Thermodynamics and folding landscapes of large proteins from a statistical mechanical model

Soundhararajan Gopi, Akashnathan Aranganathan, Athi N. Naganathan[*]

Department of Biotechnology, Bhupat & Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, India

ABSTRACT

Statistical mechanical models that afford an intermediate resolution between macroscopic chemical models and all-atom simulations have been successful in capturing folding behaviors of many small single-domain proteins. However, the applicability of one such successful approach, the Wako-Saitô-Muñoz-Eaton (WSME) model, is limited by the size of the protein as the number of conformations grows exponentially with protein length. In this work, we surmount this size limitation by introducing a novel approximation that treats stretches of 3 or 4 residues as blocks, thus reducing the phase space by nearly three orders of magnitude. The performance of the 'bWSME' model is validated by comparing the predictions for a globular enzyme (RNase H) and a repeat protein (IκBα), against experimental observables and the model without block approximation. Finally, as a proof of concept, we predict the free-energy surface of the 370-residue, multi-domain maltose binding protein and identify an intermediate in good agreement with single-molecule force-spectroscopy measurements. The bWSME model can thus be employed as a quantitative predictive tool to explore the conformational landscapes of large proteins, extract the structural features of putative intermediates, identify parallel folding paths, and thus aid in the interpretation of both ensemble and single-molecule experiments.

1. Introduction

Structure-based models of protein folding have revealed rich insights into the conformational behavior of numerous proteins, enzymes, protein–protein and protein-DNA complexes (Mirny and Shakhnovich, 2001; Papoian and Wolynes, 2003; Clementi et al., 2003; Levy et al., 2005; Hyeon and Thirumalai, 2011; Chan et al., 2011; Li et al., 2011; Orozco et al., 2011; Naganathan, 2013a; Truong et al., 2015; Kmiecik et al., 2016). These models work on the general principle that only those interactions present in the native structure define the folding mechanism, as originally proposed by Gō (Taketomi et al., 1975), a phenomenology that is enshrined in the energy landscape theory of protein folding (Bryngelson et al., 1995), and validated *via* extensive analysis of long time-scale all-atom MD simulations (Best et al., 2013). The successes of these models have spawned an array of coarse-grained treatments of protein folding to enable efficient sampling and thus minimizing the time taken to generate well-equilibrated folding paths and ensembles. While many of the treatments are customized to the problem that is being addressed, it is common to study protein folding via models that consider only the C_α-atoms or a combination of C_α and minimal side-chain and backbone representations. The latter allows for introducing energetic

flavors via either specific residue–residue interaction energy matrices (for example, the Miyazawa-Jernigan potential (Miyazawa and Jernigan, 1985)) or electrostatics through monopole–monopole interactions.

One class of structure-based models are statistical mechanical models that treat residues as independent units of folding (Wako and Saito, 1978a, 1978b; Ikegami, 1981; Go and Abe, 1981; Abe and Go, 1981; Zwanzig, 1995; Hilser and Freire, 1996; Hilser et al., 1998, 2006; Muñoz and Eaton, 1999; Alm and Baker, 1999; Muñoz, 2002; Bruscolini and Pelizzola, 2002; Bruscolini and Naganathan, 2011; Naganathan, 2012) (Fig. 1A). Unlike all-atom or coarse-grained simulations that accumulate conformations as a function of time, statistical models pre-assume an ensemble from physical considerations. Following this, the statistical weight of every microstate and hence the overall canonical partition function is calculated from structure-based expectations enabling predictions of numerous thermodynamic features including heat capacity, residue folding probabilities, free-energy profiles and surfaces (Muñoz, 2001; Naganathan, 2016). The Wako-Saitô-Muñoz-Eaton (WSME) model is one such statistical mechanical model that was first developed by Wako and Saitô (Wako and Saito, 1978a, 1978b), discussed in detail by Gō and Abe (Go and Abe, 1981; Abe and Go, 1981), and then later independently developed by Muñoz and Eaton (1999). Originally seen as a physical tool
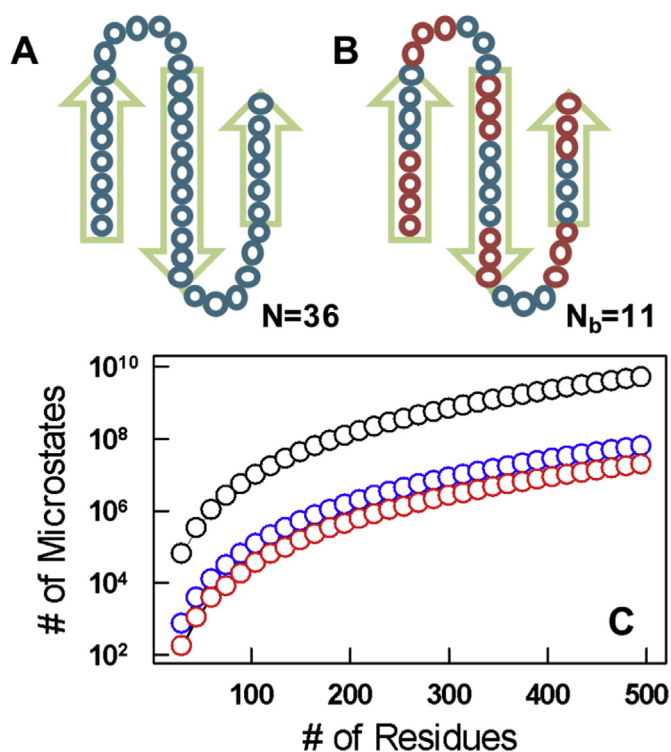

Fig. 1. The bWSME model. (A, B) Conformational units as residues (panel A) or blocks (panel B) for a three-stranded beta-hairpin. The eleven blocks are alternatively colored in red and blue in panel B. (C) Number of microstates as a function of number of residues in a protein for block size = 1 (*i.e.* residues as conformational units; black), block size = 3 (blue) and block size = 4 (red). Note that the ordinate is in logarithmic scale.

