JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

ARTICLE

pubs.acs.org/JACS

The Native Ensemble and Folding of a Protein Molten-Globule: Functional Consequence of Downhill Folding

Athi N. Naganathan*,† and Modesto Orozco*,†,‡,§

†IRB-BSC Joint Research Program in Computational Biology, Barcelona Supercomputing Center, Torre Girona, C/Jordi Girona 31, Barcelona 08034, Spain

‡IRB-BSC Joint Research Program in Computational Biology, Institute for Research in Biomedicine, Parc Cientific de Barcelona, C/Baldiri Reixac 10, Barcelona 08028, Spain

§Departament de Bioquímica, Facultat de Biologia, Avgda Diagonal 647, Barcelona 08028, Spain

Supporting Information

ABSTRACT: The continually emerging functional significance of intrinsic disorder and conformational flexibility in proteins has challenged the long-standing dogma of a well-defined structure contributing to a specific function. Molten-globular states, a class of proteins with significant secondary-structure but a fluid hydrophobic core, is one such example. They have however been difficult to characterize due to the complexity of experimental data and lack of computational avenues. Here, we dissect the folding mechanism of the α-helical molten-globular protein NCBD from three fundamentally different approaches: statistical-mechanical variable barrier model, C_α-based Gō-model and explicit water all-atom molecular dynamics (MD) simulations. We find that NCBD displays the characteristics of a one-state globally downhill folder but is significantly destabilized. Using simulation techniques, we generate a highly constrained but a heterogeneous native ensemble of the molten-globule for the first time that is consistent with experimental data including small angle X-ray scattering (SAXS), circular dichroism (CD), and nuclear magnetic resonance (NMR). The resulting native ensemble populates conformations reported in other bound-forms providing direct evidence to the mechanism of conformational selection for binding multiple partners in this domain. Importantly, our simulations reveal a connection between downhill folding and large conformational flexibility in this domain that has been evolutionarily selected and functionally exploited resulting in large binding promiscuity. Finally, the multimodel approach we employ here serves as a powerful methodology to study mechanisms and suggests that the thermodynamic features of molten-globules fall within the array of folding mechanisms available to small single-domain proteins.

INTRODUCTION

The primary sequence of proteins, evolved over millions of years of natural selection, enable them to exhibit diverse behaviors both in folding and function. Functionally, depending upon the degree of conformational flexibility and secondary/tertiary-structure content, proteins have been classified as possessing a rigid structure with minimal fluctuations, allosteric[1] (exhibiting transitions to functionally relevant states upon ligand binding), molten-globules[2] (little tertiary structure and large secondary-structure content) or intrinsically disordered[3] (IDPs; large conformational flexibility with little secondary or tertiary structure). From a folding mechanistic viewpoint, globular single-domain proteins can be categorized as one-state,[4] two-state[5] or multi-state[6,7] depending on the number of accessible macrostates in equilibrium. It is possible that folding mechanisms could be intimately tied to the eventual functional role as it establishes the number and nature of available conformational states, i.e. the folding−binding landscape.[8,9] Switching between these predetermined states (either over a free-energy barrier or in its absence) depending on the identity and magnitude of perturbation could also serve as way to optimize function. The large heterogeneity in the 'native-states' of some proteins with a well-defined 3D structure, as suggested by different experiments and simulations,[10−17] provide indirect support to this hypothesis.

In this regard, protein molten-globules represent an unique and fascinating class of proteins since the earliest works of Ptitsyn[18] and Privalov.[19] The nuclear coactivator binding domain (NCBD) of CREB binding protein (CBP) is one of the most studied of molten-globules.[20−26] CBP, a multidomain protein, is a transcriptional coactivator that plays a diverse role in many signal transduction pathways by mediating interactions through several of its domains (see ref 21 and references therein). NCBD is an α-helical domain of CBP. It is ∼50 residues in length and shows a large secondary-structure content in the free state,[21,22,25] limited dispersion in the HSQC spectra[20,21] and binds ANS.[22] It is therefore considered to be a text-book example of a molten-globule protein. Amazingly, despite its small size, it binds to at least seven different partners (for e.g. see ref 25).

Poulsen and co-workers recently obtained an experimental model structure of NCBD in the ligand-free state[25] (PDB id: 2KKJ). It is characterized by three helices (named H1, H2, and H3; Figure 1) organized in a helix-bundle topology with several low intensity long-range couplings (nuclear Overhauser effect; NOEs) between helices H1 and H2 and some between H2 and H3. This conformation resembles the structure of the complex with ACTR[21] (1KBH),

Received: May 3, 2011
Published: July 06, 2011

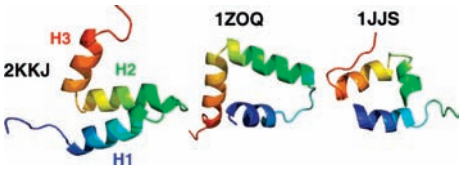

Figure 1. Structure of NCBD in the ligand-free state (2KKJ), in complex with IRF-3 (1ZOQ), and in an alternate ligand-free conformation (1JJS).

SRC-1[24] (2C52) and p53 TAD.[26] However, NCBD adopts a very different conformation in a complex with IRF-3[23] (1ZOQ; Figure 1) and a previous, if incomplete, structure of NCBD in the ligand-free state (1JJS; Figure 1) shows a drastically different organization of helices.[20] These findings hint at a very large structural plasticity and raises mechanistic questions on how NCBD modulates its function. The folding mechanism of this domain is also a matter of debate with both 'noncooperative' and 'cooperative' folding invoked to explain transitions as seen from differential scanning calorimetry (DSC) thermograms[22] and small angle X-ray scattering (SAXS) profiles,[25] respectively. Overall, these results suggest that NCBD displays a rich and complex conformational behavior, which we believe might be intimately related to the folding mechanism, and that could be of fundamental importance to its promiscuous binding.

