JACS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

Structural Characterization of the Early Events in the Nucleation–Condensation Mechanism in a Protein Folding Process

Predrag Kukic,[†] Yulia Pustovalova,[‡] Carlo Camilloni,[†,⊥] Stefano Gianni,[§] Dmitry M. Korzhnev,[‡] and Michele Vendruscolo*,[†]

[†]Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, U.K.

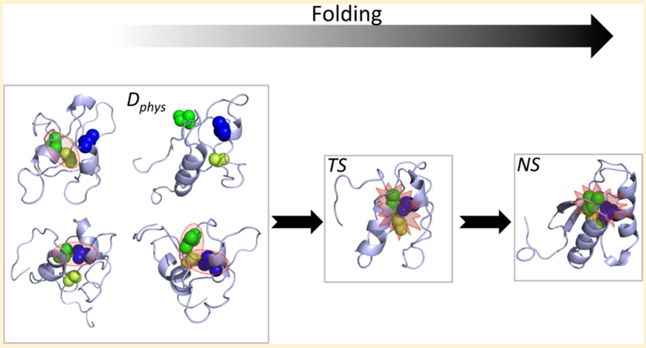
[‡]Department of Molecular Biology and Biophysics, University of Connecticut Health Center, Farmington, Connecticut 06030, United States

[§]Istituto Pasteur - Fondazione Cenci Bolognetti and Istituto di Biologia e Patologia Molecolari del CNR, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli", Sapienza Università di Roma, Rome 00185, Italy

[⊥]Technische Universität Muṅchen Institute for Advanced Study & Department of Chemistry, Lichtenbergstr. 4, 85748 Garching, Germany

Supporting Information

ABSTRACT: The nucleation–condensation mechanism represents a major paradigm to understand the folding process of many small globular proteins. Although substantial evidence has been acquired for this mechanism, it has remained very challenging to characterize the initial events leading to the formation of a folding nucleus. To achieve this goal, we used a combination of relaxation dispersion NMR spectroscopy and molecular dynamics simulations to determine ensembles of conformations corresponding to the denatured, transition, and native states in the folding of the activation domain of human procarboxypeptidase A2 (ADA2h). We found that the residues making up the folding nucleus tend to interact in the denatured state in a transient manner and not simultaneously, thereby forming incomplete and distorted versions of the folding nucleus. Only when all the contacts between these key residues are eventually formed can the protein reach the transition state and continue folding. Overall, our results elucidate the mechanism of formation of the folding nucleus of a protein and provide insights into how its folding rate can be modified during evolution by mutations that modulate the strength of the interactions between the residues forming the folding nucleus.



INTRODUCTION

A full description of a molecular reaction involves a detailed structural characterization of all the relevant states populated during the process.[1,2] In protein folding, this goal is particularly challenging as there are only small free energy differences between the different states, which are thus difficult to isolate and study on their own.[3] Furthermore, these states are often not well represented by individual conformations since they are made up by heterogeneous ensembles of rapidly interconverting structures, with the denatured state being typically the most structurally diverse state.[3−21] In fact, while transition states of different protein systems have been extensively analyzed experimentally and theoretically,[16−19] unveiling the structures of denatured states under physiological conditions has been considerably harder. In this context, it is important to emphasize the distinction between the denatured state under conditions that favor folding (D_{phys}), which is a transient species on-pathway to the folded state, and the unfolded state (U), which is populated at high concentrations of denaturant or at high temperatures.[3] Characterizing the structural and dynamic properties of denatured states under physiological conditions, D_{phys}, is essential to identify the early events in the protein folding process, but it typically requires disfavoring the population of the native state without the addition of denaturants. Reduction of disulfide bridges,[22] removal of cofactors,[23] truncation of residues at the termini of the protein,[24] or site-directed mutagenesis[25−27] are some of the approaches used so far. Overall, however, our current knowledge of the D_{phys} is still relatively incomplete, as it is limited to only a few examples.[22−26,28−31]

The recent development of relaxation dispersion nuclear magnetic resonance (NMR) spectroscopy has allowed the atomistic characterization of low populated excited states and provided an excellent tool to characterize D_{phys} under conditions favoring folding.[32−35] Although the signal of such weakly populated states cannot be directly detected in the NMR spectra, their presence leads to the broadening of peaks

Received: February 13, 2017
Published: April 12, 2017

DOI: 10.1021/jacs.7b01540

Table 1. List of RDCs Measured for *N* and D_{phys} States[a]

	$\Delta\varpi_{DN}$ no	$\Delta\varpi_{DN}$	J_{NH}	J_{NH}	D_N	ΔD_{DN}	D_D
#	align.	Pf1 19Hz	no align.	Pf1 19Hz	J(Pf1)-J		$D_N+\Delta D_{DN}$
7	-	-	-92.692	-91.587	1.105	-	-
8	**+0.50±0.02**	**0.433±0.014**	**-92.371**	**-91.914**	**0.457**	**-0.454±3.421**	**0.003±3.421**
9	**-0.86±0.02**	**-0.617±0.011**	**-93.803**	**-93.958**	**-0.155**	**-9.344±2.783**	**-9.499±2.783**
10	-0.31±0.04	-	-92.785	-92.618	0.166	-	-
11	**+3.46±0.04**	**3.430±0.037**	**-92.854**	**-88.761**	**4.092**	**4.135±9.560**	**8.227±9.560**
12	-0.28±0.04	-	-92.528	-88.361	4.167	-	-
13	**-1.42±0.02**	**-1.250±0.014**	**-93.309**	**-93.126**	**0.183**	**4.713±3.591**	**4.896±3.591**
14	**-0.65±0.02**	**-0.526±0.012**	**-91.788**	**-97.861**	**-6.073**	**2.130±2.774**	**-3.943±2.774**
15	**-3.51±0.04**	**-**	**-93.25**	**-101.594**	**-8.344**	**-**	**-**
16	-1.25±0.02	-1.286±0.017	-93.174	-103.786	-10.612	-2.838±2.669	-13.450±2.669
18	**-1.95±0.02**	**-1.645±0.015**	**-91.73**	**-100.565**	**-8.835**	**-0.995±3.431**	**-9.830±3.431**
19	+5.11±0.06	-	-93	-89.574	3.426	-	-
21	**-2.68±0.03**	**-2.305±0.023**	**-93.367**	**-90.252**	**3.115**	**-3.076±4.344**	**0.039±4.344**
22	**+1.93±0.02**	**1.885±0.018**	**-93.445**	**-91.258**	**2.187**	**4.984±3.251**	**7.171±3.251**
23	**+1.70±0.02**	**-**	**-92.486**	**-86.54**	**5.947**	**-**	**-**
24	**+2.49±0.03**	**-**	**-93.39**	**-87.983**	**5.407**	**-**	**-**
25	**+0.62±0.02**	**0.746±0.011**	**-93.631**	**-90.552**	**3.08**	**5.498±2.447**	**8.578±2.447**
26	**+0.33±0.03**	**-**	**-93.726**	**-91.81**	**1.916**	**-**	**-**
27	**+1.05±0.02**	**1.233±0.014**	**-93.565**	**-87.324**	**6.24**	**-2.471±3.598**	**3.769±3.598**
28	**+0.33±0.03**	**-**	**-93.204**	**-89.568**	**3.636**	**-**	**-**
29	**+0.82±0.02**	**0.766±0.010**	**-93.085**	**-91.578**	**1.508**	**2.181±2.301**	**3.689±2.301**
30	**+2.72±0.03**	**-**	**-92.836**	**-89.496**	**3.34**	**-**	**-**
31	**+5.27±0.06**	**-**	**-92.658**	**-88.685**	**3.972**	**-**	**-**
32	**+0.36±0.03**	**0.600±0.012**	**-93.083**	**-91.851**	**1.232**	**-1.238±3.280**	**-0.006±3.280**
35	**+2.76±0.03**	**2.751±0.035**	**-92.258**	**-87.598**	**4.66**	**3.461±5.736**	**8.121±5.736**
36	**+5.92±0.12**	**-**	**-92.695**	**-89.614**	**3.081**	**-**	**-**
37	**+6.24±0.09**	**-**	**-92.85**	**-87.091**	**5.759**	**-**	**-**
38	**-0.51±0.02**	**-**	**-93.251**	**-93.829**	**-0.579**	**-**	**-**
39	**-5.33±0.07**	**-**	**-93.104**	**-91.978**	**1.126**	**-**	**-**
40	**-8.90±0.12**	**-**	**-93.341**	**-100.005**	**-6.663**	**-**	**-**
41	**+2.93±0.03**	**-**	**-91.617**	**-102.175**	**-10.558**	**-**	**-**
44	-0.54±0.04	-0.255±0.033	-92.06	-91.949	0.111	-9.663±7.494	-9.552±7.494
45	**+4.18±0.07**	**-**	**-92.683**	**-86.134**	**6.55**	**-**	**-**
47	**-2.07±0.02**	**-1.789±0.023**	**-93.741**	**-101.358**	**-7.617**	**-3.817±3.598**	**-11.434±3.598**
48	**+0.59±0.02**	**0.486±0.014**	**-93.095**	**-88.635**	**4.46**	**0.751±2.936**	**5.211±2.936**
49	-0.83±0.02	-0.642±0.015	-93.389	-98.485	-5.096	-3.138±3.263	-8.234±3.263
50	**+2.41±0.03**	**-**	**-92.438**	**-100.972**	**-8.534**	**-**	**-**
51	**-1.33±0.02**	**-**	**-92.044**	**-96.63**	**-4.586**	**-**	**-**
52	**-1.05±0.02**	**-0.939±0.011**	**-93.699**	**-96.592**	**-2.893**	**-2.793±2.959**	**-5.686±2.959**
53	**-4.86±0.05**	**-**	**-92.74**	**-95.251**	**-2.511**	**-**	**-**
54	**-5.99±0.07**	**-**	**-92.986**	**-89.313**	**3.673**	**-**	**-**
56	**-5.06±0.06**	**-**	**-92.384**	**-90.282**	**2.102**	**-**	**-**