to predict the folding rates of proteins from three-dimensional structures (Muñoz and Eaton, 1999; Henry and Eaton, 2004), the model has expanded its scope to quantitatively analyze folding behaviors of folded globular domains (Bruscolini and Naganathan, 2011; Garcia-Mira et al., 2002; Narayan and Naganathan, 2014, 2017, 2018; Naganathan and Muñoz, 2014; Naganathan et al., 2015; Munshi and Naganathan, 2015; Rajasekaran et al., 2016; Narayan et al., 2017; Itoh and Sasai, 2006), repeat proteins (Faccin et al., 2011; Sivanandan and Naganathan, 2013; Hutton et al., 2015), disordered proteins (with appropriate controls) (Naganathan and Orozco, 2013; Gopi et al., 2015; Munshi et al., 2018a), predict and engineer thermodynamic stabilities of proteins via mutations (Naganathan, 2012, 2013b; Rajasekaran et al., 2017) and entropic effects (Rajasekaran et al., 2016), model allosteric transitions (Itoh and Sasai, 2011; Sasai et al., 2016), protein-DNA binding (Munshi et al., 2018b), quantifying folding pathways at different levels of resolution (Henry et al., 2013; Kubelka et al., 2008; Gopi et al., 2017), force-spectroscopic measurements (Imparato et al., 2007) and even crowding effects (Caraglio and Pelizzola, 2012).

In the classic WSME model, each residue is allowed to sample two sets of conformations – folded (represented as *1* in binary notation) and unfolded (*0*) – thus allowing for a maximum of 2^N conformations or microstates for an *N*-residue polypeptide. The exact solution, or the total partition function with contributions from all the 2^N microstates, can be calculated via different methods (Wako and Saito, 1978b; Go and Abe, 1981; Bruscolini and Pelizzola, 2002; Henry and Eaton, 2004). The underlying assumption in these methods is that while there can be numerous independent nucleating events or folded islands (stretches of *1*s), two different islands can interact if they are interacting in the native structure (Gō-model) and importantly only when all the intervening residues connecting the two islands are also folded. In other words, the model assumes a specific folding mechanism wherein local interactions form first while non-local interactions form later. It is of course possible

that two structured regions can interact despite an intervening unfolded region and constitute a valid microstate as long as the entropic destabilization is sufficiently offset by a gain in energy from the interacting regions. Sasai and co-workers developed the extended WSME model to address this *via* virtual loop closures (Inanami et al., 2014).

In parallel, Eaton and co-workers came up with a more realistic ensemble description by considering only single stretches of folded residues (single-sequence approximation or SSA), two stretches of folded residues (double-sequence approximation or DSA) and DSA allowing for interactions across structured islands even if the intervening residues are unfolded (DSAw/L) (Henry and Eaton, 2004; Kubelka et al., 2008). The number of model microstates for an *N* residue protein under this ensemble definition, which we call rWSME as residues (r) are the folding units and for easy reference, can be calculated using the combinatorial expression $C + 2C$. This gives an upper limit on the number of microstates, since microstates defined by DSAw/L depend on the interaction across the structured islands if and only if they are present in the folded structure. While this approach reduces the number of microstates drastically (compared to the 2^N states), it has been successful in predicting the folding mechanism of the Villin head-piece domain in quantitative agreement with experiments and all-atom MD simulations (Henry et al., 2013). A similar model but with more detailed energetics (van der Waals interactions, electrostatics, implicit solvation and excess conformational entropy) has been instrumental in providing a detailed description of folding pathway heterogeneity in five different proteins in quantitative agreement with ensemble and single-molecule data (Gopi et al., 2017).

Despite these obvious advantages, one downside is the size limitation of this method. For example, the maximum number of microstates for a 300 or 400 residue protein (as in multi-domain proteins) from the ensemble description involving rWSME is >670.5 million and >2.1 billion, respectively, making it computationally intensive (Fig. 1C). In this work, we circumvent the apparent size-limitation by introducing an approximation to reduce the accessible phase space – called the bWSME model with 'b' for block – by grouping residues into blocks (for example, see Fig. 1B and Supporting Figure S1). We show that this approximation works as well as the original approach providing a detailed view of conformational landscapes of large proteins in quantitative agreement with experiments.

2. Methods

2.1. The bWSME model

The bWSME model considers short stretches of 2, 3, 4 or 5 residues folding together and acting as a single unit. The blocks are constructed in such a way to ensure that they do not span two different secondary-structure descriptions as identified by STRIDE (Heinig and Frishman, 2004). When we say a block size consideration for a given protein is, say 3, it means that the majority of the blocks in the protein will have a length of 3. There will be exceptions specifically when the length of a specific secondary structural region is not a multiple of 3 or if there is only one residue identified as a coil but bridging two secondary structure elements. In such cases, the block will span just 2 residues or even 1 residue. The effective free energy of the microstate with a folded structure between and involving blocks *p* and *q* (*p*, *q*) can be calculated as follows,

$$\Delta G_{p,q} = \Delta G_{p,q}^{stab} - T \sum_{i=p}^{q} \Delta S_i^{conf}$$

To retain the residue-level information and provide a physically reasonable energetic description in the bWSME model variant, the stabilization free energy for the microstate (*p*, *q*) is written as,

$$\Delta G_{p,q}^{stab} = \sum_{i=p}^{q} \sum_{j \geq i}^{q} \sum_{k=L(i)} \sum_{[l=L(j)]>k} \Delta G_{k,l}^{stab}$$

where L is the set of residues comprising the protein, $L(i)$ and $L(j)$ represents the set of constituent residues of block i and j, respectively. Note that the above expression accounts for the stabilization free energy due to the interaction between residues within the same block. The stabilization free energy includes contributions from mean-field van der Waals interaction term (E_{vdW}; a uniform interaction energy ξ for the vdW contacts identified using a Gō-like approach), all-to-all electrostatics without a distance cut-off modeled using the Debye-Hückel formalism (E_{Elec}) and an implicit solvation term (ΔG_{Solv}, calculated as the heat capacity change per native contact ΔC_p^{cont}) (Naganathan, 2012, 2013b).

The entropic cost associated with the microstate (p, q) is given as,

$$\Delta S_{p,q}^{conf} = \sum_{i=p}^{q} \sum_{j=L(i)} \Delta S_j^{conf}$$

here, ΔS_j^{conf} is the entropic cost associated with fixing residue j in the native conformation, $L(i)$ represents the set of residues within block i. An excess entropic penalty ($\Delta\Delta S = -6.1$ J mol^{-1} K^{-1} per residue) is additionally assigned to residues identified as coil by STRIDE and glycine residues. Proline, which exhibits limited flexibility, is assigned an entropic cost of 0 J mol^{-1} K^{-1} per residue (Daquino et al., 1996).

