Check for updates

Markov state models and NMR uncover an overlooked allosteric loop in p53†

Emilia P. Barros, [a] Özlem Demir, [a] Jenaro Soto, [b] Melanie J. Cocco [bc] and Rommie E. Amaro *[a]

The tumor suppressor p53 is the most frequently mutated gene in human cancer, and thus reactivation of mutated p53 is a promising avenue for cancer therapy. Analysis of wildtype p53 and the Y220C cancer mutant long-timescale molecular dynamics simulations with Markov state models and validation by NMR relaxation studies has uncovered the involvement of loop L6 in the slowest motions of the protein. Due to its distant location from the DNA-binding surface, the conformational dynamics of this loop has so far remained largely unexplored. We observe mutation-induced stabilization of alternate L6 conformations, distinct from all experimentally-determined structures, in which the loop is both extended and located further away from the DNA-interacting surface. Additionally, the effect of the L6-adjacent Y220C mutation on the conformational landscape of the functionally-important loop L1 suggests an allosteric role to this dynamic loop and the inactivation mechanism of the mutation. Finally, the simulations reveal a novel Y220C cryptic pocket that can be targeted for p53 rescue efforts. Our approach exemplifies the power of the MSM methodology for uncovering intrinsic dynamic and kinetic differences among distinct protein ensembles, such as for the investigation of mutation effects on protein function.

Introduction

The transcription factor p53, known as the "guardian of the genome", is the most important tumor suppressor in humans due to its regulation of a wide range of cellular activities.[1,2] Loss of function through p53 missense mutations is associated with progression of about half of human cancers,[3,4] and reactivation of mutated p53 is emerging as an exciting possibility in cancer treatment as it has been found to lead to tumor regression.[5–9] More than 90% of the cancer mutations are found in the DNA-binding domain (DBD) of p53 [10] (Fig. 1a), but the mechanism through which a single mutation affects function is far from resolved. Moreover, the current paradigm is that p53 mutants are not equivalent proteins, but rather have distinct individual profiles in terms of loss of wildtype activity and acquisition of unique tumor-promoting gain of functions.[11,12] Oncogenic variations can be classified as contact mutations, which lead to loss of function due to disruption of the DNA interaction network,[13] or structural mutations, which cause perturbations

aDepartment of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA, 92093, USA. E-mail: ramaro@ucsd.edu; Fax: +1-858-534-9645; Tel: +1-858-534-9629

bDepartment of Pharmaceutical Sciences, University of California Irvine, Irvine, 92697, USA

cDepartment of Molecular Biology and Biochemistry, University of California Irvine, Irvine, 92697, USA

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0sc05053a

to the DBD and inactivation through destabilization of the protein structure, unfolding and aggregation.[14–18]

A strategy currently pursued for reactivation of structural mutants is the development of small molecules that bind to the folded state of the protein and restore wildtype p53 conformation and function, with promising results achieved by several groups.[19–28] Even in proof-of-concept studies, the success of small molecules in reactivating one or a few specific mutants but not others points to the unique behavior of each p53 cancer mutant. In this way, exploring and characterizing the dynamic behavior of different p53 mutants as individual entities promises to open up novel therapeutic opportunities for mutant-specific p53 reactivation.

One such mutant targeted for small molecule reactivation is Y220C, a structural mutant responsible for about 100 000 new cancer cases every year[14] and the most frequent p53 cancer mutation observed outside the DNA-binding interface of the protein. The mutation of the bulky tyrosine to the smaller cysteine induces the formation of a crevice in the protein surface that is amenable to small molecule binding,[29–31] but so far current efforts have failed to yield very high affinity binders.[32–34]

While the use of molecular dynamics (MD) simulations has allowed the successful identification of druggable pockets on the protein surface of the p53 core domain,[27,33,35] our understanding of the protein conformational ensemble and dynamics is restricted by sampling limitations. This leaves large regions of the energy landscape unexplored which may include many of
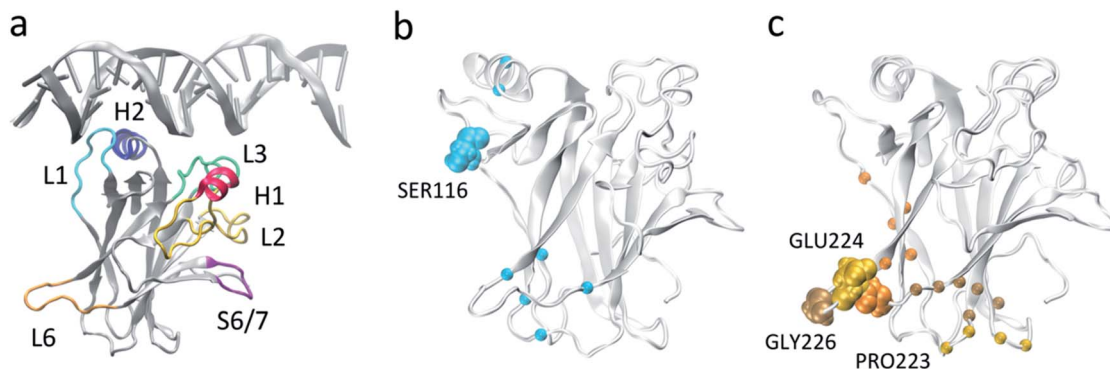
Fig. 1 p53 DNA binding domain (a) monomeric p53 DNA-binding domain in complex with DNA (from PDB 1TSR) with important functional regions highlighted. (b) and (c) Residues used for MSM construction based on pairwise distances, with L1 (b) and L6 (c) anchor residues highlighted in VDW representation. The Cα carbons of the residues that were selected as the second member of the pair with the respective anchor are represented as spheres.

the functionally important slower motions. Already, relatively short-scale MD simulations of Y220C have evidenced the flexibility of the protein and the Y220C pocket.[33] A comprehensive model of p53's conformational ensemble and the underlying free energy landscape is desirable as it will allow the understanding of the dynamics of key loops and druggable pockets and their role in the overall function and motions of the protein. To help overcome this sampling limitation, we employ here the Markov State Model (MSM) methodology in conjunction with extensive MD simulations for the investigation of the conformational dynamics of wildtype p53 and the Y220C mutant.

MSMs allow the integration of multiple MD simulations into a single model of the protein conformational ensemble that contains key thermodynamic and kinetic properties in addition to retaining the atomic level details of the system.[36–43] Because the MSM is built on the transitions between states, the information from multiple MD simulations of the same system can be combined into a single model and no single simulation has to explore all the states. Importantly, as the equilibrium distribution of states can be derived from the final model, the thermodynamics of the states can be determined, in addition to kinetics, principal motions, and transition pathways of the protein conformational ensemble.

