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

Irene Ruiz-Ortiz[1] and David de Sancho[2]
[1)] *Donostia International Physics Center, 20018, Donostia-San Sebastián, Spain*
[2)] *University of the Basque Country, Faculty of Chemistry, Paseo Manuel Lardizabal, 3, 20018 Donostia-San Sebastián, Spain*[a)]

(Dated: 21 January 2020)

Many intrinsically disordered proteins (IDPs) are involved in complex signalling networks inside the cell. Their particular binding modes elicit different types of responses and subtle regulation of biological responses. 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 molecular simulations with a coarse-grained model to provide a glimpse of the structure 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 molecular description of the emergent properties in the complex binding of these IDPs.

I. 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 initial work on IDP binding discussed the mechanism in terms of either conformational selection or induced fit[5,6], and, more recently, the "dock and coalesce" model[7]. But IDPs can be even more wildly dissimilar from folded proteins, and undergo multivalent binding through phosphorylated sites[8] or form complexes in their unfolded state with ultra-high affinity[9]. From a kinetic standpoint, binding in an expanded, disordered state may be advantageous, as explained by the "fly-casting" model[10]. But as important as characterizing the biophysics of binding processes, is understanding how these new modes 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]. An interesting case, involving two different IDPs, is that of the α-subunit of the hypoxia inducible transcription factor HIF-1α and its negative feedback regulator CITED2[16–23], a system that is highly relevant for the development of cancer therapies[24]. Both proteins are able to bind to the TAZ1 domain of the CREB binding protein (CBP). Structures for the binary complexes of both IDPs with CBP and its homolog p300 have been resolved using NMR[16,17,19,20] (see Figure 1a). Recent work has shown that in binary mixtures, the HIF-1α and CITED2 complexes with TAZ1 have very much the same binding affinities ($K_D = 10 \pm 2$ nM)[21]. A particularly important role in this high affinity is played by the conserved LPQL or LPEL motifs, for HIF-1α and CITED2 , respectively (see Figure 1c). Both IDPs share part of their interacting surface with their partner TAZ1. Surprisingly, in the ternary mixture increasing CITED2 concentrations were found to be able to displace HIF-1α bound to TAZ1, while the reverse process was not observed[21]. The displacement of HIF-1α was coupled to a 50-fold decrease of the apparent K_D (0.2 ± 0.1 nM) of CITED2. To explain this counterintuitive result, Berlow et al invoked a transient intermediate state[21], although a detailed molecular description of this state is still lacking. This process involving two IDPs and their shared partner may be critical for the hypoxic response, as it acts as a negative feedback loop to control the response induced by HIF-1α[21].

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[25–28]. However, simulating full protein-protein binding with all atom MD simulations is extremely demanding even for binary systems[29,30]. Instead, we resort to simple structure based, coarse grained models[31]. This type of models have consistently made successful, semi-quantitative predictions on protein folding and binding, both for folded proteins[32] and for IDPs[33–35]. In fact one of us already studied the effects of the model parametrization on the simulated binding for the HIF-1α:TAZ1 complex[35]. In this paper we expand that work to study the CITED2:TAZ1 complex and the properties emerging from the simulation model for the ternary mixture. Our main focus is on resolving a molecular picture

[a)]Donostia International Physics Center, 20018, Donostia-San Sebastián, Spain; Electronic mail: david.desancho@ehu.eus

