JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Local Fluctuations and Conformational Transitions in Proteins

Rocco Caliandro,[†,#] Giulia Rossetti,[‡,§,∥,#] and Paolo Carloni*[,∥,⊥]

[†]CNR—Institute of Crystallography, via Amendola 122/o, I-70126, Bari, Italy

[‡]Institute for Research in Biomedicine and Barcelona Supercomputing Center, Joint Research Program on Computational Biology, Baldiri I Reixac 10, 08028, Barcelona, Spain

[§]Jülich Supercomputing Centre, Institute for Advanced Simulation, Forschungszentrum Jülich, D-52425 Jülich, Germany

[∥]Computational Biophysics, German Research School for Simulation Sciences 1, D-52425 Jülich, Germany, and Institute for Advanced Simulation, Forschungszentrum Jülich, D-52425 Jülich, Germany

[⊥]Statistical and Biological Physics Sector, International School for Advanced Studies (SISSA), Trieste, Italy

Supporting Information

ABSTRACT: The intrinsic plasticity of protein residues, along with the occurrence of transitions between distinct residue conformations, plays a pivotal role in a variety of molecular recognition events in the cell. Analysis aimed at identifying both of these features has been limited so far to protein-complex structures. We present a computationally efficient tool (T-pad), which quantitatively analyzes protein residues' flexibility and detects backbone conformational transitions. T-pad is based on directional statistics of NMR structural ensembles or molecular dynamics trajectories. T-pad is here applied to human ubiquitin (hU), a paradigmatic cellular interactor. The calculated plasticity is compared to hU's Debye−Waller factors from the literature as well as those from experimental work carried out for this paper. T-pad is able to identify most of the key residues involved in hU's molecular recognition, also in the absence of its cellular partners. Indeed, T-pad identified as many as 90% of ubiquitin residues interacting with their cognate proteins. Hence, T-pad might be a useful tool for the investigation of interactions between proteins and their cellular partners at the genome-wide level.

INTRODUCTION

Protein plasticity plays a key role for molecular recognition in a variety of cellular processes, including signaling, metabolism, protein aggregation, and gene expression.[1−4] The traditional view of proteins as rigid objects binding their ligands in the "lock-and-key" model[5] was long ago replaced by mechanisms incorporating the idea that proteins are able to explore multiple conformational states.[6] In the "induced-fit" (IF) mechanism,[7] the ligand induces a conformational rearrangement of its target upon binding. Hence, protein plasticity is here mainly induced by the ligand. In "conformational selection (CS)",[8−10] the ligand binds one of the pre-existing conformations of its target. Therefore, plasticity is an intrinsic property of proteins, and the ligand only stabilizes one of the conformations relative to the others. IF and CS are by no means mutually exclusive, and they can play a joint role.[11−15]

Key residues for IF- and CS-based molecular recognition processes ('hot spots') are currently being successfully identified by computational analysis of protein complex structures at the atomic level. These analyses include alanine scanning mutagenesis,[16,17] computational[18] and energy-based methods,[19−21] MD-based methods,[22,23] bioinformatics,[24−27] combined solvent accessibility/conservation-based approaches,[28] and disrupted solvent hydrogen bond networks[29] evaluation. These approaches have proven themselves to be very valuable. However, the paucity of high resolution protein complex structures relative to high resolution structures of single proteins[30] calls for strategies that allow the identification of hot spots from protein structures in the absence of their cellular partners. Indeed, the number of the latter structures is currently 1 order of magnitude larger than the number of protein/protein and nucleic acid/protein complexes (information taken from PDB: http://www.pdb.org/pdb/home/home.do).

T-pad gives a quantitative description of the intrinsic plasticity of each residue. It detects, residue-by-residue, high plasticity sites in proteins, as well as relatively rigid residues (i.e., residues with similar backbone conformations across MD and/or NMR structures), also when located on flexible protein surfaces. Both may play a role in IF and CS mechanisms.[31−36] In addition, T-pad reveals backbone transitions between two conformations of the Ramachandran plot from MD simulations. This may be crucial for structural adaptations governed by the CS mechanism.[37−41] Finally, T-pad identifies the so-called "hinge" points for conformational changes (Table 1). These allow the movement of protein domains during their binding to interactors. Hence, they may play a pivotal role in molecular recognition.[31,42−44] The tool is based on directional statistics[45] applied to the Ramachandran angles[46] in molecular dynamics (MD) trajectories and/or NMR families of structures.

As an example of the power of this approach, we present here an application to human ubiquitin (hU). hU specifically recognizes a large variety of proteins[47] (currently estimated to be more than 150[48] with 32 344 binary interactions found in the IntAct database[49] from EMBL-EBI) using mostly[50] the CS mechanism for specific binding.[51] We apply the tool to an NMR family of structures of hU recently deposited in the

Received: July 18, 2012
Published: October 4, 2012

Table 1. Residues Tagged as T by the *T-pad* Analysis[a]

PDB ID	no. of conformers	no. of hU	T7	T9	K11	E12	R72	R74	G75
1YX5	18	1	A (18)	B(18)	B(18)	A(18)	B(5),C(3),D(10)	A(3),B(7), X(8)	A(1), B(4), C(6), D(1), E(5), F(1)
1YX6	18	1	A(18)	B(18)	B(18)	A(18)	A(1), B(7), C(6), D(4)	A(3) B(15)	A(3), B(5), C(1), D(4), E(1), F(4)
2BGF	10	2	hU_1: A(1), B(9)	hU_1: B(10)	hU_1: A(4), B(6)	hU_1: A(10)	hU_1: A(2),C(4),D(4)	hU_1: B(3),C(2),D(4),X(1)	hU_1: C(4), D(4), E(2)
			hU_2: A(1), B(9)	hU_2: B(10)	hU_2: A(4), B(6)	hU_2: A(10)	hU_2: A(3),D(7)	hU_2: A(3),B(3),D(1),X(3)	hU_2: C(3), D(4), E(1),F(2)
2JRI	20	2	hU_1: A(2), B(18)	hU_1: A(16), B(2), X(2)	hU_1: A(20)	hU_1: A(20)	hU_1: B(19),D(1)	hU_1: A(19), D(1)	hU_1: C(20)
			hU_2: A(7), B(13)	hU2: A(4), B(16)	hU_2: B(20)	hU_2: A(20)	hU_2: A(5), B(10), C(1), D(3), E(1)	hU_2: A(16), B(4)	hU_2: A(8), B(1),E(10), F(1)
2JY6	10	1	A(1), B(9)	B(10)	A(2), B(8)	A(10)	C(2), B(8)	D(10)	A(5), E(5)
2K8B	10	1	B(10)	B(10)	A(3), B(7)	A(8), B(2)	B(10)	A(10)	C(10)
2K8C	10	1	B(10)	B(10)	A(3), B(7)	A(10)	A(5), B(5)	A(7), B(3)	C(9), F(1)
2KDF	7	2	hU_1: A(7)	hU_1: A(6), B(1)	hU_1: A(7)	hU_1: A(7)	hU_1: B(7)	hU_1: A(5), C(1), X(1)	hU_1: A(1),B(1), C(3), E(1), F(1)
			hU_2: A(7)	hU_1: A(5), B(2)	hU_2: A(7)	hU_2: A(7)	hU_2: B(7)	hU_2: A(1), C(1), X(5)	hU_2: A(1), B(4), E(1), F(1)
2KTF	20	1	B(20)	B(20)	A(20)	A(2), B(18)	D(2), E(3), X(15)	A(16), B(4)	C(18), D(2)
2L0F	20	1	B(20)	B(20)	B(20)	B(20)	E(20)	A(2), D(12), X(6)	C(7), D(13)
2L0T	10	1	B(10)	A(2), B(8)	A(5), B(5)	A(10)	A(2),B(8)	A(5), B(2), X(3)	A(1), B(1), C(6), D(2)
2RR9	20	2	hU_1: B(20)	hU_1: B(20)	hU_1: A(20)	hU_1: A(20)	hU_1: A(6), C(1), X(13)	hU_1: A(4), B(4), C(1), D(1), X(10)	hU_1: A(1), B(3), C(5), D(5), E(1), F(5)
			hU_2: B(20)	hU_2: B(20)	hU_2: A(20)	hU_2: A(20)	hU_2: A(8), B(4), X(8)	hU_2: A(3), B(3), C(1), D(6), X(7)	hU_2: A(2), C(4), D(9), E(1), F(4)

[a]The conformations of these residues in the MD and NMR structures are reported. Conformations in the NMR structures are always present in the MD. The table reports (i) the PDB ID of the NMR complexes, (ii) the number of conformers present in the NMR structures, (iii) the number of hU structures present in the complex, and (iv) the regions of the RP for each residue tagged as T (see Figure 3). When the letter "X" is present, it means that the conformations are in regions of RP not present in NMR and MD of free hU in aqueous solution. Note that the references are reported in section S6.1.