Our computational models, followed by experimental validation with NMR relaxation studies, allow for the first time a thorough exploration of the conformational ensemble of p53 DBD and uncovers the involvement of a loop located away from the DNA binding interface, L6 (residues 221–230, also termed S7/S8 loop), in the slowest dynamics of the wildtype protein. This loop is adjacent to the Y220C mutation, but our models indicate that the mutation affects the conformational landscape of not only L6 but also of the essential DNA-interacting L1 loop. The existence of allosteric communication between the two loops is suggested and provides a mechanistic rationale to the effect of the mutation on the activity of p53. Moreover, analysis of the conformational diversity of L6 evidences the existence of very distinct loop conformations than previously observed experimentally, and the identification of a novel cryptic pocket

nestled in the extended conformation of L6 that could be exploited for mutant-specific drug design efforts. Our work emphasizes the ability of MSMs to explore in detail protein conformational landscapes, uncover hidden states inaccessible to experiments and inform on mutation or other environmental effects (such as ligand binding or pH) on protein function.

Results and discussion

L6 is the slowest loop in p53 DBD dynamics

Markov state models provide a framework for exploring protein dynamics with atomic resolution beyond the timescales typically accessed by molecular dynamics simulations. A crucial step when integrating molecular dynamics trajectories for model building is the selection of features used to discretize the protein conformations sampled, which decreases the dimensionality of the conformational space while still allowing for discrimination between distinct states and appropriate representation of the relevant motions. For a general understanding of the protein conformational ensemble, the task can become challenging due to the conflict between the large degrees of freedom required to describe the protein ensemble and the need to limit the feature dimension to a small, tractable number for model building.

To investigate the basal dynamics of wildtype p53, we employed an unbiased method that started from computing all possible pairwise distances (18 336 features), and iteratively performed time-lagged Independent Components Analysis (tICA)[44] to identify the linear combination of features that describe the slowest motions of the system, followed by elimination of the features with low tICA correlation (ESI Table 1†). Using this methodology we arrived at a final number of 24 pairs (ESI Table 2†). tICA is useful in the data processing for MSM construction as it maximizes the feature combination to yield kinetically relevant independent components (tICs) representing the slowest degrees of freedom in the system. Despite starting from all possible pairwise distances and including no directed selection of features besides the elimination of pairs that involve the clipped terminal residues or that are

consistently too close (<3 Å) or too far (>10 Å) throughout the whole simulations, the final set consisted of interacting pairs centered around loops L1 (residues 113–123) and L6 (residues 221–230): all pairs involved at least one residue located in either L1 (Ser116) or L6 (Pro223, Glu224, Gly226), hereafter referred to as L1 and L6 anchor residues, respectively (Fig. 1b and c, ESI Table 2†). While methods are available for the identification of relevant features among different input options,[45] these still rely on user-defined feature candidates. The iterative method applied here provides an alternative approach to arrive at kinetically-relevant input features without any *a priori* knowledge of the system's dynamics which could be relevant for many other systems and applications of MSMs.

The presence of the repeated anchor residues in the final feature pairs suggests that loops L1 and L6 are involved in the slowest and most significant motions of the protein. Loop L1 is known as a dynamic and biologically important motif for p53 function, having been observed experimentally and computationally in two very distinct extended (Fig. 1a) and recessed conformations.[46–48] The identification of the relevance of loop L6, on the other hand, sheds a light on a relatively unexplored region of p53. Not much attention has been given to the role of this structural motif, possibly because of its distance from the DNA-binding surface. However, elevated B factors in p53 crystal structures and NMR NOE values[49] point to its intrinsic dynamics, and flexibility in this loop was observed in an early short simulation of wildtype p53, even though implications for functionality were not explored as the motion was deemed to stem from a lack of crystal packing.[50] Additionally, comparison of our wildtype ensemble's dihedral angle and distance distributions to the J-coupling and NOE derived restraints from the single solution NMR structure (PDB ID 2FEJ[51]) evidenced that the great majority of distances do not violate the NOE upper distance restraints (ESI Fig. 1†) and that average dihedral angles differ from the experimental values for no more than 23% of the dihedrals, further supporting the accuracy of our simulations.

For a comparison of the conformational landscapes of wildtype and Y220C, the conformations explored by each of the simulations and represented by the 24 features were jointly used as input for tICA, and the resulting free energy landscapes are shown in Fig. 2a. Processing the trajectories together (as described by the selected features) with tICA ensures that the novel coordinate space is equivalent for the wildtype and mutant data, allowing for a direct comparison of the conformational ensemble explored by each system. The wildtype simulation presents two preferred states, corresponding to the minima in the free energy landscape. The main distinction between them are the conformations of L1 and indicate the
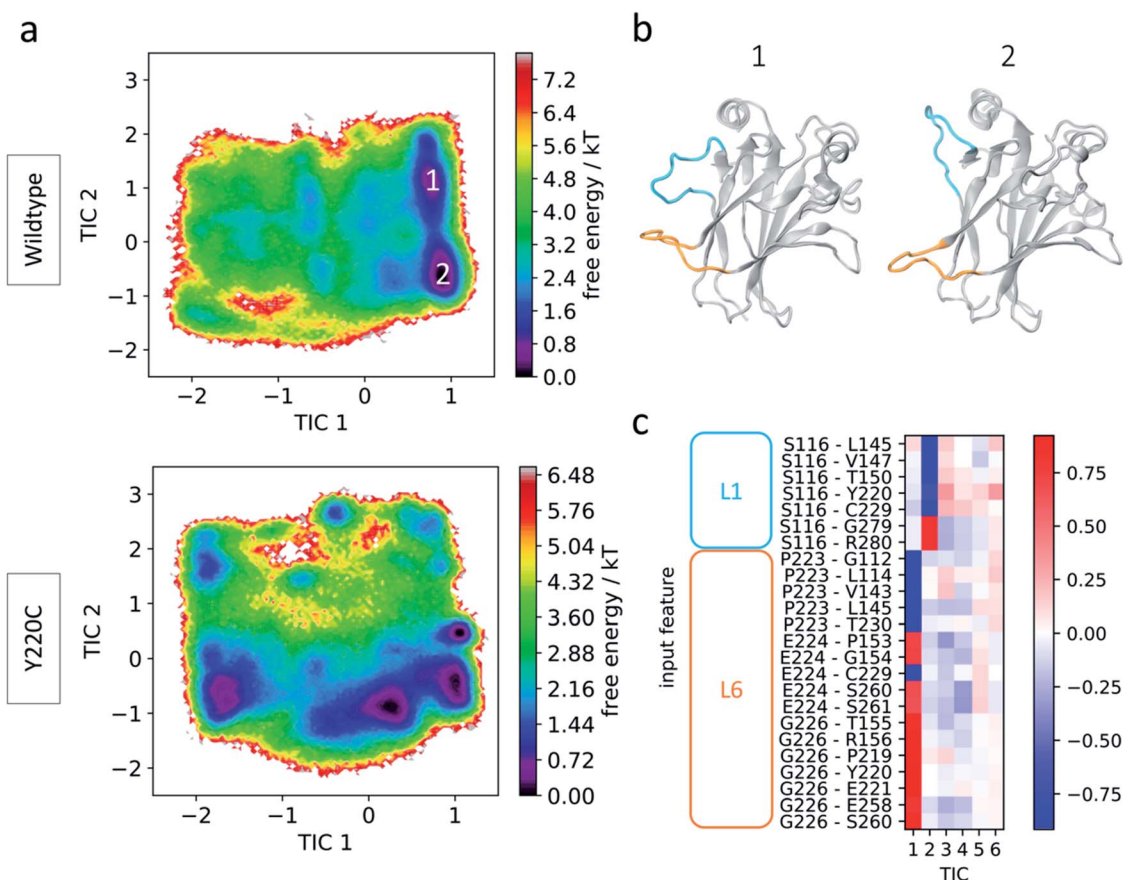