Table 1. continued

#	$\Delta\varpi_{DN}$ no align.	$\Delta\varpi_{DN}$ Pf1 19Hz	J_{NH} no align.	J_{NH} Pf1 19Hz	D_N J(Pf1)-J	ΔD_{DN}	D_D D_N+ΔD_{DN}
57	**+6.91±0.08**	-	**-92.835**	**-94.047**	**-1.212**	-	-
58	+4.73±0.05	-	-91.809	-96.584	-4.775	-	-
59	**-1.28±0.02**	**-1.158±0.017**	**-93.36**	**-84.403**	**8.957**	**-13.839±2.908**	**-4.882±2.908**
60	**+2.20±0.02**	**2.172±0.026**	**-93.396**	**-87.409**	**5.987**	**2.015±4.330**	**8.002±4.330**
61	**+2.17±0.02**	-	**-93.553**	**-87.377**	**6.177**	-	-
62	**-0.29±0.04**	-	**-93.604**	**-84.683**	**8.921**	-	-
63	**+2.12±0.02**	-	**-93.696**	**-86.283**	**7.413**	-	-
64	**+1.81±0.02**	**1.970±0.020**	**-92.569**	**-87.513**	**5.056**	**-2.859±4.232**	**2.197±4.232**
65	**+0.73±0.02**	**0.740±0.015**	**-93.685**	**-86.667**	**7.018**	**-5.198±2.859**	**1.820±2.859**
66	**+2.52±0.03**	**2.642±0.029**	**-93.508**	**-84.771**	**8.737**	**8.700±6.838**	**17.437±6.838**
67	**+1.79±0.02**	**1.893±0.017**	**-93.397**	**-87.219**	**6.178**	**6.955±3.250**	**13.133±3.250**
68	**-0.89±0.02**	**-0.825±0.013**	**-93.31**	**-87.841**	**5.47**	**-6.065±2.514**	**-0.595±2.514**
70	**-0.54±0.02**	**-0.185±0.031**	**-94.353**	**-95.302**	**-0.949**	**-3.639±7.088**	**-4.588±7.088**
71	**-2.14±0.02**	-	**-92.818**	**-94.85**	**-2.032**	-	-
72	**-4.56±0.05**	-	**-93.374**	**-92.265**	**1.109**	-	-
73	**+3.16±0.03**	-	**-93.519**	**-101.829**	**-8.31**	-	-
74	**+2.69±0.03**	-	**-92.915**	**-99.872**	**-6.958**	-	-
75	-1.19±0.02	-1.095±0.016	-92.737	-96.577	-3.84	-9.651±3.892	-13.491±3.892
76	**-4.84±0.05**	-	**-93.154**	**-91.061**	**2.093**	-	-
77	+3.29±0.04	-	-92.227	-93.04	-0.814	-	-
78	**-0.81±0.02**	**-0.664±0.011**	**-92.722**	**-96.028**	**-3.306**	**0.091±2.799**	**-3.215±2.799**
79	**+1.30±0.02**	-	**-92.409**	**-96.49**	**-4.081**	-	-
80	**-2.03±0.02**	**-1.808±0.017**	**-92.712**	**-95.842**	**-3.13**	**3.798±4.216**	**0.668±4.216**
81	**-1.11±0.02**	**-0.911±0.009**	**-92.302**	**-94.983**	**-2.68**	**0.395±1.955**	**-2.285±1.955**

[a]Chemical shift differences $\Delta\varpi_{DN}$ between the D_{phys} and N states (with and without alignment), J_{NH} couplings (with and without alignment), RDC differences ΔD_{DN} between the D_{phys} and N states, RDCs of the N state, D_N, and RDCs of the D_{phys} state, D_D, for the $^{15}N/^{2}H$ labeled I71V ADA2h weakly aligned in Pf1 phage solution obtained from ^{15}N spin-state selective CPMG data[40] and IPAP experiments.[52] For the reference, column no. 2 shows (signed) $\Delta\varpi_{DN}$ values obtained previously for the $^{15}N/^{13}C/^{2}H$ labeled I71V ADA2h sample without alignment.[42] Shown in **bold** are residues with $\Delta\varpi_{DN}$ signs determined experimentally, as described elsewhere.[42] In all other cases, signs of $\Delta\varpi_{DN}$ were assumed to be the same as $\varpi_{RC}-\varpi_N$, where ϖ_{RC} is predicted random coil chemical shift, as described in detail elsewhere.[42] ^{15}N $\Delta\varpi_{DN}$ and ΔD_{DN} values measured in Pf1 phage solution are shown for a subset of 31 residues whose ΔD_{DN} can be determined with uncertainties <10 Hz if 800 and 500 MHz CPMG data were fitted together (marked in red) or <5 Hz if only 800 MHz CPMG data were available. For the most conservatively selected set of 14 residues (underlined), (i) spin-state selective CPMG data were available at two magnetic fields, (ii) errors in both $\Delta D_{DN,800}$ and $\Delta D_{DN,500}$ obtained from independent fits of 500 and 800 MHz CPMG data were less than 15 Hz (see Materials and Methods for details), (iii) $\Delta D_{DN,800}$ and $\Delta D_{DN,500}$ values were within one standard deviation from each other, and (iv) experimental $\Delta\varpi_{DN}$ signs (and thus ΔD_{DN} signs) were available.

of the ground state that can be quantified in NMR relaxation dispersion experiments to elucidate structure and thermodynamics of these elusive excited conformers.[33,36−41] Here, we used this methodology to study the structural features of D_{phys} for the destabilizing I71V mutant of the activation domain of human procarboxypeptidase A2 (ADA2h) under conditions that favor folding.[42] ADA2h provides a convenient and interesting model system because of the extensive experimental characterization of its unfolding and refolding kinetics[43−46] and its ability to form amyloid fibrils with aggregation kinetics related to the structural properties in its denatured state.[47,48] Furthermore, this protein represents a prototypical example of the so-called nucleation-condensation mechanism, a general mechanism for protein folding implying the concurrent formation of secondary and tertiary structure around a specific nucleus formed by a small number of key residues.[19,49,50]

To provide a complete structural description of a protein folding process via two-state nucleation−condensation mechanism, here we report the characterization of all the major states along the folding pathway of ADA2h. By measuring $^{1}H^{N}-^{15}N$ residual dipolar couplings (RDCs) in combination with extensive sets of ^{15}N, $^{1}H^{N}$, $^{13}C^{\alpha}$, $^{1}H^{\alpha}$, $^{13}C'$ NMR chemical shifts in both the native (N) and D_{phys} states of I71V ADA2h at physiological conditions, we were able to determine structural ensembles of these states at nearly atomic resolution. Furthermore, by using previously published kinetic data on a series of site-directed mutants[45] as restrains in molecular dynamics simulations,[19] we determined the structure of the transition state for folding. The direct comparison of structural

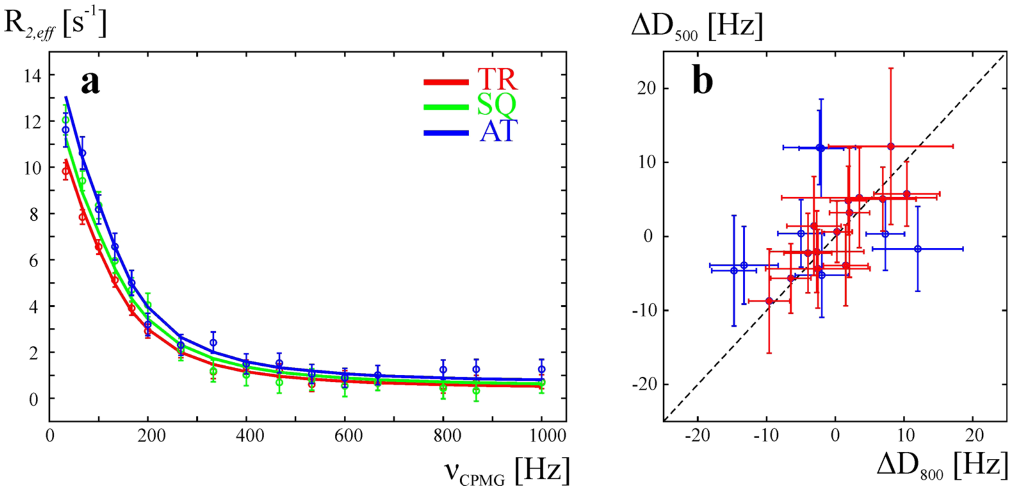


Figure 1. Measurements of RDCs of I71V ADA2h in the *N* and D_{phys} states. (a) Example of relaxation profiles recorded for Val 59. ^{15}N TR, SQ, and AT CPMG dispersion profiles recorded at 800 MHz (^{1}H) spectrometer (circles) and their best fits to the two-state exchange model (lines) for the residue #59 of the I71V ADA2h weakly aligned in Pf1 phage solution. Relaxation dispersion profiles recorded at 500 and 800 MHz (^{1}H) spectrometers were fit together with k_{ex} and p_D values fixed to 522 1/s and 4.31%, respectively. (b) Correlation of ΔD_{DN} values obtained from fits of ^{15}N TR/AT CPMG data recorded at 500 MHz (*y* axis) and 800 MHz (*x* axis) spectrometers (^{15}N SQ CPMG data at both magnetic fields were included in the fits). We show 22 residues with (i) data available at two magnetic fields and (ii) errors in both $\Delta D_{DN,800}$ and $\Delta D_{DN,500}$ less than 15 Hz. For all these residues $\Delta D_{DN,800}$ and $\Delta D_{DN,500}$ values fall within two standard deviations from each other. Shown in red is a subset of 14 residues with (i) data available at two magnetic fields, (ii) errors in both $\Delta D_{DN,800}$ and $\Delta D_{DN,500}$ less than 15 Hz, (iii) $\Delta D_{DN,800}$ and $\Delta D_{DN,500}$ within one standard deviation from each other, and (iv) experimental ΔD_{DN} signs available. The remaining eight residues are shown in blue. Note that two residues (16 and 49) satisfy criteria i−iii, but have $\Delta\varpi_{DN}$ signs assigned according to $\varpi_{RC}-\varpi_N$ (see Materials and Methods, and Table 1).

features of the D_{phys} and transition states reported here highlights the role of the residual structure of D_{phys} in governing unfolding and refolding kinetics of ADA2h and allows us to depict the characteristic structural features of the early events in the nucleation−condensation mechanism of ADA2h folding.

■ MATERIALS AND METHODS

Protein Samples. Uniformly ^{15}N/^{2}H-labeled I71V ADA2h was expressed and purified as described previously.[42] The final NMR samples contained 0.5−0.7 mM protein, 50 mM sodium phosphate, 90% H_2O/10% D_2O, pH 7.6 with or without addition of 16 mg/mL of Pf1 phage cosolvent (Alsa Biotech).