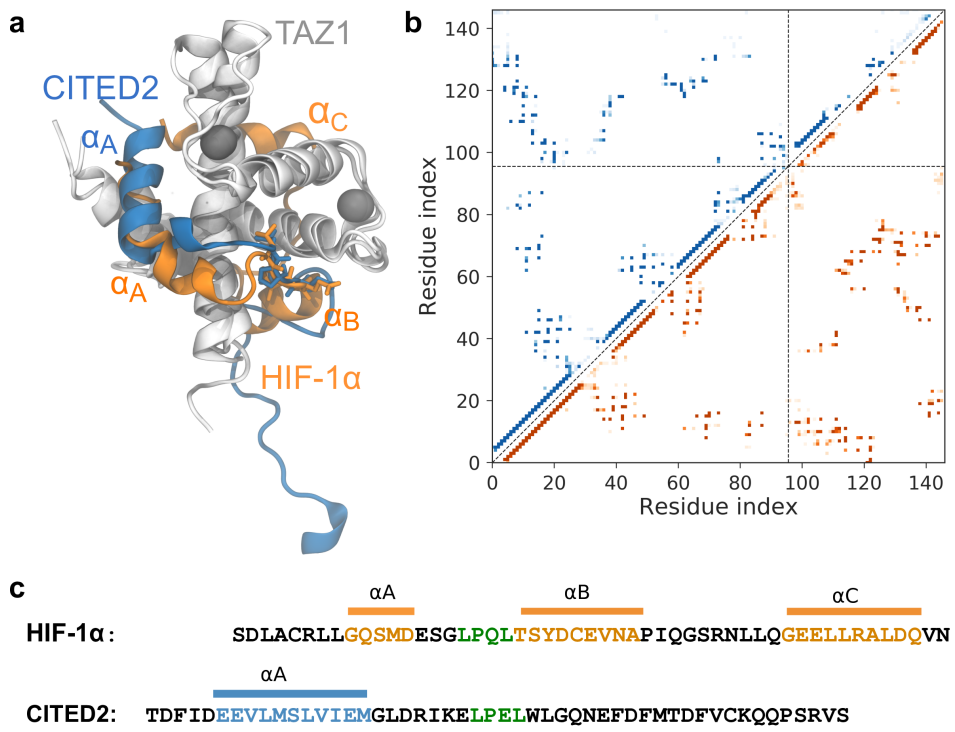

FIG. 1. (a) Cartoon representation of the complexes formed by HIF-1α (PDB id: 1l8c, orange) and CITED2 (1r8u, blue) with the TAZ1 domain of CBP (white). (b) Corresponding contact maps from the 20 experimental NMR models. The upper and lower triangles correspond to the CITED2:TAZ1 and HIF-1α:TAZ1 complexes, 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.

of the ternary intermediate proposed by Berlow et al[21].

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 attention to the influence of the binding on the dynamics of TAZ1, which has been recently investigated using NMR[22,23]. Finally, we show the emergent properties that we predict for the ternary complex and discuss the implications of our findings.

II. MATERIALS AND METHODS

A. Coarse-grained simulation model

In this work we start from the structure-based Karanicolas and Brooks model[36], which is coarse grained to the level of the C$^{\alpha}$ atoms. The potential energy function is defined as a sum of terms

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

The first energy term in this equation is a harmonic native-centric potential for all bonds between beads in the model, with the equilibrium distance being that in the reference (experimental) structure. The terms for angles and torsions are both derived from PDB statistics. Finally, favourable nonbonded interactions are defined among amino acid pairs that are in contact in the native conformation[36], and are described by the following expression,

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

where ε_{ij} is the strength of the interaction between residues i and j, proportional to the Mijazawa and Jernigan contact energies[37], 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

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

where ε_{rep} is a generic repulsive energy term related to the energy scale in the model and σ_{ij} is the repulsive radius[36].

B. Calibration of the model

Using this description, we have built our computational models based on the most representative conformer of the HIF-1α:TAZ1 and CITED2:TAZ1 experimental structures (PDB id 1l8c[16] and 1r8u[20] respectively, see Figure 1a). 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. The angular and dihedral terms in the Karanicolas and Brooks model are statistical[36], 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[38]. 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 TAZ1:HIF-1α and TAZ1:CITED2 complexes. This involved decreasing the interaction strength of native contacts by 2% for TAZ1:CITED2 and increasing it by 4% for TAZ1:HIF-1α. In the simulations of the ternary complex, nonbonded interactions between HIF-1α and CITED2 are defined simply using the repulsive term from Equation 3.

Finally, in the experimental structures of the TAZ1 complexes with HIF-1α and CITED2, there are three Zn^{2+} cations, each of them coordinated to a His and three Cys residues. These metal cations are essential for TAZ1 to retain a stable fold[16,20]. 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.

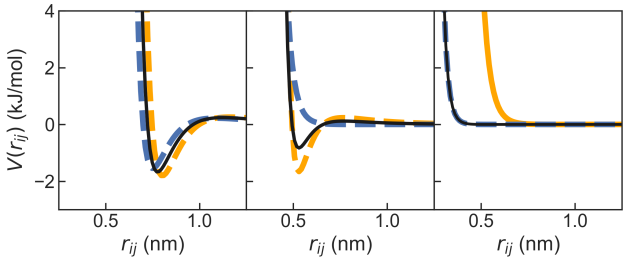

FIG. 2. Intra-TAZ1 nonbonded potentials for the HIF-1α:TAZ1 (orange), CITED2:TAZ1 (blue) and consensus (black) models. We show three representative cases, corresponding to a contact shared between both experimental structures (left), a contact present in one and absent in the other structure (center),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

C. 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[39]). 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}.

D. Analysis of the simulations

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

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

where 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[40]. Because the unbound state collapses onto a single bin when histogramming on Q, we also calculate the d_{RMS}, defined as[41]

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

Results for different runs were combined to estimate potentials of mean force using the weighted histogram analysis method (WHAM)[42]. 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

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

as before[35]. In this expression, the bound population p_b is calculated integrating the exponential of the potential of mean force on Q, $F(Q)$, as $p_{\text{b}} = \int_{Q_{\text{TS}}}^{1} \exp(-\beta F(Q))dQ$, where $\beta = 1/k_{\text{B}}T$ is the inverse thermal energy and Q_{TS} is the value corresponding to the barrier top in the potential of mean force. Then, the unbound population is simply $p_{\text{u}} = 1 - p_{\text{b}}$.