Fig. 2 Wildtype and Y220C simulation results. (a) Free energy landscape of wildtype (top) and Y220C (bottom) in terms of tICA components (tICs). (b) Representative conformations from the wildtype preferred states. Loops L1 and L6 are highlighted in blue and orange, respectively. (c) Feature correlation with the first five tICA components. Pairwise distances involving L1 or L6 loop anchor residues are indicated.

same recessed and extended L1 conformations that have been previously observed (Fig. 2b). Interestingly, the pairwise features used for construction of the map align with the tICA components in this transformed coordinate space, permitting a direct interpretation in terms of protein conformation: tIC1 is closely correlated with features that include L6 anchors, and tIC2 is more closely correlated with features involving the L1 anchor, Ser116 (Fig. 2c), such that tIC1 and 2 are associated with the relative motions of L6 and L1, respectively. Moreover, visual inspection of the conformations distributed on the free energy landscape evidence that smaller values of each of the tICs describe conformations with extended loops, while larger values describe recessed loop conformations (novel conformations are discussed in more detail in following sections).

Since the tICs are ordered in terms of slowest to fastest motions, the correspondence of L6 anchor features with the first of the components indicates that transitions involving loop L6 are slower than those for loop L1. To further check the importance of these loops in the relevant motions of the

protein, we performed additional tICA analysis specifically incorporating other motifs known to play significant roles in p53 function: helices H1 and H2 and loops L2 and L3, which together with L1 make up the DNA interaction surface, and loop S6/7, recently identified as a flexible region in p53 mutants[52] (Fig. 1a). Even though several of these loops show pronounced flexibility in the simulations as indicated by Cα RMSF values (ESI Fig. 2†), loops L1 and particularly L6 still dominate the slowest transitions (ESI Fig. 3†). This suggests that, while other regions such as loops L2 and S6/7 may be highly flexible as evidenced by their high RMSF values, they display fast dynamics and act as further evidence to the important role of L6 on the slow dynamics of p53.

Allosteric communication between L1 and L6

In Fig. 2a it can be seen that the Y220C mutation affects not only the conformational landscape of loop L6, where it is located, but also of loop L1 (as represented by tIC 2). This loop is essential
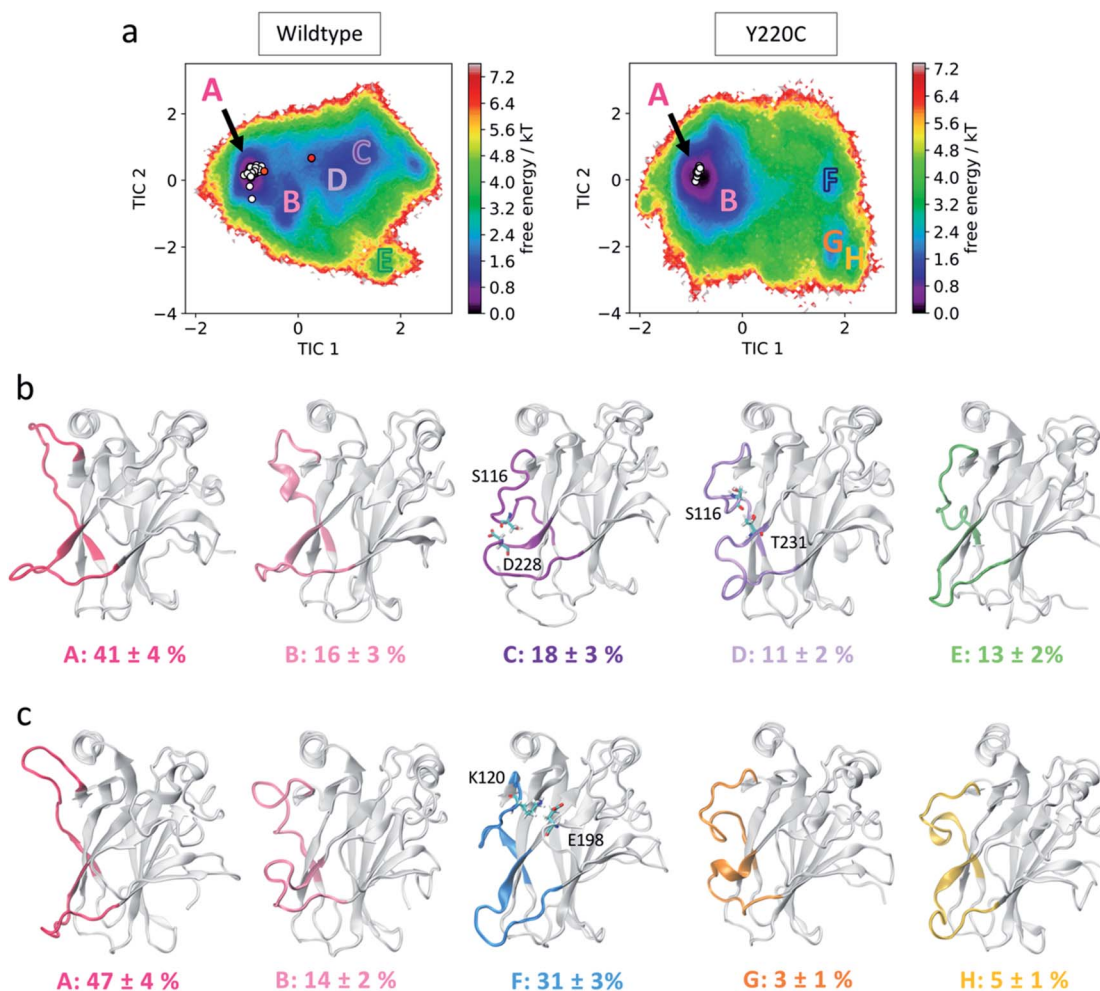

