyson and P. E. Wright, *J. Biol. Chem.*, 2004, **279**, 3042–3049.
23 R. B. Berlow, H. J. Dyson and P. E. Wright, *Nature*, 2017, **543**, 447.
24 I. Nyqvist, E. Andersson and J. Dogan, *J. Phys. Chem. B*, 2019, **123**, 2882–2888.
25 R. B. Berlow, M. A. Martinez-Yamout, H. Dyson and P. E. Wright, *Biochemistry*, 2019, **58**, 1354–1362.
26 G. L. Semenza, *Nat. Rev. Cancer*, 2003, **3**, 721–732.
27 R. B. Best, W. Zheng and J. Mittal, *J. Chem. Theory Comput.*, 2014, **10**, 5113–5124.
28 J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller and A. D. MacKerell Jr, *Nat. Methods*, 2017, **14**, 71.
29 P. Robustelli, S. Piana and D. E. Shaw, *Proc. Natl. Acad. Sci. U.S.A.*, 2018, **115**, E4758–E4766.
30 P. S. Shabane, S. Izadi and A. V. Onufriev, *J. Chem. Theory Comput.*, 2019, **15**, 2620–2634.
31 N. Plattner, S. Doerr, G. De Fabritiis and F. Noé, *Nat. Chem.*, 2017, **9**, 1005.
32 A. C. Pan, D. Jacobson, K. Yatsenko, D. Sritharan, T. M. Weinreich and D. E. Shaw, *Proc. Natl. Acad. Sci. U.S.A.*, 2019, **116**, 4244–4249.
33 C. Clementi, *Curr. Opin. Struct. Biol.*, 2008, **18**, 10 – 15.
34 J. Bryngelson, J. Onuchic, N. Socci and P. Wolynes, *Proteins*, 1995, **21**, 167–195.
35 L. L. Chavez, J. N. Onuchic and C. Clementi, *J. Am. Chem. Soc.*, 2004, **126**, 8426–8432.
36 R. B. Best, *J. Phys. Chem. B*, 2013, **117**, 13235–13244.
37 Y. Levy, J. N. Onuchic and P. G. Wolynes, *J. Am. Chem. Soc.*, 2007, **129**, 738–739.
38 A. G. Turjanski, J. S. Gutkind, R. B. Best and G. Hummer, *PLoS Comput. Biol.*, 2008, **4**, e1000060.
39 D. Ganguly and J. Chen, *Proteins*, 2011, **79**, 1251–1266.
40 D. De Sancho and R. B. Best, *Mol. BioSyst.*, 2012, **8**, 256–267.
41 J. Karanicolas and C. L. Brooks, *Protein Sci*, 2002, **11**, 2351–2361.
42 S. Miyazawa and R. L. Jernigan, *J. Mol. Biol.*, 1996, **256**, 623 – 644.
43 M. F. Rey-Stolle, M. Enciso and A. Rey, *J. Comput. Chem.*, 2009, **30**, 1212–1219.
44 A. M. Rubio and A. Rey, *Phys. Chem. Chem. Phys.*, 2019, **21**, 6544–6552.
45 A. Schug, P. C. Whitford, Y. Levy and J. N. Onuchic, *Proc. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 17674–17679.
46 B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.
47 T. G. W. Graham and R. B. Best, *J. Phys. Chem. B*, 2011, **115**, 1546–1561.
48 R. B. Best and G. Hummer, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 1088–1093.
49 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 1011–1021.
50 I. Lindström, E. Andersson and J. Dogan, *Sci. Rep.*, 2018, **8**, 7872.
51 D. Ganguly, W. Zhang and J. Chen, *PLoS Comput. Biol.*, 2013, **9**, 1–12.
52 C. Clementi, H. Nymeyer and J. N. Onuchic, *J. Mol. Biol.*, 2000, **298**, 937 – 953.
53 J. K. Noel, M. Levi, M. Raghunathan, H. Lammert, R. L. Hayes, J. N. Onuchic and P. C. Whitford, *PLoS Comput. Biol.*, 2016, **12**, 1–14.
54 M. Gao, J. Yang, S. Liu, Z. Su and Y. Huang, *Biophys. J.*, 2019, **117**, 1301 – 1310.
55 J. M. Rogers, V. Oleinikovas, S. L. Shammas, C. T. Wong, D. De Sancho, C. M. Baker and J. Clarke, *Proc. Natl. Acad. Sci. U.S.A.*, 2014, **111**, 15420–15425.
56 J. M. Wojciak, M. A. Martinez-Yamout, H. J. Dyson and P. E. Wright, *EMBO J.*, 2009, **28**, 948–958.
57 S. P. Mukherjee, M. Behar, H. A. Birnbaum, A. Hoffmann, P. E. Wright and G. Ghosh, *PLOS Biology*, 2013, **11**, 1–20.
58 I. Nyqvist, E. Andersson and J. Dogan, *J. Phys. Chem. B*, 2019, **123**, 2882–2888.
59 Y. Gu, D.-W. Li and R. Brüschweiler, *J. Chem. Theory Comput.*, 2014, **10**, 2599–2607.
60 I. Nyqvist and J. Dogan, *Sci. Rep.*, 2019, **9**, 1–9.
61 M. Ubbink, *FEBS Lett.*, 2009, **583**, 1060–1066.
62 G. Schreiber, G. Haran and H.-X. Zhou, *Chem. Rev.*, 2009, **109**, 839–860.
63 H. F. Kyle, K. F. Wickson, J. Stott, G. M. Burslem, A. L. Breeze, C. Tiede, D. C. Tomlinson, S. L. Warriner, A. Nelson, A. J. Wilson and T. A. Edwards, *Mol. BioSyst.*, 2015, **11**, 2738–2749.
64 A. J. Baldwin and L. E. Kay, *Nat. Chem. Biol.*, 2009, **5**, 808.
65 M. Brucale, B. Schuler and B. Samori, *Chem. Rev.*, 2014, **114**, 3281–3317.
66 W.-T. Chu, X. Chu and J. Wang, *Proc. Natl. Acad. Sci. U.S.A.*, 2020, In press.
67 Y. Wang and C. L. Brooks III, *J. Phys. Chem. Lett.*, 2020, **11**, 864–868.