III. RESULTS AND DISCUSSION

A. 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[33]. The sharp barrier at low values of Q is qualitatively consistent with the results from experimental Φ-value analysis[43], which indicated 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[35,44]. In this intermediate HIF-1α binds only through one of its helices to TAZ1[35], primarily the C-terminal α_C-helix, which forms more contacts with TAZ1 than α_A or α_B (see Fig. 1b). The prediction of an intermediate state and a highly dynamic α_A helix is consistent with experimental NMR measurements on this system[21], in particular with the low $\{^1H\} - ^{15}N$ values reported for the region encompassing the α_A helix and the LPQL motif, which imply high flexibility[21]. The consistency with previous simulation results for HIF-1α:TAZ1 binding[35,44] 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[36]. We note that for the similar coarse-grained SMOG model[45,46], the binding of HIF-1α to TAZ1 does not involve any intermediate states[47]. In the case of the CITED2:TAZ1 complex, both the Karanicolas and Brooks model and the SMOG model result in two-state binding[47].

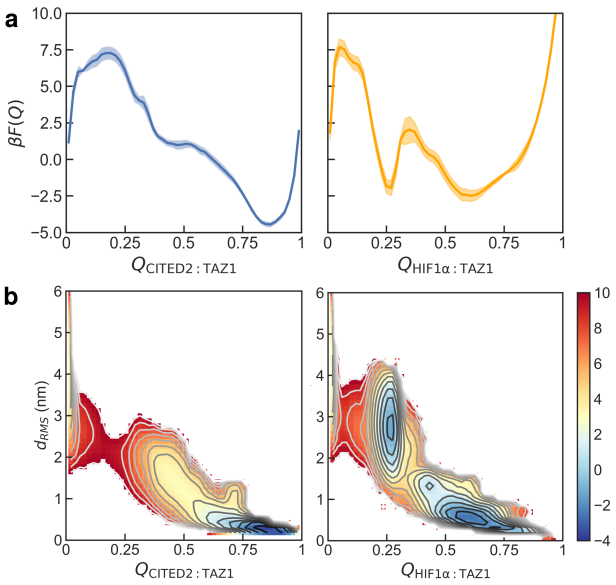

FIG. 3. Free energy landscapes for binding of HIF-1α (orange) and CITED2 (blue) to TAZ1. (a) Free energy landscapes for the projection on the fraction of native contacts (Q). Errors are shown as bands. (b) Two dimensional free energy landscapes for the projection on both Q and d_{RMS}. Free energies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

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 mechanisms between HIF-1α:TAZ1 and CITED2:TAZ1 are unsurprising given that the topologies of these complexes are also distinct (see contact maps in Fig. 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.

B. Binding mechanism of the IDPs to TAZ1

The umbrella sampling simulations that we have used to derive the free energy landscapes do not contain any valuable information about the dynamics. In order to monitor the binding transitions for both IDPs, we have run a set of equilibrium simulations starting from conformations where both proteins are separated by randomly selected distances within the unbound state. In Figure 4 we show data from two representative binding trajectories. In the case of for HIF-1α, the initial binding step may occur via either helices α_A or α_C, but we find the latter to be the dominant pathway. In the case of CITED2, we find that the initial approach involves the LPEL residues, and then the remainder of the contacts involving the α_A helix binds to TAZ1. The C-terminal tail of the protein, which does not involve intermolecular native contacts, remains unstructured in the bound state. Importantly, one of the predictions of the simulation model is that the LPEL from CITED2 and the LPQL from HIF-1α, which interact with an overlapping region of TAZ1, have very different roles in the binding process.