2.2. Partition function and predicting experimental observables

The total partition function (Z) is calculated as the sum of statistical weights associated with the microstates defined within the model framework,

$$Z = \sum_{i=1}^{\mu} \exp(-\Delta G_i / RT)$$

here, μ is the total number of microstates, ΔG_i is the effective free energy of the microstate i, R is the gas constant and T is temperature. Heat capacity profiles can be directly calculated through derivatives of total partition function as a function of temperature:

$$C_P \cong C_V = 2RT \left(\frac{d \ln Z}{dT}\right) + RT^2 \left(\frac{d^2 \ln Z}{dT^2}\right)$$

The folding probability of residue i (χ_i) is calculated by algorithmically accumulating the statistical weights of microstates in which residue i is folded,

$$\chi_i = Z^{-1} \sum_{k} \exp(-\Delta G_k / RT)$$

Here, ΔG_k is the effective free energy of microstate k, and k runs over all the microstates with residue i folded. The mean residue folding probability ($\langle\chi_i\rangle_T$) as a function of temperature can serve as a proxy for global unfolding curve, say as monitored by far-UV CD or fluorescence experiments. One-dimensional free energy profiles and two-dimensional free energy surfaces are constructed by appropriately summing up the statistical weights of the microstates as a function of the number of structured blocks, either globally or in a specific part of the structure.

In the current work, heavy-atom contacts are identified using a 6 Å distance cut-off excluding the nearest neighbors, and charges are assigned based on the protonation state at pH 7. The value of dielectric constant in the Debye-Hückel formalism is fixed to 29 derived from our previous works involving comparison of unfolding curves of homologous proteins and mutations involving charged residues (Naganathan, 2012, 2013b). The model is independently parameterized for RNase H (1F21), IκBα (1NFI; with contact map correction similar to a previous work (Sivanandan and Naganathan, 2013)) and Maltose-Binding protein (MBP; 1OMP) to reproduce the experimental thermal unfolding profiles (Table S1). 1D free-energy profiles and 2D free-energy surfaces are constructed at 298 K, unless otherwise mentioned.

3. Results and discussion

3.1. Free-energy profiles and intermediates are insensitive to block definition

We validate the bWSME predictions using the 213-residue repeat protein IκBα as a model system (Fig. 2A), the folding and dynamics of
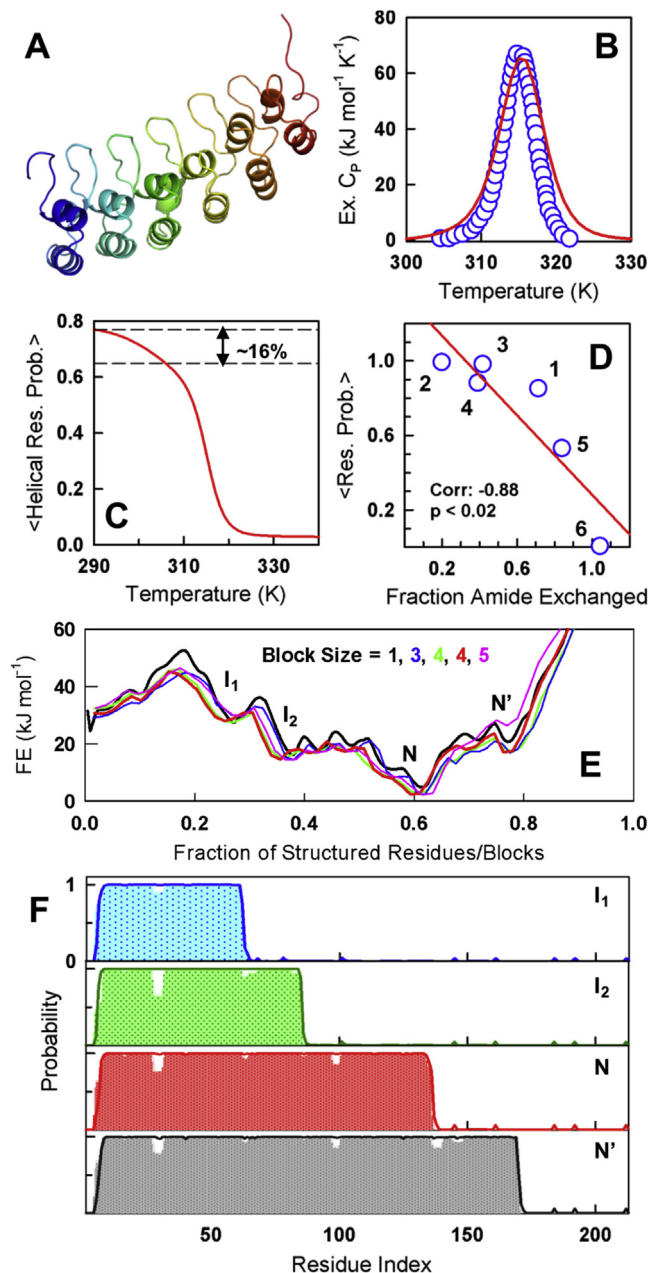

Fig. 2. Conformational landscape of IκBα. (A) The holo-structure of the six-repeat α-helical protein IκBα. (B) The experimental excess heat capacity (blue) used for calibrating the model parameters together with the fit (red). (C) The average probability of finding a helical residue folded as a function of temperature with a 16% 'pre-transition' amplitude as observed in far-UV CD experiments. (D) Fraction of amide exchanged correlates well with the folded fraction within individual repeats. The numbers within the figure represent the repeat identity. (E) One-dimensional free energy profiles for various block definitions. Two slightly different block definitions of length 4 (green and red) were also studied. (F) The folded probability as a function of residue index for different states as observed from the free-energy profile in panel E. Solid lines represent the residue probabilities while the bWSME model predicted block probabilities are shown in the shaded regions.

which is driven by a combination of order and disorder (Lamboy et al., 2011, 2013). For this protein, the number of possible microstates is >169 million from the rWSME while it reduces to just >900,000 by the bWSME treatment and on considering a block-size of 4 (Figure S1). As a first step, we adjust the model parameters to simultaneously reproduce the excess heat capacity curve and the unfolding pre-transition seen in experiments (Ferreiro et al., 2007) (Fig. 2B, C) by eliminating the non-local contacts (but maintaining local contacts) involving the 5th and 6th repeats as described before (Sivanandan and Naganathan, 2013). The effective entropic penalty is estimated to be −14.5 J mol^{-1} K^{-1} per residue, which is of the same order expected from size-scaling arguments on large proteins (−16.5 J mol^{-1} K^{-1} per residue). Moreover, the van der Waals interaction energy is estimated to be −58.6 J mol^{-1} per atom-level contact within 6 Å, which is very similar to the expectation from atomic-level force–field parameters (−46.1 J mol^{-1} interaction energy for two carbon atoms separated by 6 Å; Table S1).