PDB[52] and the MD trajectories performed here. Anticipating our results, T-pad detects almost all (90%) of the identified hU key residues for molecular recognition, without any false positives and in a fully automatic and fast procedure.

MATERIALS AND METHODS

X-Ray Structure Solution. Eight hU crystals were obtained under the same conditions and soaked in the presence of different metals. Diffraction data were taken at the synchrotron ELETTRA (Trieste, Italy), and the crystal structures were solved using the same crystallographic software. The refined Debye−Waller factors of the C_α atoms were taken as representative of the backbone flexibility, and they were averaged residue-by-residue over all the structures.

hU crystals were grown in hanging drop plates at 293 K under the following conditions: ubiquitin, 22.5 mg/mL and PEG3350, 30%. Seeding was used to improve crystal quality. Ubiquitin crystals were used to perform soaking experiments with several ligands containing metal ions. The data sets for the soaked crystals, which lead to satisfactory diffraction patterns, are reported in Table S4. All the data collections were carried out under cryogenic conditions. The data sets were indexed by MOSFLM.[53] Data reduction was accomplished by means of the POINTLESS, SCALA, and TRUNCATE programs,[54] and then XTRIAGE from the PHENIX suite[55] was run to detect the anomalous signal. For all of the data sets, the characteristic $P2_12_12_1$ symmetry of the human ubiquitin apo crystal was found. A summary of the relevant crystallographic data is reported in Table S5.

The data sets were phased by using the following protocol:

- The program REMO[56] from the package ILMILIONE[57] was run to perform molecular replacement. The known crystal structure of hU (PDB code 1UBQ[58]) was used as a search model. In all the cases, only one solution was found, with a very high figure of merit.
- The resulting structural model was rebuilt in-place using ARP-WARP[59] with the observed amplitudes and phases from the molecular replacement solution. This step was accomplished to reduce the model bias from molecular replacement.
- The rebuilt model was inspected against the $2F_o-F_c$ and F_o-F_c Fourier maps by using the program Coot.[60] It was modified by means of the local fitting and inspection tools in Coot.
- The structural model was refined by REFMAC[61] against the observed amplitudes.
- The presence of heavy atoms from the ligands was checked by inspection of the anomalous difference map calculated from the refined models with REFMAC.

A summary of the results obtained is given in Table S6.

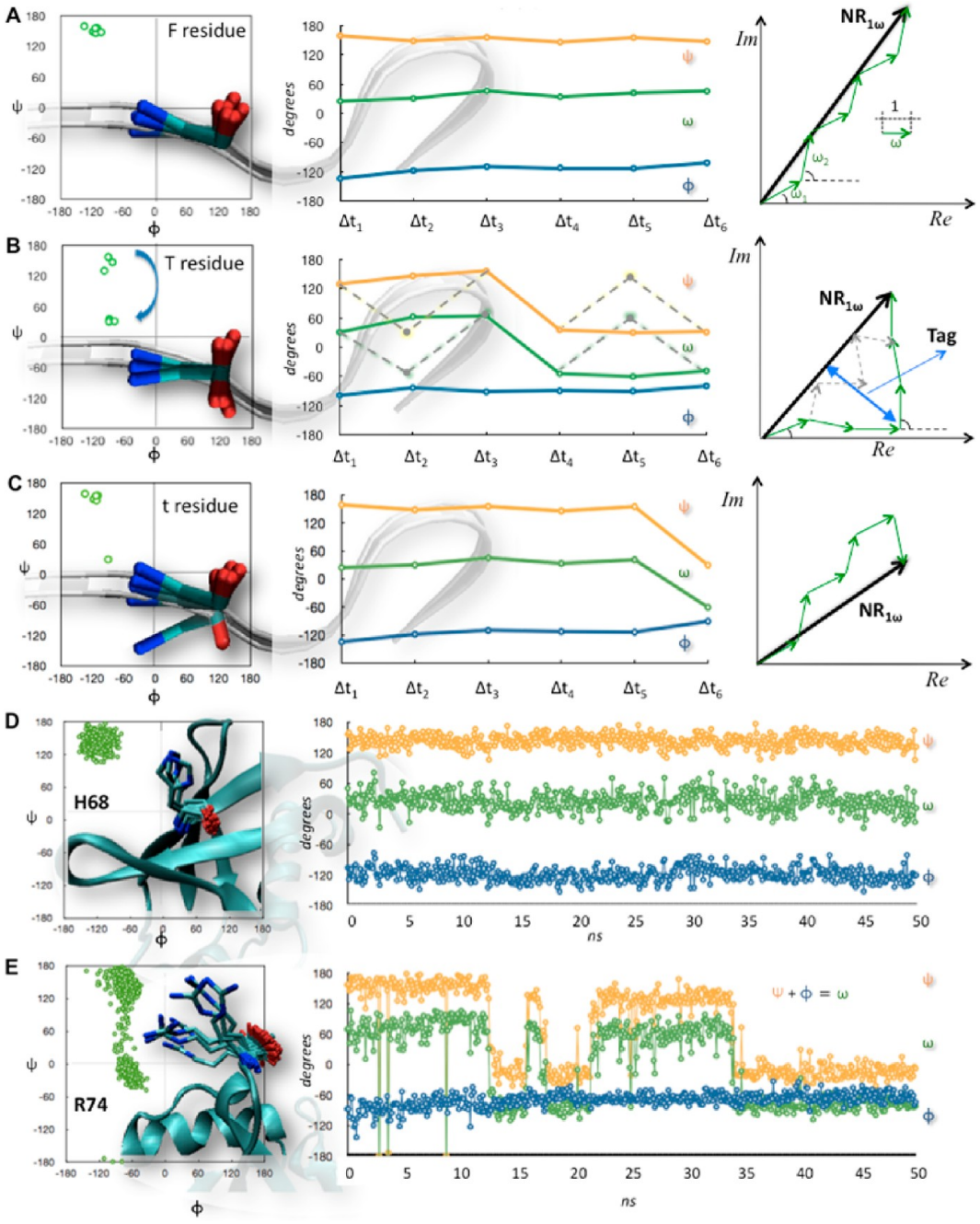

Figure 1. (A–C) Protein residues may either fluctuate around a configuration (A) or undergo a "long" conformational transition (B) or a "short" one (C). The three cases, tagged F, T, and t, respectively, are described in detail in the text. Each case displays, from left to right, a cartoon of a hypothetical protein residue superimposed with its Ramachandran plot, the correspondent Ramachandran angles Φ and ψ along with $\omega = \Phi + \psi$ plotted as a function of simulated time, and the representation in the complex plane of six unit vectors $e^{i\omega_j}$ (green). The resultant $NR_{1\omega}$ (black vector) is used to calculate PAD_ω. In the case of long transition (B), we also show with dotted lines how it is possible to obtain the same value of $NR_{1\omega}$ and hence PAD_ω by ordering the vector in a different way, i.e., an ordering resembling a fluctuation. (D, E) Left: cartoon representation of H68 (C) and of R74 (D) residues over 50 ns extracted from the 0.2 μs MD of hU in water performed here. The corresponding Ramachandran plots are superimposed with the cartoons. Right: Ramachandran angles along with ω plotted as a function of the simulated time. T-pad tags H68 and R74 as F and T, respectively. The same tag is obtained for the overall trajectory.

No metal ion was found in the model despite the fact that an anomalous signal was detected for some data sets. This can be ascribed to the presence of anomalous scatterers in the bulk solvent which are not anchored to the protein structure. In the cases of the CisCryI, CuCryI, and CuCryII data sets, we verified that the additional nonprotein atoms expected for the rest of the ligand were not present.

Computational Details. *a. Molecular Dynamics.* The 1UBQ[58] hU structure was solvated in a box of explicit water, ensuring that the solvent shell would extend for at least 1.6 nm around them. Periodic boundary conditions were applied. The all-atom AMBER ff99SB-ILDN force field[62] was used for the protein with Åqvist parameter force fields[63] for the sodium counterion. The TIP3P[64] model was used for water molecules. Long-range electrostatic interactions were treated with the

dx.doi.org/10.1021/ct300610y | J. Chem. Theory Comput. 2012, 8, 4775–4785

particle mesh Ewald method.[65] A Fourier spacing of 0.12 nm combined with a fourth-order cubic spline interpolation was used. A 1.2 nm cutoff was used for van der Waals interactions as well as the real-space part of the electrostatic interactions.[66] All bond lengths were constrained with the LINCS algorithm, and the time step was set to 2 fs. NPT ensemble (T = 298 K, P = 1 bar) MD simulations were carried out by coupling the systems with a Nosè−Hoover thermostat[67] and Andersen−Parrinello−Rahman barostat.[68,69] The systems underwent 1000 steps of steepest-descent energy minimization with 5000 kcal mol^{-1} Å^{-2} harmonic position restraints on the protein, followed by 2500 steps of steepest-descent and 2500 steps of conjugate-gradient minimization without restraints. The systems were then gradually heated from 0 K up to 298 K in 12 steps of 100 ps simulation. After that, a 200 ps equilibration was performed. Finally, 200-ns-long productive MD simulations were run in the NPT ensemble (298 K, 1 atm, and 2 fs time-step).

b. T-pad. We applied the PAD analysis (for a definition of "PAD analysis" see Theory section) to the following data sets: (i) 20 000 frames from 200 ns of MD simulations of hU in explicit solvent (see above) and (ii) 640 conformations of hU from NMR measurements (PDB ID: 2KOX[52]).