C. 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[48]. Partner proteins like TAZ1 do

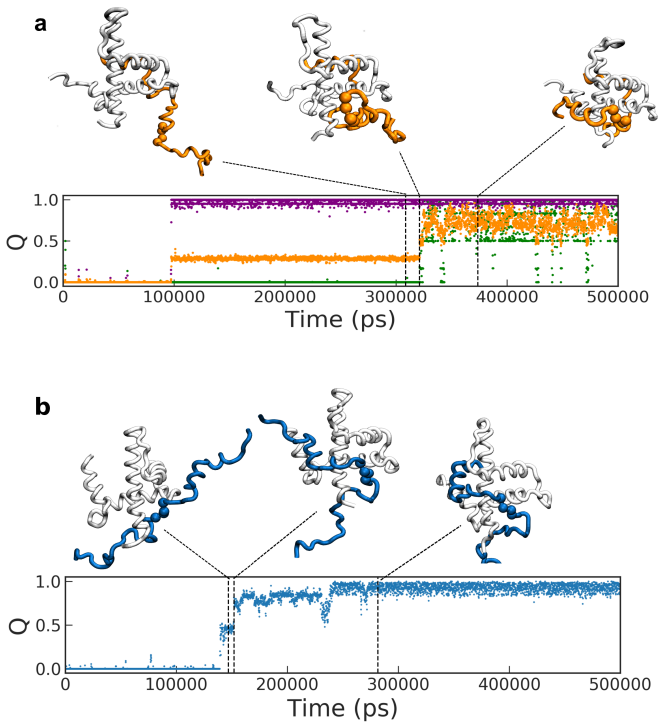

FIG. 4. Time series data for the fraction of intermolecular native contacts (Q) and corresponding snapshots for two representative binding trajectories for binary complexes HIF-1α:TAZ1 (a) and CITED2:TAZ1 (b). In (a) we show Q for all the contacts (orange), together with that for helices α_A (green)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

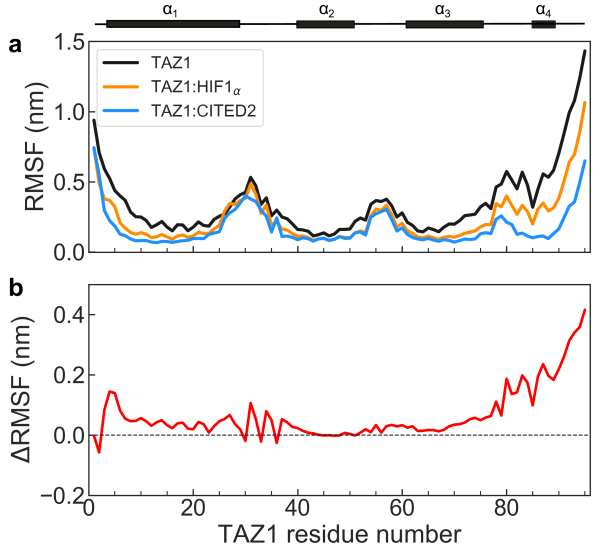

FIG. 5. Fluctuations in equilibrium simulations of TAZ1. (a) Values of the *RMSF* in the unbound state of TAZ1 (black), and in its complexes with HIF-1α (orange) and CITED2 (blue). At the top, we show schematically the DSSP assignment of secondary structure, with bars corresponding to α-helices. (b) Differences in the RMSF between the HIF-1α and CITED2 complexes. Positive values indicate more fluctuations in HIF-1α.

not behave like mere scaffolds that disordered ligands bind to. Instead, they can adjust their flexibility in relation to the interaction with ligands. This conclusion has been drawn for TAZ1 from studies of a collection of IDPs, including HIF-1α, CITED2, RelA-TA2 or TAD-STAT2[23,49].

We check these effects in a set of equilibrium simulations of TAZ1 and its complexes with both HIF-1α and CITED2. In Figure 5a we report the fluctuations in the different stable states of TAZ1 in the consensus models. 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. This result is qualitatively consistent with NMR experiments. In particular, S^2 order parameters had high values for the four amphipathic helices, indicating restricted mobility, and low values in most of the loops[23], indicating large amplitude fluctuations. This agreement with experiment validates our coarse description of the stabilizing effects of Zn^{2+} cations (see Methods).

In the bound states with either HIF-1α or CITED2, the calculated *RMSF* for TAZ1 decreases considerably with respect to the free form of the protein (see Figure 5a). These rigidification effects were also derived from NMR experiments, with higher S^2 order parameters in the bound forms of TAZ1. Experimentally, an exception to the rigidification effects in the case of HIF-1α-bound TAZ1 is the loop between α_1 and α_2[23,50]. This exception is also recovered in the simulations. Finally, we also capture the rank order of rigidification between the two complexes. The differences in the order parameters between both complexes indicate that CITED2:TAZ1 forms the least fluctuating complex[23], as we also see in the simulations (Figure 5b). The main differences between the bound states are found in helices α_1 and, more importantly, α_4 of TAZ1. 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[23].

