The Volumetric Properties of the Transition State Ensemble for Protein Folding

Samvel Avagyan[a,c] and George I. Makhatadze[a,b,c*]

[a] *Department of Biological Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180, USA.*
[b] *Department on Chemistry and Chemical Biology, Rensselaer Polytechnic Institute, Troy, NY 12180, USA.*
[c] *Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY 12180,*

ABSTRACT: Hydrostatic pressure together with the temperature is an important environmental variable that plays an essential role in biological adaptation of extremophilic organisms. In particular, the effects of hydrostatic pressure on the rates of the protein folding/unfolding reaction are determined by the magnitude and sign of the activation volume changes. Here we provide computational description of the activation volume changes for folding/unfolding reaction, and compare them with the experimental data for six different globular proteins. We find that the volume of the transition state ensemble is always in-between the folded and unfolded states. Based on this, we conclude that hydrostatic pressure will invariably slow down protein folding and accelerate protein unfolding.

Life on Earth exists under a wide range of environmental conditions including high salinity, high and low pH, high and low temperatures and a range of hydrostatic pressures [1]. Importantly, the total biomass distribution is highly skewed towards environments with high hydrostatic pressure. According to recent estimates, over 90% of biomass on Earth is associated with the high pressure environments [2-3]. Thus, understanding the effects of pressure on structure, function and dynamics of biomacromolecules is of a particular interest [1]. However, since the realization that vast majority of life on Earth exists under high pressure conditions it has become evident that there is a significant lag in experimental and computational studies of the effects of pressure on the biophysics of biomacromolecules. In particular studies of the effects of pressure on energy landscape of proteins have been limited.

The effects of perturbations such as increase in temperature or high denaturant concentrations on the rates of protein folding/unfolding reaction are analyzed within the framework of the transition state theory. In the case when perturbation is high hydrostatic pressure, the pressure derivative of the rate constant, k, reflects the difference between the volume of the ground state (folded, V_F, or unfolded, V_U, state ensembles) and the volume of the transition state ensemble, V_{TSE}:

$$-RT\left(\frac{\partial k_F}{\partial P}\right)_T = \Delta V_F^{\#} = V_{TSE} - V_U$$

and

$$-RT\left(\frac{\partial k_U}{\partial P}\right)_T = \Delta V_U^{\#} = V_{TSE} - V_F$$

Knowledge of the value of V_{TSE} relative to the values V_F and V_U provides additional information on the structural ensemble of the TS. The activation volume of folding, $\Delta V_F^{\#}$, and unfolding, $\Delta V_U^{\#}$, is equally important to understand how the rates of protein folding and unfolding will be affected by high hydrostatic pressures. If for example, the volume of transition state ensembles is greater than the folded state volume, the rate of protein unfolding will decrease at higher pressures, in effect imparting kinetic pressure stability onto the protein [4]. However, if the volume of the transition state is less than the native state volume, the rate of unfolding will increase at high pressure.

Experimental data on the activation volumes of folding/unfolding of six proteins have been reported to date. Tendamistat (Protein Data Bank structure PDB:1OK0) is a small globular protein of 74 amino acid residues. Equilibrium and kinetic studies of this protein have shown that its folding/unfolding reaction is closely approximated by a two-state transition. The equilibrium unfolding studies of tendamistat performed at 35°C as a function of pressures up to 100 MPa showed that the protein is destabilized by increase in hydrostatic pressure [5]. This decrease in stability was well described by a negative equilibrium volume change of unfolding, ΔV_{Exp}=-41.6±2.7 cm^3/mol. Analysis of kinetics of GdmCl-induced folding/unfolding reactions at different pressures was done using Chevron plots. It was found that the activation volume of folding is $\Delta V_F^{\#}$ = 25.0±1.2 cm^3/mol while activation volume of unfolding is $\Delta V_U^{\#}$= -16.4±1.4 cm^3/mol [5]. There was an excellent agreement for the overall volume of unfolding as determined from equilibrium (-41.6±2.7 cm^3/mol) and kinetic (-41.4±2.0 cm^3/mol) analysis.

Thermodynamic stability and kinetics of folding of ubiquitin has been extensively characterized and shown to closely resemble a two-state folding mechanism. Ubiquitin is a small globular protein of 76 amino acid residues (PDB:1UBQ). Heberhold & Winter [6], used FTIR spectroscopy to characterize the effects of hydrostatic pressure on the stability of this protein. Experimental measurements were done on broad range of temperatures (from -10°C to 100°C) and pressures (up to 900 MPa). The equilibrium volume change obtained from pressure-induced unfolding was found to be negative at ΔV_{Exp}= -50±20 cm^3/mol. The pressure jump experiments performed at 21°C were used to obtain the activation volumes of unfolding, reported at $\Delta V_U^{\#}$= -38 cm^3/mol. Considering that both equilibrium and kinetic unfolding are two-state, the activation volume for folding of 12 cm^3/mol was calculated as $\Delta V_F^{\#}$=$\Delta V_U^{\#}$-ΔV_{Exp}.