The *tag* analysis (for a definition of "*tag* analysis" see Theory section) was applied to i. We first divided the trajectory into 10 intervals of 20 ns (2000 frames). For each of these, we detected backbone fluctuations and transitions (see the Theory section). We exclude G76, the last hU residue, from the investigations, since it is not always present in the X-ray/NMR structures. The complexes have an identical sequence for hU, namely, MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQ-QRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRG(G). NMR structural families also have the same range of pH, pressure, and temperature. All these factors may strongly influence the plasticity of a protein.

c. Debye−Waller Factors. Comparisons were made between the trend of the PAD$_\omega$ and of the RMSF values extracted from our simulation with those of the Debye−Waller factors of the C$_\alpha$ taken from hU X-ray structures. We considered only crystal structures of free hU, crystallized under room conditions, to avoid introducing flexibility features due to the presence of ligands or other proteins in complex. Only two entries of this kind were found in the Protein Data Bank (1UBQ[58] and 1UBI);[70] The structure by 2ZCC[71] was obtained at high pressure (500 bar) and in the presence of Zn^{2+} ions. It has not been considered here. They are very similar to each other with an RMSD for their C$_\alpha$ atoms of only 0.09 Å. To increase the statistics of the Debye−Waller data, the set was extended with eight new X-ray structures of free hU under room conditions, obtained under the same crystallization and data-taking conditions (see the Supporting Information for details).

The RMSF values on the C$_\alpha$ atoms read

$$\mathrm{RMSF} = \sqrt{\frac{\sum_{j=1}^{N} (x(j) - \langle x\rangle)^2 + (y(j) - \langle y\rangle)^2 + (z(j) - \langle z\rangle)^2}{N}} \quad (1)$$

where $x(j)$, $y(j)$, and $z(j)$ are the Cartesian coordinates of the C$_\alpha$ atom at frame j; $\langle x\rangle$, $\langle y\rangle$, and $\langle z\rangle$ are their averages in time; and N is the total number of frames.

The Debye−Waller factors can be put in relation with the RMSF values through the following equation:[72]

$$B_{\mathrm{calcd}} = \frac{8}{3}\pi^2 \mathrm{RMSF}^2 \quad (2)$$

Theoretical Basis. *Theory.* Concepts from statistics such as the mean or standard deviation cannot be used for angular values. One may use in this case notions from directional statistics.[73] In particular, the circular mean and circular spread replace, respectively, the average and the standard deviation calculated for nonangular variables. The latter may quantify proteins' backbone plasticity. Consider N values ω_j of an angular variable ω. In directional statistics, these values are associated with N unit vectors of the type $e^{i\omega_j}$ (Figure 1).[45] The circular spread of the angle ω (CS$_\omega$) may be calculated from the resultant of such unit vectors:[45]

$$\mathrm{CS}_\omega = \frac{1 - R_{2\omega}}{2R_{1\omega}^2} \quad (3)$$

where

$$R_{k\omega} = \frac{1}{N}\left|\sum_{j=1}^{N} e^{ik\omega_j}\right|;\ k = 1, 2 \quad (4)$$

are the lengths of the so-called 'sample moments' of first (k = 1) and second (k = 2) order. $NR_{k\omega}$ represents the resultant of the N unit vectors $e^{ik\omega_j}$ (see Figure 1C). Therefore, CS$_\omega$ ranges between 0 (no dispersion, $R_{1\omega} = R_{2\omega} = 1$, the vectors $e^{ik\omega_j}$ are aligned along a unique direction) and infinity (maximum dispersion, $R_{1\omega} = R_{2\omega} = 0$, the vectors $e^{ik\omega_j}$ are randomly distributed).

Protein Plasticity. The circular spread of the Ramachandran angles Φ or ψ (CS$_\Phi$ or CS$_\psi$) has been used in ref 74 to analyze the plasticity of the prion protein backbone [in this reference, it was called the *Angular Dispersion Index*]. However, CS$_\Phi$ or CS$_\psi$ alone cannot provide information on the conformation of the backbone, characterized by the two Ramachandran angles. In addition, CS$_\Phi$ and CS$_\psi$ range from 0 to infinity. This does not allow a comparison of circular spreads across different residues in the same protein, or the same residue across different proteins. Here, we introduce a new quantity (Protein Angular Dispersion of the angle ω, PAD$_\omega$) that overcomes these problems and is well suited to quantify the plasticity of protein backbone residues. Notice that ω, as defined here, does not refer to the torsion angle C$_\alpha$−C−N−C$_\alpha$ in a peptide bond. PAD$_\omega$ differs from CS$_\omega$ in two aspects: (1) it is function of $\omega = \Phi + \psi$, hence dependent on both Ramachandran angles; (2) PAD$_\omega$ is formulated so as to range between 0° and 180°, allowing for quantitative comparisons among residues and among proteins (notice that ω can instead range between 0° and 360°). PAD$_\omega$ reads

$$\mathrm{PAD}_\omega = \frac{180}{\pi}\cos^{-1}\left(\frac{1 - \mathrm{CS}_\omega}{1 + \mathrm{CS}_\omega}\right) \quad (5)$$

PAD$_\omega$ is 0° when the Φ and ψ dihedral angles of the residue do not change across a set of protein structures. It is 180° when the Φ and ψ dihedral angles assume random values. It can be applied to a set of structures determined by NMR or simulated by MD.

The evaluation of CS$_\omega$ and PAD$_\omega$ quantifies the protein backbone plasticity but does not allow discrimination between fluctuations and transitions (from one region of the Ramachandran plot to another in MD simulations) when they occur with the same amplitude. We can see this graphically