The bWSME is able to capture the fraction of amide exchanged (averaged over all residues in a specific repeat) employing the mean residue unfolding probability per repeat as a proxy (Fig. 2D). The predicted 1D free energy profiles are characterized by multiple intermediates and partially structured states with the fully folded native state never populated. This can be seen from the fact that the minimum of the free energy profile occurs at a reaction coordinate value of 0.6 that corresponds to just ~128 folded residues (213*0.6, Fig. 2E). Importantly, the basic features of the free energy profiles and the nature of the intermediates are relatively insensitive to the block definition (Fig. 2E, F). On comparison, it is clear that the native ensemble is defined by structure only in the first four repeats with a state N' representing a conformation in which the 5th repeat is also folded. The free energy profiles however appear slightly smoother in the block definitions, which is in agreement with the expectation from block averaging that further coarse-grains the landscape. The overall picture suggests a complex native ensemble that is intimately determined by disorder in the 5th and 6th repeats. Our results are therefore consistent not only with experiments (Lamboy et al., 2011) but also with the previous attempts at reproducing the conformational landscape using the WSME model with 2^{213} microstates (Sivanandan and Naganathan, 2013).

Similarly, the basic energetics of the model was parameterized by using the thermal unfolding curve of RNase H, a globular enzyme of 152 residues (Fig. 3A, B, Figure S1). The number of microstates as per the rWSME model is >43 million that reduces to just >200,000 for the bWSME variant (block size = 4; Table S1). The predicted free energy profile with and without block definition are in agreement with each other (Fig. 3C) and with the WMSE model that accounts for 2^{152} microstates (Narayan and Naganathan, 2014). The identity of the predicted intermediates is also in agreement between the two block definitions (Fig. 3D). HX-MS experiments point to a likely folding mechanism wherein the helices A/D form first, followed by helix BC and strand 4, and then finally the strands 1/2/3/5 and helix E (Hu et al., 2013). The bWSME model predicts a mechanism with the helices BC and D forming first (residues 70–110, intermediate I$_1$), followed by helix A (residues 42–56, I$_2$) and strands 4/5 (residues 61–67, 112–118, I$_2$), strands 1/2/3 (residues 1–40, I$_3$) and finally helix E (residues 126–140) (Fig. 3D, Figure S2). The differences in the nature of the initial nucleating event between the two approaches (experiments and bWSME model) likely originates from the intrinsic secondary structure propensity of helix A (Narayan and Naganathan, 2014), which is not captured in the current model. The late intermediate I$_3$ is identical to the high free-energy excited state observed in independent HX experiments (Chamberlain et al., 1996). Effectively, the agreement of the block model predictions with experiments and the internal consistency with both the rWSME and the WSME model with 2^N states strongly support our reduced approach that maximizes the information with a fewer number of microstates but without losing the strong underlying physical basis.
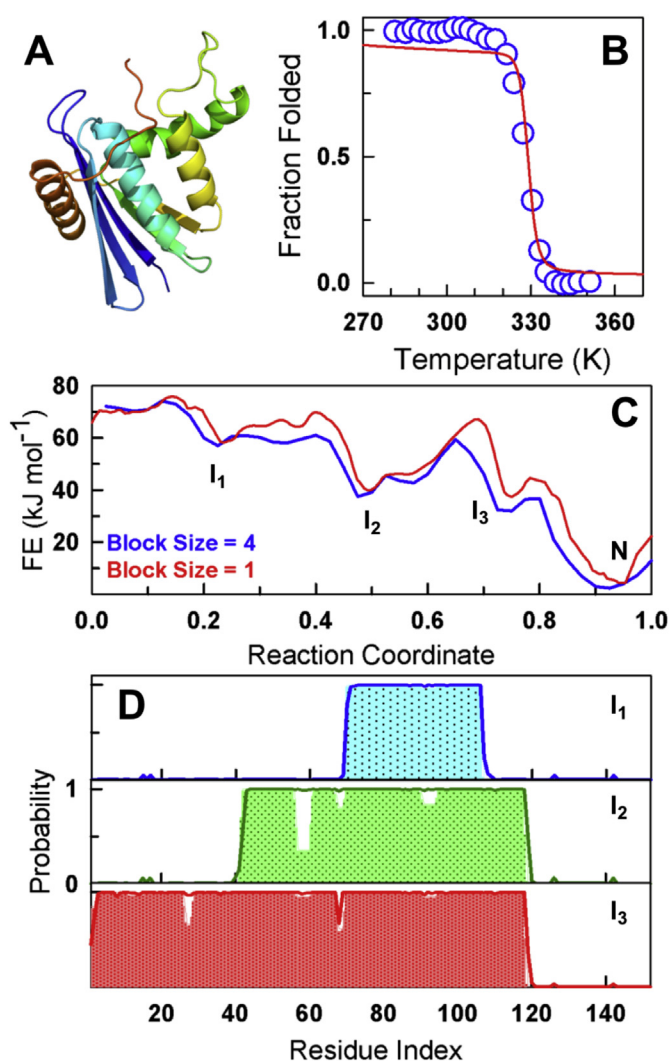

Fig. 3. Predicted folding mechanism of RNase H. (A) Structure of the 152-residue RNase H. (B) Experimental fraction folded (blue) used to calibrate the model parameters and the resulting fit (red). (C) Predicted one-dimensional free-energy profile as a function of reaction coordinate, the fraction of structured residues or blocks. (D) The identity of intermediates extracted from the rWSME (lines) and bWSME (shaded regions).