In the current work, we employ a combination of methods from pure macroscopic analysis to atomistic simulations to characterize the folding behavior of NCBD. We show that this domain unfolds in a continuous manner without crossing a free-energy barrier, i.e. one-state downhill folding. Combining atomistic and coarse-grained simulations with experimental data we generate a constrained native-ensemble of the molten-globule that confirms our hypothesis on the connection between folding complexity and functional binding promiscuity.

METHODS

Variable Barrier Model. The model is implemented as discussed before.[27] The best fit parameters to the calorimetry data of NCBD are $\Sigma\alpha = 40$ kJ/mol, $T_0 = 342$ K, $f = 1$ and $\beta = 0.01$ kJ/mol. The thermogram of NCBD was not reported in absolute units and hence was up-shifted by 22.15 kJ/mol to agree with the Freire native-baseline[27] assuming a molecular weight of 6561 g/mol. This shift is done merely for visualization purposes and does not affect the results. $\Sigma\alpha$ is referred to as unfolding enthalpy for the sake of simplicity.

Gō-Model. The C_α-based Gō-model was parametrized along the same lines as Onuchic and co-workers[28,29] and run in GROMACS 4.0.2. The default parameters were employed while contacting residues from 2KKJ, 1ZOQ or Im9 (1IMQ) were identified with a heavy-atom cutoff distance of 5 Å excluding up to $i-i+3$ sequential neighbors. A Langevin dynamics approach with a damping coefficient of 1 ps^{-1} and a time-step of 0.5 fs was used to simulate unfolding.

MD Simulations. All-atom simulations were run in graphical processing units (GPUs) employing the ACEMD molecular simulation software.[30] Residues 1−51 of 2KKJ was solvated in the pH 7 protonation state in a box of dimensions $55 \times 61 \times 46$ $Å^3$ with 4815 TIP3P water molecules using the Amber99 SB* force-field.[31] Six Cl^- atoms were added to neutralize the charge of the system, thus amounting to a total of 15,283 atoms. The system was energy-minimized and then relaxed for 200 ps at the required temperature following which 5×0.1 μs runs at 2 fs time-step were initiated with different starting velocities at each temperature. A langevin thermostat with a damping coefficient of 0.1 ps^{-1} was employed. Simulations were performed at constant volume with long-range electrostatics calculated using the particle mesh Ewald (PME) procedure at a grid-spacing of 1.1 Å and with a 9 Å cutoff for nonbonded interactions.

Structure Calculation from the All-atom MD Ensemble at 370 K. An average structure was generated from the all-atom MD ensemble at 370 K using the Xplor-NIH software[32] employing an NOE-like calculation. The required average distance between two protons i and j ($d_{i,j}$) was approximated from[33,34]

$$d_{i,j} = \langle r_{i,j}^{-6} \rangle^{-1/6}$$

Here, the brackets indicate averaging over the snapshots generated in the MD run. For simplicity, intraresidue proton distances were ignored. This resulted in a total of more than 90,000 pairwise distances. Applying a distance cutoff of 5 Å (as done in a typical NOE structure calculation protocol), we ended up with 2478 distance constraints which were used as inputs to the Xplor program. Mean backbone dihedral angles (ϕ, ψ) were also provided as inputs.

Simulation of SAXS Profiles. The SAXS profiles were generated using the CRYSOL module in the software package ATSAS.[35] Individual snapshots from the C_α-based coarse-grained simulations were the inputs. The final calculated SAXS profile (I_{calc}) was minimized against the experiment (I_{exp}) using a chi-square like potential defined as:

$$\chi^2 = \sum_{q_{min}}^{q_{max}} (I_{exp} - (\mu I_{calc} + \Delta))^2$$

where q_{min} and q_{max} are the lower and upper limits of the scattering-vector range, μ is a scaling factor, and Δ is a shifting function to account for small experimental baseline subtraction errors.

Simulation of Far-UV CD Spectra. An empirical approach developed by Chen and co-workers[36] was used to calculate the helical spectrum from the coarse-grained ensemble. In this methodology, the mean-residue ellipticity ($[\theta_\lambda]$) can be represented as:

$$[\theta_\lambda] = f_H[\theta_\lambda^{\infty,H}](1 - k_\lambda/\langle l_H \rangle) + (1 - f_H)[\theta_\lambda^{Coil}]$$

where f_H is the fraction of residues in helical conformation (H), $\langle l_H \rangle$ is the mean helix-length (in a protein with multiple helical blocks), and $[\theta_\lambda^{\infty,H}]$, $[\theta_\lambda^{Coil}]$ are the basis spectra for an infinite-length helix and coil, respectively. k_λ accounts for the wavelength dependence of the helical spectrum. The basis spectra are obtained directly from Chen et al.[36] while the k_λ is further tuned to reproduce the 278 K spectra of NCBD[22] using the helical assignments from the PDB file 2KKJ.

Since we employ a Gō-model to study the folding properties, the probability of finding a residue i folded (p_i^{res}) can be directly calculated from:

$$p_{i,j}^{cont} = \sum m_{i,j}/n_{snapshots}$$

and

$$p_i^{res} = \langle p_{i,x}^{cont} \rangle$$

where $p_{i,j}^{cont}$ is the contact probability between two residues i and j, $m_{i,j}$ assumes a value of 0 or 1 depending upon whether that particular contact is within the distance threshold of 1.2 times the native-distance, $n_{snapshots}$ is the total number of snapshots in the simulated ensemble, and with x being the contacts involving residue i. From the residue probabilities it is straightforward to calculate the mean helix-length and fraction helix as:

$$\langle l_H \rangle = \sum_H p_i^{res}/n_H$$

and

$$f_H = \sum_H p_i^{res}/N$$

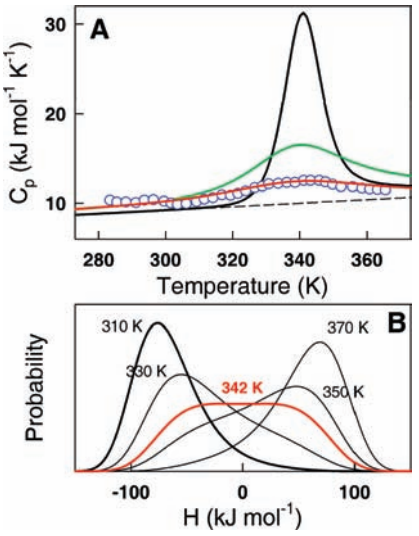

Figure 2. Quantitative predictions from the variable-barrier model. (A) Variable barrier model fit (red curve) to the experimental thermogram of NCBD (blue circles) resulting in a destabilized one-state scenario together with the native baseline (dashed line). Simulations of two-state (black curve) and one-state (green curve) scenarios with larger unfolding enthalpy are also shown for comparison. (B) Predicted probability density of states for NCBD at different temperatures with enthalpy as the order parameter.

wherein the summation runs over the regions identified as helical (H) from the PDB structure, n_H is the number of helical blocks (=3), and N is the protein length (=51).

RESULTS

Predictions from a Statistical Mechanical Model. DSC thermograms carry critical information on the partition function[37,38] and hence provide valuable insights into the statistical nature of the unfolding transition. Therefore, as a first step, we analyzed the heat capacity profile of NCBD at a quantitative level employing the phenomenological variable barrier (VB) model.[27] It bases its description on the Landau theory of phase transitions and employs a quartic free-energy functional at a reference temperature (T_0) to predict thermodynamic barriers. The predicted barriers scale extremely well with the experimental folding relaxation rates[39] attesting to the ability of the model to estimate thermodynamic barriers.

The experimental thermogram of NCBD is unusually broad, featureless and lacks a well-defined excess heat capacity peak (blue circles in Figure 2A). This observation is at odds with the expectation for a two-state protein with a large barrier and unfolding enthalpy that should exhibit a sharp excess heat capacity peak and a well-defined pretransition baseline (black curve in Figure 2A; simulated with a barrier (β) of 15 kJ/mol and an unfolding enthalpy ($\Sigma\alpha$) of 320 kJ/mol; see Methods). The experimental heat capacity profile of NCBD is also very different to that of a stable one-state folder (i.e., $\beta \approx 0$ kJ/mol) with an intermediate unfolding enthalpy value ($\Sigma\alpha$ = 120 kJ/mol; green curve in Figure 2A). On comparison, it is clear that the unfolding thermodynamics of NCDB is characterized by neither a large unfolding enthalpy nor a significant free-energy barrier. In fact, the best fit to the VB model results in zero barriers ($\beta \approx 0$ kJ/mol) under all conditions indicative of one-state downhill folding (Figure 2B) and with a $\Sigma\alpha$ of just 40 kJ/mol. The very small unfolding enthalpy is also evident in the chemical and thermal denaturation curves of this protein which are significantly broad and oftentimes do not show a sigmoidal transition.[22,25] The thermodynamic signatures are suggestive of a heterogeneous ensemble that is populated even under physiological temperatures (see the probability distribution at 310 K in Figure 2B). The results from the VB fit therefore suggest that NCBD resembles a 'destabilized one-state folder', i.e. it exhibits all the characteristics of downhill folding but is significantly destabilized stemming from weak long-range interactions within the ensemble.

Insights from a Gō-like Model. While statistical models allow for exact fits to experimental data they do not provide structural information on the unfolding process. We therefore employed the simplified C_α-based representation of Onuchic and co-workers[28] to understand the folding mechanism of NCBD at a coarse-grained level (see Methods). The resulting simulated heat capacity profile is broad and with a very small excess heat capacity peak (circles in Figure 3A) in contrast to that obtained for a classical two-state-like helical protein Im9 simulated using an identical protocol (lines in Figure 3A). In fact, the simulations parallel the experimental observation in Figure 2A hinting at the non-two-state nature of the transition. This is clearly observable in the single-molecule behavior of NCBD at the midpoint temperature (green in Figure 3B) that shows no sharp transition between folded- and unfolded-like states required of a two-state behavior (black in Figure 3B for Im9).

One-dimensional projections as a function of various reaction coordinates reveal a feature that is the hallmark of one-state downhill folding: a unimodal probability density under all conditions that progressively shifts to less structured distribution upon increasing thermal stress (Figure 3C−F). Moreover, the mean potential energy exhibits weak temperature dependence and the individual residue probabilities show a broad distribution of midpoint temperatures (Figure S1, Supporting Information [SI]) confirming that this domain folds following a one-state mechanism.[40] Since there is a single well under all conditions, the variance of structural coordinates (rmsd and fraction of native-contacts) serves as a direct measure of the conformational heterogeneity. This is shown in Figure 3G where the variance (σ^2) increases continuously with temperature, reaches a peak at the midpoint, after which it flattens out but with a higher basal value for both the variables. This finding suggests that structural heterogeneity and downhill folding go hand-in-hand and that the ensemble is broadest at the midpoint as previously observed in experiments.[27,40] Any mutation or change in solvent conditions that destabilize a one-state protein will make it less structured (Figure 3C,D) and as a rule more flexible since the protein will be pushed into more unfolding-like conditions in Figure 3G. Note that such a variance analysis is applicable only for a one-state folder as a unimodal density is available. Gō-model simulations starting from the 1ZOQ model (Figure S2, SI) also result in a globally downhill behavior indicating that our conclusions are robust and independent of the starting structure.