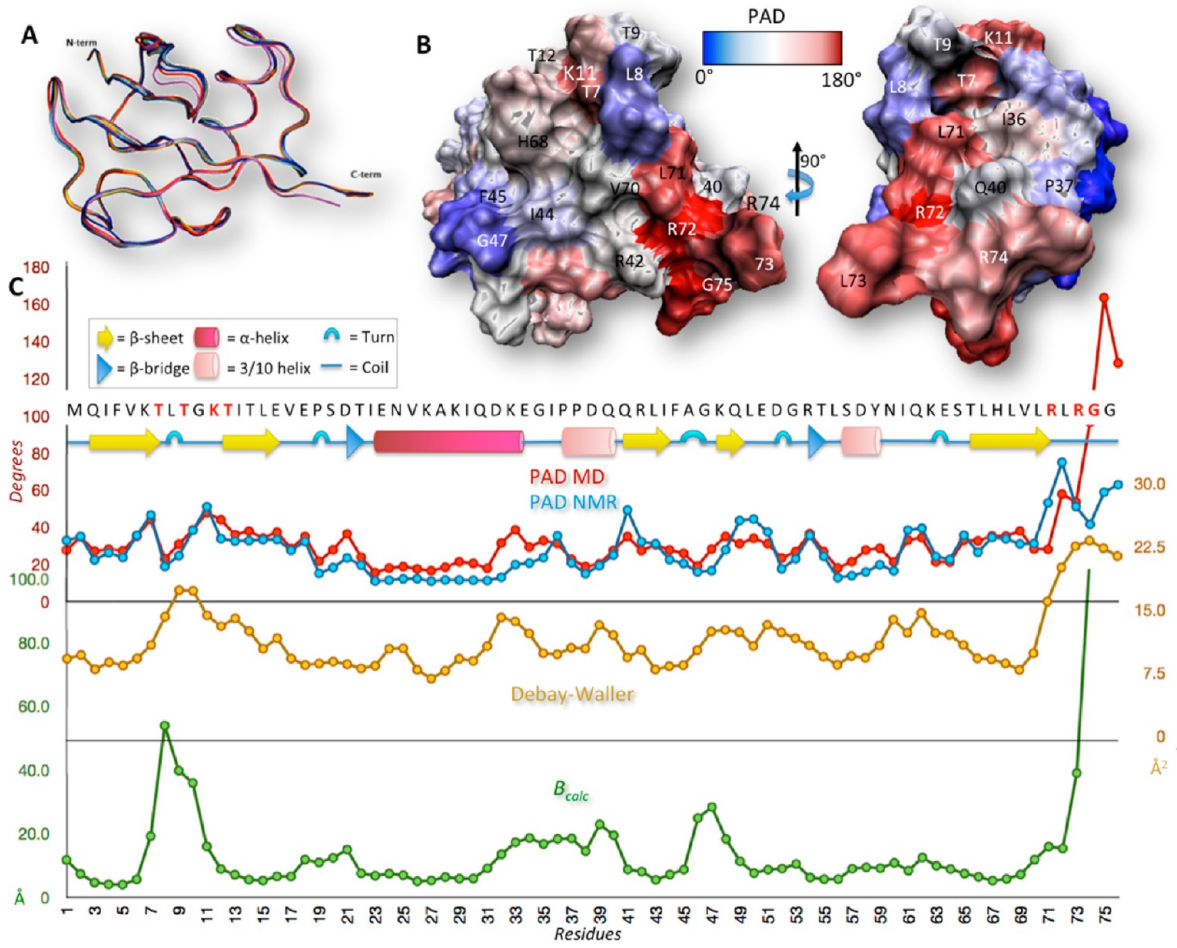

Figure 2. (A) Superposition of the backbone of hU X-ray structures from this work and from the Protein Data Bank (1UBQ[58] and 1UBI[70]). Each color corresponds to a different structure. hU has a mixed α/β secondary structure, the major feature of which is a five-stranded antiparallel β sheet that appears to "grasp" an α-helix (β-grasp fold family).[91] (B) Surface representation of hU, colored according their PAD values from NMR structures.[52] The relevant residues for molecular recognition identified by the *T-pad* tool are indicated with their sequence numbers. (C) PAD values (in degrees) are calculated over our MD simulation and over the NMR structural ensemble of hU in solution.[52] The experimentally observed Debye–Waller factors for the C_α (in Å^2) are averaged over the X-ray structures mentioned above, and those calculated from our MD simulations (B_{calc}, in Å^2) are obtained from the RMSF values through eq 2. The secondary structure scheme of the protein according to DSSP[92] of hU is also shown.

in Figure 1: the same $NR_{1\omega}$ value—and hence the same values of CS_ω and PAD_ω—are due to the same set of unit vectors $e^{ik\omega_j}$, but they are distributed in two different ways: one fluctuating around a given direction (Figure 1B, dotted lines), i.e., identified as a fluctuation, and the other fluctuating around two separate directions (Figure 1B, continuous lines), i.e., identified as a transition [we can say that a transition occurs when two consecutive subsets of unit vectors $e^{ik\omega_j}$ show different common directions and the change in direction is greater than the intrinsic fluctuations within each subset]. In the Supporting Information, section 1, we report a numerical example where transitions with different periods (τ) and amplitudes (A) are introduced on a series of ω_i ordered as a function of simulated time (Figure S1). In that example, CS_ω does not correlate with transitions and/or fluctuations of ω_I. For instance, CS_ω reaches its maxima both in the presence and in the absence of transitions (Figure S3).

Discriminating Fluctuations from Transitions. In MD simulations, the ω_j values can be ordered as a time series. By exploiting this fact, here we show that it is possible to identify a mathematical condition involving two quantities (the cumulative protein angular dispersion $(PAD_{cum})_\omega$ and the phase Angular Index, PAI_ω) that allow one to distinguish between "T" (Figure 1B), "t" (Figure 1C), and "F" (Figure 1A) residues, where "T" indicates "long transition" (i.e., contributing more than 30% of the simulation time), "t" short transition (i..e. contributing less than 30% of the simulation time), and "F" fluctuations. This mathematical condition is identified by analyzing the correlation plots of $(PAD_{cum})_\omega$ and PAI_ω against τ and A in the numerical example (see the Supporting Information, section 1). Notice that the two quantities per se are not sufficient for a complete discernment of transitions among fluctuations: their simultaneous evaluation is required. Indeed, $(PAD_{cum})_\omega$ discriminates fluctuations from transitions with respect to their amplitude while PAI_ω discriminates with respect to their frequency.

$(PAD_{cum})_\omega$ introduces the memory of the time series, perhaps in the most straightforward way. This quantity is the same as that of PAD_ω except that it is not simply a function of ω_j but on a time series of $\bar{\omega}_i$ values, representing the angular values of the vectors $\sum_{j=1}^{i} e^{i\omega_j}$:

$$(\mathrm{PAD}_{\mathrm{cum}})_{\omega} = \frac{180}{\pi} \cos^{-1}\left(\frac{1 - \mathrm{CS}_{\bar{\omega}}}{1 + \mathrm{CS}_{\bar{\omega}}}\right) \quad (6)$$

$(\mathrm{PAD}_{\mathrm{cum}})_{\omega}$ varies in the range [0°, 180°]. It increases with the deviations of the $\sum_{j=1}^{i} e^{i\omega_j}$ vectors from the direction of the resultant vector (as in Figure 1B). Graphically, we can see that the $\sum_{j=1}^{i} e^{i\omega_j}$ vectors in the case of transitions may be different than in case of fluctuations, even if the distributions of $e^{ik\omega_j}$ vectors are the same (see Figure S5). Therefore, this quantity can in principle distinguish fluctuations from transitions. Indeed, this is the case if $(\mathrm{PAD}_{\mathrm{cum}})_{\omega} < 10°$: the residue fluctuates with an amplitude comparable to that of the intrinsic fluctuations of the baseline, and therefore it can be unambiguously tagged as "F" (Figure S3a). However, for larger values, this quantity is not useful, as it does not distinguish short transitions from fluctuations, as seen from the correlation plots in Figure S3a.

PAI_{ω} is a function of CS_{ω} and the Angular Transition Index ATI_{ω}, developed by adapting the studies of Hurst[75] to time series of angular data (see the Supporting Information, Figure S4 and section 2.4 for details):

$$\mathrm{PAI}_{\omega} = \frac{180}{\pi} \tan^{-1}\left(\frac{\mathrm{ATI}_{\omega}}{\mathrm{CS}_{\omega}}\right) \quad (7)$$

PAI_{ω} varies in the range [0°, 90°]. On the basis of the correlation plots of our numerical example, PAI_{ω} does not depend significantly on A but does distinguish fluctuations from short and long transitions based on the value of τ (Figure S2b). Values of $\mathrm{PAI}_{\omega} < 30°$ are associated with fluctuations. Those comprised between 30° and 60° are associated with long transitions. Finally, large values ($60° < \mathrm{PAI}_{\omega} < 90°$) occur for short transitions.

On the basis of the above considerations, we tag each residue as F if $(\mathrm{PAD}_{\mathrm{cum}})_{\omega} < 10°$ and/or $\mathrm{PAI}_{\omega} < 30°$. The residue undergoes a conformational transition if $(\mathrm{PAD}_{\mathrm{cum}})_{\omega} > 10°$ and $\mathrm{PAI}_{\omega} > 30°$. It is tagged T or t depending of the value of PAI_{ω}, as observed in our numerical example:

$$\begin{cases} \text{F if } (\mathrm{PAD}_{\mathrm{cum}})_{\omega} > 10° \cap \mathrm{PAI}_{\omega} \in [0°, 30°] \cup (\mathrm{PAD}_{\mathrm{cum}})_{\omega} < 10° \\ \text{T if } (\mathrm{PAD}_{\mathrm{cum}})_{\omega} > 10° \cap \mathrm{PAI}_{\omega} \in [30°, 60°] \\ \text{t if } (\mathrm{PAD}_{\mathrm{cum}})_{\omega} > 10° \cap \mathrm{PAI}_{\omega} \in [60°, 90°] \end{cases} \quad (8)$$

The reader is referred to Supporting Information, section 2 for more details on all of the quantities defined in this section.

RESULTS

Application to Human Ubiquitin. The previous section showed that T-pad calculates two properties for each residue of a protein. The first is the plasticity of the residue from MD and/or from NMR structures. This is obtained by the PAD_{ω} values: the closer it is to its maximum value, the larger the plasticity is. The second is a tagging analysis, which detects eventual transitions from one conformation to another. The residue is tagged F if it experiences only fluctuations, T for long transitions, and t for shorter ones. In the last case, the transition occurs for less than 30% of the simulated time.

T-pad is applied here to an ensemble of NMR structures of hU in solution[52] and for a 0.2 μs MD trajectory from the same system. The MD was explicitly carried out for this research. The PAD_{ω} parameters are compared with the experimentally derived Debye–Waller factors. The latter have been reported so far for two free hU structures under room conditions (1UBQ[58] and 1UBI[70]). To provide a relatively large set of Debye–Waller factors, we have determined in this work eight new X-ray structures of hU. In the following section, we describe those structures briefly and our MD simulations of hU in water. Next, we present a T-pad analysis of the protein. We close the section by discussing the performance of T-pad.