Chemical Shifts of the *N* and D_{phys} States of I71V ADA2h. Nearly complete sets of the backbone ^{15}N, ^{1}H^N, ^{13}C$^\alpha$, ^{1}H$^\alpha$, ^{13}C′ chemical shifts in the *N* and D_{phys} states were reported in our previous work.[42]

RDCs of the *N* and D_{phys} States of I71V ADA2h. RDCs for the backbone amide groups of the *N* state of I71V ADA2h were obtained from the ^{1}H^N−^{15}N IPAP experiments[51,52] recorded with and without partial alignment of the protein in Pf1 phage solution at 40 °C on 800 MHz (^{1}H) Agilent VNMRS spectrometer (Table 1). The backbone amide RDCs of the low-populated D_{phys} state of I71V ADA2h weakly aligned in Pf1 phage solution were obtained from single-quantum (SQ) ^{15}N CPMG experiment[53] and spin-state selective ^{15}N TROSY (TR) and anti-TROSY (AT) CPMG experiments[40] recorded at 40 °C on 500 and 800 MHz (^{1}H) Agilent VNMRS spectrometers equipped with cold probes. Each of the six CPMG relaxation dispersion experiments (SQ, TR, and AT at 500 and 800 MHz (^{1}H) spectrometers) included a series of 2D spectra collected at 16 CPMG frequencies, $\nu_{CPMG} = 1/(2\delta)$ (where δ is the delay between 180° refocusing CPMG pulses), ranging from 33−1000 Hz. Peak intensities in 2D spectra were converted into the effective relaxation rates as $R_{2,eff}(\nu_{CPMG}) = -1/T_{relax} \ln(I_1(\nu_{CPMG})/I_0)$, where T_{relax} = 30 ms is a constant relaxation period, and I_0 represents peak intensities obtained with relaxation delay T_{relax} omitted.[54] Errors of $R_{2,eff}$ rates were estimated from duplicate spectra recorded at the same ν_{CPMG}. Minimal errors of 2% or 0.2 s^{-1} and 3% or 0.3 s^{-1}, respectively, were assumed for ^{15}N SQ and ^{15}N TR/AT CPMG data collected from 800 MHz (^{1}H) spectrometer, and 3% or 0.3 s^{-1} and 5% or 0.5 s^{-1}, respectively, for ^{15}N SQ and ^{15}N TR/AT CPMG data recorded at 500 MHz (^{1}H) spectrometer. Additionally, to model magnetization evolution during ^{15}N SQ, TR and AT CPMG sequences (see further),[40] we have measured relaxation rates for ^{15}N longitudinal magnetization $R(N_z)$ and longitudinal two-spin order $R(2H_zN_z)$.[55,56]

RDC differences between the D_{phys} and *N* states, $|\Delta D_{DN}|$, were obtained from least-squares fits of experimental ^{15}N SQ, TR, and AT CPMG data recorded at two magnetic fields to theoretical values calculated by numerical modeling of magnetization evolution in corresponding pulse sequences, as described elsewhere.[40] In the first step, ^{15}N SQ CPMG dispersion profiles for a subset of 43 NH groups with all the data available at two magnetic fields were fit together to a model of two-state exchange between the *N* and denatured D_{phys} states to extract exchange rate constant k_{ex} (522 ± 7 1/s) and population of the denatured state p_D (4.31 ± 0.04%). In the second step, ^{15}N SQ, TR, and AT CPMG data for individual amide groups were fit with k_{ex} and p_D fixed to the previously determined values to extract chemical shift and RDC differences between the states, $|\Delta\varpi_{DN}|$ and $|\Delta D_{DN}|$. Figure 1a shows examples of ^{15}N SQ, TR, and AT CPMG profiles for a selected residue recorded at 800 MHz (^{1}H) spectrometer and their best fits. Note that the combined analysis of ^{15}N SQ, TR, and AT CPMG data provides information about relative signs of $\Delta\varpi_{DN}$ and ΔD_{DN}.[55,56] Therefore, ΔD_{DN} signs are available for the same residues that $\Delta\varpi_{DN}$ signs were determined as described previously[42] (bold in Table 1). Residues with high per-residue χ^2 target function obtained in ^{15}N SQ, TR, and AT CPMG data fits (2.5-times greater than the number of experimental data points in CPMG profiles; that is, χ^2 > 240 for the residues with 500 and 800 MHz data, χ^2 > 120 for the residues with 800 MHz data only) were excluded from the analysis. Table 1 shows $\Delta\omega_{DN}$ and ΔD_{DN} values for a subset of 31 residues whose ΔD_{DN} can be determined with reasonable uncertainties (<10 Hz if both 800 and 500 MHz data are available, red in Table 1; <5 Hz if only 800 MHz data are available, black in Table 1). RDCs of the D_{phys} state, D_D, were calculated from RDCs of the *N* state, D_N, and the measured ΔD_{DN} values as $D_D = D_N + \Delta D_{DN}$ (Table 1).

To illustrate the accuracy of the obtained ΔD_{DN} values, for those residues with SQ, TR, and AT CPMG data available at both magnetic strengths, we have calculated ΔD_{DN} from independent fits of ^{15}N TR/

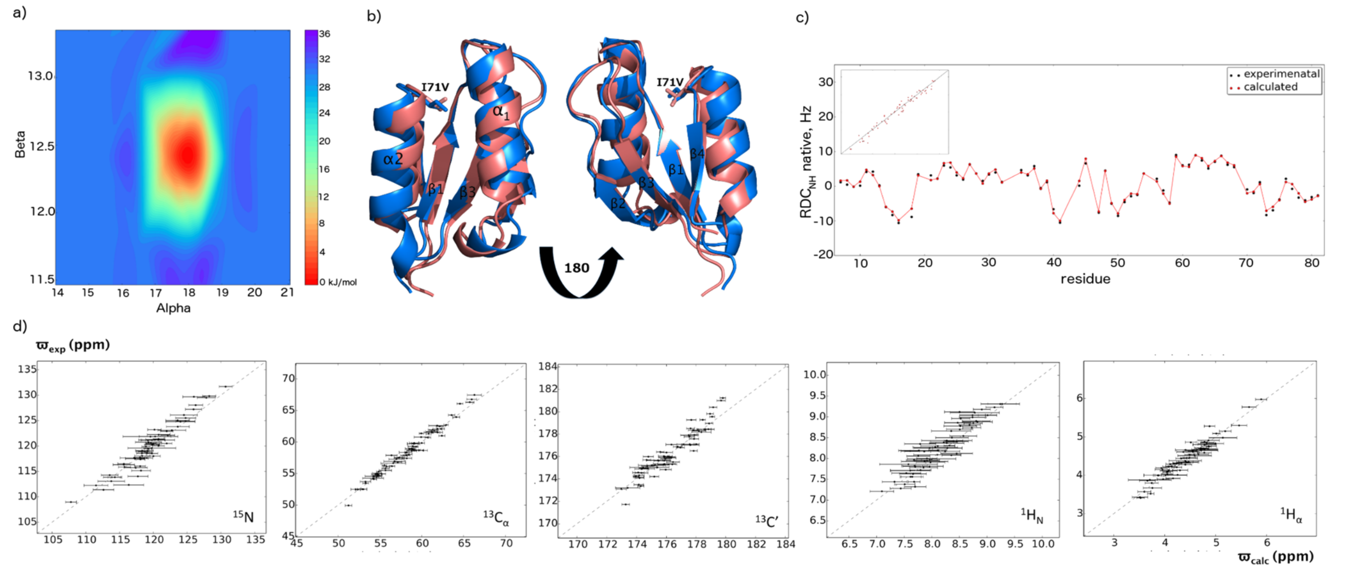


Figure 2. Determination of the structure and dynamics of the *N* state of I71V ADA2h using chemical shifts as restraints in RAM simulations. (a) Free energy landscape of the *N* state of I71V ADA2h. The free energy landscape is represented as a function of two collective variables that encode the fraction of α-helical (*x*-axis) and β-sheet (*y*-axis) secondary structure. (b) Lowest energy structure of I71V ADA2h (red) superimposed onto a X-ray structure of wild-type ADA2h (PDB ID: 1AYE, blue). (c) Validation of the *N* state ensemble by comparing 65 experimentally measured (black) and back-calculated (red) ^{1}H^{N}-^{15}N RDCs (*Q*-factor = 0.18). None of the experimental ^{1}H^{N}-^{15}N RDCs was used as a restraint in the RAM simulation. (d) Validation of the *N* state ensemble by correlating experimental (*y*-axis) and back-calculated (*x*-axis) ^{15}N, ^{13}Cα, ^{13}C′, ^{1}H$_N$, and ^{1}Hα chemical shifts (used as experimental restraints in the RAM simulation).

AT CPMG data recorded at 500 and 800 MHz (^{1}H) spectrometers (^{15}N SQ CPMG data at both magnetic fields were included in the fits). Figure 1b shows the comparison of ΔD_{DN} values obtained from 500 and 800 MHz TR/AT CPMG data for a subset of 22 residues with ΔD_{DN} uncertainties <15 Hz at both magnetic field strengths. For all 22 residues in this subset, ΔD_{DN} values obtained from 500 and 800 MHz data were within two standard deviations from each other, while for 16 of 22 residues 500 and 800 MHz ΔD_{DN} were within one standard deviation from each other. For a smaller subset of 16 residues, the uncertainties of ΔD_{DN} obtained from 800 MHz data were within 5 Hz. For 12 of these 16 residues, 500 and 800 MHz ΔD_{DN} values matched within one standard deviation.

Molecular Dynamics Simulations. Molecular dynamics simulations of I71V ADA2h were performed using the Amber ff99SB* force field[57] with the TIP3P[58] and TIP4P/2005[59] water models. All simulations were run using GROMACS 4.5[60] modified with PLUMED2.[61] The starting conformation was taken from an NMR-obtained structure, PDB ID: 1O6X[62] modified with I71V mutation. This structure was solvated with 26 090 water molecules and neutralized with 7 Na$^+$ ions in a water box of 800 nm^3 of volume. A high-temperature (500 K) 30 ns preliminary unfolding simulation was used to select four starting conformations for the D_{phys} state (Figure S1). Each conformation was then subsequently relaxed at 300 K for 10 ns. A time step of 2 fs was used together with LINCS constraints for all simulations.[63] The van der Waals interactions were implemented with a cutoff at 0.9 nm, and long-range electrostatic effects were treated with the particle mesh Ewald method.[64] All simulations were carried out in the canonical ensemble at constant volume and by thermosetting the system with the modified Berendsen thermostat.[65]

The RAM simulations[66] were carried out by combining two advanced sampling methods, replica exchange[67] and metadynamics.[68] First, replica exchange is particularly effective in overcoming the multiple minima problem on a rugged energy surface through the exchange of conformations between multiple replicas. Second, metadynamics, efficiently computes free energies and explores the reaction pathways in the space of specific functions of atomic coordinates, called collective variables (CVs).[68] In the present case, four CVs have been employed: the total α-helical content, the total β-sheet content, similarity of ψ and φ dihedral angles to a reference value, and the total number of contacts between side chains of Ile 15, Leu 26, and Val 52 (three residues that form the folding nucleus). The secondary structure elements and dihedral angles were chosen as CVs because they have been shown to capture to a good extent the relevant dynamics of both folded and disordered proteins.[69] The additional choice of the total number of contacts between the folding nucleus residues was chosen to capture the presence of the folding nucleus in the *N* and D_{phys} state.