D. 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. 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. We recall that the only additional tweak of the interaction parameters is the inclusion of a repulsive core that avoids overlaps between the two IDPs (see Methods). The inclusion of the three molecules results in a more complex 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 $Q_{\mathrm{CITED2:TAZ1}} = 0$. Con-

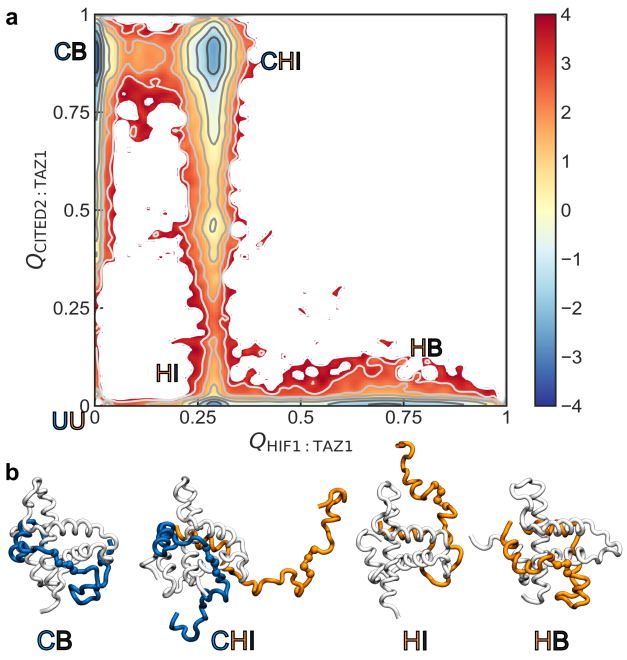

FIG. 6. (a) Two dimensional free energy landscape for the projection on the fraction of intermolecular native contacts, Q, from replica exchange simulations of the ternary complex. Free energies are shown in units of k_BT. Names of relevant bound states are overlaid on the figure. (b) Representative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

versely, for the CITED2:TAZ1 complex, the bound state (CB) appears in the potential of mean force at $Q_{HIF1\alpha:TAZ1} = 0$. 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 Fig. 6a and b). The intermediate appears at $Q_{HIF1:TAZ1} \simeq 0.3$ and values of $Q_{CITED2:TAZ1}$ corresponding to the bound state. We note that there is an additional, more shallow basin at $Q_{HIF1:TAZ1} \simeq 0.3$ and $Q_{CITED2:TAZ1} \simeq 0.4$, 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[21]. 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 of CITED2 and the homologous LPQL motif from HIF-1α, with the latter being within 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 LP(Q/E)L competition.

E. 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 Fig. 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 charges, 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. 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

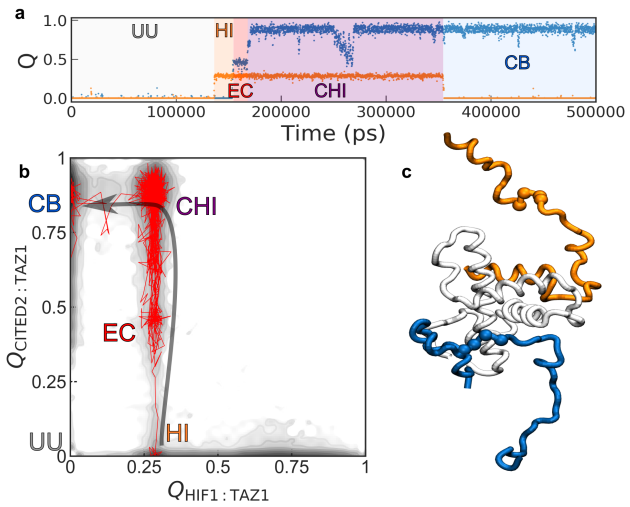

FIG. 7. Binding mechanism for the ternary complex. (a) Time series data for the fraction of native contacts for a trajectory where CITED2 displaces HIF-1α. Coloured bands indicate segments in the trajectory corresponding to different conformational states (labelled). Note the "encounter complex" (EC) 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".

are 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[51,52] (see Fig. 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 Fig. 6b) closely resembles the one proposed from the experimental results[21].

IV. 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[16,17,19–21,23,50], have resulted in a detailed description of the behaviour of the individual 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[23,50], 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[21]. 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 used molecular simulations with a coarse grained model to study this system, resulting in a detailed characterization of the binding of both IDPs. One of the predictions from our simulations on HIF-1α:TAZ1 is the presence of a binding intermediate where the α_C helix is bound to TAZ1. Although one of us has shown that the weight of this intermediate in the ensemble varies with details of the model[35], 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 only is responsible for the formation of a 200 μM complex[53]. Also, the α_A and LPQL motif have low $\{^1$H$\}$-^{1}5N NOE values in NMR experiments, indicating that these are the most dynamic regions of the complex[21]. Finally, Berlow and her co-workers proposed that differences in the binding cooperativity of CITED2 and HIF-1α may be key to enable the regulatory role of this hypoxic switch[21]. The results from our simulations make us concur on this conclusion: we find strong coupling between the LPEL and the α-helix for CITED2, and weak coupling between the LPQL-α_B motif and the terminal α_C for HIF-1α.