X-Ray Structures and MD Simulations of hU. Eight hU structures were obtained by using the same crystallization and structure solution protocols. A perturbation was introduced by soaking the crystals in solutions containing metal ions. Despite the existence of an anomalous signal detected during diffraction, no metal ions were found in the model, indicating that metal ions were present in the bulk solvent but not anchored to the protein structure. Therefore, the reconstructed structures may represent hU packed in crystal but subjected to a stimulus residing in the solvent. All the structural models preserve the secondary structure arrangement (Figure 2A) from the α-helix (I23–G34) to the two small 3_{10} helices (P38–Q40 and S57–Y59), the five β-sheets (Q2–T7, T12–E16, Q41–F45, K48–Q49, and T66–L71), the two β-bridges (D21–T22 and R54–T55), and the five turns (L8–T9, P19–S20, A46–G47, D52–G53, and K63–E64). The RMSDs of the C_{α} atoms among the structures vary from 0.1 to 0.3 Å, except for one structure obtained by soaking with chromium. This shows higher deviations (RMSD ≈ 0.6 Å) because of a distortion of the helix and of loop 8–10 (Figure 2A). This set is extended with two other X-ray structures of free hU (1UBQ[58] and 1UBI[70]).

In 0.2 μs of MD simulations of hU in water solutions, the protein preserves its folding and is very stable (RMSD_{1-71} = 0.58 ± 0.1 Å), with the exception of the flexible C-terminal tail (R72–L73–R74–G75–G76), which exhibits a high mobility ($\mathrm{RMSD}_{\text{C-term}}$ = 4.73 ± 1.75 Å, $\mathrm{RMSD}_{\mathrm{ALL}}$ = 1.3 ± 0.4 Å). The secondary structure is also retained (Figure S6).

T-pad Analysis of hU. Most PAD_{ω} values calculated from our MD simulations are in quantitative agreement with those calculated from the NMR ensemble of structures.[52] The values for the residues at the C-terminus (from R72 to G76, Figure 2 BC) are larger. The PAD_{ω} values for each residue along with the Debye–Waller factors of the C_{α} atoms (averaged over the 10 X-ray structures) are plotted in Figure 2C. The PAD_{ω} and the Debye–Waller factors of the C_{α} atoms are different quantities. The first measures the fluctuations of the torsion angles Φ and ψ of proteins in solution, while the second measures the atomic fluctuations of C_{α} of proteins in the solid state. However, our indices are totally related to local fluctuations, whereas the B-factor may be related in part to local fluctuations. Therefore, similar trends of the two quantities in few specific regions of the protein might exist. In the case of hU, the trend of the two quantities is indeed not too dissimilar in β1, β2, and the C-term. Next, we compare the measured Debye–Waller values with those calculated from the RMSF (B_{calc}, see Methods, eq 2). The trends of the β1–β2 loop, α1, and the α3–β5 loop are similar (Figure 2C). The differences in trends in most of the rest of the plot may be ascribed, at least in part, to the fact that RMSFs are calculated for hU in solution and B-factors are measured for hU in the solid state.

The relatively rigid residues, characterized by PAD_{ω} less than 28° in both MD and NMR structures, include T22–Q31 and L56–N60 in the α-helix and the 3_{10} helix, respectively, along with residues I44–F45, G47, and V70 in strands β3, β4, and β5,

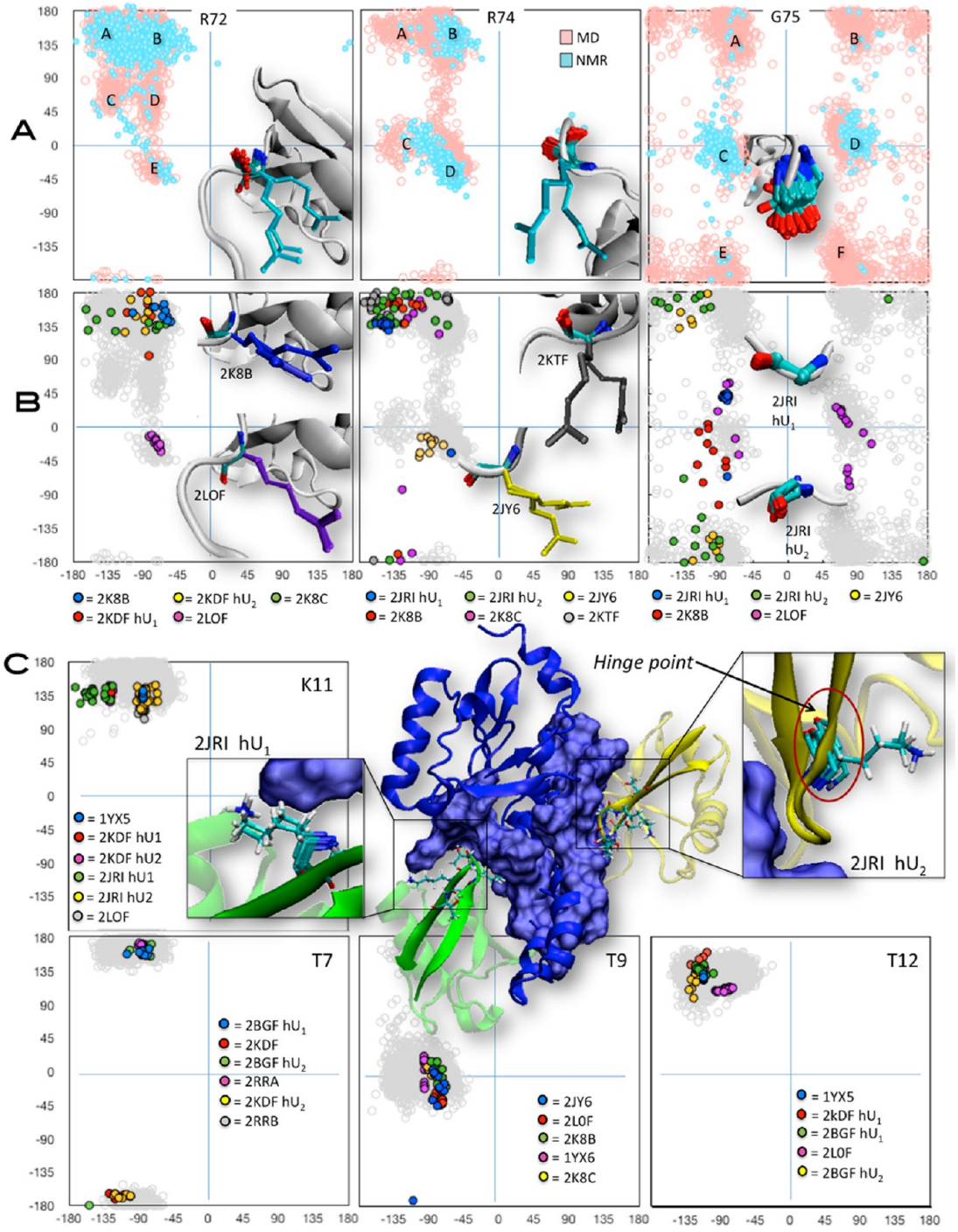

Figure 3. Conformational transitions in hU. R72, R74, and G75 at the C-terminus (A, B) and T7, T9, K11, and T12 in the $\beta1-\beta2$ region (C) experience transitions in MD of hU in aqueous solution, according to the *T-pad* analysis. (A) Ramachandran plots (RPs) in the MD (cyan) and in the NMR (pink)[52] structures. Despite the regions explored by each residue being the same in NMR and MD, the population of these is different. R72 visits the regions centered in A = [−135°; +140°], B = [−65°; +140°], C = [−135°; +65°], D = [−65°;+65°], E = [−65°; −22°]; R74 the regions centered in A = [−135°; +135°], B = [−67°; +135°], C = [−135°; 0°], D = [−67°; −22°]; G75 the regions centered in A = [−65°; +155°], B = [+75°; +155°], C = [−67°; 0°], D = ∼[+75°; 10°], E = [−67°; −155°], and F = [+75°; −155°]. (B) The RPs of the MD structures (empty gray dots) are compared with those of NMR structures of hU in complex with its cellular partners in Table 1 (colored dots). Single residues in cartoon representation from NMR are also reported as an insert in each RP in both A and B sections of the figure. (C) RPs of T7, T9, K11, and T12 in the MD (gray). The regions explored and their populations are the same among NMR and MD. Hence, NMR has not been reported. The RPs are compared with those of NMR structures of hU complexes with cellular partners (Table 1, full colored points). T7 visits the regions centered in A = [−105°; +170°] and B = [−90; +170°]; T9, the regions centered in A = [−70°; 0°] and B = [−120°; 0°]; T12, the regions centered in A = [−125°; +135°] and B = [−70°; +135°]; and K11, the regions centered in A = [−115°; +160°] and B = [−65°; +160°]. A close up of the hU·Ataxin-3·hU ternary complex (2JRI)[93] is shown as an inset. Residue K11 at the two interfaces between Ataxin-3 and the two hU molecules (hU_1 and hU_2 in the figure) is maximized. Ataxin is shown as a blue mixed surface/cartoon representation, while the two hU structures are shown as green and yellow cartoons. K11 is in licorice representation.

dx.doi.org/10.1021/ct300610y | *J. Chem. Theory Comput.* 2012, 8, 4775−4785

respectively, and residues L8, P37−G40, D52−G53, and K63−E64 in the loops. The C-terminal residues feature the largest plasticity (>60°). Residues belonging to the β-sheets (K6−T7, K11−G16, Q41−R42, K48−L51, and S65), the $\beta1-\beta2$ loop (T9−G10−K11), loop $\alpha1-\beta3$ (D35−G36), res R54 (turn 4), and Q61−K62 from the $\beta4-\beta5$ loop exhibit intermediate plasticity (from 28° to 60°).