By using the RAM scheme, four replicas of the two systems (*N* and D_{phys}) were simulated in parallel at 300 K with a restraint applied on the average value of the CamShift[70] back-calculated NMR chemical shifts:[66,69]

$$E_{CS} = \alpha \sum_{k=1}^{81} \sum_{l=1}^{5} \left(\delta_{kl}^{exp} - \frac{1}{M} \sum_{m=1}^{M} \delta_{klm}^{calc} \right)^2$$

where α is the force constant set to 24 kJ/(mol ppm^2), *k* runs over 81 residues in the protein, *l* runs over the five backbone nuclei for which the experimental chemical shifts have been measured (C$_\alpha$, C′, H$_\alpha$, H$_N$, and N) for *N* and D_{phys} states, and *m* runs over *M* = 4 replicas. In this way, the system evolves with a force field perturbed in such a way to increase agreement with the experimental chemically measured shifts and at the same time to satisfy the maximum entropy principle.[71] The number of replicas can be increased at expense of an increasing computational cost, but, in our experience, four replicas are sufficient to recover protein dynamics from chemical shifts with high accuracy. Each replica has been evolved for 250 ns, with exchange trials every 50 ps.

The convergence of the sampling was assessed by monitoring the differences of the free energies at increasing simulation length during the simulations. After the first 250 ns, the free energy landscapes were stable within 2.5 kJ/mol, which suggested that all the relevant minima in the landscape have been found, and the average changes in the free energy landscapes over the last 80 ns of simulations were below 1 kJ/mol. These results suggest that the free energies that we obtained from the RAM simulations are on average correct within 1.5 kJ/mol. The free energy landscapes of the *N* and D_{phys} states were reconstructed using a standard weighted histogram analysis. Lastly, the free energy

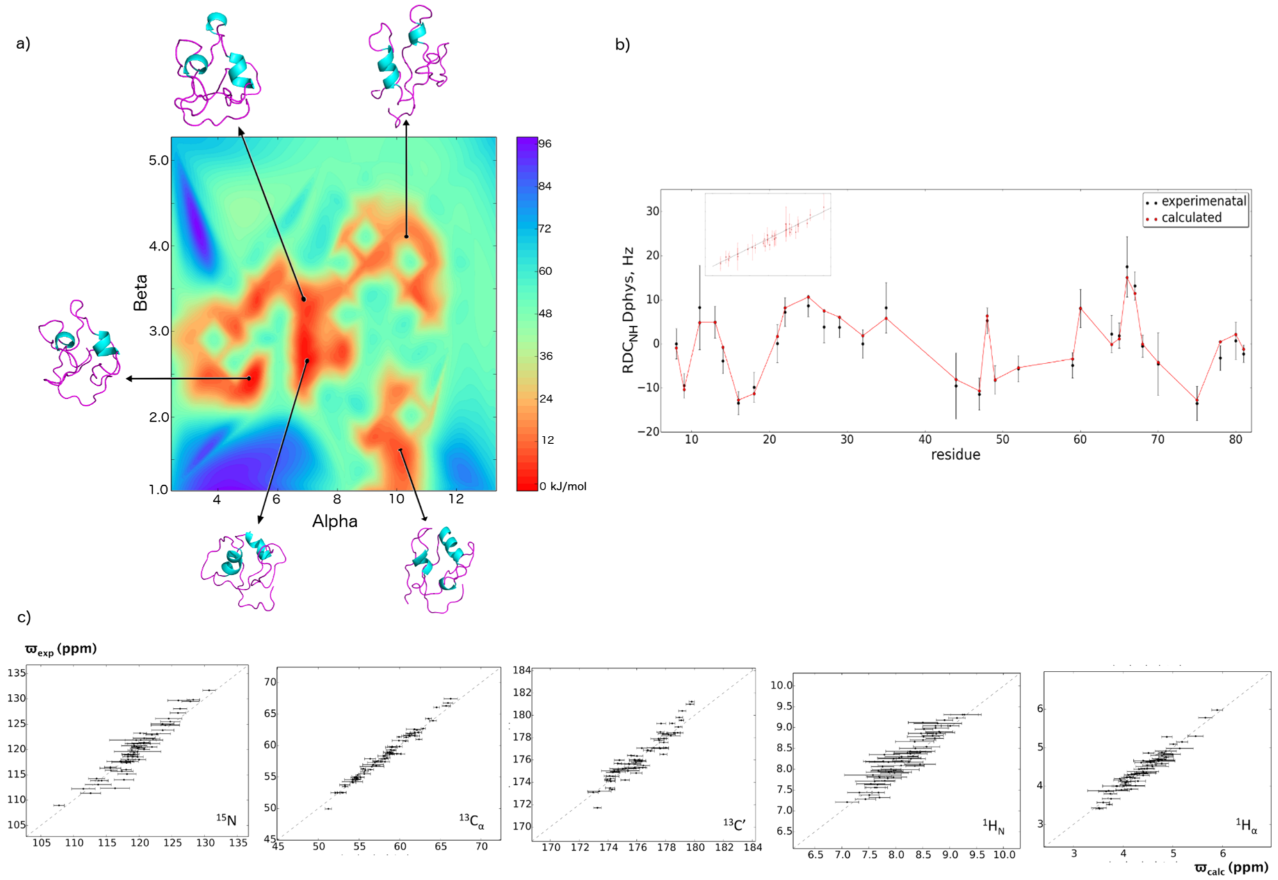


Figure 3. Determination of the structure and dynamics of the D_{phys} state of I71V ADA2h using chemical shifts and RDCs as structural restraints in RAM simulations. (a) Free energy landscape of the D_{phys} state of I71V ADA2h. The free energy landscape is represented as a function of two collective variables that encode the fraction of α helical (x-axis) and β sheet (y-axis) protein structure. (b) Validation of the D_{phys} state ensemble by comparing experimental (black) and back-calculated (red) $^1H^N$-^{15}N RDCs (Q-factor = 0.25). All 31 RDCs were used as restraints in the RAM simulation. (c) Validation of the D_{phys} state ensemble by correlating experimental (y-axis) and back-calculated (x-axis) ^{15}N, $^{13}C\alpha$, $^{13}C'$, ^1H_N, and $^1H\alpha$ chemical shifts (used as experimental restraints in the RAM simulation).

landscapes of the N and D_{phys} state as a function of two CVs are shown in Figures 2 and 3, respectively.

Calculation of S^2 Order Parameters from Simulations. Microstate ensembles were derived from the RAM simulation using the standard WHAM approach.[72] S^2 order parameters were calculated as averages over values predicted for each microstate ensemble using the following equation:[73]

$$S_k^{2,\text{ens}} = \frac{3}{2(r_k^{\text{eff}})^4}\left(\sum_{i=1}^{3}\sum_{j=1}^{3}\left[\frac{1}{N_{\text{struc}}}\sum_{l=1}^{N\text{struc}} r_{i,k,l}r_{j,k,l}\right]^2 - 1\right)$$

where $r_{i,k}$ is the ith Cartesian component of bond vector k, and N_{struc} is the number of structures in the microstate ensemble. In total, 75 S^2 order parameters were determined for the D_{phys} state, and they are shown in Figure S2.

Control RAM Simulations. As a stringent test of the RAM method for the characterization of D_{phys} of ADA2h, we first carried out four sets of RAM simulations that used four different subsets of experimental data as restraints and then cross-validated the resulting ensembles with experimental data not used as restraints in the simulations. These four control RAM simulations were restrained with (a) the NMR chemical shifts and the first subset of 16 RDC values (out of 31); (b) the NMR chemical shifts and the second subset containing the remaining 15 RDC values; (c) the complete set of the measured RDCs; and (d) the complete set of the measured RDCs and a subset of two-thirds of the measured NMR chemical shifts that was selected randomly. These simulations were then cross-validated with the RDC subsets that were not used as experimental restraints (control simulations a and b), and the NMR chemical shifts that were not used as experimental restraints (simulations c and d). The purpose of these tests was to examine whether the NMR chemical shifts, which are sensitive reporters of the secondary structure, and RDCs, which are sensitive reporters of the relative orientation of different structural elements, can yield an ensemble of structures that can reproduce structure and dynamics of the D_{phys} state with high accuracy. We found that the combination of the NMR chemical shifts and RDCs used as restrains resulted in a high correlation between predicted and control experimental RDCs that were not used as restraints (Figure S3), with a coefficient of correlation of 0.87 (p < 0.001) for the first RDC subset and 0.94 (p < 0.001) for the second subset. These correlations were significantly better than those obtained using the same RAM method without the experimental chemical shift and RDC restraints (coefficients of correlation of 0.08 and 0.5, respectively, Figure S3). Moreover, by using the measured RDCs as sole restraints in the RAM simulation, we were able to improve the agreement between the experimental and back-calculated chemical shifts (Figure S4; 1.85 ppm, ^{15}N; 0.76 ppm, $^{13}C\alpha$; 0.44 ppm, $^{13}C'$; 0.31 ppm, ^1H_N; 0.27, ppm $^1H\alpha$) in comparison to the values obtained using the same RAM method without the experimental chemical shift and the RDC restraints (Figure S4; 1.88 ppm, ^{15}N; 0.8 ppm, $^{13}C\alpha$; 0.75 ppm, $^{13}C'$; 0.39 ppm, ^1H_N; 0.42, ppm $^1H\alpha$). The improvement was even more significant when the measured RDCs and two-thirds of the randomly selected NMR chemical shifts were used as restraints, and the rest one-third of the NMR chemical shifts were used for benchmarking (Figure S4; 1.65 ppm, ^{15}N; 0.76 ppm, $^{13}C\alpha$; 0.4 ppm, $^{13}C'$; 0.3 ppm, ^1H_N; 0.24, ppm $^1H\alpha$).

DOI: 10.1021/jacs.7b01540
J. Am. Chem. Soc. 2017, 139, 6899−6910

Comparison of Simulations Using Two Different Water Models. As a further validation of the RAM simulations in studying the D_{phys} state, we examined a possible influence of the water model used in the molecular dynamics simulations on the properties of the D_{phys} ensemble of ADA2h I71V. This control is relevant since the choice of water model has been shown to be significant in structural characterizations of intrinsically disordered proteins,[13] but its role in structural analysis of denatured states under native conditions has not been explored yet. We hence generated two D_{phys} ensembles using the identical RAM method restrained with the experimentally measured NMR chemical shifts and the complete set of RDCs, and using two different water models: TIP3P[58] and TIP4P/2005.[59] As a control, we generated another D_{phys} ensemble using unrestrained RAM simulation with the TIP3P water model. The free energy surfaces of conformational ensembles generated in the three simulations are shown in Figure S5. Only relatively subtle differences were observed in the free energy surfaces obtained in the two restrained simulations with two different water models, while the free energy surface generated in the unrestrained simulation deviated significantly from those obtained in restrained ones. Thus, even though the two water models are significantly different, the chemical shift and RDC restraints were able to drive the RAM trajectories of the D_{phys} state of ADA2h I71V toward the same region of the conformational space, as required by the enforcement of the agreement with the experiments.