3.2. Conformational landscape of maltose binding protein

To further explore the applicability of the model, we study the folding of Maltose Binding Protein (MBP), a mixed α/β protein with a complex topology comprising 370 residues (Fig. 4A, S1). Single-molecule force spectroscopy experiments indicate that MBP unfolds via an intermediate but through two parallel paths (Aggarwal et al., 2011). It is challenging to employ the rWSME model for this protein given that this would involve algorithmically accumulating the statistical weights of >1.5 billion microstates. However, the block approximation (block size = 4, 104 sequential blocks) makes this problem more amenable by reducing the number of microstates to just over 9.5 million (Table S1). We reproduce the excess heat capacity profile of MBP to calibrate the model parameters that determine the unfolding sharpness (entropic penalty per residue), melting temperature (van der Waals interaction energy per contact) and the higher unfolding heat capacity (heat capacity change on forming a native contact) (Fig. 4B).

The resulting one-dimensional free-energy profile highlights a single intermediate at ~40 structured blocks following which a downhill gradient is seen towards the native state (Fig. 4C). A structural view of
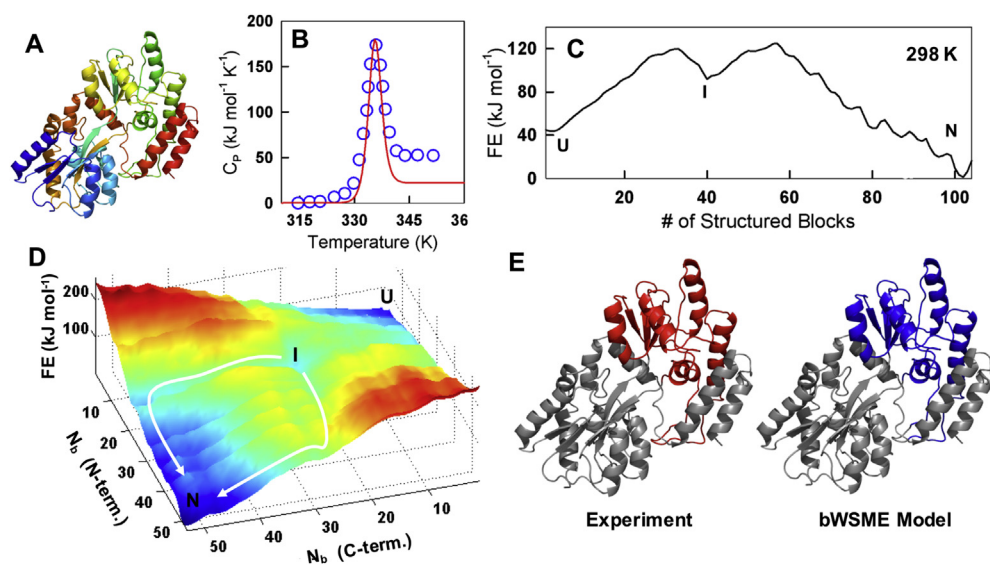
Fig. 4. Intermediates and parallel folding paths in MBP conformational landscape. (A) Structure of the 370-residue multi-domain MBP. (B) The experimental excess heat capacity curve (blue) and the bWSME model fit (red). The fit was primarily employed to estimate the thermodynamic cooperativity in the system. (C) The one-dimensional free-energy profile as a function of the number of structured blocks highlighting an intermediate-like state (I). (D) A free-energy surface generated by partitioning the structure into two equal halves involving 52 blocks in the N- and C-termini, respectively. The arrows highlight the two likely folding paths from the intermediate. (E) The identity of the folded regions in the intermediate as obtained from the bWSME model prediction (blue) compared against the experiments (red). The gray regions represent unfolded regions.

the intermediate can be obtained from a three-dimensional free energy surface with the number of structured blocks at the N- and C-termini as the x- and y-coordinates (52 blocks each) - it represents to a structure with 23 and 19 structured blocks at the C- and N-terminus, respectively (Fig. 4D). In terms of the residue level information, this corresponds to a fully folded C-terminal domain (residues 114–257 or blocks 34–74 based on the MBP domain definition) in exact agreement with single-molecule experiments (Fig. 4E) (Aggarwal et al., 2011). In other words, the bWSME model predicts a folding mechanism in which C-terminal domain of MBP forms first (Figure S3), following which two parallel paths can be populated. In one macroscopic high free-energy path, the entire N-terminal region of the protein forms (including the residues constituting the N-terminal domain, residues 1–113 or blocks 1–33), following which the terminal helices (residues 286–370 or blocks 82–104) fold. In a second but low free-energy path, the terminal helices form first and then mechanically weak N-terminal domain folds in a gradual manner contributing to the shallow gradient towards the folded state in the 1D free energy profile. The exact reverse sequence of events is observed in the single-molecule force spectroscopy unfolding experiments (Aggarwal et al., 2011) highlighting the advantages of bWSME method for large proteins. We would like to emphasize that the surface representations and free-energy profiles provide only the most probable macroscopic folding paths and not the microscopic residue- or block-level routes that could be more complex, as previously shown for several small single-domain proteins (Gopi et al., 2017).

4. Conclusions

We develop and validate a simplified version of the WSME model by switching the fundamental conformational units from residues to blocks that can be 2 to 5 residues in length. This re-formatting reduces the combinations drastically and therefore the number of conformational states by nearly three orders of magnitude. The model, termed the block WSME method or bWSME, can be employed as a predictive tool in the same vein as the original version developed by Wako and Saitô or in the sequence approximations of Muñoz and Eaton. The fewer number of microstates enables the generation of conformational landscapes of large proteins (~300–400 residues in length) without compromising on the predictive ability or the physical underpinnings. The bWSME model does not assume any folding mechanism and allows for stabilization of conformations *via* both local and non-local interactions. Therefore, it can be employed as a first step to probe regions of protein that are thermodynamically less (more) stable contributing to unfolding

(folding) and in identifying the number and nature of macroscopic folding paths.