The tagging analysis shows that eight residues (R72, R74, G75, D21, T7, T9, K11, and T12) experience long transitions during the time scale investigated by the MD simulations (the first three residues also experience short ones). Inspection of the MD trajectory confirms that this is indeed the case. During these transitions, both of the Ramachandran angles change by more than 60° with respect to the initial conformation for the first three residues (Figure 3AB), exploring four to six regions of the Ramachandran plot. For the latter four residues, only Φ changes, and it changes by less than 60° relative to the initial conformation (Figure 3C). Each of these residues explores two regions of the Ramachandran plot. Interestingly, in the NMR structures of free hU, these residues are in all of the conformations identified by the MD. A careful check of the MD trajectory shows that no other residue undergoes transitions in the simulated time scale.

Performance. T-pad is very fast: it takes 10 s on a single-core laptop for the above analysis and less than 1 min for a 120 ns trajectory of a protein with 400 residues. The code (freely available upon request) has been organized as a scripting extension of the VMD package.[76] By reading the structures from MD or NMR, VMD colors the residues according to their PAD_ω (Figures 2B) or their tags and produces a flexibility profile by plotting residue-by-residue the PAD_ω values colored according to their residue tagging (see Figure S7).

DISCUSSION

We have presented a computational tool—T-pad—for the investigation of protein plasticity of single residues from MD simulations or NMR structures. The tool is tailored for the investigation of protein plasticity of single residues from MD simulations or NMR structures. As an illustrative example, T-pad is applied here to hU NMR structures and MD data obtained here

The calculation of the PAD_ω values allows the identification of highly flexible regions and rigid residues, also when located on flexible surfaces (Figure 2). In particular, the trend of the calculated PAD_ω is similar to that of the B-factors in the $\beta2$ and C-term (Figure 2).

Both highly flexible regions and rigid residues may play a crucial role for molecular recognition and the specificity of binding.[31−36,48,77] The tag analysis, which is also within the T-pad tool, detects backbone transitions for each residue in MD simulations. The resulting distinct conformations may play a key role in CS.[37−41,51,52,78]

All of the highly flexible regions identified by the tool (see Results section) play a role for molecular recognition in hU. Namely, (i) the C-terminal tail (residues R72−G76) is a target for the regulation of conjugation and deubiquitination as well as the selective binding of polyubiquitin chains.[79] (ii) K6−T7 ($\beta1$), T12−G16 ($\beta2$), T9−G10-K11 ($\beta1-\beta2$ loop), and S65−R72 ($\beta5$) belong to the region peripheral to the hydrophobic patch (L8, I44, V70), the most common site of interaction,[47] and moreover the residues belonging to $\beta2$ may also participate in the $\beta_2-\alpha_1$ groove,[47] another site of binding. Finally, (iii) the D35−G37 region is centered on I36 between $\alpha1$ and $\beta3$, another site of hU recognition.[80]

The relatively rigid residues identified in flexible surfaces of the protein are L8, I44, and V70 (hydrophobic patch) and G47, R42, H68, and G40 (Figure 2B). The first six residues form more contacts than any other residue of the protein.[48,77] Some of them (I44, G47, R42, and H68) may also be responsible for hU binding specificity.[51,77,81] We note that V70 and L8 have so far been classified as flexible[77] from an analysis of an ensemble of hU X-ray complexes. This was a false-positive due to the Cartesian-coordinates-based descriptor used. These residues indeed belong to flexible and unstructured parts of the proteins, and despite the fact that they preserve the same backbone conformation overall during simulations, they are characterized by a high RMSF. The latter descriptor (like other Cartesian coordinates descriptors) is biased by the problematic alignment of extremely flexible structures, besides being affected by the movements of neighboring residues[74] (see the Supporting Information, section 5, Figures S8−S9, Table S1). T-pad avoids such limits being based upon dihedral angles (i.e., internal coordinates).

However, the RMSF calculates the overall atomic fluctuations; instead, T-pad quantifies local fluctuations. Therefore, the T-pad analysis cannot be used to study large-scale fluctuations of the protein, as the RMSF.

It is worth noticing that the T-pad analysis was able to correctly identify the critical residues for hU molecular recognition discussed above, requiring here only structural information on the free protein.

Finally, residues T7, T9, K11, T12, D21, R72, R74, and G75 were identified as undergoing conformational transitions. Strikingly, an analysis of the Protein Data Bank shows that all of them form direct or metal ion mediated interactions with hU's target proteins,[82,83] and/or they are "hinge" points for conformational changes (Tables S2 and S3 and Figure 3).[84−86] Moreover, the finding that three (T7, T9, and K11) of the four residues belonging to the $\beta1-\beta2$ (T7−K11) loop undergo conformational transitions is in agreement with the well-established role of the $\beta1-\beta2$ loop in molecular recognition.[51] Notably, all the identified residues are confined to a few of the possible configurations sampled during MD, when hU is complexed with cellular partners (Table 1). This fact may corroborate the well-established[37−39,41] key role of the CS mechanism for molecular recognition processes involving hU.

Let us take the hU/ataxin-3/hU ternary complex (2JRI, Figure 3C) as an example. Here, ataxin-3 binds to two hU molecules (hU_1 and hU_2 hereafter) through distant and structurally different binding sites. All of the residues belonging to the binding surfaces (Table S2) except L8 (identified with PAD_ω), A46, G47, and K48 hU_2 are tagged as T; i.e., they undergo transitions, and they are involved in the molecular recognition process. Here, we discuss two particularly interesting cases.

The first is that of R72. In the free hU in aqueous solution, it undergoes transitions in five different regions of the Ramachandran plot (A−E in Figure 3A). In the complex, R72@hU_1 binds to Ataxin-3, and it adopts one conformation [among 20 conformers, 19 are in B and 1 is in D (Table 1); hence we assume the residue's conformation is B], "B" (Table 1). We may expect therefore that Ataxin-3 binding to hU_1 stabilizes this conformation in the CS-based mechanism. In contrast, R72@hU_2 does not bind to Ataxin-3 (Table S2). It indeed assumes all the conformations assumed by the free

protein in aqueous solution (Table 1), consistently with the fact that none of them has been selected by its cellular partner. The second is K11, which explores conformations A and B in hU in water (Table 1). K11@hU_1 binds Ataxin-3 and its conformation is A. K11@hU_2 does not bind directly to Ataxin-3 even though it adopts only the conformation B (Table 1). Notably, it may play a role as a hinge point. Indeed, the β2@hU_2 appears twisted with respect to β2@hU_1, and the twist starts on the K11@hU_2 backbone (Figure 3C).

As a final remark, we note that our tagging analysis is based purely on geometrical criteria. This contrasts with kinetic approaches, which define the global conformational states of the protein by calculating transition times with Markov state models[87] and/or free-energy barrier heights with landscape theories.[88] The kinetic approach, contrary to the geometric one, allows interpreting of the overall movements of the protein, separating their time scales, and predicting long-time scale motions.[89] T-pad detects instead local fluctuations and/or transitions of single residues characterizing their time-dependent properties.

CONCLUSION

Our computational tool T-pad is tailored for the analysis of local plasticity of proteins, which may play a role in molecular recognition. It identifies almost all (90%) of the residues involved in molecular recognition of hU complexes (Table S2 and S3) with no false positives, based on analysis of the protein in the absence of interactors.

To understand specificity in signaling, as well as many other molecular-recognition-based cellular mechanisms, we need to know how proteins interact with particular targets and how the correspondent interaction pathway can in turn be affected. Applying T-pad to NMR structural families and to MD trajectory databases such as MODEL[90] may help identify key residues for protein/protein interactions in the interactome. This information may be valuable for structural predictions of protein/protein interaction pathways genome-wide.

ASSOCIATED CONTENT

Supporting Information

Section 1: The numerical example, Figures S1, S2, S3, S4. Section 2: Mathematical derivations of quantities defined in the Theory section, Figure S5. Section 3: Details from the MD simulation, Figure S6. Section 4: Graphical outputs of T-pad in the VMD program, Figure S7. Section 5: Advantages of PAD_ω relative to RMSF, Figures S8 and S9, Table S1. Section 6: hU interactors, Tables S2 and S3. Section 7: The X-ray diffraction experiment, Tables S4, S5, and S6. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: p.carloni@grs-sim.de.

Author Contributions

#The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. These authors contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

R.C. thanks ELETTRA for provision of synchrotron beam time for structural studies on hU and D.B. Belviso for crystal structure determination.