RESULTS

Measurements of RDCs of I71V ADA2h in *N* and D_{phys} States. We have recently reported an analysis of the NMR relaxation dispersion data for a destabilizing I71V mutant of ADA2h, which provided a nearly complete set of the backbone ^{15}N, ^{13}C and ^{1}H NMR chemical shifts in the *N* and D_{phys} states.[42] In addition to the extensive backbone chemical shift data, which enable modeling the structures of I71V ADA2h in the *N* and D_{phys} states, here we measured the RDCs for the backbone amide groups of the *N* and D_{phys} states of I71V ADA2h. The RDCs of the *N* state, D_N, were obtained from ^{1}H^{N}-^{15}N IPAP experiments[51,52] performed under conditions of fractional protein alignment in Pf1 phage solution. High-quality data resulted in a set of D_N values for 65 residues (Table 1). The RDCs of the D_{phys} state, D_D, obtained from D_N and RDC differences between the *N* and D_{phys} states, ΔD_{DN}, were measured using a combination of single-quantum[53] and spin-state selective TROSY and anti-TROSY[40] ^{15}N Carr−Purcell−Meiboom−Gill (CPMG) dispersion experiments performed on ^{15}N/^{2}H I71V ADA2h (see Materials and Methods). High-quality CPMG dispersion data resulted in a set of D_D values for 31 residues determined with uncertainties less than 10 Hz that were used in structure calculation of the D_{phys} state. As an example, the recorded ^{15}N relaxation dispersion profiles for Val59 are shown in Figure 1, whereas a complete set of ^{1}H^{N}-^{15}N RDCs in the *N* and D_{phys} states is listed in Table 1.

The dispersion data for I71V ADA2h weakly aligned in Pf1 phage solution were well fitted using a two-state exchange model, consistent with our previous analysis.[42] The values k_{ex} = 522 ± 7 s^{-1} and p_D = 4.31 ± 0.04% obtained from ^{15}N single-quantum CPMG dispersion data for ^{15}N/^{2}H I71V ADA2h at 40 °C in Pf1 phage solution were somewhat different from k_{ex} = 731.9 ± 4.5 s^{-1} and p_D = 1.64 ± 0.01% previously measured for ^{15}N/^{13}C/^{2}H I71V ADA2h at 40 °C without the alignment medium.[42] Higher p_D and slower k_{ex} in Pf1 phage suggest a weak preferential interaction of the D_{phys} state with the alignment medium. This interaction resulted in a stronger alignment of the D_{phys} state relative to *N* state, and somewhat broader range of ^{1}H^{N}-^{15}N RDCs obtained for D_{phys} (Table 1). However, an excellent agreement between ^{15}N chemical shift differences, $\Delta\varpi_{DN}$, obtained with and without Pf1 phage (Figure S5) suggests that interaction with the alignment medium had only limited effects on the structural ensemble of the D_{phys} state.

When the RDCs of individual residues in the D_{phys} state were plotted against the corresponding values in the *N* state, a statistically significant correlation was detected with a Pearson's coefficient of correlation of 0.76 ($p < 0.005$, Figure S7). This relatively high correlation between the RDCs of the *N* and D_{phys} states implies that, at least on average, the overall spatial positioning of different chain segments in D_{phys} of I71V ADA2h is to some degree similar to that of the *N* state. Correlated RDC patterns were also observed for denatured and native states of a fragment of *Staphylococcal nuclease* and were interpreted as being indicative of a closer resemblance between the structures of the two states.[74] Among different secondary structure regions in ADA2h, the highest correlation between the RDCs of the *N* and D_{phys} states was obtained for the β-sheet region (coefficient of correlation $r = 0.84$, $p < 0.05$), while the loop region ($r = 0.64$, $p < 0.05$) and the α-helical region ($r = 0.4$, $p > 0.05$) exhibited lower correlations (see Figure S7). The outliers belonged to the disordered N-terminal region (Gly9), β2 strand (Thr44), α2 helix (Val59), and β4 strand (Ile75). No major outliers were detected for residues on β1 and β3 strands and α1 helix, whose three residues Ile15 (β1), Leu26 (α1), and Val52 (β3) were previously suggested to form the folding nucleus of ADA2h.[45] Despite the high correlation between the RDC values for the β1-β3 strands, a fully ordered β-sheet is almost absent in the D_{phys} (Figure S8), consistent with the notion that full formation of β-structure is requires the firm establishment of a network of tertiary context.[75] The RDCs in the α-helical region show lower level of correlation between the D_{phys} and *N* states, indicating that, although the α-helical structure in the D_{phys} state is partially formed (Figure S8), the α-helices tend to be in different orientations in the *N* and D_{phys} states.

Determination of Structure and Dynamics of the *N* state of I71V ADA2h Using Chemical Shifts as Restraints in RAM Simulations. To analyze the information about the structure and dynamics of the *N* and D_{phys} states of I71V ADA2h provided by the measured NMR chemical shifts and RDCs, we incorporated them as replica-averaged structural restraints in molecular dynamics simulations, using the RAM method[14,66] (see Materials and Methods). This approach combines the sampling efficiency of metadynamics with the on-the-fly modification of the force field with experimental data, and as a result generates ensembles of conformations consistent with the maximum entropy principle.[15,71,76] In this view, the generated ensemble is the most probable one, given the force field and the experimental data included. This RAM method has been shown to accurately characterize the free energy surface of folded proteins[69,77] and, what is more challenging, the free energy surface of protein denatured states[14,78] and intrinsically disordered proteins.[79]

We first characterized the structure and dynamics of the *N* state by only using the experimentally determined ^{15}N, ^{1}H^{N}, ^{13}C$^{\alpha}$, ^{1}H$^{\alpha}$, ^{13}C′ chemical shifts as restraints in the RAM simulations. The free energy landscape of the resulting structural ensemble (Figure 2a) reveals the existence of a distinct free energy minimum with a radius of gyration (R_{gyr} = 1.25 nm) almost identical to that of the wild-type protein. Furthermore, a superposition of the I71V and wild-type ADA2h native structures[80] revealed no significant conformational perturbations induced by the I71V mutation, with a heavy atom root−mean−square deviation (RMSD) of 1.09 Å (Figure 2b). We then validated the *N* state ensemble by back-calculating chemical shifts, secondary structure population and RDCs. We first verified that the back-calculated NMR chemical shifts (obtained using Sparta+), which were used as restraints in the RAM simulation (where they were calculated using CamShift), are in agreement with the experimental values. As expected, a high level of agreement was found (Figure 2d) with RMSD values of 1.4 ppm (^{15}N), 0.59 (^{13}Cα), 0.67 (^{13}C′), 0.22 (^{1}H$_N$), and 0.13 (^{1}Hα); these values are comparable to the reported errors of Sparta+ and CamShift chemical shift predictors.[70,81] We also compared chemical shift-derived and ensemble-calculated secondary structure populations and found high coefficient of correlations both for the α-helical content (0.98, $p < 0.001$) and for the β-sheet content (0.94, $p < 0.001$, Figure S9). The *N* state ensemble was in addition validated against the 65 experimentally measured ^{1}H^{N}-^{15}N RDCs that were not used as restraints in the ensemble determination. The level of agreement was very high, with a *Q* factor of 0.18 (Figure 2c).

Determination of Structure and Dynamics of D_{phys} State of I71V ADA2h Using Chemical Shifts and RDCs as Structural Restraints in RAM Simulations. We then analyzed the structure and dynamics of the D_{phys} state of I71V ADA2h using chemical shifts and RDCs as structural restraints in RAM simulations. Since denatured states of proteins are conformationally highly heterogeneous, the structural characterization of these states represents a challenging problem. As a stringent test of the RAM method for the characterization of D_{phys} of ADA2h, we first carried out four control RAM simulations that use four different subsets of the experimentally measured chemical shifts and RDCs as restraints and then cross-validated the resulting ensembles with chemical shift and RDC values not used as restraints in the simulations (see Materials and Methods). We found that the combination of the NMR chemical shifts and RDCs used as restrains in RAM simulations consistently improved the correlation between predicted and experimental NMR chemical shifts and RDCs that were not used as restraints. Furthermore, even when we used TIP4*P*/2005[59] instead of the originally used TIP3P[58] water model, we showed that the experimental NMR chemical shifts and RDCs were still effective in driving the RAM trajectories of D_{phys} to the region of the conformational space in agreement with the experiments (see Materials and Methods).

After showing the robustness of the RAM method restrained with NMR chemical shifts and RDCs in reproducing a D_{phys} state of a protein, we generated a final D_{phys} ADA2h I71V ensemble by using the complete set of measured NMR chemical shifts and RDCs as structural restraints. It is important to emphasize here that even though this ensemble was used for further analysis, it showed essentially identical properties in relation to the measured kinetic and thermodynamic parameters (see below) as the control D_{phys} ensembles generated with the different subsets of the experimental restraints (see Materials and Methods).

The D_{phys} ensemble was first tested using the back-calculated NMR chemical shifts (obtained using Sparta+), which were used as restraints in the RAM simulations (where they were calculated using CamShift) and gave low RMSD values of 1.55 ppm (^{15}N), 0.69 (^{13}Cα), 0.38 (^{13}C′), 0.19 (^{1}H$_N$), and 0.14 (^{1}Hα) (Figure 3c). When the D_{phys} ensemble was validated against the 31 ^{1}H^{N}-^{15}N RDCs that were used as restraints in the RAM procedure, the level of agreement was very high with a coefficient of correlation of 0.96 ($p < 0.001$) and a *Q* factor of 0.25 (Figure 3b). In addition, we observed a good agreement between secondary structure populations derived from the chemical shifts and calculated directly from the D_{phys} ensemble with a coefficient of correlation of 0.92 ($p < 0.001$) for the α-helical content and 0.88 for the β-sheet content, suggesting that transient β-sheet content was captured rather successfully in our D_{phys} ensemble (Figure S10). Finally, we cross-validated the D_{phys} ensemble against the S^2 order parameters for the backbone amide groups derived from the chemical shifts using the RCI approach[82,83] (Figure S2a). Simulation-derived S^2 order parameters were obtained as averages of S^2 values calculated for individual microstates in the D_{phys} ensemble (see Materials and Methods) and showed good agreement with the RCI-derived S^2 ($r = 0.77$, $p < 0.005$), indicating that the D_{phys} ensemble correctly reflects amplitudes of angular fluctuations of the amide groups.