All-Atom MD Simulations. Though Gō-like potentials are able to reproduce several experimental features of NCBD folding, we cannot ignore the fact that they are topology-based which might yield biased results. Therefore, we decided to analyze the degree of structural heterogeneity in higher detail by employing state-of-the-art atomistic molecular dynamics simulations (MD) in explicit solvent (see Methods). Since the effective temperature of a molecular simulation is unclear and largely depends on the force-field[41] (i.e., current force-fields tend to overstabilize proteins;

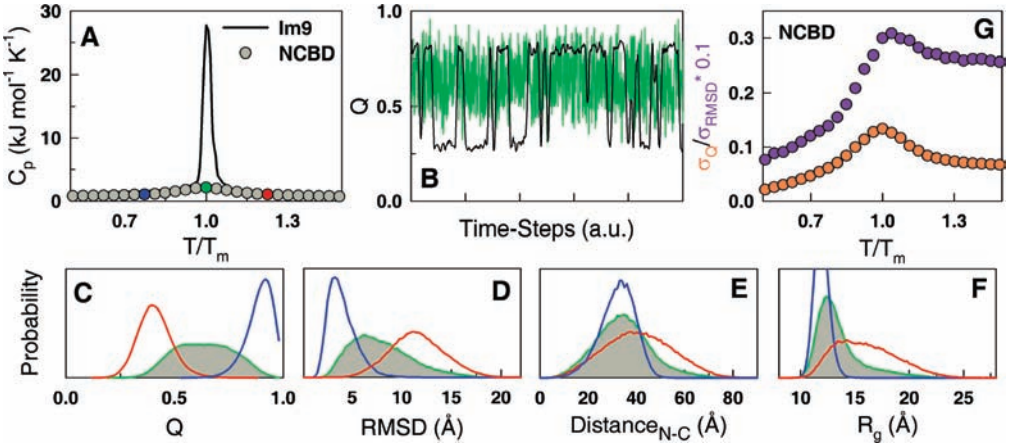

Figure 3. Coarse-grained Gō-model analysis. (A) Simulated heat capacity thermogram of NCBD (gray circles) compared to that of Im9 (a two-state folder; black curve). (B) Time-dependent single-molecule behavior of NCBD (green) and Im9 (black) at the respective midpoint temperatures. (C−F) Probability density from one-dimensional projections as a function of fraction of native contacts (C), root-mean-square deviation (D), end-to-end distance (E) and radius of gyration (F) at three different temperatures spanning the midpoint: $0.77T_m$ (blue), $1.0T_m$ (green) and $1.23T_m$ (red). (G) Structural heterogeneity of NCBD measured as the standard deviation in the fraction of native contacts (orange) or root-mean-square deviation (in Å; violet).

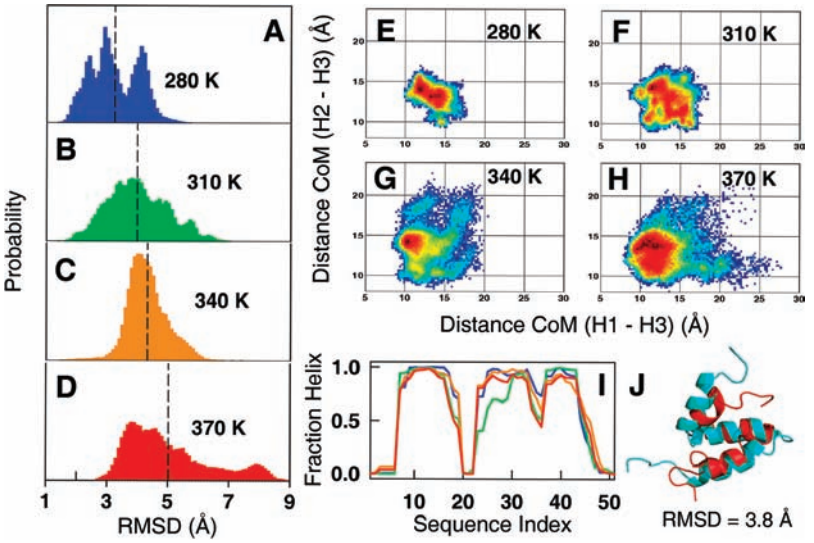

Figure 4. All-atom MD simulations. (A-D) Histogram of the root-mean-square deviation between the NMR structure (2KKJ) and MD snapshots at the indicated temperatures. The vertical dashed line signals the mean. (E-H) Two-dimensional logarithmic probability density plot with abscissa being the distance between the center-of-mass (CoM) of helices 1 and 3 and ordinate being the distance between CoM of helices 2 and 3. Color coding goes from highly populated (dark red) to low population (dark blue). (I) Fractional helicity at temperatures 280 (blue), 310 (green), 340 (orange), and 370 K (red). (J) Structure calculated from the simulated 370 K ensemble (red) superimposed onto the experimental 2KKJ model (cyan).

T_m from simulations are on average higher than experimental numbers), simulations were performed at four different temperatures (280, 310, 340, and 370 K) for a total of 2 microsecond of aggregated simulation time (five replicas of 0.1 μs for each temperature). We observe several unique features that highlight the folding complexity: 1) Even at the lowest temperatures, there are large changes in rmsd with respect to the NMR model (Figure 4A,B and Figure S3A [SI]). Note that such a behavior is not seen in a small two-state-like protein simulated under similar conditions (Figure S3B, SI). 2) With increasing temperatures the mean of the distribution moves to higher rmsd values. Moreover, the width of the ensemble hints at an increasing trend with the ensemble being the most diverse at 370 K (Figure 4A−D). 3) A two-dimensional density map constructed using the distances between the center-of-mass of helices H1/H2 and H3 reveals a characteristic behavior wherein the protein increasingly samples 'open' conformations upon increasing the temperature (Figure 4E-4H). It is unlikely that when using force-fields that tend to overstabilize proteins[41] complete unfolding will be detected in 0.1 μs individual trajectories. But the unique conformational behavior highlighted in points 2 and 3 are strikingly similar to the expectation for a one-state downhill folding system from the Gō-model (see Figure 3C-3G).