One of the limitations of block approximation is that the populations of the partially structured states (for example, Figs. 2E or 3C) can vary slightly depending on the block approximation employed. Second, many large proteins unfold irreversibly on temperature changes that cannot be captured by the current model. Therefore, it is challenging to obtain a perfect agreement between experimental and predicted unfolding curve (for example, see Figs. 2B, 3B and 4B). The fitting procedures should therefore be seen as avenues to capture the overall sharpness or cooperativity of the unfolding transition. Third, while the model currently incorporates contributions from van der Waals interactions, electrostatics, and implicit solvation, it does not include intrinsic conformational preferences of amino acids to be in specific secondary structure elements, a feature that can be incorporated in future versions. Despite these, we expect the bWSME model to be successful in not only exploring conformational landscapes but also to quantify the effect of mutations in large multi-domain proteins that frequently underlie many diseases. It should also be possible to capture the effect of DNA, RNA or even ligand binding on the conformational landscapes of larger proteins by extending a recently developed protocol that maps the protein-ligand interactions on to the protein (Munshi et al., 2018b). Similarly, post-translational modifications, particularly those that introduce or remove charges can be introduced in a straightforward manner as before (Gopi et al., 2015). The bWSME model thus stands on the cusp of addressing and exploring numerous questions on the conformational behavior of large proteins.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crstbi.2019.10.002.

References

Abe, H., Go, N., 1981. Noninteracting local-structure model of folding and unfolding transition in globular proteins. II. Application to two-dimensional lattice proteins. Biopolymers 20, 1013–1031. https://doi.org/10.1002/bip.1981.360200512.

Aggarwal, V., Kulothungan, S.R., Balamurali, M.M., Saranya, S.R., Varadarajan, R., Ainavarapu, S.R., 2011. Ligand-modulated parallel mechanical unfolding pathways of maltose-binding proteins. J. Biol. Chem. 286, 28056–28065. https://doi.org/10.1074/jbc.M111.249045.

Alm, E., Baker, D., 1999. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. Proc. Natl. Acad. Sci. U.S.A. 96, 11305–11310. https://doi.org/10.1073/pnas.96.20.11305.

Best, R.B., Hummer, G., Eaton, W.A., 2013. Native contacts determine protein folding mechanisms in atomistic simulations. Proc. Natl. Acad. Sci. U.S.A. 110, 17874–17879. https://doi.org/10.1073/pnas.1311599110.

Bruscolini, P., Naganathan, A.N., 2011. Quantitative prediction of protein folding behaviors from a simple statistical model. J. Am. Chem. Soc. 133, 5372–5379. https://doi.org/10.1021/ja110884m.

Bruscolini, P., Pelizzola, A., 2002. Exact solution of the Muñoz-Eaton model for protein folding. Phys. Rev. Lett. 88, 258101. https://doi.org/10.1103/PhysRevLett.88.258101.

Bryngelson, J.D., Onuchic, J.N., Socci, N.D., Wolynes, P.G., 1995. Funnels, pathways, and the energy landscape of protein-folding - a synthesis. Proteins 21, 167–195. https://doi.org/10.1002/prot.340210302.

Caraglio, M., Pelizzola, A., 2012. Effects of confinement on thermal stability and folding kinetics in a simple Ising-like model. Phys. Biol. 9, 016006. https://doi.org/10.1088/1478-3975/9/1/016006.

Chamberlain, A.K., Handel, T.M., Marqusee, S., 1996. Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. Nat. Struct. Biol. 3, 782–787. https://doi.org/10.1038/nsb0996-782.

Chan, H.S., Zhang, Z., Wallin, S., Liu, Z., 2011. Cooperativity, local-nonlocal coupling, and nonnative interactions: principles of protein folding from coarse-grained models. In: Leone, S.R., Cremer, P.S., Groves, J.T., Johnson, M.A. (Eds.), Ann. Rev. Phys. Chem, Annual Review of Physical Chemistry, 62, pp. 301–326. https://doi.org/10.1146/annurev-physchem-032210-103405.

Clementi, C., Garcia, A.E., Onuchic, J.N., 2003. Interplay among tertiary contacts, secondary structure formation and side-chain packing in the protein folding mechanism: all-atom representation study of protein L. J. Mol. Biol. 326, 933–954. https://doi.org/10.1016/s0022-2836(02)01379-7.

Daquino, J.A., Gomez, J., Hilser, V.J., Lee, K.H., Amzel, L.M., Freire, E., 1996. The magnitude of the backbone conformational entropy change in protein folding. Proteins 25, 143–156. https://doi.org/10.1002/(SICI)1097-0134(199606)25:2<143::AID-PROT1>3.0.CO;2-J.

Faccin, M., Bruscolini, P., Pelizzola, A., 2011. Analysis of the equilibrium and kinetics of the ankyrin repeat protein myotrophin. J. Chem. Phys. 134, 075102. https://doi.org/10.1063/1.3535562.

Ferreiro, D.U., Cervantes, C.F., Truhlar, S.M.E., Cho, S.S., Wolynes, P.G., Komives, E.A., 2007. Stabilizing I kappa B alpha by "consensus" design. J. Mol. Biol. 365, 1201–1216. https://doi.org/10.1016/j.jmb.2006.11.044.

Garcia-Mira, M.M., Sadqi, M., Fischer, N., Sanchez-Ruiz, J.M., Muñoz, V., 2002. Experimental identification of downhill protein folding. Science 298, 2191–2195. https://doi.org/10.1126/science.1077809.

Go, N., Abe, H., 1981. Noninteracting local-structure model of folding and unfolding transition in globular proteins. I. Formulation. Biopolymers 20, 991–1011. https://doi.org/10.1002/bip.1981.360200511.

Gopi, S., Rajasekaran, N., Singh, A., Ranu, S., Naganathan, A.N., 2015. Energetic and topological determinants of a phosphorylation-induced disorder-to-order protein conformational switch. Phys. Chem. Chem. Phys. 17, 27264–27269. https://doi.org/10.1039/c5cp04765j.

Gopi, S., Singh, A., Suresh, S., Paul, S., Ranu, S., Naganathan, A.N., 2017. Toward a quantitative description of microscopic pathway heterogeneity in protein folding. Phys. Chem. Chem. Phys. 19, 20891–20903. https://doi.org/10.1039/c7cp03011h.

Heinig, M., Frishman, D., 2004. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. Nuc. Acids Res. 32, W500–W502. https://doi.org/10.1093/nar/gkh429.