When the D_{phys} ensemble was compared with the *N* ensemble, a few characteristics were the most noticeable. First, the free energy landscape of the D_{phys} ensemble plotted as a function of the secondary structure content (Figure 3a) markedly differs from that of the *N* ensemble (Figure 2a). While the *N* state exhibits a well-defined free energy minimum, the free energy landscape of D_{phys} exhibits several local minima. The left region of the free energy landscape on Figure 3a includes microstates with marginally formed α-helix 1 and α-helix 2, whereas in the microstates of the right region, α-helix 1 and α-helix 2 have higher structural content. All microstates contain only transiently formed β-sheet elements, in agreement with secondary structural populations inferred from the experimentally measured chemical shift[84] (Figure S10). Despite the relatively low populations of the secondary structure elements, D_{phys} is relatively compact, with a radius of gyration (R_{gyr} = 1.32 ± 0.04 nm) slightly larger than that (R_{gyr} = 1.25 nm) of the *N* state.

Even though the D_{phys} ensemble of ADA2h I71V samples a range of interconverting microstates, we were able to identify some common structural properties. Perhaps the most notable characteristic of the D_{phys} state is the formation of the native-like tertiary contacts between the preformed N−terminus of α-helix 1 and the preformed C-terminus of α-helix 2 (residues Leu29 and Phe65). These interhelical tertiary contacts involve hydrophobic side chains from the two α-helices (Leu27, Val62, and Leu66) accessible for transient interactions with residues from β-strands 1−4 that form hydrophobic contacts in the *N* state (Figure S2b). Unlike the two α-helices, the regions of native β-strands 1−4 do not populate well-defined secondary structures in the D_{phys} state as confirmed by the measured chemical shifts (Figures S8b and S10). Among the β-strand, the regions of native β-strands 1 and 4 show the greatest conformational diversity owning to the relatively dynamic loops between β-strand 1 and α-helix 1 (average S^2 order parameter of 0.38) and between α-helix 2 and β-strand 4 (average S^2 order parameter of 0.49). On the other side, the loop region between α helix 1 and β strand 2 (residues Ala31−Asp38) displays unusually restricted mobility (average S^2 order parameter range 0.63−0.75). In D_{phys} state, this region positions β-strand 2 away from the core of the protein, unlike in the *N* state, which engages β-strand 3 in a range of transient side chain interactions (Figure S2b).

Folding Mechanism of I71V ADA2h at Nearly Atomic Resolution. To provide a complete description of the folding

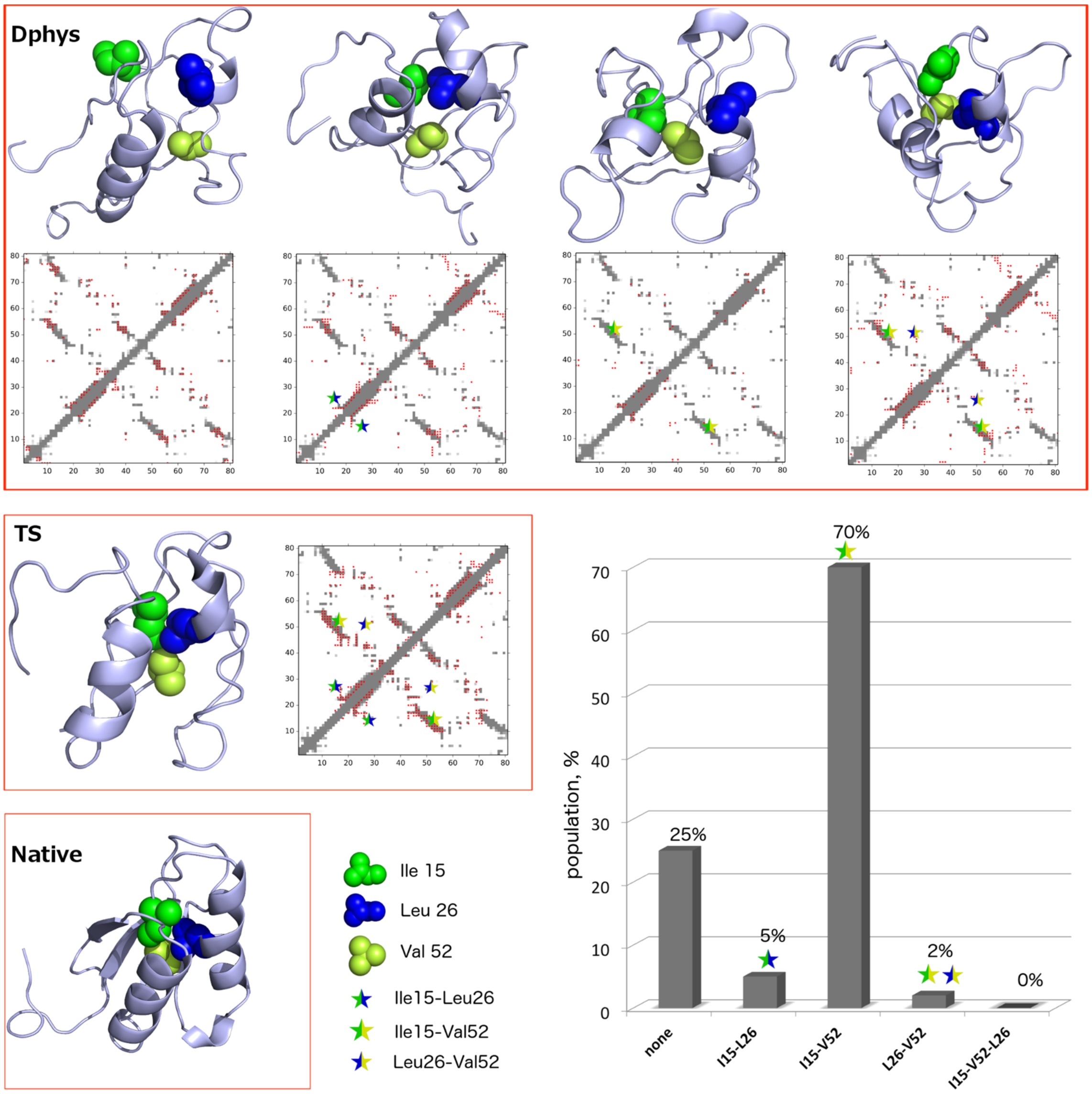


Figure 4. Characterization of the major states, and of the early events, in the nucleation−condensation mechanism of ADA2h. (Top) D_{phys} ensemble divided into four clusters, each with different inter-residue contacts (<5 Å) formed between the folding nucleus residues (Ile15, Leu26, and Val52). (Middle) In all structures of the TS ensemble, the folding nucleus is formed. (Bottom) In the N state, the folding nucleus residues are always in contact. Contacts in D_{phys} and TS (red) are depicted in relation to the N state contacts of I71V ADA2h (gray). (Right) Populations of the four clusters that describe inter-residue contacts within the D_{phys} ensemble.

pathway of ADA2h, we modeled the structural ensemble of the transition state (TS) using previously determined experimental Φ-values[45] of the wild-type protein as restrains in molecular dynamics simulations, a well-established computational procedure that has been successfully employed on different protein systems.[85−87] We assumed that the transition state of the I71V variant is very similar to that of the wild-type protein, owing to its relatively low Φ-value of 0.21. The average conformation of the resulting TS ensemble and the comparison of contact maps of the TS and N states are shown in Figure 4. As expected, the transition state shares similar topological properties with the N state, TM_score = 0.52.[88] This finding is consistent with earlier experimental studies[45] that suggested ADA2h folds via a nucleation−condensation mechanism, whereby folding occurs via a global condensation concomitant with nucleus formation, with the transition state resembling a distorted version of the native state.

A structural analysis of the TS ensemble revealed a relatively compact conformation (R_{gyr} = 1.27 ± 0.02 nm) with the α-helix 2 and N-terminus of α-helix 1 fully folded and in contact with the two central β-strands (β1 and β3), as previously suggested by Serrano and co-workers.[45] A compact arrangement formed by α-helices 1 and 2, and β strands β1 and β3 in the generated TS ensemble is further confirmed by the analysis presented in Figure S11a, suggesting that relatively high back-calculated Φ_i-values are present only in the regions previously identified by Serrano and co-workers.[45] Among the residues that form long-range contacts between the α-helical and β-sheet region, we identified Ile15 (belonging to β1-strand), Leu26 (α-helix 1), and Val52 (β3-strand) as possible folding nucleus of ADA2h (Figure 4). These three residues display the highest measured Φ-values, 1.0 for Ile15, 0.59 for Val52, and 0.55 for Leu26.[45] They are simultaneously in contact in all conformations of the TS and N ensemble and they are considered as the folding nucleus residues in the analysis presented in this work.

DOI: 10.1021/jacs.7b01540
J. Am. Chem. Soc. 2017, 139, 6899−6910

Table 2. Inter-residue Contact Probabilities in D_{phys} for Residues Considered as Candidate Members of the Folding Nucleus[a]

res1	res2	res3	res1−2 (%)	res1−3 (%)	res2−3 (%)	none (%)	res1−2−3 (%)
Leu13	Leu26	Val62	11.9	7.2	42.8	50.0	0
Leu13	Val52	Val62	20.3	7.2	0	78.3	0
Leu13	Leu26	Val52	11.9	20.3	2.0	74.0	0
Ile15	Leu26	Val52	4.8	70.2	2.0	25.0	0
Ile15	Leu26	Val62	4.8	0.1	42.8	55.3	0
Ile15	Val52	Val62	70.2	0.1	0	34.7	0
Leu26	Val52	Phe65	2.0	17.9	0	82.3	0

[a] Probabilities do not add up to 100%, as two pairs of residues can sometimes be in contact simultaneously. However, in none of the structures of the D_{phys} ensemble was the folding nucleus fully formed by making contacts between all three residues.

■ DISCUSSION

The analysis of enzymatic reactions suggests that catalysis is optimal when intermediates are minimally populated.[3,89] In analogy with this view, it has been argued that concurrent formation of secondary and tertiary structure from a highly disordered D_{phys} is an efficient strategy to optimize folding rate constants.[49] Consequently, it has been proposed that there is an evolutionary pressure opposing the accumulation of nucleation sites in the denatured states, where structure formation should be only flickering. The recent development of NMR relaxation dispersion methods that enable measurements of chemical shifts and RDCs in sparsely populated non-native states, in synergy with restrained molecular dynamics simulations, offers the tantalizing possibility of providing a description of the protein folding reaction at nearly atomic resolution and, therefore, testing these hypotheses from a structural perspective.