The small oncogenic product P13 protein, consist of 117 amino acid residues (PDB:1QTU), and shows unfolding transition that can be closely approximated by a two-state model [7]. Changes in the intrinsic fluorescence intensities as a function of pressure at 21°C were analyzed to obtain the total volume change of unfolding ΔV_{Exp}= -105±15 cm^3/mol. The pressure jump unfolding experiments were closely approximated by a single-exponential fit which allowed to compute the activation volume of unfolding $\Delta V_U^{\#}$=79±35 cm^3/mol [7]. Considering that both equilibrium and kinetic unfolding are two-state, the activation volume for folding of -26 cm^3/mol was calculated as $\Delta V_F^{\#}$=$\Delta V_U^{\#}$-ΔV_{Exp}.

Azurin from *Pseudomonas aeruginosa*, a single chain polypeptide of 128 amino acid residues (PDB:5AZU), was studied by Cioni et al [8]. The kinetics of folding was monitored by changes in fluorescence intensity during pressure jumps at 50°C. Somewhat different values for $\Delta V^{\#}_U$ and $\Delta V^{\#}_F$ were obtained from the experiments performed in the upward p-jump ($\Delta V_U^{\#}$= -17.1±1.2 cm^3/mol and $\Delta V_F^{\#}$= 39.5±1.2 cm^3/mol) and the downward p-jump ($\Delta V_U^{\#}$= -11.7±2.9 cm^3/mol and $\Delta V_F^{\#}$= 48.6±2.4 cm^3/mol). However, overall these values are consistent with the results of independent experiments to obtain the equilibrium volume changes of unfolding ΔV_{Exp} (50°C)=-54.5±0.5 cm^3/mol [8].

The 23-kDa protein from the spinach photosystem II (PII23kDa) is a monomer of 175 amino acid residues (PDB:4RTI), and pressure induced fluorescence measurements suggest that both pressure-induced equilibrium unfolding and kinetics of folding/unfolding reactions are well approximated by a two-state model [9]. The equilibrium volume changes of unfolding is reported to be ΔV_{Exp}(20°C)=-157.6 cm^3/mol. The corresponding activation volume of unfolding $\Delta V_U^{\#}$= -66.2 cm^3/mol and folding $\Delta V_F^{\#}$= 84.1 cm^3/mol are consistent with the equilibrium measurements [9].

Trp-repressor is a dimer of 105 amino acid residues per monomer (PDB:3WRP). The effects of high hydrostatic pressure on the folding/unfolding reaction of this protein have been monitored by fluorescence spectroscopy and infra-red absorption techniques [10]. It was found that unfolding of Trp-repressor follows a biomolecular two-state unfolding, whereby the dimer dissociation leads to unfolding of monomers. The equilibrium volume of unfolding is reported to be ΔV_{Exp}(21°C)=-162 cm^3/mol per monomer. The activation volumes, measured with pressure jump experiments are $\Delta V_U^{\#}$= -65±6 cm^3/mol and folding $\Delta V_F^{\#}$= 114±8 cm^3/mol [10].

These experiments provide a comprehensive dataset to benchmark our computational work. The goal of this work is to use computer simulations to characterize the volumetric properties of the transition state ensemble for protein folding. It relies on two computational methods.

The first is a coarse-grained simulation of protein folding/unfolding reactions. Energy landscape theory of proteins, and in particular the principle of minimal frustration, allows the development of an effective computational approach to map energy landscapes of individual proteins [11-14]. To this end structure-based models (SBM) of protein folding have been widely explored to rationalize the experimental ϕ-value analysis of protein transition states, effects of charged residues on the folding energy landscape, and dynamics within folded state ensemble [15-24].

The second is the recently developed semi-empirical computational framework to calculate volumetric properties of proteins in solution, the so-called ProteinVolume (PV) approach. This method has been benchmarked against experimental data and shown to reproduce well the total volume changes upon protein unfolding [25-26]. It uses structural information obtained from all-atom explicit solvent molecular dynamics simulations starting with x-ray coordinates in order to compute the volume changes upon protein unfolding [26-28].

Here we combine the SBM and PV to map volumetric properties of transition states upon protein unfolding. The results of these calculations are compared to the experimental data available for the six aforementioned proteins that unfold according to a two-state model. To further validate the PV algorithm for the six proteins used in SBM, we compared the results of the calculations, ΔV_{Tot}, with the experimentally measured equilibrium volume changes upon unfolding of these proteins, ΔV_{Exp} (Figure 1*). It is evident, that there is a very good correspondence between ΔV_{Tot} and ΔV_{Exp}, thus providing rationale for applying the PV algorithm to the analysis of volumetric properties of structural ensembles from SBM.