In view of the force-field stability issues, a simple procedure to identify the native ensemble is to calculate the temperature-dependence of helicity that decreases linearly in experiments for

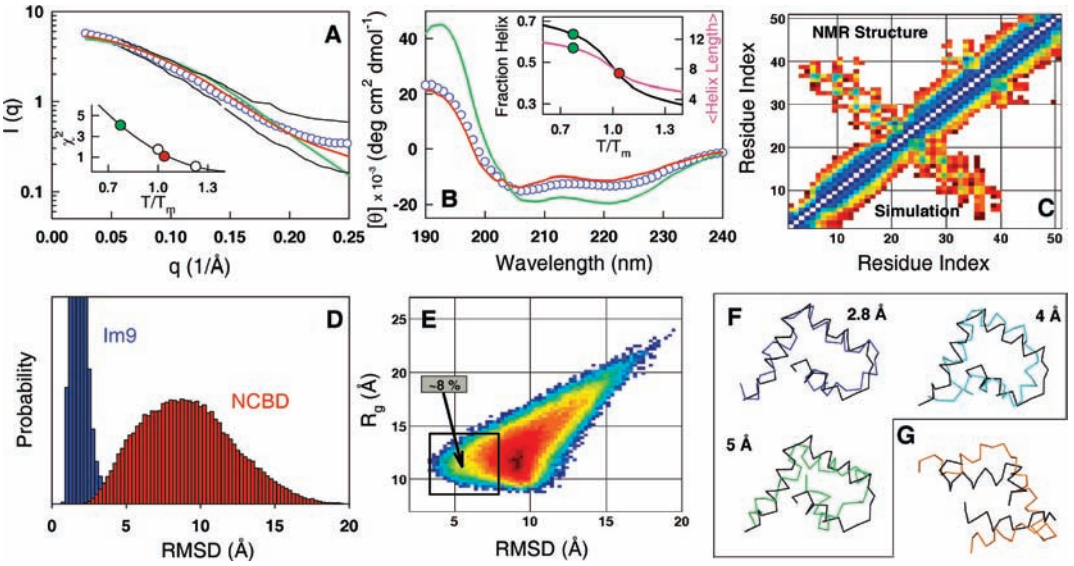

Figure 5. Native ensemble of NCBD. (A) Small-angle X-ray scattering profile under folding conditions (blue circles; black curves signal the maximal error in the data). The green and red curves are the fits to the experimental data using the ensemble generated by the Gō-model at temperatures of $0.77T_m$ and $1.04T_m$, respectively. Inset: Sum-of-least-squares (χ^2) between the experimental SAXS data and fit at various temperatures. (B) The predicted far-UV CD signal compared with experiments following the color code of panel A. Inset: The fraction helical content (black) and mean helix-length (pink) as a function of temperature. (C) Long-range correlations between the alpha-carbons at $1.04T_m$ from simulations (bottom right triangle) and from the NMR structure (top left triangle). The color coding goes from dark blue (~4 Å) to green (~8 Å) to dark red (12 Å). (D) The difference in the structural heterogeneity of the native ensembles of one-state NCBD and two-state Im9. (E) A two-dimensional logarithmic probability density plot with the abscissa being the rmsd of the native ensemble of NCBD with respect to 1ZOQ. The black square highlights the conformational space with a high similarity to the 1ZOQ conformation. (F) Superimposition of selected snapshots from the simulated native ensemble onto the 1ZOQ model (black). (G) An example of the sampled conformation that resembles the 1JJS model.

this domain.[25] Unfortunately, all three helices are highly structured and very similar to the NMR model even at 370 K (Figure 4I) possibly due to limited sampling combined with a potential lack of cooperativity in the force-field employed which might lead to the overstabilization of helices.[42] In fact, when we use the atomistic MD-ensemble at 370 K to derive NOE signals and back-compute an average structure (mimicking the procedure used in solving structures from NMR spectra), we obtain a structure that is very close to the NMR model (Figure 4J; see Methods). This result directly shows that the existence of an NMR structure does not preclude a broad ensemble in the native-state[43] (see below), an issue that might be critical in small proteins with weak long-range interactions.

The Native Ensemble of NCBD. Despite the remarkable agreement between three models in predicting the overall folding mechanism, questions on the properties of the NCBD native ensemble still remains. In principle this information can be directly obtained from the simulated ensemble at the experimental temperature. In reality this is however challenging due to: a) the limited sampling of MD simulations and b) the lack of knowledge on the effective simulation temperature. While the first problem affects only atomistic simulations that requires several microseconds to milliseconds to converge, the second affects both the simulation methods as they have been shown to be overstabilizing in nature particularly for proteins/peptides with weak interactions.[41,44] A simple solution to this problem is to scan the different simulated ensembles for the best match with the available experimental data. Fortunately, the availability of experimental SAXS profile,[25] far-UV CD[21,22,25] and NOE data[25] — three important observables that provide information on the overall dimensions of the ensemble, secondary and tertiary-structure content, respectively, required to generate any ensemble - under native conditions enables us to do precisely that. We decided to look into the ensemble predicted by the Gō-model as even the high temperature MD ensemble is unable to reproduce the experimental SAXS profile (Figure S4, SI), pointing to the extreme structural diversity of this protein under native conditions.