Furthermore, it has been proposed that the structure of proteins is to some extent imprinted in the residual structure of their D_{phys}.[90,91] It is therefore of interest to analyze the structural behavior of the folding nucleus in the D_{phys} ensemble of I71V ADA2h. In fact, among all D_{phys} microstates, we did not find a single conformation with all the three key residues of the folding nucleus Ile15-Leu26-Val 52 being in contact simultaneously (Figure 4). Most of the time, contacts were only formed between the β-strand residues Ile15 and Val52 (70% population). Twenty-five percent of the time, the three key residues were far apart forming transient non-native interactions. In 5% of D_{phys} population, the contacts were formed between Ile15 and Leu26 exclusively, whereas in 2% population the contacts were made simultaneously between Leu26 and Val52, and between Ile15 and Val52, but not between Ile15 and Leu26 (Figure 4).

To rule out possible influences in our analysis caused by the definition of the ADA2h folding nucleus that we adopted here, we extended our analysis to other residues for which the predicted Φ_i-values were higher than 0.5 as possible candidates for the folding nucleus (Figure S11a), that is, Leu13 and Ile15 (β1-strand); Asn19, Asn 25, and Leu26 (α-helix 1); Val52 (β3-strand); Ala61, Val62, and Phe65 (α-helix 2). Among these residues, we identified five additional residue triads that can form simultaneous contacts in the TS and N ensembles as alternative candidates for the folding nucleus of ADA2h, namely Leu13-Leu26-Val62, Leu13-Val52-Val62, Leu13-Leu26-Val52, Ile15-Leu26-Val62, Ile15-Val52-Val62. Then we considered the structural behavior of these five alternative candidates for the folding nucleus in the D_{phys} ensemble of I71V ADA2h (Table 2). Similar to the behavior of our original choice for the folding nucleus Ile15-Leu26-Val52, in all cases, the three residues of the possible alternative folding nuclei were never found in simultaneous contact in any of the structures of the D_{phys} ensemble. Moreover, when a possible folding nucleus was considered as formed by the three residues with the highest folding barrier $\Delta\Delta G_{TS\text{-}Dphys}$ (Leu26-Val52-Phe65), the identical property was observed (Figure S11b and Table 2). Thus, it appears that while the residues belonging to the folding nucleus (however defined) have some tendency to be close in the space in D_{phys}, they interact only in pairs. These findings, together with the observation that the three residues were always in contact in the transition state ensemble (Figure 4), reinforce the role of nucleation in protein folding: while D_{phys} is characterized by embryonic unstable native-like segments, consolidation of the nucleus occurs only downstream the transition state.

To elucidate the role that the D_{phys} state plays in the unfolding and refolding kinetics of I71V ADA2h, we further analyzed the fraction of native inter-residue contacts with respect to the total number of contacts formed in D_{phys}, and compared these fractions with the effects of single-point mutations on the folding kinetics of ADA2h.[45] This relationship shows a statistically significant correlation (Figure 5a), with a Pearson's coefficient of correlation of 0.66 ($p < 0.005$). An even better correlation (a Pearson's coefficient of correlation of 0.78, $p < 0.005$) was obtained when the average number of native inter-residue contacts formed in D_{phys} was related to the effects of single-point mutations on the folding kinetics of ADA2h (Figure 5b). As a control of this correlation, we verified that the change in the unfolding rate of ADA2h upon mutation is not correlated with the fraction of native contacts in D_{phys} (Figure S12). A similar correlation was present for the D_{phys} ensemble generated with an alternative water model (TIP4P/2005) and the identical RAM procedure (Figure S13). In particular, Pearson's coefficients of correlation of 0.63 ($p < 0.005$) and 0.72 ($p < 0.005$) were obtained when the fraction and total number of inter-residue native contacts were related to the effects of single point mutations in this TIP4P/2005 D_{phys} ensemble, respectively.

The formation of stable structural elements in the D_{phys} ensemble is expected to lower its free energy and, therefore, to slow down folding.[49] Conversely, we observed here that the majority of mutated residues[45] predominately make native-like contacts in the D_{phys} of ADA2h and their mutations decelerate, rather than accelerate the folding process (up to 20%). A possible scenario to explain these observations may imply that these mutations increase the free energy of the TS more than that of the D_{phys} in ADA2h, thus further reinforcing the importance of evaluating the effects of mutations in the D_{phys} state to fully understand the equilibrium and kinetics of folding. These results, however, also suggest that mutations that affect positions in which there is a large fraction of non-native

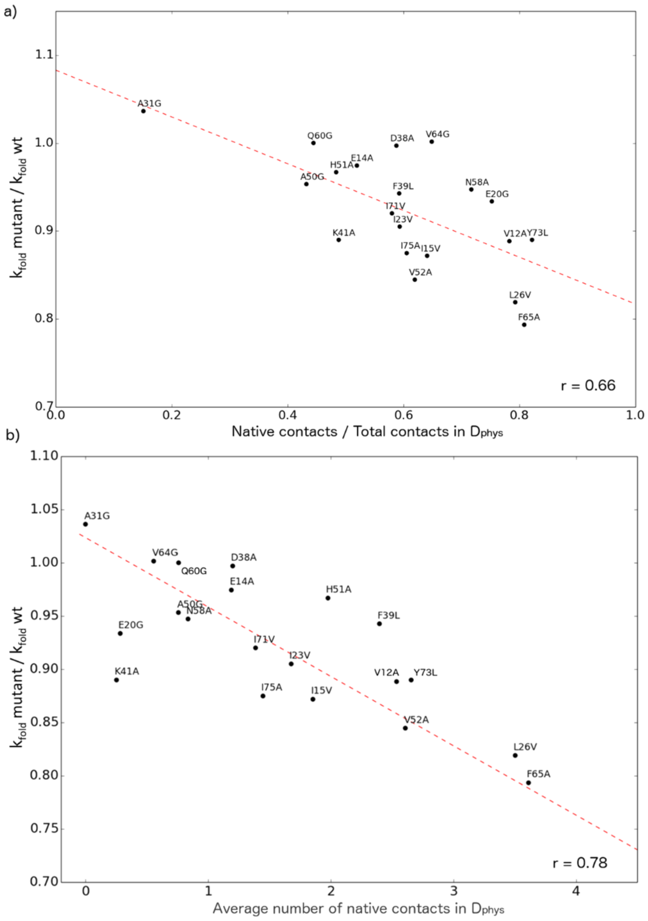


Figure 5. Influence on the folding rate of transiently formed contacts in the D_{phys} state. We found a statistically significant correlation between the fraction of native contacts in D_{phys} and the folding rate of ADA2h mutants. (a) When we considered the fraction of native contacts with respect to the total number of contacts, the Pearson's coefficient of correlation was 0.66. (b) When we considered the average number of native contacts, the Pearson's coefficient of correlation was 0.78.

contacts in D_{phys} can speed up the folding process. Indeed, this effect was found for Ala31 positioned on the unusually restricted loop region in D_{phys}, which is the only mutated residue that mainly engages in non-native interactions in D_{phys} (more than 80% of time) and is the only mutant of the pool that speeds up the folding reaction, relative to the wild-type protein. While the effects of mutations on the folding rate constants described here are relatively small, the observed statistically significant correlations support the conclusions of our analysis.

CONCLUSIONS

By using a synergy between computational and experimental techniques, we have described all the major states on the folding pathway of ADA2h at nearly atomic resolution. These results have offered the possibility to infer the early events in the nucleation-condensation mechanism of ADA2h, which precede the formation of the folding nucleus. We found that the residues that constitute the folding nucleus form often transient contacts in the denatured state, but never simultaneously, thus showing how the folding nucleus in ADA2h is only flickering prior to the formation of the transition state, in a process amenable to a fine-tuning of the folding rate by mutational selection.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b01540.

Starting conformations for RAM simulation; validation of RAM ensemble; analysis of experimental RDCs; validation of secondary structural content; folding nucleus analysis; folding rate dependence on transiently formed contacts in D_{phys} (PDF)
3D renderings of RAM structures (ZIP)

AUTHOR INFORMATION

Corresponding Author

*mv245@cam.ac.uk

ORCID

Carlo Camilloni: 0000-0002-9923-8590
Michele Vendruscolo: 0000-0002-3616-1610

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by UCHC startup funds, a NSF grant (MCB 1615866), and a CNPq Science without Boarders grant to D.M.K., the Italian Ministero dell'Istruzione dell'Università e della Ricerca (Progetto di Interesse "Invecchiamento" to S.G.), and by Sapienza Università di Roma (C26A155S48 and B52F16003410005 to S.G).

REFERENCES

(1) Polanyi, J. C. *Acc. Chem. Res.* **1972**, *5*, 161.
(2) Levine, R. D. *Molecular Reaction Dynamics*; Cambridge University Press, 2005.
(3) Fersht, A. *Structure and Mechanism in Protein Science: Guide to Enzyme Catalysis and Protein Folding*; Freeman: New York, 1999.
(4) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins: Struct., Funct., Genet.* **1995**, *21*, 167.
(5) Dill, K. A.; Chan, H. S. *Nat. Struct. Biol.* **1997**, *4*, 10.
(6) Henzler-Wildman, K.; Kern, D. *Nature* **2007**, *450*, 964.
(7) Mittag, T.; Forman-Kay, J. D. *Curr. Opin. Struct. Biol.* **2007**, *17*, 3.
(8) Boehr, D. D.; Nussinov, R.; Wright, P. E. *Nat. Chem. Biol.* **2009**, *5*, 789.
(9) Jensen, M. R.; Zweckstetter, M.; Huang, J.-R.; Blackledge, M. *Chem. Rev.* **2014**, *114*, 6632.
(10) Varadi, M.; Kosol, S.; Lebrun, P.; Valentini, E.; Blackledge, M.; Dunker, A. K.; Felli, I. C.; Forman-Kay, J. D.; Kriwacki, R. W.; Pierattelli, R.; et al. *Nucleic Acids Res.* **2014**, *42*, D326.
(11) Habchi, J.; Tompa, P.; Longhi, S.; Uversky, V. N. *Chem. Rev.* **2014**, *114*, 6561.
(12) Bhowmick, A.; Brookes, D. H.; Yost, S. R.; Dyson, H. J.; Forman-Kay, J. D.; Gunter, D.; Head-Gordon, M.; Hura, G. L.; Pande, V. S.; Wemmer, D. E.; et al. *J. Am. Chem. Soc.* **2016**, *138*, 9730.
(13) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. *J. Phys. Chem. B* **2015**, *119*, 5113.
(14) Camilloni, C.; Vendruscolo, M. *J. Am. Chem. Soc.* **2014**, *136*, 8982.
(15) Bonomi, M.; Heller, G. T.; Camilloni, C.; Vendruscolo, M. *Curr. Opin. Struct. Biol.* **2017**, *42*, 106.
(16) Fersht, A.; Daggett, V. *Cell* **2002**, *108*, 573.
(17) Clementi, C.; Nymeyer, H.; Onuchic, J. N. *J. Mol. Biol.* **2000**, *298*, 937.
(18) Onuchic, J. N.; Wolynes, P. *Curr. Opin. Struct. Biol.* **2004**, *14*, 70.