Henry, E.R., Eaton, W.A., 2004. Combinatorial modeling of protein folding kinetics: free energy profiles and rates. Chem. Phys. 307, 163–185. https://doi.org/10.1016/j.chemphys.2004.06.064.

Henry, E.R., Best, R.B., Eaton, W.A., 2013. Comparing a simple theoretical model for protein folding with all-atom molecular dynamics simulations. Proc. Natl. Acad. Sci. U.S.A. 110, 17880–17885. https://doi.org/10.1073/pnas.1317105110.

Hilser, V.J., Freire, E., 1996. Structure-based calculation of the equilibrium folding pathway of proteins. Correlation with hydrogen exchange protection factors. J. Mol. Biol. 262, 756–772. https://doi.org/10.1006/jmbi.1996.0550.

Hilser, V.J., Dowdy, D., Oas, T.G., Freire, E., 1998. The structural distribution of cooperative interactions in proteins: analysis of the native state ensemble. Proc. Natl. Acad. Sci. U.S.A. 95, 9903–9908. https://doi.org/10.1073/pnas.95.17.9903.

Hilser, V.J., Garcia-Moreno, B., Oas, T.G., Kapp, G., Whitten, S.T., 2006. A statistical thermodynamic model of the protein ensemble. Chem. Rev. 106, 1545–1558. https://doi.org/10.1021/cr040423+.

Hu, W.B., Walters, B.T., Kan, Z.Y., Mayne, L., Rosen, L.E., Marqusee, S., Englander, S.W., 2013. Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. Proc. Natl. Acad. Sci. U.S.A. 110, 7684–7689. https://doi.org/10.1073/pnas.1305887110.

Hutton, R.D., Wilkinson, J., Faccin, M., Sivertsson, E.M., Pelizzola, A., Lowe, A.R., Bruscolini, P., Itzhaki, L.S., 2015. Mapping the topography of a protein energy landscape. J. Am. Chem. Soc. 137, 14610–14625. https://doi.org/10.1021/jacs.5b07370.

Hyeon, C., Thirumalai, D., 2011. Capturing the essence of folding and functions of biomolecules using coarse-grained models. Nat. Commun. 2, 487. https://doi.org/10.1038/ncomms1481.

Ikegami, A., 1981. Statistical thermodynamics of proteins and protein denaturation. Adv. Chem. Phys. 46, 363–413. https://doi.org/10.1002/9780470142653.ch6.

Imparato, A., Pelizzola, A., Zamparo, M., 2007. Ising-like model for protein mechanical unfolding. Phys. Rev. Lett. 98, 148102. https://doi.org/10.1103/PhysRevLett.98.148102.

Inanami, T., Terada, T.P., Sasai, M., 2014. Folding pathway of a multidomain protein depends on its topology of domain connectivity. Proc. Natl. Acad. Sci. U.S.A. 111, 15969–15974. https://doi.org/10.1073/pnas.1406244111.

Itoh, K., Sasai, M., 2006. Flexibly varying folding mechanism of a nearly symmetrical protein: B domain of protein A. Proc. Natl. Acad. Sci. U.S.A. 103, 7298–7303. https://doi.org/10.1073/pnas.0510324103.

Itoh, K., Sasai, M., 2011. Statistical mechanics of protein allostery: roles of backbone and side-chain structural fluctuations. J. Chem. Phys. 134, 125102. https://doi.org/10.1063/1.3565025.

Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A.E., Kolinski, A., 2016. Coarse-grained protein models and their applications. Chem. Rev. 116, 7898–7936. https://doi.org/10.1021/acs.chemrev.6b00163.

Kubelka, J., Henry, E.R., Cellmer, T., Hofrichter, J., Eaton, W.A., 2008. Chemical, physical, and theoretical kinetics of an ultrafast folding protein. Proc. Natl. Acad. Sci. U.S.A. 105, 18655–18662. https://doi.org/10.1073/pnas.0808600105.

Lamboy, J.A., Kim, H., Lee, K.S., Ha, T., Komives, E.A., 2011. Visualization of the nanospring dynamics of the I kappa B alpha ankyrin repeat domain in real time. Proc. Natl. Acad. Sci. U.S.A. 108, 10178–10183. https://doi.org/10.1073/pnas.1102226108.

Lamboy, J.A., Kim, H., Dembinski, H., Ha, T., Komives, E.A., 2013. Single-molecule FRET reveals the native-state dynamics of the IκBα ankyrin repeat domain. J. Mol. Biol. 425, 2578–2590. https://doi.org/10.1016/j.jmb.2013.04.015.

Levy, Y., Cho, S.S., Onuchic, J.N., Wolynes, P.G., 2005. A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes. J. Mol. Biol. 346, 1121–1145. https://doi.org/10.1016/j.jmb.2004.12.021.

Li, W., Wolynes, P.G., Takada, S., 2011. Frustration, specific sequence dependence, and nonlinearity in large-amplitude fluctuations of allosteric proteins. Proc. Natl. Acad. Sci. U.S.A. 108, 3504–3509. https://doi.org/10.1073/pnas.1018983108.

Mirny, L., Shakhnovich, E., 2001. Protein folding theory: from lattice to all-atom models. Ann. Rev. Biophys. Biomol. Struct. 30, 361–396. https://doi.org/10.1146/annurev.biophys.30.1.361.

Miyazawa, S., Jernigan, R.L., 1985. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. Macromolecules 18, 534–552. https://doi.org/10.1021/ma00145a039.

Muñoz, V., 2001. What can we learn about protein folding from Ising-like models? Curr. Opin. Struct. Biol. 11, 212–216. https://doi.org/10.1016/S0959-440X(00)00192-5.

Muñoz, V., 2002. Thermodynamics and kinetics of downhill protein folding investigated with a simple statistical mechanical model. Int. J. Quant. Chem. 90, 1522–1528. https://doi.org/10.1002/qua.10384.

Muñoz, V., Eaton, W.A., 1999. A simple model for calculating the kinetics of protein folding from three-dimensional structures. Proc. Natl. Acad. Sci. U.S.A. 96, 11311–11316. https://doi.org/10.1073/pnas.96.20.11311.