Fig. 3 L1-centered MSM (a) free energy landscape of wildtype (left) and Y220C (right) in terms of the features that describe L1 relative dynamics. Location of metastable states are indicated with letters from A to H. Experimentally resolved DBD structures are indicated as white (extended L1 conformation) and red (recessed L1) circles. (b) Conformations from each of the wildtype metastable states. Equilibrium populations and standard deviations are indicated. Residues identified in key interactions are highlighted. (c) Conformations from each of the Y220C metastable states.

for p53 activity as it is involved in key interactions with DNA through hydrogen bonds formed by Lys120 and Ser121.[48] Wildtype p53 shows important intrinsic L1 flexibility, but the effect of the mutation on loop L1's dynamics indicates the existence of possible long-range communication between L1 and L6.

To look into this in more detail, we constructed MSMs for the wildtype and mutant system using only the above identified features that include the L1 anchor, Ser116. The conformational landscape in terms of these 7 features, following tICA transformation, is shown in Fig. 3a, and includes the coordinates of experimentally-characterized wildtype and Y220C p53 structures for comparison (ESI Tables 3 and 4†). Coarse-graining of the structures using Hidden Markov state models identifies the presence of 5 metastable states in each case based on the relative separation of the slowest relaxation timescales in the implied timescale plot (see Methods section, ESI†). Two wildtype metastable states, states A and B, are retained in the mutant system with slight changes to their equilibrium populations (Fig. 3b and c). State A is the most populated state in both systems, and shows loop L1 in the most extended-like conformations, in agreement with the experimentally-determined structures. In wildtype state B, we see a previously-identified 3–10 helix in the L1 loop, absent in the corresponding Y220C state.

The second, shallower wildtype minima, centered at TIC1 = 1, is absent in the Y220C sampled conformations. Indeed, we find that two wildtype metastable states are abrogated by the mutation (states C and D), being substituted by a single state in the Y220C system (state F). These wildtype states show L1 in recessed conformations, and jointly account for 29% of the equilibrium population. Interestingly, in both cases we find that L6 is also organized in a recessed conformation, such that both loops are located in close proximity to each other. Investigation of the loop residues indicates inter-loop hydrogen bonds formed between the side-chain oxygen of Ser116 in L1 and backbone nitrogen of Asp228 (in state C) or side-chain oxygen of Thr231 (state D) in L6 (Fig. 3b and ESI Fig. 4†).

Loop L1 in the corresponding Y220C state F, on the other hand, is found to be more collapsed into the protein surface, in a conformation that does not allow for interaction with L6. Rather, a salt bridge between Lys120 in L1 and Glu198 in loop S5/S6 seems to promote the stabilization of this alternate conformation, which accounts for 31% of the Y220C equilibrium population and is the second most populated Y220C state (Fig. 3c). The sequestering of the DNA-interacting Lys120 in this significant metastable state could provide a mechanistic explanation to the p53 inactivation effect of the mutation. Furthermore, the conformation-dependent interaction between L1 and L6 identified here suggests the existence of an allosteric communication between them in functional p53, which is disrupted by the mutation.

Finally, we observe a slight destabilization of states located at low values of TIC 2 in the Y220C system, which display loop L1 in extremely-recessed conformations (equilibrium population of 13% for wildtype state E and 8% for Y220C states G and H). There are no persistent L1–L6 interactions in these states. A helical content in loop L6 of Y220C state G seems to be promoted by an inter-L6 hydrogen bond between Ser227 and Thr231.

Dynamics and druggability of loop L6

The significance of L6 dynamics suggested by the tICA analysis and its effect on the L1 conformational ensemble of wildtype and Y220C prompted us to consider its conformational plasticity in more detail. Fig. 4a shows the free energy landscape of the wildtype and Y220C systems now in terms of the tICA components calculated from the 17 L6 anchor features. Again for comparison, we overlay the corresponding coordinates of the X-ray structures of wildtype and Y220C p53. It is striking how all the previously identified structures are confined to a small area of the map, and the simulations suggest the existence of novel significant protein conformations that remain unexplored to date and could be potentially targeted for drug discovery.

While the experimental structures align with the wildtype low energy well, the mutation leads to the stabilization of multiple alternative loop L6 conformations, including one mutant-exclusive well at high values of tIC2 that is distinct from the experimentally characterized structures. Five metastable states can be identified from the Markov state models for each of the wildtype and Y220C systems. Two populated wildtype metastable states at equilibrium remain significant states in the Y220C ensemble, albeit with changes to their relative equilibrium population and rates of transitions. Three wildtype states are abrogated by the mutation, while we observe the formation of two Y220C-exclusive metastable states. The conformational differences between the highly populated states and potential for drug discovery are explored in more detail below.

The mutation induces stabilization of extended L6 conformations. The most populated metastable state in the wildtype ensemble, accounting for over 58% of the population at equilibrium, corresponds to L6 in a recessed conformation similar to that observed experimentally (Fig. 4b and c, pink R state). This organization of the loop allows for the formation of a crevice in between loops L6 and S3/S4 upon the substitution of the bulky tyrosine for the much smaller cysteine residue in the Y220C mutant, which results in the pocket currently targeted for p53 rescue.[29–34]

In several of the mutant frames belonging to this metastable state we observed the opening of a transient channel through loop L6, connecting the crevice to another area of the protein surface. This cryptic pocket has been identified previously by Fersht and co-workers using molecular dynamics simulations,[33] and in agreement with their studies, we find it to exhibit promising druggable characteristics (as suggested by FTMap[53] solvent mapping analysis, ESI Fig. 5†). Exploitation of this channel by small molecules could improve the potency of rescue drugs and increase specificity towards mutant p53, as the channel is unavailable in the wildtype simulations due to the larger volume occupied by the tyrosine residue.