The agreement between the simulated SAXS profile (Methods) and the experimental data is best for temperatures equal to or greater than the midpoint in the Gō-model (Figure 5A). The calculated far-UV CD signal resembles the experimental spectrum at temperatures close to the midpoint (~43% helicity) while at higher temperatures the predicted helicity is much smaller (Figure 5B). The long-range coupling between C-αs calculated either from a NOE-like calculation or from fractional occurrences (Figures 5C and S5A [SI]) reproduces the expectations for the experimental structure well until around the midpoint after which the correlations are lost (Figure S5B, SI). Taken together, these results suggest that the simulated ensemble matches *all* the experimental observables of the native ensemble only at temperatures close to the midpoint predicted by the Gō-model (the best agreement is ~$1.04\ T_m$). The derived native ensemble is indeed quite diverse with a broad distribution of rmsd (Figure 5D) and R_g (Figure 5E) thus explaining the status of this domain as a molten-globule. Interestingly, the predicted large distribution of R_g (9−22 Å) under native conditions is in striking agreement with the previous independent analysis of the SAXS profile.[25]

DISCUSSION

Evidence for One-State Downhill Folding in a Molten-Globule. The molten-globular nature of NCBD makes the

characterization of its folding transition quite challenging. We have therefore sought to detail out the folding mechanism of this domain using a multimodel approach that includes the statistical-mechanical VB model, C_α-based Gō-model and all-atom explicit water MD simulations. In spite of the approximations inherent to each of the method, we identify characteristic behaviors of one-state downhill folding consistently − broad thermograms and thermal unfolding curves, single-molecule trajectory lacking sharp transitions, unimodal probability distributions, and native-state heterogeneity (Figures 2−4). The interpretation we provide here is also consistent with studies on other molten-globule-like proteins that concluded either folding over a marginal barrier (bovine α-lactalbumin[45]) or noncooperative folding (apomyoglobin,[19] human α-lactalbumin[46]) from both experiments and simulations. In parallel, we capture the signature behavior of molten-globules highlighted by the large secondary structure and diverse ensemble from both the atomic-simulation protocols.

Experimentally, there is strong evidence against the mechanism of two-state/cooperative folding in NCBD: (a) NCBD lacks a well-defined excess heat capacity peak[22] that is necessary condition for a two-state folder.[5] This is even more evident in a stabilizing variant of NCBD that displays a smoothly increasing heat capacity term as a function of temperature.[22] (b) Chemical and thermal denaturation experiments show no sign of a pretransition baseline that does not appear even upon the addition of a stabilizing osmolyte.[25] (c) Urea-temperature double perturbation experiments monitored by far-UV CD suggest that the 'native-ensemble' moves continuously upon destabilization.[25] In summary, there is a bulk of experimental evidence that cannot be explained by a two-state mechanism but that is in agreement with the behavior expected for downhill folders.[4,47,48] It is further difficult to reconcile the predicted broad distribution of R_g and two-state folding as the latter usually assumes a well-defined folded state with only minimal fluctuations (for example, see Figure S3, SI).

Implications for Folding Mechanisms. What structural features make NCBD fold in a one-state fashion? In this regard, NCBD satisfies most of the criteria common to downhill folders: small size (∼50 residues), α-helical, the only aromatic group within the structured region (F43) does not contribute to the hydrophobic core, and possesses a loosely packed structure. As a consequence, the helices are free to occasionally sample larger conformational space with the mean agreeing to that of the NMR model (Figures 5C, 4J). With increasing temperature, the conformational sampling increases proportionately (Figures 3, 4E−H) and the helices start to melt thus shifting the ensemble to a progressively more unstructured space, explaining the one-state-like behavior in NCBD.

Where does the molten-globule stand in the array of folding mechanisms available to small single-domain proteins? Folding properties of natural proteins can be classified as ranging one-state to two-state depending on the magnitude of free-energy barriers[49,50] that is determined by the relative number of local and nonlocal contacts in the structure[51−53] and by the precise location of electrostatic residues.[45] As seen from Figure 2A, NCBD has neither a large unfolding enthalpy nor a macroscopic free-energy barrier. It is also apparent from Gō-model simulations that the native ensemble of NCBD is significantly destabilized when using the same model parameters used for the simulation of one-state/two-state-like systems. Therefore it is tempting to speculate that molten-globules are a special case of one-state or marginal barrier proteins but with a lower stability, i.e. 'destabilized one-state folders'. By same token, the IDPs would also be expected to fall within this category as they have even smaller stability and little secondary/tertiary structure. This expectation is in line with classical statistical thermodynamic theories on the 'unfolding' of denatured states of proteins.[54,55] Strong experimental evidence to this behavior in IDPs comes from single-molecule FRET studies on the intrinsically disordered 235-residue yeast prion protein Sup35 that displays unimodal probability density at various denaturant concentrations constituting a diverse ensemble of structures.[56] A counter-example is the 119-residue P protein from *Bacillus subtilis* that shows multiphasic kinetics; barrier-limited folding through an on-pathway intermediate has been invoked to explain its folding mechanism in the presence of stabilizing osmolyte.[57] Therefore, even if it is too early to conclude that all disordered proteins indeed fold one-state-like, our results suggest that small single-domain proteins can exhibit a variety of folding behaviors ranging from IDP to molten-globule, to one-state and two-state, and that can be simply tuned by the strength and sequence separation between interactions.

The Native Ensemble and Functional Implications. Experimental evidence demonstrates that NCBD exhibits significant binding promiscuity. The question is then which of the mechanisms, induced-fit[58] or conformational selection,[8] describes best the binding in this domain? The experimentally constrained native ensemble we obtain here for the first time enables us to test for these hypotheses directly. In their earlier paper, Poulsen and co-workers show that the NMR structure of NCBD is consistent with the SRC-1/ACTR-bound conformation[25] thus supporting conformational selection for this partner. Our current results show that not only is this structure highly populated (as expected from a Gō-model) but alternate structures might have functional significance as well. For example, we find that for ∼8% of the time (an upper limit) structures similar to that reported for the complex between NCBD with IRF-3 (1ZOQ) is populated (Figure 5E,F). This conformation is characterized by H1 and H2 in similar orientation to that of 2KKJ but with H3 rotated clockwise by ∼90°. Another reported NMR structure for the free-state is also observed (NMR model: 1JJS) where H1 and H3 are antiparallel and oriented at ∼90° to H2 (Figure 5G). These results validate the generated native ensemble and suggest that conformational selection is the dominant mechanism to bind multiple partners in this domain. The methodology we have developed here can be seen as a stringent procedure to generate structural ensembles from simulations that are consistent with multiple experimental observables.