(19) Vendruscolo, M.; Paci, E.; Dobson, C. M.; Karplus, M. *Nature* **2001**, *409*, 641.

(20) Friel, C. T.; Smith, D. A.; Vendruscolo, M.; Gsponer, J.; Radford, S. E. *Nat. Struct. Mol. Biol.* **2009**, *16*, 318.

(21) Oliveberg, M.; Wolynes, P. G. *Q. Rev. Biophys.* **2005**, *38*, 245.

(22) Staley, J.; Kim, P. *Protein Sci.* **1994**, *3*, 1822.

(23) Ptitsyn, O. *Trends Biochem. Sci.* **1995**, *20*, 376.

(24) Alexandrescu, A.; Gittis, A.; Abeygunawardana, C.; Shortle, D. *J. Mol. Biol.* **1995**, *250*, 134.

(25) Religa, T.; Markson, J.; Mayor, S.; Freund, V.; Fersht, A. *Nature* **2005**, *437*, 1053.

(26) Mayor, U.; Guydosh, N.; Johnson, C.; Grossmann, G.; Sato, S.; Jas, G.; Freund, S.; Alonso, D.; Daggett, V.; Fersht, A. *Nature* **2003**, *421*, 863.

(27) Scaloni, F.; Federici, L.; Brunori, M.; Gianni, S. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 5447.

(28) Mok, Y.-K.; Kay, C.; Kay, L.; Forman-Kay, J. *J. Mol. Biol.* **1999**, *289*, 619.

(29) Cho, J.-H.; Meng, W.; Sato, S.; Kim, E. Y.; Schindelin, H.; Raleigh, D. P. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 12079.

(30) Choy, W.-Y.; Mulder, F. A.; Crowhurst, K. A.; Muhandiram, D.; Millett, I. S.; Doniach, S.; Forman-Kay, J. D.; Kay, L. E. *J. Mol. Biol.* **2002**, *316*, 101.

(31) Dyson, H. J.; Wright, P. E. *Chem. Rev.* **2004**, *104*, 3607.

(32) Korzhnev, D. M.; Salvatella, X.; Vendruscolo, M.; Di Nardo, A. A.; Davidson, A. R.; Dobson, C. M.; Kay, L. E. *Nature* **2004**, *430*, 586.

(33) Korzhnev, D. M.; Kay, L. E. *Acc. Chem. Res.* **2008**, *41*, 442.

(34) Korzhnev, D. M.; Religa, T. L.; Banachewicz, W.; Fersht, A. R.; Kay, L. E. *Science* **2010**, *329*, 1312.

(35) Neudecker, P.; Robustelli, P.; Cavalli, A.; Walsh, P.; Lundstroem, P.; Zarrine-Afsar, A.; Sharpe, S.; Vendruscolo, M.; Kay, L. E. *Science* **2012**, *336*, 362.

(36) Palmer, A. G., 3rd; Kroenke, C. D.; Loria, J. P. *Methods Enzymol.* **2001**, *339*, 204.

(37) Neudecker, P.; Lundstrom, P.; Kay, L. E. *Biophys. J.* **2009**, *96*, 2045.

(38) Palmer, A. G., 3rd *J. Magn. Reson.* **2014**, *241*, 3.

(39) Hansen, F.; Neudecker, P.; Vallurupalli, P.; Mulder, F.; Kay, L. *J. Am. Chem. Soc.* **2010**, *132*, 42.

(40) Vallurupalli, P.; Hansen, F.; Stollar, E.; Meirovitch, E.; Kay, L. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 18473.

(41) Igumenova, T.; Brath, U.; Akke, M.; Palmer, A., III *J. Am. Chem. Soc.* **2007**, *129*, 13396.

(42) Pustovalova, Y.; Kukic, P.; Vendruscolo, M.; Korzhnev, D. M. *Biochemistry* **2015**, *54*, 4611.

(43) Viguera, A. R.; Villegas, V.; Aviles, F. X.; Serrano, L. *Folding Des.* **1997**, *2*, 23.

(44) Villegas, V.; Azuaga, A.; Catasus, L.; Reverter, D.; Mateo, P. L.; Aviles, F. X.; Serrano, L. *Biochemistry* **1995**, *34*, 15105.

(45) Villegas, V.; Martinez, J. C.; Aviles, F. X.; Serrano, L. *J. Mol. Biol.* **1998**, *283*, 1027.

(46) Fernandez, A. M.; Villegas, V.; Martinez, J. C.; Van Nuland, N. A.; Conejero-Lara, F.; Aviles, F. X.; Serrano, L.; Filimonov, V. V.; Mateo, P. L. *Eur. J. Biochem.* **2000**, *267*, 5891.

(47) Cerda-Costa, N.; Esteras-Chopo, A.; Aviles, F. X.; Serrano, L.; Villegas, V. *J. Mol. Biol.* **2007**, *366*, 1351.

(48) Villegas, V.; Zurdo, J.; Filimonov, V. V.; Aviles, F. X.; Dobson, C. M.; Serrano, L. *Protein Sci.* **2000**, *9*, 1700.

(49) Fersht, A. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 10869.

(50) Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. *Biochemistry* **1994**, *33*, 10026.

(51) Ottiger, M.; Delaglio, F.; Bax, A. *J. Magn. Reson.* **1998**, *131*, 373.

(52) Yao, L.; Ying, J.; Bax, A. *J. Biomol. NMR* **2009**, *43*, 161.

(53) Hansen, D. F.; Vallurupalli, P.; Kay, L. E. *J. Phys. Chem. B* **2008**, *112*, 5898.

(54) Mulder, F. A.; Skrynnikov, N. R.; Hon, B.; Dahlquist, F. W.; Kay, L. E. *J. Am. Chem. Soc.* **2001**, *123*, 967.

(55) Farrow, N. A.; Muhandiram, R.; Singer, A. U.; Pascal, S. M.; Kay, C. M.; Gish, G.; Shoelson, S. E.; Pawson, T.; Forman-Kay, J. D.; Kay, L. E. *Biochemistry* **1994**, *33*, 5984.

(56) Hansen, D. F.; Yang, D.; Feng, H.; Zhou, Z.; Wiesner, S.; Bai, Y.; Kay, L. E. *J. Am. Chem. Soc.* **2007**, *129*, 11468.

(57) Best, R.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004.

(58) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926.

(59) Abascal, J.; Vega, C. *J. Chem. Phys.* **2005**, *123*, 234505.

(60) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M.; Smith, J.; Kasson, P.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013**, *29*, 845.

(61) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. *Comput. Phys. Commun.* **2014**, *185*, 604.

(62) Jimenez, M. A.; Villegas, V.; Santoro, J.; Serrano, L.; Vendrell, J.; Aviles, F. X.; Rico, M. *Protein Sci.* **2003**, *12*, 296.

(63) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435.

(64) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

(65) Berendsen, H. J.; Postma, J. v.; van Gunsteren, W. F.; DiNola, A.; Haak, J. *J. Chem. Phys.* **1984**, *81*, 3684.

(66) Camilloni, C.; Cavalli, A.; Vendruscolo, M. *J. Chem. Theory Comput.* **2013**, *9*, 5610.

(67) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.

(68) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562.

(69) Kukic, P.; Leung, H. T. A.; Bemporad, F.; Aprile, F. A.; Kumita, J. R.; De Simone, A.; Camilloni, C.; Vendruscolo, M. *Structure* **2015**, *23*, 745.

(70) Kohlhoff, K. J.; Robustelli, P.; Cavalli, A.; Salvatella, X.; Vendruscolo, M. *J. Am. Chem. Soc.* **2009**, *131*, 13894.

(71) Cavalli, A.; Camilloni, C.; Vendruscolo, M. *J. Chem. Phys.* **2013**, *138*, 094112.

(72) Marinelli, F.; Pietrucci, F.; Laio, A.; Piana, S. *PLoS Comput. Biol.* **2009**, *5*, e1000452.

(73) Gsponer, J.; Christodoulou, J.; Cavalli, A.; Bui, J. M.; Richter, B.; Dobson, C. M.; Vendruscolo, M. *Structure* **2008**, *16*, 736.

(74) Shortle, D.; Ackerman, M. *Science* **2001**, *293*, 487.

(75) Minor, D. L., Jr.; Kim, P. S. *Nature* **1994**, *371*, 264.

(76) Roux, B.; Weare, J. *J. Chem. Phys.* **2013**, *138*, 138.

(77) Kukic, P.; Lundström, P.; Camilloni, C.; Evenäs, J.; Akke, M.; Vendruscolo, M. *Biochemistry* **2016**, *55*, 19.

(78) Granata, D.; Camilloni, C.; Vendruscolo, M.; Laio, A. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 6817.

(79) Toto, A.; Camilloni, C.; Giri, R.; Brunori, M.; Vendruscolo, M.; Gianni, S. *Sci. Rep.* **2016**, DOI: 10.1038/srep21994.

(80) Garcia-Saez, I.; Reverter, D.; Vendrell, J.; Avilés, F. X.; Coll, M. *EMBO J.* **1997**, *16*, 6906.

(81) Shen, Y.; Bax, A. *J. Biomol. NMR* **2010**, *48*, 13.

(82) Berjanskii, M. V.; Wishart, D. S. *J. Am. Chem. Soc.* **2005**, *127*, 14970.

(83) Berjanskii, M. V.; Wishart, D. S. *J. Biomol. NMR* **2008**, *40*, 31.

(84) Camilloni, C.; De Simone, A.; Vranken, W. F.; Vendruscolo, M. *Biochemistry* **2012**, *51*, 2224.

(85) Salvatella, X.; Dobson, C.; Fersht, A.; Vendruscolo, M. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 12389.

(86) Paci, E.; Vendruscolo, M.; Dobson, C.; Karplus, M. *J. Mol. Biol.* **2002**, *324*, 151.

(87) Paci, E.; Clarke, J.; Steward, A.; Vendruscolo, M.; Karplus, M. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 394.

(88) Zhang, Y.; Skolnick, J. *Proteins: Struct., Funct., Genet.* **2004**, *57*, 702.

(89) Pauling, L. *Chem. Eng. News* **1946**, *24*, 1375.

(90) Giri, R.; Morrone, A.; Travaglini-Allocatelli, C.; Jemth, P.; Brunori, M.; Gianni, S. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17772.

(91) Morrone, A.; McCully, M. E.; Bryan, P. N.; Brunori, M.; Daggett, V.; Gianni, S.; Travaglini-Allocatelli, C. *J. Biol. Chem.* **2011**, *286*, 3863.