Besides this well-characterized state, the MSMs indicate an additional common metastable state in the wildtype and Y220C
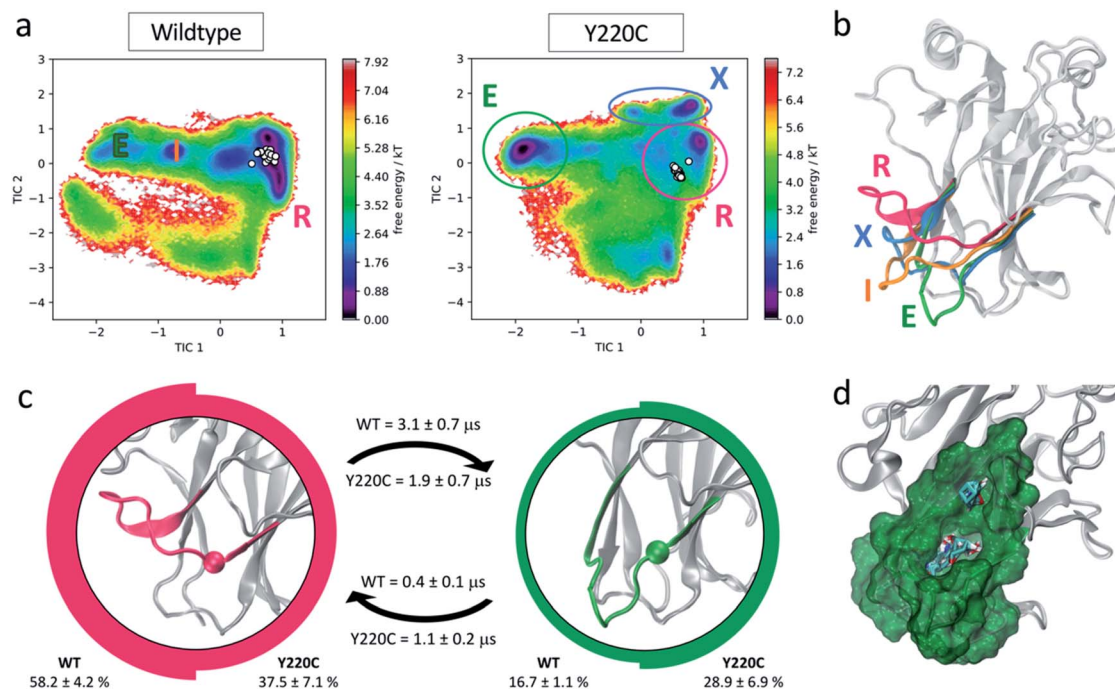
Fig. 4 L6-centered MSM. (a) Free energy landscape of wildtype and Y220C systems in terms of L6 features. Experimentally resolved DBD structures are indicated as white circles. Populated metastable states are identified. (b) Structural representation of the metastable states identified in the free energy landscapes: recessed (R, pink), intermediate (I, orange), extended (E, green) and mutant exclusive (X, blue) conformations. (c) Equilibrium population and mean first passage times (MFPT) for the two major wildtype and Y220C metastable states. The images at the center of the circles represent the respective state L6 conformation, with the α carbon of the mutated residue highlighted. Thickness of the circle edge is proportional to the equilibrium population in the respective system (wildtype on the left, Y220C on the right). MFPTs of the transitions are indicated above (for wildtype) and below (for Y220C) the respective arrows. (d) Representation of the novel pocket found in the L6 extended conformation and solvent mapping results.

ensemble at equilibrium. This metastable state, corresponding to 16.7% of the wildtype population and 28.9% of the Y220C ensemble, exhibits loop L6 in a dramatically different extended conformation (green E state in Fig. 4b and c). In this conformation, the crevice underneath L6 typically targeted by small molecules for Y220C reactivation is disrupted. However, visual inspection identified the formation of another cryptic pocket nestled within this loop, promoted by the extended conformation of L6 (Fig. 4d). Similar to the mutant-induced crevice, this pocket is only evident in the Y220C simulations due to the presence of the less bulky cysteine in its center. The entrance of the cavity in this case faces the opposite side of the loop relative to the known Y220C crevice, in the direction of the DNA binding surface, and the cavity corresponds to a relatively deep hydrophobic pocket with average volume of 333.5 ± 57.6 Å^3 with opportunities for hydrogen bonding interaction, as well as other polar interactions in the more solvent-exposed region above L6. Analysis of the cryptic pocket using FTMap,[53] a computational solvent mapping software that distributes small organic molecule probes in the protein surface, finds that the novel pocket is a consensus site, binding 30 different clusters of drug-like probes and characterizing as a binding energy hot spot (Fig. 4d). FTMap has been benchmarked against experimental data and is thus taken as a good measure of the druggability potential of a protein pocket.[54] Indeed, cryptic pockets found

close to binding energy hotspots, which can accommodate diverse small molecule fragments, have increased potential of being druggable and affect protein function when targeted.[55]

Several hydrogen bonds between loops L6 and S3/S4 (residues 146–155) are found to be established for longer fractions of the simulation in the mutant state, with increases of up to 100× in persistence time, and suggest possible interactions promoting the extended conformation (ESI Table 5†). Further indication of the stabilization of the extended conformation promoted by the mutation is given by the calculation of mean first passage times (MFPT) between these metastable states: the mutation decreases the mean first passage time from the recessed to the extended L6 conformation by a factor of 1.6, resulting in a faster transition in the Y220C mutant compared to the wildtype, while the MFPT out of the extended conformation and into the recessed increases by more than 2 in the Y220C mutant (Fig. 4c).

Finally, the third significantly-populated state in the wildtype ensemble, with a stationary population of 18.3% and corresponding to an intermediate state between the extended and recessed conformations (Fig. 4b) is completely abrogated in the Y220C ensemble, such that the recessed–extended transition occurs without an intermediate state for the mutant.

Characterization of mutant-exclusive metastable states. The long-timescale exploration of the Y220C mutant dynamics

evidenced the existence of two mutant-exclusive states (Fig. 4a). Jointly, these metastable states account for 24.7% of the relative Y220C ensemble population, a significant portion of the conformational ensemble that opens up promising avenues for mutant-specific therapeutic opportunities. In these states the loop L6 shows a similar extended conformation to the novel metastable state E described above, but with a "sideways" bend likely promoted by a Thr149–Pro222 interaction (Fig. 4b, ESI Fig. 6a†). This bend slightly disrupts the cryptic pocket identified in the fully extended L6 conformation, resulting in a smaller and shallower cavity, but also leads to the formation of a channel across loop L6 and underneath the mutation which reaches across to the protein surface at a different site (ESI Fig. 6b†). Transitions into or out of these states constitute the slowest process in the Y220C MSM, with a timescale of approximately 7.3 ± 2.8 µs.