Given these results, it is very appealing to link the one-state folding mechanism and resultant heterogeneous ensemble that helps in binding multiple partners. A functional model for downhill folding that exploits the large inherent conformational fluctuations called the 'molecular rheostat' hypothesis—a protein that changes its dimensions/structure smoothly upon perturbation to bind multiple partners—has been proposed for the one-state BBL in view of its rich conformational behavior in equilibrium.[4,59,60] Similarly, 'fly cast' models have been shown to help low-barrier proteins to search for and bind their cognate DNA sequences.[61,62] NCBD, with its one-state-like folding thermodynamics, serves as an alternate test-case. The presence of diverse functional conformations (Figure 5E,F) in its native ensemble and whose population can be tuned by modifying the stability conditions (Figure 3) provides strong support to this hypothesis. It explains the promiscuous binding nature of this domain and signals that one-state folding is a mechanistic feature evolutionarily selected in this

domain for precise functional reasons, thus highlighting the direct connection between folding mechanism and function. Further experimental work on other molten-globules and IDPs in close association with simulations is needed to further explore this idea and to understand its molecular origins and significance.

■ ASSOCIATED CONTENT

Supporting Information. Figures that highlight the broadness of unfolding curve of NCBD and dispersion in melting temperatures from the Gō-model on 2KKJ, Gō-model simulations starting from the 1ZOQ structure, individual all-atom MD trajectories as a function of time at low temperature for NCBD and NTL9, comparison of the experimental and simulated SAXS curve from the MD ensemble at 370 K, NOE-like calculations from the Gō-ensemble around the midpoint temperature. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
anarayan@bsc.es; modesto.orozco@irbbarcelona.org

■ ACKNOWLEDGMENT

We are indebted to Eva de Alba and Xavier Salvatella for discussions on NMR and Pau Bernado for discussions on SAXS. We thank Victor Muñoz and Jose Manuel Sanchez-Ruiz for comments on the manuscript. This work was supported by the Spanish Ministry of Science and Innovation (BIO2009-10964 and Consolider E-Science), Instituto de Salud Carlos III (INB-Genoma España and COMBIOMED RETICS) and Fundación Marcelino Botín. A.N.N acknowledges Juan de la Cierva fellowship from the Spanish Ministry of Science and Innovation.