ABBREVIATIONS

CS, Conformational Selection; IF, Induced Fit; hU, human Ubiquitin; RP, Ramachandran Plot; CS_ω, Circular Spread of the angle ω; PAD_ω, Protein Angular Dispersion of the angle ω; $(PAD_{cum})_\omega$, cumulative Protein Angular Dispersion of the angle ω; PAI_ω, Phase Angular Index of the angle ω; ATI_ω, Angular Transition Index of the angle ω

REFERENCES

(1) Huber, R. Conformational flexibility in protein molecules. *Nature* **1979**, *280*, 538−539.

(2) Teilum, K; Olsen, J. G.; Kragelund, B. B. Protein stability, flexibility and function. *Biochim. Biophys. Acta* **2011**, *1814*, 969−976.

(3) Lin, J.-H. Accommodating protein flexibility for structure-based drug design. *Curr. Top. Med. Chem.* **2011**, *11*, 171−178.

(4) Nocker, M.; Cozzini, P. Induced fit simulations on nuclear receptors. *Curr. Top Med. Chem.* **2011**, *11*, 133−147.

(5) Fischer, E. Einfluss der Configuration auf die Wirkung der Enzyme. *Ber Dtsch Chem Ges*; WILEY-VCH Verlag: New York, 1894; Vol. *27*, pp 2985−2993.

(6) Tsai, C.-J.; Kumar, S.; Ma, B.; Nussinov, R. Folding funnels, binding funnels, and protein function. *Protein Sci.* **1999**, *8*, 1181−1190.

(7) Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **1958**, *44*, 98−104.

(8) Monod, J.; Wyman, J.; Changeux, J. P. On the Nature of Allosteric Transitions: A Plausible Model. *J. Mol. Biol.* **1965**, *12*, 88−118.

(9) Ma, B.; Kumar, S.; Tsai, C. J.; Nussinov, R. Folding funnels and binding mechanisms. *Protein Eng. Des. Sel.* **1999**, *12*, 713−720.

(10) Freire, E. Statistical Thermodynamic Linkage Between Conformational and Binding Equilibria. *Advances in Protein Chemistry*; Academic Press: New York, 1998; Vol. *51*, pp 255−279.

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

(12) Okazaki, K. i.; Takada, S. Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 11182−11187.

(13) Grant, B. J.; Gorfe, A. A.; McCammon, J. A. Large conformational changes in proteins: signaling and other functions. *Curr. Opin. Struct. Biol.* **2010**, *20*, 142−147.

(14) Boehr, D.; Nussinov, R.; Wright, P. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **2009**, *5*, 789−796.

(15) Hammes, G. G.; Chang, Y.-C.; Oas, T. G. Conformational selection or induced fit: A flux description of reaction mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 13737−13741.

(16) Clackson, T.; Wells, J. A. A hot spot of binding energy in a hormone-receptor interface. *Science* **1995**, *267*, 383−386.

(17) Thorn, K. S.; Bogan, A. A. ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics* **2001**, *17*, 284−285.

(18) DeLano, W. L. Unraveling hot spots in binding interfaces: progress and challenges. *Curr. Opin. Struct. Biol.* **2002**, *12*, 14−20.

(19) Gao, Y.; Wang, R.; Lai, L. Structure-based method for analyzing protein-protein interfaces. *J. Mol. Modeling* **2004**, *10*, 44−54.

(20) Guerois, R.; Nielsen, J. E.; Serrano, L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.* **2002**, *320*, 369−387.

(21) Kortemme, T.; Baker, D. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 14116−14121.
(22) González-Ruiz, D.; Gohlke, H. Targeting protein-protein interactions with small molecules: challenges and perspectives for computational binding epitope detection and ligand finding. *Curr. Med. Chem.* **2006**, *13*, 2607−2625.
(23) Rajamani, D.; Thiel, S.; Vajda, S.; Camacho, C. J. Anchor residues in protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 11287−11292.
(24) Halperin, I.; Wolfson, H.; Nussinov, R. Protein-Protein Interactions: Coupling of Structurally Conserved Residues and of Hot Spots across Interfaces. Implications for Docking. *Structure* **2004**, *12*, 1027−1038.
(25) Keskin, O.; Ma, B.; Nussinov, R. Hot regions in protein–protein interactions: the organization and contribution of structurally conserved hot spot residues. *J. Mol. Biol.* **2005**, *345*, 1281−1294.
(26) Ma, B.; Elkayam, T.; Wolfson, H.; Nussinov, R. Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 5772−5777.
(27) Ofran, Y.; Rost, B. Protein-protein interaction hotspots carved into sequences. *PLoS Comput. Biol.* **2007**, *3*, e119.
(28) Guney, E.; Tuncbag, N.; Keskin, O.; Gursoy, A. HotSprint: database of computational hot spots in protein interfaces. *Nucleic Acids Res.* **2008**, *36*, D662−666.
(29) Fernández, A. Epistructural Tension Promotes Protein Associations. *Phys. Rev. Lett.* **2012**, *108*, 188102.
(30) Bonvin, A. M. J. J. Flexible protein−protein docking. *Curr. Opin. Struct. Biol.* **2006**, *16*, 194−200.
(31) Sinha, N.; Nussinov, R. Point mutations and sequence variability in proteins: redistributions of preexisting populations. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 3139−3144.
(32) Luque, I.; Freire, E. Structural stability of binding sites: consequences for binding affinity and allosteric effects. *Proteins* **2000**, *Suppl 4*, 63−71.
(33) Schlessinger, A.; Yachdav, G.; Rost, B. PROFbval: predict flexible and rigid residues in proteins. *Bioinformatics* **2006**, *22*, 891−893.
(34) Yuan, Z.; Zhao, J.; Wang, Z.-X. Flexibility analysis of enzyme active sites by crystallographic temperature factors. *Protein Eng.* **2003**, *16*, 109−114.
(35) Fülöp, V.; Jones, D. T. β Propellers: structural rigidity and functional diversity. *Curr. Opin. Struct. Biol.* **1999**, *9*, 715−721.
(36) Bartlett, G. J.; Porter, C. T.; Borkakoti, N.; Thornton, J. M. Analysis of catalytic residues in enzyme active sites. *J. Mol. Biol.* **2002**, *324*, 105−121.
(37) Aleksandrov, A.; Simonson, T. Molecular dynamics simulations show that conformational selection governs the binding preferences of imatinib for several tyrosine kinases. *J. Biol. Chem.* **2010**, *285*, 13807−13815.
(38) Changeux, J.-P.; Edelstein, S. Conformational selection or induced fit? 50 years of debate resolved. *F1000 Biol. Rep.* **2011**, *3*, 19−34.
(39) Silva, D.-A.; Bowman, G. R.; Sosa-Peinado, A.; Huang, X. A role for both conformational selection and induced fit in ligand binding by the LAO protein. *PLoS Comput. Biol.* **2011**, *7*, e1002054.
(40) Stein, A.; Rueda, M.; Panjkovich, A.; Orozco, M.; Aloy, P. A systematic study of the energetics involved in structural changes upon association and connectivity in protein interaction networks. *Structure* **2011**, *19*, 881−889.
(41) Velazquez, H. A.; Hamelberg, D. Conformational selection in the recognition of phosphorylated substrates by the catalytic domain of human Pin1. *Biochemistry* **2011**, *50*, 9605−9615.
(42) Stanley, W. A.; Pursiainen, N. V.; Garman, E. F.; Juffer, A. H.; Wilmanns, M; Kursula, P. A previously unobserved conformation for the human PexSp receptor suggests roles for intrinsic flexibility and rigid domain motions in ligand binding. *BMC Struct. Biol.* **2007**, *7*, 24.
(43) Gunasekaran, K.; Nussinov, R. How Different are Structurally Flexible and Rigid Binding Sites? Sequence and Structural Features Discriminating Proteins that Do and Do not Undergo Conformational Change upon Ligand Binding. *J. Mol. Biol.* **2007**, *365*, 257−273.
(44) Yan, B.; Takahashi, T.; Johnson, R.; Spudich, J. L. Identification of signaling states of a sensory receptor by modulation of lifetimes of stimulus-induced conformations: the case of sensory rhodopsin II. *Biochemistry* **1991**, *30*, 10686−10692.
(45) Fisher, N. *Statistical Analysis of Circular Data*; Cambridge University Press: Cambridge, U. K., 1993; pp 1−277.
(46) Wüthrich, K. *NMR in Structural Biology: A Collection of Papers by Kurt Wüthrich*; World Scientific: River Edge, NJ, 1995; pp 1−760.
(47) Winget, J. M.; Mayor, T. The diversity of ubiquitin recognition: hot spots and varied specificity. *Mol. Cell* **2010**, *38*, 627−635.
(48) Dikic, I.; Wakatsuki, S.; Walters, K. J. Ubiquitin-binding domains [mdash] from structures to functions. *Nat. Rev. Mol. Cell. Biol.* **2009**, *10*, 659−671.
(49) Kerrien, S.; Aranda, B.; Breuza, L.; Bridge, A.; Broackes-Carter, F.; Chen, C.; Duesbury, M.; Dumousseau, M.; Feuermann, M.; Hinz, U.; Jandrasits, C.; Jimenez, R. C.; Khadake, J.; Mahadevan, U.; Masson, P.; Pedruzzi, I.; Pfeiffenberger, E.; Porras, P.; Raghunath, A.; Roechert, B.; Orchard, S.; Hermjakob, H. The IntAct molecular interaction database in 2012. *Nucleic Acids Res.* **2012**, *40*, D841−846.
(50) Wlodarski, T.; Zagrovic, B. Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 19346−19351.
(51) Lange, O. F.; Lakomek, N.-A.; Farès, C.; Schröder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmüller, H.; Griesinger, C.; de Groot, B. L. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* **2008**, *320*, 1471−1475.
(52) Fenwick, R. B.; Esteban-Martín, S.; Richter, B.; Lee, D.; Walter, K. F. A.; Milovanovic, D.; Becker, S.; Lakomek, N. A.; Griesinger, C.; Salvatella, X. Weak Long-Range Correlated Motions in a Surface Patch of Ubiquitin Involved in Molecular Recognition. *J. Am. Chem. Soc.* **2011**, *133*, 10336−10339.
(53) Leslie, A. G. W. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*; 1992; No. *26*.
(54) Evans, P. An introduction to stereochemical restraints. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2007**, *63*, 58−61.
(55) Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L-W; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr, Sect D: Biol Crystallogr.* **2002**, *58*, 1948−1954.
(56) Caliandro, R.; Carrozzini, B.; Cascarano, G. L.; De Caro, L.; Giacovazzo, C; Mazzone, A. M.; Siliqi, D. Molecular replacement: the approach of the program REMO. *J. Appl. Crystallogr.* **2006**, *39*, 185−193.
(57) Burla, M. C.; Caliandro, R.; Camalli, M.; Carrozzini, B.; Cascarano, G. L.; De Caro, L.; Giacovazzo, C.; Polidori, G.; Siliqi, D; Spagna, R. IL MILIONE: a suite of computer programs for crystal structure solution of proteins. *J. Appl. Crystallogr.* **2007**, *40*, 609−613.
(58) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.* **1987**, *194*, 531−544.
(59) Cohen, S. X.; Ben Jelloul, M.; Long, F.; Vagin, A.; Knipscheer, P.; Lebbink, J.; Sixma, T. K.; Lamzin, V. S.; Murshudov, G. N.; Perrakis, A. ARP/wARP and molecular replacement: the next generation. *Acta Crystallogr, Sect. D* **2008**, *64*, 49−60.
(60) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126−2132.
(61) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1997**, *53*, 240−255.
(62) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials

for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950−1958.
(63) Aaqvist, J. Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.* **1990**, *94*, 8021−8024.
(64) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(65) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10089.
(66) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(67) Nose, S.; Klein, M. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983**, *50*, 1055−1076.
(68) Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **1980**, *72*, 2384−2393.
(69) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(70) Ramage, R.; Green, J.; Muir, T. W.; Ogunjobi, O. M.; Love, S.; Shaw, K. Synthetic, structural and biological studies of the ubiquitin system: the total chemical synthesis of ubiquitin. *Biochem. J.* **1994**, *299* (Pt 1), 151−158.
(71) Kitahara, R.; Tanaka, T.; Yamashita, M.; Araya, K.; Yokoyama, S.; Akasaka, K.; Taniguchi, Y.; Kato, M. Structure of Ubiquitin crystallized under high pressure. To be published.
(72) Willis, B. T. M.; Pryor, A. W. *Thermal Vibrations in Crystallography*; Cambridge University Press: London, 1975.
(73) Mardia, K.; Jupp, P. Directional Statistics. *Fundamental Theorems and Distribution Theory*; John Wiley & Sons, Inc.: New York, 2008; pp 57−82.
(74) Rossetti, G.; Cong, X.; Caliandro, R.; Legname, G.; Carloni, P. Common Structural Traits across Pathogenic Mutants of the Human Prion Protein and Their Implications for Familial Prion Diseases. *J. Mol. Biol.* **2011**, *411*, 700−712.
(75) Hurst, H. Long Term Storage Capacity of Reservoirs. *Am. Soc. Civ. Eng.* **1951**, *116*, 770−799.
(76) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14* (33−38), 27−38.
(77) Perica, T.; Chothia, C. Ubiquitin–molecular mechanisms for recognition of different structures. *Curr. Opin. Struct. Biol.* **2010**, *20*, 367−376.
(78) Osborne, M. J.; Schnell, J; Benkovic, S. J.; Dyson, H. J.; Wright, P. E. Backbone dynamics in dihydrofolate reductase complexes: role of loop flexibility in the catalytic mechanism. *Biochemistry.* **2001**, *40*, 9846−9859.
(79) Komander, D.; Reyes-Turcu, F.; Licchesi, J. D. F.; Odenwaelder, P.; Wilkinson, K. D.; Barford, D. Molecular discrimination of structurally equivalent Lys 63-linked and linear polyubiquitin chains. *EMBO Rep.* **2009**, *10*, 466−473.
(80) Kamadurai, H. B.; Souphron, J.; Scott, D. C.; Duda, D. M.; Miller, D. J.; Stringer, D.; Piper, R. C.; Schulman, B. A. Insights into ubiquitin transfer cascades from a structure of a UbcH5B approximately ubiquitin-HECT(NEDD4L) complex. *Mol. Cell* **2009**, *36*, 1095−1102.
(81) Friedland, G. D.; Lakomek, N.-A.; Griesinger, C.; Meiler, J.; Kortemme, T. A Correspondence Between Solution-State Dynamics of an Individual Protein and the Sequence and Conformational Diversity of its Family. *PLoS Comput. Biol.* **2009**, *5*, e1000393.
(82) Weeks, S. D.; Grasty, K. C.; Hernandez-Cuebas, L; Loll, P. J. Crystal structures of Lys-63-linked tri- and di-ubiquitin reveal a highly extended chain architecture. *Proteins: Struct., Funct., Bioinf.* **2009**, *77*, 753−759.
(83) Wang, M.; Cheng, D.; Peng, J.; Pickart, C. M. Molecular determinants of polyubiquitin linkage selection by an HECT ubiquitin ligase. *EMBO J.* **2006**, *25*, 1710−1719.
(84) Gerstein, M.; Chothia, C. Analysis of protein loop closure. Two types of hinges produce one motion in lactate dehydrogenase. *J. Mol. Biol.* **1991**, *220*, 133−149.
(85) Joseph, D.; Petsko, G. A.; Karplus, M. Anatomy of a conformational change: hinged ″lid″ motion of the triosephosphate isomerase loop. *Science* **1990**, *249*, 1425−1428.
(86) Kumar, S.; Ma, B.; Tsai, C. J.; Wolfson, H.; Nussinov, R. Folding funnels and conformational transitions via hinge-bending motions. *Cell Biochem. Biophys.* **1999**, *31*, 141−164.
(87) Keller, B. G.; Prinz, J.-H.; Noé, F. Markov models and dynamical fingerprints: Unraveling the complexity of molecular kinetics. *Chem. Phys.* **2012**, *396*, 92−107.
(88) Yu, H.; Gupta, A. N.; Liu, X.; Neupane, K.; Brigley, A. M.; Sosova, I.; Woodside, M. T. Energy landscape analysis of native folding of the prion protein yields the diffusion constant, transition path time, and rates. *Proc. Natl. Acad. Sci. U. S. A.* **2012**.
(89) Chiang, T.-H.; Hsu, D.; Latombe, J.-C. Markov dynamic models for long-timescale protein motion. *Bioinformatics* **2010**, *26*, i269−277.
(90) Meyer, T.; D'Abramo, M.; Hospital, A.; Rueda, M.; Ferrer-Costa, C.; Pèrez, A.; Carrillo, O.; Camps, J.; Fenollosa, C.; Repchevsky, D.; Gelpì, J. L.; Orozco, M. MoDEL (Molecular Dynamics Extended Library): A†Database of Atomistic Molecular Dynamics Trajectories. *Structure* **2010**, *18*, 1399−1409.
(91) Burroughs, A. M.; Balaji, S.; Iyer, L. M.; Aravind, L. Small but versatile: the extraordinary functional and structural diversity of the beta-grasp fold. *Biol. Direct* **2007**, *2*, 18−30.
(92) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577−2637.
(93) Nicastro, G.; Masino, L.; Esposito, V.; Menon, R. P.; De Simone, A.; Fraternali, F.; Pastore, A. Josephin domain of ataxin-3 contains two distinct ubiquitin-binding sites. *Biopolymers* **2009**, *91*, 1203−1214.