Taken together, our models suggest a molecular explanation for the reactivation of the Y220C mutant achieved by small molecules:[30–34] since in the mutant the recessed L6 conformation is destabilized (37.5% of the Y220C population *versus* 58.2% for wildtype) with a preference for the novel E and X extended conformations, binding of a small molecule into the crevice underneath L6 should prevent the transition towards these extended conformations and could lead to a shift in the equilibrium towards a wildtype-like, recessed loop conformational ensemble. Additionally, as the investigation of the full p53 conformational flexibility suggests a high degree of correlation between L1 and L6 dynamics (Fig. 2 and 3), this could further indicate a previously uncharacterized functional link between L6 conformation and p53 rescue.

NMR relaxation analysis

As an external validation of the loop dynamics identified by the MSMs, we performed NMR relaxation studies to determine flexibility of backbone atoms of the wildtype protein. Measurement of ^{15}N NOE, longitudinal (R_1), and transverse (R_2) relaxation rates were used to obtain generalized order parameters (S^2). The list of relaxation rates and NOEs are found in ESI Table 6.† Relaxation rates for some residues could not be obtained due to rapid signal decay (not enough points to fit) or significant signal overlap. We used the program *Modelfree* to determine backbone flexibility based on heteronuclear NOE, R_1, R_2 measurements.[56,57] Using the *quadric_diffusion* program[58] we found that the best fitted diffusion tensor model was an axial symmetry model. The rotational correlation time (τ_m) was calculated to be 14.7 ns with an axially symmetric tensor ($D_{\parallel/\perp}$) = 0.29.

Calculated order parameters show that the most flexible regions are in the loop regions (Fig. 5a), and that L6 is the most flexible loop. A similar trend in wildtype backbone order parameters has been previously observed,[49] although our values are larger in magnitude, possibly due to differences in magnetic field strength in which the experiments were performed. Calculation of generalized order parameters from the MD simulations was not possible as µs-timescale, continuous sampling is required for accurate estimation of values[59,60] and our MSM approach consisted of multiple, short-timescale
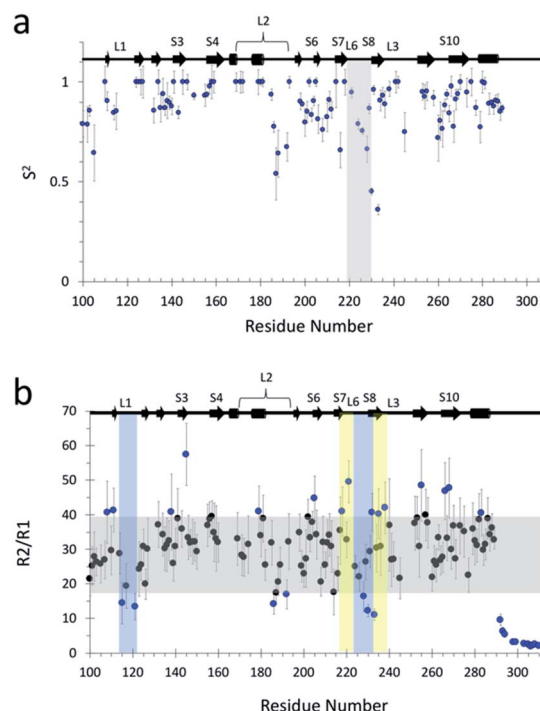

Fig. 5 (a) Quantitative characterization of fast dynamics of wildtype p53 DBD. Generalized order parameter (S^2) for wildtype obtained *via* *Modelfree* analysis[56,57] of relaxation rates R_1, R_2, and $^1H-^{15}N$ NOEs at 800 MHz. Vertical boxed area highlights L6. (b) R_2/R_1 plot for qualitative analysis of backbone dynamics. Grey rectangles highlight mean values ± 1 SD of all R_2/R_1. Blue data points highlight the residues that are outside the mean ± 1 SD. Yellow vertical shaded rectangle highlights regions of interest of slow dynamics (µs–ms). Similarly, blue shaded rectangle highlights regions of interest of fast dynamics (ps–ns).

simulations. Additionally, macrostate-specific generalized order parameters would be difficult to calculate as, even though new simulations could be seeded from the MSM macrostates, there is no guarantee that other macrostates would not be visited during the course of µs-long simulations, and thus the results would likely be inaccurate and affected by noise.

Regarding back-calculating chemical shifts from the simulations and comparing to NMR results, we do not believe that they are a reliable metric for comparison with the MD ensemble – despite being exquisitely sensitive to structure, they are not unique solutions and thus in principle thousands of conformations and arrangements of atoms could give the same value for the chemical shift. Instead, additional experimental validation is available in the form of the measured R_2/R_1 ratio, which provides a qualitative indication of the timescales of motions involving backbone residues. NMR R_2/R_1 ratio approximates the correlation time of the ^{15}N nucleus for each backbone position; these times are fast, usually in the tens of ns for proteins.[61] Residues that surpass the mean and standard deviation indicate areas of motion with µs–ms timescale, while residues that deviate negatively from the mean by more than the standard deviation indicate areas with ps–ns timescales.[49,62] The R_2/R_1 results indicate that L6 dynamics contain a slow-motion

component that occurs at longer timescales (μs–ms) than that of L1 (ps–ns) (Fig. 5b), in agreement with our findings.

Conclusions

Our combined tICA and MSM approach, validated by NMR relaxation measurements, highlights a functional role to the dynamic loop L6, which exhibits motions at longer timescales than other characterized structural motifs and presents potential for mutant-rescue therapeutic opportunities. The conformational landscape suggests some degree of allostery between L6 and the functionally-important loop L1, likely promoted by hydrogen bonds formed when both loops are in the recessed conformation and thus in close proximity to each other.

The Y220C mutation, which characterizes one of the most common cancer mutants, is located at the N terminus of L6, and we find that the mutation promotes the stabilization of novel protein conformations which exhibit loop L6 in extended states instead of the only currently characterized and targeted recessed conformation. The stabilization of the extended conformation induces the formation of a deep hydrophobic pocket within L6 due to the removal of the bulky tyrosine, as well as the population of two mutant-exclusive L6 states that could be explored for mutant-specific therapies. Computational simulations and virtual screening methods are powerful approaches to aid the discovery and design of allosteric ligands,[63–65] and will be explored to target the identified cryptic pocket. An interesting approach, Boltzmann docking, takes advantage of the different representative metastable state conformations identified in MSMs to obtain Boltzmann-weighted averages of the docking score.[66] The allosteric communication within the DBD can be also further investigated using other recent computational approaches such as the energy decomposition method[67–69] and the dynamical network analysis.[70]

In summary, the comparison of the dynamics of wildtype and mutant p53 DBD's using MD simulations and Markov state models evidenced for the first time the existence of functionally-relevant motions involving loop L6 and presents applications for mutant-specific rescue efforts. We anticipate that this approach will be useful in the study of the conformational ensembles of other p53 cancer mutants or protein targets, as a way to provide atomic-level information on these proteins' motions combined with thermodynamic and kinetic details in tandem with experimental observations.