We also provide a detailed description of this binding process, together with snapshots for the binding mechanism. In our binding trajectories, the LPEL motif of CITED2 has a critical role in the initial encounter with TAZ1, taking advantage of the flexibility of the HIF-1α complex in that region. From this encounter complex, the system progresses to the ternary binding intermediate (CHI) that closely resembles the one proposed in the original work. The relative weights of the different metastable states *en-route* to the CB state may be resolved in future experiments using methods as relaxation-dispersion NMR for the detection of "invisible states"[54] or smFRET[55].

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 TAZ1:CITED2 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 despite the combination in a ternary mixture. Also, we cannot figure out a thermodynamic argument for the values of the binary equilibrium constants to change. One possibility is that, in the displacement experiments, the decreased K_D for CITED2:TAZ1 corresponds to the combined populations of the CB and CHI intermediate.

While completing this work, we have become aware of another study with the alternative coarse-grained SMOG model with the same level of resolution[56]. The details of the coarse-grained model[45] are slightly different: for example, in the model by Clementi et al angular and dihedral terms in the potential energy function are structure-based[45], while in the Karanicolas and Brooks model they are based on PDB statistics[36]. Also there are inevitable design decisions for building the ternary model that diverge between both works, and in their case electrostatics were considered explicitly in the interaction energy function[56]. Regardless of these differences, both works agree qualitatively in the prediction of a ternary intermediate state, which makes this finding robust to parametrization details. Our model results in an additional intermediate state (HI) in the binding of HIF-1α to TAZ1, that is very important for the ternary equilibria. However, the relative weights of the different states are sensitive to details in the parametrization, as shown by one of us[35]. Further simulation work with a high resolution description of the proteins involved and detailed experiments will help clarify atomistic details of this state and its role in the regulation of the hypoxic response.

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, "Intrinsic disorder and protein function," Biochemistry **41**, 6573–6582 (2002).

[2] V. N. Uversky, J. R. Gillespie, and A. L. Fink, "Why are "natively unfolded" proteins unstructured under physiologic conditions?" Proteins **41**, 415–427 (2000).

[3] H. Y. J. Fung, M. Birol, and E. Rhoades, "Idps in macromolecular complexes: the roles of multivalent interactions in diverse assemblies," Curr. Opin. Struct. Biol. **49**, 36 – 43 (2018).

[4] J. Gsponer, M. E. Futschik, S. A. Teichmann, and M. M. Babu, "Tight regulation of unstructured proteins: From transcript synthesis to protein degradation," Science **322**, 1365–1368 (2008).

[5] P. Csermely, R. Palotai, and R. Nussinov, "Induced fit, conformational selection and independent dynamic segments: an extended view of binding events," Trends Biochem. Sci. **35**, 539 – 546 (2010).

[6] M. Knott and R. B. Best, "Discriminating binding mechanisms of an intrinsically disordered protein via a multi-state coarse-grained model," J. Chem. Phys. **140**, 175102 (2014).

[7] H.-X. Zhou, X. Pang, and C. Lu, "Rate constants and mechanisms of intrinsically disordered proteins binding to structured targets," Phys. Chem. Chem. Phys. **14**, 10466–10476 (2012).

[8] A. Bah, R. M. Vernon, Z. Siddiqui, M. Krzeminski, R. Muhandiram, C. Zhao, N. Sonenberg, L. E. Kay, and J. D. Forman-Kay, "Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch," Nature **519**, 106 (2015).

[9] 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.*, "Extreme disorder in an ultrahigh-affinity protein complex," Nature **555**, 61 (2018).

[10] B. A. Shoemaker, J. J. Portman, and P. G. Wolynes, "Speeding molecular recognition by using the folding funnel: The fly-casting mechanism," Proc. Natl. Acad. Sci. U.S.A. **97**, 8868–8873 (2000).

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

[12] A. Cumberworth, G. Lamour, M. Babu, and J. Gsponer, "Promiscuity as a functional trait: intrinsically disordered regions as central players of interactomes," Biochemical Journal **454**, 361–369 (2013).

[13] A. C. M. Ferreon, J. C. Ferreon, P. E. Wright, and A. A. Deniz, "Modulation of allostery by protein intrinsic disorder," Nature **498**, 390 (2013).