Munshi, S., Naganathan, A.N., 2015. Imprints of function on the folding landscape: functional role for an intermediate in a conserved eukaryotic binding protein. Phys. Chem. Chem. Phys. 17, 11042–11052. https://doi.org/10.1039/c4cp06102k.

Munshi, S., Rajendran, D., Naganathan, A.N., 2018. Entropic control of an excited folded-like conformation in a disordered protein ensemble. J. Mol. Biol. 430, 2688–2694. https://doi.org/10.1016/j.jmb.2018.06.008.

Munshi, S., Gopi, S., Asampille, G., Subramanian, S., Campos, L.A., Atreya, H.S., Naganathan, A.N., 2018. Tunable order-disorder continuum in protein-DNA interactions. Nucleic Acids Res 46, 8700–8709. https://doi.org/10.1093/nar/gky732.

Naganathan, A.N., 2012. Predictions from an ising-like statistical mechanical model on the dynamic and thermodynamic effects of protein surface electrostatics. J. Chem. Theory Comput. 8, 4646–4656. https://doi.org/10.1021/ct300676w.

Naganathan, A.N., 2013. Coarse-grained models of protein folding as detailed tools to connect with experiments. WIREs Comput. Mol. Sci. 3, 504–514. https://doi.org/10.1002/wcms.1133.

Naganathan, A.N., 2013. A rapid, ensemble and free energy based method for engineering protein stabilities. J. Phys. Chem. B 117, 4956–4964. https://doi.org/10.1021/jp401588x.

Naganathan, A.N., 2016. Predictive modeling of protein folding thermodynamics, mutational effects and free-energy landscapes. Proc. Indian Natl. Sci. Acad. 82, 1211–1228. https://doi.org/10.16943/ptinsa/2016/48570.

Naganathan, A.N., Muñoz, V., 2014. Thermodynamics of downhill folding: multi-probe analysis of PDD, a protein that folds over a marginal free energy barrier. J. Phys. Chem. B 118, 8982–8994. https://doi.org/10.1021/jp504261g.

Naganathan, A.N., Orozco, M., 2013. The conformational landscape of an intrinsically disordered DNA-binding domain of a transcription regulator. J. Phys. Chem. B 117, 13842–13850. https://doi.org/10.1021/jp408350v.

Naganathan, A.N., Sanchez-Ruiz, J.M., Munshi, S., Suresh, S., 2015. Are protein folding intermediates the evolutionary consequence of functional constraints? J. Phys. Chem. B 119, 1323–1333. https://doi.org/10.1021/jp510342m.

Narayan, A., Naganathan, A.N., 2014. Evidence for the sequential folding mechanism in RNase H from an ensemble-based model. J. Phys. Chem. B 118, 5050–5058. https://doi.org/10.1021/jp500934f.

Narayan, A., Naganathan, A.N., 2017. Tuning the continuum of structural states in the native ensemble of a regulatory protein. J. Phys. Chem. Lett. 8, 1683–1687. https://doi.org/10.1021/acs.jpclett.7b00475.

Narayan, A., Naganathan, A.N., 2018. Switching protein conformational substates by protonation and mutation. J. Phys. Chem. B 122, 11039–11047. https://doi.org/10.1021/acs.jpcb.8b05108.

Narayan, A., Campos, L.A., Bhatia, S., Fushman, D., Naganathan, A.N., 2017. Graded structural polymorphism in a bacterial thermosensor protein. J. Am. Chem. Soc. 139, 792–802. https://doi.org/10.1021/jacs.6b10608.

Orozco, M., Orellana, L., Hospital, A., Naganathan, A.N., Emperador, A., Carrillo, O., Gelpi, J.L., 2011. Coarse-grained representation of protein flexibility. Foundations, successes, and shortcomings. Adv. Protein Chem. Struct. Biol. 85, 183–215. https://doi.org/10.1016/B978-0-12-386485-7.00005-3.

Papoian, G.A., Wolynes, P.G., 2003. The physics and bioinformatics of binding and folding-an energy landscape perspective. Biopolymers 68, 333–349. https://doi.org/10.1002/bip.10286.

Rajasekaran, N., Gopi, S., Narayan, A., Naganathan, A.N., 2016. Quantifying protein disorder through measures of excess conformational entropy. J. Phys. Chem. B 120, 4341–4350. https://doi.org/10.1021/acs.jpcb.6b00658.

Rajasekaran, N., Suresh, S., Gopi, S., Raman, K., Naganathan, A.N., 2017. A general mechanism for the propagation of mutational effects in proteins. Biochemistry 56, 294–305. https://doi.org/10.1021/acs.biochem.6b00798.

Sasai, M., Chikenji, G., Terada, T.P., 2016. Cooperativity and modularity in protein folding. Biophys. Physicobiol. 13, 281–293. https://doi.org/10.2142/biophysico.13.0_281.

Sivanandan, S., Naganathan, A.N., 2013. A disorder-induced domino-like destabilization mechanism governs the folding and functional dynamics of the repeat protein IκBα. PLOS Comput. Biol. 9, e1003403 https://doi.org/10.1371/journal.pcbi.1003403.

Taketomi, H., Ueda, Y., Go, N., 1975. Studies on protein folding, unfolding and fluctuations by computer-simulation .1. Effect of specific amino-acid sequence represented by specific inter-unit interactions. Inter. J. Prot. Pep. Res. 7, 445–459. https://doi.org/10.1111/j.1399-3011.1975.tb02465.x.

Truong, H.H., Kim, B.L., Schafer, N.P., Wolynes, P.G., 2015. Predictive energy landscapes for folding membrane protein assemblies. J. Chem. Phys. 143, 243101. https://doi.org/10.1063/1.4929598.

Wako, H., Saito, N., 1978. Statistical mechanical theory of protein conformation .1. General considerations and application to homopolymers. J. Phys. Soc. Jpn 44, 1931–1938. https://doi.org/10.1143/JPSJ.44.1931.

Wako, H., Saito, N., 1978. Statistical mechanical theory of protein conformation .2. Folding pathway for protein. J. Phys. Soc. Jpn 44, 1939–1945. https://doi.org/10.1143/JPSJ.44.1939.

Zwanzig, R., 1995. Simple model of protein folding kinetics. Proc. Natl. Acad. Sci. U.S.A. 92, 9801–9804. https://doi.org/10.1073/pnas.92.21.9801.