Author contributions

E. P. B built Markov state models. E. P. B. and O. D. performed simulations and interpreted the models. J. S. performed NMR data collection and relaxation analysis. M. J. C oversaw the experimental aspects of the work, and R. E. A. devised and oversaw the computational components. E. P. B. wrote the paper with edits from all authors.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 K. T. Bieging and L. D. Attardi, *Trends Cell Biol.*, 2012, **22**, 97–106.
2 A. Lujambio, L. Akkari, J. Simon, D. Grace, D. F. Tschaharganeh, J. E. Bolden, Z. Zhao, V. Thapar, J. A. Joyce, V. Krizhanovsky and S. W. Lowe, *Cell*, 2013, **153**, 449–460.
3 M. Olivier, R. Eeles, M. Hollstein, M. A. Khan, C. C. Harris and P. Hainaut, *Hum. Mutat.*, 2002, **19**, 607–614.
4 T. Soussi and C. Béroud, *Nat. Rev. Cancer*, 2001, **1**, 233–240.
5 A. Ventura, D. G. Kirsch, M. E. Mclaughlin, D. A. Tuveson, J. Grimm, L. Lintault, J. Newman, E. E. Reczek, R. Weissleder and T. Jacks, *Nature*, 2007, **445**, 661–665.
6 A. Parrales and T. Iwakuma, *Front. Oncol.*, 2015, **5**, 288.
7 C. P. Martins, L. Brown-Swigart and G. I. Evan, *Cell*, 2006, **127**, 1323–1334.
8 G. Selivanova and K. G. Wiman, *Oncogene*, 2007, **26**, 2243–2254.
9 W. Xue, L. Zender, C. Miething, R. A. Dickins, E. Hernando, V. Krizhanovsky, C. Cordon-Cardo and S. W. Lowe, *Nature*, 2007, **445**, 656–660.
10 W. A. Freed-Pastor and C. Prives, *Genes Dev.*, 2012, **26**, 1268–1286.
11 P. A. J. Muller and K. H. Vousden, *Cancer Cell*, 2014, **25**, 304–317.
12 K. Sabapathy and D. P. Lane, *Nat. Rev. Clin. Oncol.*, 2018, **15**, 13–30.
13 A. Eldar, H. Rozenberg, Y. Diskin-posner, R. Rohs and Z. Shakked, *Nucleic Acids Res.*, 2013, **41**, 8748–8759.
14 A. C. Joerger and A. R. Fersht, *Oncogene*, 2007, **26**, 2226–2242.
15 A. N. Bullock, J. Henckel and A. R. Fersht, *Oncogene*, 2000, **19**, 1245–1256.
16 Ö. Demir, R. Baronio, F. Salehi, C. D. Wassman, L. Hall, G. W. Hatfield, R. Chamberlin, R. H. Lathrop and R. E. Amaro, *PLoS Comput. Biol.*, 2011, 7, e1002238.
17 R. Wilcken, G. Z. Wang, F. M. Boeckler and A. R. Fersht, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 13584–13589.
18 G. Z. Wang and A. R. Fersht, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, E2634–E2643.
19 V. J. N. Bykov, N. Issaeva, N. Zache, A. Shilov, M. Hultcrantz, J. Bergman, G. Selivanova and K. G. Wiman, *J. Biol. Chem.*, 2005, **280**, 30384–30391.
20 N. Beraza and C. Trautwein, *Hepatology*, 2007, **45**, 1578–1579.
21 N. Zache, J. M. R. Lambert, K. G. Wiman and V. J. N. Bykov, *Cell. Oncol.*, 2008, **30**, 411–418.

22 N. Zache, J. M. R. Lambert, N. Rökaeus, J. Shen, P. Hainaut, J. Bergman, K. G. Wiman and V. J. N. Bykov, *Mol. Oncol.*, 2008, **2**, 70–80.

23 C. J. Brown, S. Lain, C. S. Verma, A. R. Fersht and D. P. Lane, *Nat. Rev. Cancer*, 2009, **9**, 862–873.

24 X. Yu, A. Vazquez, A. J. Levine and D. R. Carpizo, *Cancer Cell*, 2012, **21**, 614–625.

25 S. Lehmann, V. J. N. Bykov, D. Ali, O. Andrén, U. Tidefelt, B. Uggla, J. Yachnin, G. Juliusson, A. Moshfegh, C. Paul, K. G. Wiman and P. O. Andersson, *J. Clin. Oncol.*, 2012, **30**, 3633–3639.

26 X. Liu, R. Wilcken, A. C. Joerger, I. S. Chuckowree, J. Amin, J. Spencer and A. R. Fersht, *Nucleic Acids Res.*, 2013, **41**, 6034–6044.

27 C. D. Wassman, R. Baronio, Ö. Demir, B. D. Wallentine, C.-K. Chen, L. V Hall, F. Salehi, D. Lin, B. P. Chung, G. W. Hatfield, A. R. Chamberlin, H. Luecke, R. H. Lathrop, P. Kaiser and R. E. Amaro, *Nat. Commun.*, 2013, **4**, 1407.

28 D. Russo, L. Ottaggio, G. Foggetti, M. Masini, P. Masiello, G. Fronza and P. Menichini, *Biochim. Biophys. Acta*, 2013, **1833**, 1904–1913.

29 A. C. Joerger, H. C. Ang and A. R. Fersht, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 15056–15061.

30 F. M. Boeckler, A. C. Joerger, G. Jaggi, T. J. Rutherford, D. B. Veprintsev and A. R. Fersht, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 10360–10365.

31 N. Basse, J. L. Kaar, G. Settanni, A. C. Joerger, T. J. Rutherford and A. R. Fersht, *Chem. Biol.*, 2010, **17**, 46–56.

32 R. Wilcken, X. Liu, M. O. Zimmermann, T. J. Rutherford, A. R. Fersht, A. C. Joerger and F. M. Boeckler, *J. Am. Chem. Soc.*, 2012, **134**, 6810–6818.

33 A. C. Joerger, M. R. Bauer, R. Wilcken, F. M. Boeckler, J. Spencer and A. R. Fersht, *Struct. Des.*, 2015, **23**, 2246–2255.