[14] L. Dahal, T. O. Kwan, J. J. Hollins, and J. Clarke, "Promiscuous and selective: How intrinsically disordered bh3 proteins interact with their prosurvival partner mcl-1," J. Mol. Biol. **430**, 2468 – 2477 (2018).

[15] H. N. Motlagh, J. O. Wrabl, J. Li, and V. J. Hilser, "The ensemble nature of allostery," Nature **508**, 331–339 (2014).

[16] S. A. Dames, M. Martinez-Yamout, R. N. De Guzman, H. J. Dyson, and P. E. Wright, "Structural basis for HIF-1α/CBP recognition in the cellular hypoxic response," Proc. Natl. Acad. Sci. U.S.A. **99**, 5271–5276 (2002).

[17] S. J. Freedman, Z.-Y. J. Sun, F. Poy, A. L. Kung, D. M. Livingston, G. Wagner, and M. J. Eck, "Structural basis for recruitment of CBP/p300 by hypoxia-inducible factor," Proc. Natl. Acad. Sci. U.S.A. **99**, 5367–5372 (2002).

[18] G. L. Semenza, "Physiology meets biophysics: Visualizing the interaction of hypoxia-inducible factor α with p300 and CBP," Proc. Natl. Acad. Sci. U.S.A. **99**, 11570–11572 (2002).

[19] S. J. Freedman, Z.-Y. J. Sun, A. L. Kung, D. S. France, G. Wagner, and M. J. Eck, "Structural basis for negative regulation of hypoxia-inducible factor-1α by CITED2," Nat. Struct. Mol. Biol. **10**, 504 (2003).

[20] R. N. De Guzman, M. A. Martinez-Yamout, H. J. Dyson, and P. E. Wright, "Interaction of the taz1 domain of the creb-binding protein with the activation domain of cited2: Regulation by competition between intrinsically unstructured ligands for non-identical binding sites," J. Biol. Chem. **279**, 3042–3049 (2004).

[21] R. B. Berlow, H. J. Dyson, and P. E. Wright, "Hypersensitive termination of the hypoxic response by a disordered protein switch," Nature **543**, 447 (2017).

[22] I. Nyqvist, E. Andersson, and J. Dogan, "Role of conformational entropy in molecular recognition by taz1 of cbp," J. Phys. Chem. B **123**, 2882–2888 (2019), https://doi.org/11.1021/acs.jpcb.9b01343.

[23] R. B. Berlow, M. A. Martinez-Yamout, H. Dyson, and P. E. Wright, "Role of backbone dynamics in modulating the interactions of disordered ligands with the taz1 domain of the creb-binding protein," Biochemistry **58**, 1354–1362 (2019).

[24] G. L. Semenza, "Targeting hif-1 for cancer therapy," Nat Rev Cancer **3**, 721–732 (2003).

[25] R. B. Best, W. Zheng, and J. Mittal, "Balanced protein–water interactions improve properties of disordered proteins and non-specific protein association," J. Chem. Theory Comput. **10**, 5113–5124 (2014).

[26] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, and A. D. MacKerell Jr, "Charmm36m: an improved force field for folded and intrinsically disordered proteins," Nat. Methods **14**, 71 (2017).

[27] P. Robustelli, S. Piana, and D. E. Shaw, "Developing a molecular dynamics force field for both folded and disordered protein states," Proc. Natl. Acad. Sci. U.S.A. **115**, E4758–E4766 (2018).

[28] P. S. Shabane, S. Izadi, and A. V. Onufriev, "General purpose water model can improve atomistic simulations of intrinsically disordered proteins," J. Chem. Theory Comput. **15**, 2620–2634 (2019), pMID: 30865832.

[29] N. Plattner, S. Doerr, G. De Fabritiis, and F. Noé, "Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and markov modelling," Nat. Chem. **9**, 1005 (2017).

[30] A. C. Pan, D. Jacobson, K. Yatsenko, D. Sritharan, T. M. Weinreich, and D. E. Shaw, "Atomic-level characterization of protein–protein association," Proc. Natl. Acad. Sci. U.S.A. **116**, 4244–4249 (2019).

[31] C. Clementi, "Coarse-grained models of protein folding: toy models or predictive tools?" Curr. Opin. Struc. Biol. **18**, 10 – 15 (2008).

[32] Y. Levy, J. N. Onuchic, and P. G. Wolynes, "Fly-casting in protein-dna binding: Frustration between protein folding and electrostatics facilitates target recognition," J. Am. Chem. Soc. **129**, 738–739 (2007).