■ REFERENCES

(1) Monod, J.; Wyman, J.; Changeux, J. P. *J. Mol. Biol.* **1965**, *12*, 88–118.
(2) Ptitsyn, O. B. *Adv. Protein Chem.* **1995**, *47*, 83–229.
(3) Wright, P. E.; Dyson, H. J. *J. Mol. Biol.* **1999**, *293*, 321–331.
(4) Garcia-Mira, M. M.; Sadqi, M.; Fischer, N.; Sanchez-Ruiz, J. M.; Muñoz, V. *Science* **2002**, *298*, 2191–2195.
(5) Jackson, W. M.; Brandts, J. F. *Biochemistry* **1970**, *9*, 2294–2301.
(6) Maity, H.; Maity, M.; Krishna, M. M. G.; Mayne, L.; Englander, S. W. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 4741–4746.
(7) Jha, S. K.; Udgaonkar, J. B. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 12289–12294.
(8) Tsai, C. J.; Kumar, S.; Ma, B. Y.; Nussinov, R. *Protein Sci.* **1999**, *8*, 1181–1190.
(9) Levy, Y.; Cho, S. S.; Onuchic, J. N.; Wolynes, P. G. *J. Mol. Biol.* **2005**, *346*, 1121–1145.
(10) Margittai, M.; Widengren, J.; Schweinberger, E.; Schroder, G. F.; Felekyan, S.; Haustein, E.; Konig, M.; Fasshauer, D.; Grubmuller, H.; Jahn, R.; Seidel, C. A. M. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15516–15521.
(11) Li, H.; Frieden, C. *Biochemistry* **2005**, *44*, 2369–2377.
(12) Whitten, S. T.; Garcia-Moreno, B.; Hilser, V. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 4282–4287.
(13) Rueda, M.; Ferrer-Costa, C.; Meyer, T.; Perez, A.; Camps, J.; Hospital, A.; Gelpi, J. L.; Orozco, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 796–801.
(14) Hilser, V. J.; Thompson, E. B. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 8311–8315.
(15) Lange, O. F.; Lakomek, N. A.; Fares, C.; Schroder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmuller, H.; Griesinger, C.; de Groot, B. L. *Science* **2008**, *320*, 1471–1475.
(16) Wani, A. H.; Udgaonkar, J. B. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 20711–20716.
(17) Li, W.; Wolynes, P. G.; Takada, S. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 3504–3509.
(18) Dolgikh, D. A.; Gilmanshin, R. I.; Brazhnikov, E. V.; Bychkova, V. E.; Semisotnov, G. V.; Venyaminov, S. Y.; Ptitsyn, O. B. *FEBS Lett.* **1981**, *136*, 311–315.
(19) Griko, Y. V.; Privalov, P. L. *J. Mol. Biol.* **1994**, *235*, 1318–1325.
(20) Lin, C. H.; Hare, B. J.; Wagner, G.; Harrison, S. C.; Maniatis, T.; Fraenkel, E. *Mol. Cell* **2001**, *8*, 581–590.
(21) Demarest, S. J.; Martinez-Yamout, M.; Chung, J.; Chen, H. W.; Xu, W.; Dyson, H. J.; Evans, R. M.; Wright, P. E. *Nature* **2002**, *415*, 549–553.
(22) Demarest, S. J.; Deechongkit, S.; Dyson, H. J.; Evans, R. M.; Wright, P. E. *Protein Sci.* **2004**, *13*, 203–210.
(23) Qin, B. Y.; Liu, C.; Srinath, H.; Lam, S. S.; Correia, J. J.; Derynck, R.; Lin, K. *Structure* **2005**, *13*, 1269–1277.
(24) Waters, L.; Yue, B. G.; Veverka, V.; Renshaw, P.; Bramham, J.; Matsuda, S.; Frenkiel, T.; Kelly, G.; Muskett, F.; Carr, M.; Heery, D. M. *J. Biol. Chem.* **2006**, *281*, 14787–14795.
(25) Kjaergaard, M.; Teilum, K.; Poulsen, F. M. *Proc. Natl. Acad. Sci. U.S.A* **2010**, *107*, 12535–12540.
(26) Lee, C. W.; Martinez-Yamout, M. A.; Dyson, H. J.; Wright, P. E. *Biochemistry* **2010**, *49*, 9964–9971.
(27) Muñoz, V.; Sanchez-Ruiz, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17646–17651.
(28) Noel, J. K.; Whitford, P. C.; Sanbonmatsu, K. Y.; Onuchic, J. N. *Nucleic Acids Res.* **2010**, *38*, W657–W661.
(29) Clementi, C.; Nymeyer, H.; Onuchic, J. N. *J. Mol. Biol.* **2000**, *298*, 937–953.
(30) Harvey, M. J.; Giupponi, G.; De Fabritiis, G. *J. Chem. Theory Comput.* **2009**, *5*, 1632–1639.
(31) Best, R. B.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.
(32) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 65–73.
(33) Tang, K. E. S.; Dill, K. A. *Int. J. Quantum Chem.* **1999**, *75*, 147–164.
(34) Zagrovic, B.; van Gunsteren, W. F. *Proteins* **2006**, *63*, 210–218.
(35) Svergun, D. I.; Barberato, C.; Koch, M. H. J. *J. App. Crystallogr.* **1995**, 28.
(36) Chen, Y. H.; Yang, J. T.; Chau, K. H. *Biochemistry* **1974**, *13*, 3350–3359.
(37) Freire, E.; Biltonen, R. L. *Biopolymers* **1978**, *17*, 463–479.
(38) Kaya, H.; Chan, H. S. *Proteins* **2000**, *40*, 637–661.
(39) Naganathan, A. N.; Sanchez-Ruiz, J. M.; Muñoz, V. *J. Am. Chem. Soc.* **2005**, *127*, 17970–17971.
(40) Sadqi, M.; Fushman, D.; Muñoz, V. *Nature* **2006**, *442*, 317–321.
(41) Best, R. B.; Mittal, J. *Proteins* **2011**, *79*, 1318–1328.
(42) Best, R. B.; Buchete, N. V.; Hummer, G. *Biophys. J.* **2008**, *95*, L07–L09.
(43) Francis, C. J.; Lindorff-Larsen, K.; Best, R. B.; Vendruscolo, M. *Proteins* **2006**, *65*, 145–152.
(44) Ganguly, D.; Chen, J. *Proteins* **2011**, *79*, 1251–1266.
(45) Halskau, O.; Perez-Jimenez, R.; Ibarra-Molero, B.; Underhaug, J.; Muñoz, V.; Martinez, A.; Sanchez-Ruiz, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 8625–8630.
(46) Paci, E.; Smith, L. J.; Dobson, C. M.; Karplus, M. J. *Mol. Biol.* **2001**, *306*, 329–347.
(47) Fung, A.; Li, P.; Godoy-Ruiz, R.; Sanchez-Ruiz, J. M.; Muñoz, V. *J. Am. Chem. Soc.* **2008**, *130*, 7489–7495.
(48) Naganathan, A. N.; Li, P.; Perez-Jimenez, R.; Sanchez-Ruiz, J. M.; Muñoz, V. *J. Am. Chem. Soc.* **2010**, *132*, 11183–11190.
(49) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins* **1995**, *21*, 167–195.

(50) Naganathan, A. N.; Doshi, U.; Muñoz, V. *J. Am. Chem. Soc.* **2007**, *129*, 5673–5682.
(51) Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. *J. Mol. Biol.* **1995**, *252*, 460–471.
(52) Zuo, G.; Wang, J.; Wang, W. *Proteins* **2006**, *63*, 165–173.
(53) Knott, M.; Chan, H. S. *Proteins* **2006**, *65*, 373–391.
(54) Ikegami, A. *Adv. Chem. Phys.* **1981**, *46*, 363–413.
(55) Dill, K. A.; Stigter, D. *Adv. Protein Chem.* **1995**, *46*, 59–104.
(56) Mukhopadhyay, S.; Krishnan, R.; Lemke, E. A.; Lindquist, S.; Deniz, A. A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2649–2654.
(57) Chang, Y. C.; Oas, T. G. *Biochemistry* **2010**, *49*, 5086–5096.
(58) Koshland, D. E. *Proc. Natl. Acad. Sci. U.S.A* **1958**, *44*, 98–104.
(59) Naganathan, A. N.; Doshi, U.; Fung, A.; Sadqi, M.; Muñoz, V. *Biochemistry* **2006**, *45*, 8466–8475.
(60) Desai, T. M.; Cerminara, M.; Sadqi, M.; Muñoz, V. *J. Biol. Chem.* **2010**, *285*, 34549–34556.
(61) Levy, Y.; Onuchic, J. N.; Wolynes, P. G. *J. Am. Chem. Soc.* **2007**, *129*, 738–739.
(62) Murugan, R. *Biophys. J.* **2010**, *99*, 353–359.