34 M. R. Bauer, R. N. Jones, R. K. Tareque, B. Springett, F. A. Dingler, L. Verduci, K. J. Patel, A. R. Fersht, A. C. Joerger and J. Spencer, *Future Med. Chem.*, 2019, **11**, 2491–2504.

35 Ö. Demir, P. U. Ieong and R. E. Amaro, *Oncogene*, 2017, **36**, 1451–1460.

36 D. Shukla, C. X. Hernández, J. K. Weber and V. S. Pande, *Acc. Chem. Res.*, 2015, **48**, 414–422.

37 V. S. Pande, K. Beauchamp and G. R. Bowman, *Methods*, 2010, **52**, 99–105.

38 G. R. Bowman, V. S. Pande and F. Noe, *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, Springer, 2014.

39 J. D. Chodera and F. Noe, *Curr. Opin. Struct. Biol.*, 2014, **25**, 135–144.

40 G. R. Bowman, E. R. Bolin, K. M. Hart, B. C. Maguire and S. Marqusee, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 2734–2739.

41 F. Pontiggia, D. V. Pachov, M. W. Clarkson, J. Villali, M. F. Hagan, V. S. Pande and D. Kern, *Nat. Commun.*, 2015, **6**, 7284.

42 P. Wapeesittipan, A. S. J. S. Mey, M. D. Walkinshaw and J. Michel, *Commun. Chem.*, 2019, **2**, 41.

43 J. Juárez-Jiménez, A. A. Gupta, G. Karunanithy, A. S. J. S. Mey, C. Georgiou, H. Ioannidis, A. De Simone, P. N. Barlow, A. N. Hulme, M. D. Walkinshaw, A. J. Baldwin and J. Michel, *Chem. Sci.*, 2020, **11**, 2670–2680.

44 G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis and F. Noé, *J. Chem. Phys.*, 2013, **139**, 015102.

45 H. Wu and F. Noé, *J. Nonlinear Sci.*, 2020, **30**, 23–66.

46 T. J. Petty, S. Emamzadah, L. Costantino, I. Petkova, E. S. Stavridi, J. G. Saven, E. Vauthey and T. D. Halazonetis, *EMBO J.*, 2011, **30**, 2167–2176.

47 S. Emamzadah, L. Tropia and T. D. Halazonetis, *Mol. Cancer Res.*, 2011, **9**, 1493–1500.

48 S. Lukman, D. P. Lane and C. S. Verma, *PLoS One*, 2013, **8**, e80221.

49 J. A. Rasquinha, A. Bej, S. Dutta and S. Mukherjee, *Biochemistry*, 2017, **56**, 4962–4971.

50 Q. Lu, Y. H. Tan and R. Luo, *J. Phys. Chem. B*, 2007, **111**, 11538–11545.

51 J. M. Pérez Cañadillas, H. Tidow, S. M. V. Freund, T. J. Rutherford, H. C. Ang and A. R. Fersht, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 2109–2114.

52 M. R. Pradhan, J. W. Siau, S. Kannan, M. N. Nguyen, Z. Ouaray, C. K. Kwoh, D. P. Lane, F. Ghadessy and C. S. Verma, *Nucleic Acids Res.*, 2019, **47**, 1637–1652.

53 D. Kozakov, L. E. Grove, D. R. Hall, T. Bohnuud, S. Mottarella, L. Luo, B. Xia, D. Beglov and S. Vajda, *Nat. Protoc.*, 2015, **10**, 733–755.

54 D. Kozakov, D. R. Hall, R. L. Napoleon, C. Yueh, A. Whitty and S. Vajda, *J. Med. Chem.*, 2015, **58**, 9063–9088.

55 A. Kuzmanic, G. R. Bowman, J. Juarez-Jimenez, J. Michel and F. L. Gervasio, *ACS Chem. Res.*, 2020, **53**, 654–661.

56 A. M. Mandel, M. Akke and A. G. Palmer, *J. Mol. Biol.*, 1995, **246**, 144–163.

57 A. G. Palmer, M. Rance and P. E. Wright, *J. Am. Chem. Soc.*, 1991, **113**, 4371–4380.

58 L. K. Lee, M. Rance, W. J. Chazin and A. G. Palmer III, *J. Biomol. NMR*, 1997, **9**, 287–298.

59 P. Maragakis, K. Lindorff-Larsen, M. P. Eastwood, R. O. Dror, J. L. Klepeis, I. T. Arkin, M. Jensen, H. Xu, N. Trbovic, R. A. Friesner, A. G. Palmer and D. E. Shaw, *J. Phys. Chem. B*, 2008, **112**, 6155–6158.

60 G. R. Bowman, *J. Comput. Chem.*, 2016, **37**, 558–566.

61 L. E. Kay, D. A. Torchia and A. Bax, *Biochemistry*, 1989, **28**, 8972–8979.

62 A. Friedler, B. S. DeDecker, S. M. V. Freund, C. Blair, S. Rüdiger and A. R. Fersht, *J. Mol. Biol.*, 2004, **336**, 187–196.

63 R. Wagner, C. T. Lee, J. D. Durrant, R. D. Malmstrom, V. A. Feher and R. E. Amaro, *Chem. Rev.*, 2016, **116**, 6370–6390.

64 S. A. Serapian and G. Colombo, *Chem.–Eur. J.*, 2020, **26**, 4656–4670.

65 A. Paladino, M. R. Woodford, S. J. Backe, R. A. Sager, P. Kancherla, M. A. Daneshvar, V. Z. Chen, D. Bourboulia, E. F. Ahanin, C. Prodromou, G. Bergamaschi, A. Strada, M. Cretich, A. Gori, M. Veronesi, T. Bandiera, R. Vanna, G. Bratslavsky, S. A. Serapian, M. Mollapour and G. Colombo, *Chem.–Eur. J.*, 2020, **26**, 9459–9465.

66 K. M. Hart, C. M. W. Ho, S. Dutta, M. L. Gross and G. R. Bowman, *Nat. Commun.*, 2016, **7**, 12965.

67 G. Tiana, F. Simona, G. M. S. De Mori, R. A. Broglia and G. Colombo, *Protein Sci.*, 2004, **13**, 113–124.

68 G. Morra and G. Colombo, *Proteins: Struct., Funct., Genet.*, 2008, **72**, 660–672.

69 M. Montefiori, S. Pilotto, C. Marabelli, E. Moroni, M. Ferraro, S. A. Serapian, A. Mattevi and G. Colombo, *J. Chem. Inf. Model.*, 2019, **59**, 3927–3937.

70 M. C. R. Melo, R. C. Bernardi, C. De La Fuente-Nunez and Z. Luthey-Schulten, *J. Chem. Phys.*, 2020, **153**, 134104.