[33] A. G. Turjanski, J. S. Gutkind, R. B. Best, and G. Hummer, "Binding-induced folding of a natively unstructured transcription factor," PLoS Comput. Biol. **4**, e1000060 (2008).

[34] D. Ganguly and J. Chen, "Topology-based modeling of intrinsically disordered proteins: Balancing intrinsic folding and intermolecular interactions," Proteins **79**, 1251–1266 (2011).

[35] D. De Sancho and R. B. Best, "Modulation of an idp binding mechanism and rates by helix propensity and non-native interactions: association of hif1α with cbp," Mol. BioSyst. **8**, 256–267 (2012).

[36] J. Karanicolas and C. L. Brooks, "The origins of asymmetry in the folding transition states of protein l and protein g," Protein Sci **11**, 2351–2361 (2002).

[37] S. Miyazawa and R. L. Jernigan, "Residue - residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading," J. Mol. Biol. **256**, 623 – 644 (1996).

[38] A. Schug, P. C. Whitford, Y. Levy, and J. N. Onuchic, "Mutations as trapdoors to two competing native conformations of the rop-dimer," Proc. Natl. Acad. Sci. U.S.A. **104**, 17674–17679 (2007).

[39] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, "Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation," J. Chem. Theory Comput. **4**, 435–447 (2008).

[40] T. G. W. Graham and R. B. Best, "Force-induced change in protein unfolding mechanism: Discrete or continuous switch?" J. Phys. Chem. B **115**, 1546–1561 (2011).

[41] R. B. Best and G. Hummer, "Coordinate-dependent diffusion in protein folding," Proc. Natl. Acad. Sci. U. S. A. **107**, 1088–1093 (2010).

[42] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, "The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method," J. Comput. Chem. **13**, 1011–1021 (1992).

[43] I. Lindström, E. Andersson, and J. Dogan, "The transition state structure for binding between TAZ1 of CBP and the disordered HIF-1α CAD," Sci. Rep. **8**, 7872 (2018).

[44] D. Ganguly, W. Zhang, and J. Chen, "Electrostatically accelerated encounter and folding for facile recognition of intrinsically disordered proteins," PLoS Comput. Biol. **9**, 1–12 (2013).

[45] C. Clementi, H. Nymeyer, and J. N. Onuchic, "Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? an investigation for small globular proteins," J. Mol. Biol. **298**, 937 – 953 (2000).

[46] J. K. Noel, M. Levi, M. Raghunathan, H. Lammert, R. L. Hayes, J. N. Onuchic, and P. C. Whitford, "Smog 2: A versatile software package for generating structure-based models," PLoS Comput. Biol. **12**, 1–14 (2016).
[47] M. Gao, J. Yang, S. Liu, Z. Su, and Y. Huang, "Intrinsically disordered transactivation domains bind to taz1 domain of cbp via diverse mechanisms," Biophys .J.l **117**, 1301 – 1310 (2019).
[48] J. M. Rogers, V. Oleinikovas, S. L. Shammas, C. T. Wong, D. De Sancho, C. M. Baker, and J. Clarke, "Interplay between partner and ligand facilitates the folding and binding of an intrinsically disordered protein," Proc. Natl. Acad. Sci. U.S.A. **111**, 15420–15425 (2014).
[49] I. Nyqvist, E. Andersson, and J. Dogan, "Role of conformational entropy in molecular recognition by taz1 of cbp," J. Phys. Chem. B **123**, 2882–2888 (2019).
[50] I. Nyqvist and J. Dogan, "Characterization of the dynamics and the conformational entropy in the binding between TAZ1 and CTAD-HIF-1α," Sci. Rep. **9**, 1–9 (2019).
[51] M. Ubbink, "The courtship of proteins: Understanding the encounter complex," FEBS Lett. **583**, 1060–1066 (2009).
[52] G. Schreiber, G. Haran, and H.-X. Zhou, "Fundamental aspects of protein–protein association kinetics," Chemical Reviews **109**, 839–860 (2009).
[53] 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, "Exploration of the HIF-1α/p300 interface using peptide and Adhiron phage display technologies," Mol. BioSyst. **11**, 2738–2749 (2015).
[54] A. J. Baldwin and L. E. Kay, "NMR spectroscopy brings invisible protein states into focus," Nat. Chem. Biol. **5**, 808 (2009).
[55] M. Brucale, B. Schuler, and B. Samori, "Single-molecule studies of intrinsically disordered proteins," Chem. Rev. **114**, 3281–3317 (2014).
[56] W.-T. Chu, X. Chu, and J. Wang, "Investigations of the underlying mechanisms of HIF-1α and CITED2 binding to TAZ1," bioRxiv , 755074 (2019).