Molecular dynamic simulations using all-atom structure-based model AA-SBM were performed in Gromacs [29] (details are given in the supplementary data, ESI†). To accelerate equilibration, Replica Exchanged Molecular Dynamics (REMD) at 20 different temperatures was employed. Temperatures were spaced by 0.5 K and trajectories were combined in the Weighted Histogram Analysis Method (WHAM) to calculate relevant thermodynamic parameters including the free energy and the constant volume heat capacity profiles from the simulations [30]. The fraction of native contacts, Q, was used as a reaction coordinate.

Figure 2 shows the results of analysis of SBM simulations of six different proteins in terms of free energy profiles as a function of Q, computed at the corresponding transition temperatures. In all cases, the transitions closely resemble a two-state with a single maximum corresponding to the TSE. For larger proteins the transition state appears to be more diffused (i.e. spanning wider range of Q-values) than for smaller proteins. Also notable is that the position of the TSE is different for different proteins, in agreement with previous observations for other proteins [15-16, 22]. The heat map of native contacts formed in the TSE is also shown in Figure 2. Again, depending on the protein, there is a unique set of contacts that remains populated in the TSE.

Most importantly, the free energy profiles as a function of Q, allows us to perform volumetric analysis of all states, i.e. unfolded, folded and TS. To this end, a set of 200 structures corresponding to each of these states was extracted from the trajectories and volumes of each structure was calculated using the PV algorithm. The ensemble-averaged volumes for each protein are compared on the top right plot of each panel in Figure 2. The same panel shows the experimental data, plotted with the unfolded state set as a reference. The relative (to the folded and unfolded state volumes) positions of the volume of TSE in experiments and in calculations (based on SBM) are in a good agreement. To further facilitate the comparison, we introduce a parameter β_V defined as the ratio of the activation volume to the total volume of unfolding [31]:

$$\beta_{V,F} = 1 - \beta_{V,U} = \frac{\left(\partial \Delta G_f^{\#} / \partial \mathrm{P}\right)_T}{\left(\partial \Delta G_{eq} / \partial \mathrm{P}\right)_T} = \frac{\Delta V_F^{\#}}{\Delta V_{Eq}}$$

The $\beta_{V,F}$ parameter is similar to β_T, Tanford beta-parameter used in the analysis of the position of the transition state relative to the native and unfolded states in denaturant-induced kinetic experiments [32], and expressed as the ratio of the activation to equilibrium Gibbs energy, ΔG:

$$\beta_T = \frac{\partial \Delta G_f^{\#} / \partial [den.]}{\partial \Delta G_{eq} / \partial [den.]} = \frac{m_F^{\#}}{m_{Eq}}$$

The $\beta_{V,F}$ values larger than 1 will indicate that the volume of the transition state is larger than the volume of the folded state. In this case increase in hydrostatic pressure will slow down the rates of unfolding, thus making a protein kinetically more stable at higher pressures. The $\beta_{V,F}$ values less than 1 will indicate that the volume of the transition state is in-between the volumes of the folded and unfolded states. Furthermore, the values of $\beta_{V,F}$ that are smaller than 0.5 will suggest that the volume of the transition state is closer to the unfolded state, while values larger than 0.5 will indicate that the TSE is volumetrically closer to the native state.

Comparison of experimental and computed $\beta_{V,F}$ parameters is shown in Figure 3. In all cases, $\beta_{V,F}$ is less than 1, suggesting the volume of TSE for all six studied proteins is larger than the volume of unfolded state but smaller than the volume of the native state. It is also evident that for two proteins, ubiquitin (1UBQ) and photosystem II 23kDa protein (4PTI) the transition state is closer to the unfolded state. The remaining four proteins, judging by their $\beta_{V,F}$ values, have their TSE closer to the native state.

CONCLUSIONS

It is rather remarkable, that the computed and experimentally derived β_V values are rather similar. Based on this, one can argue, that the method presented here can be valuable for gaining additional insight into transition state ensembles, through the lenses of the volumetric properties of TSE. However, it also implies that because the volume of TSE is always in-between the volumes of folded and unfolded states, hydrostatic pressure will always impair protein kinetics stability by increasing the rates of unfolding. Thus proteins from piezophilic organisms that live under high hydrostatic pressure will need to employ adaptation mechanisms that counteract it. These mechanisms remain to be discovered.

AUTHOR INFORMATION

Corresponding Author

* George Makhatadze, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY 12180, USA makhag@rpi.edu

Author Contributions

G.I.M. initiated and designed the project, S.A. and G.I.M. performed computation, analyzed the data and wrote the manuscript.

Funding Sources

This work was supported by a grant CHEM/CLP-1803045 (to G.I.M.) from the US National Science Foundation (NSF).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

This work used the Extreme Science and Engineering Discovery Environment (XSEDE) comet (SDSC) and stampede2 (TACC) using allocation TG-MCB140107, which is supported by the US National Science Foundation grant number ACI-1548562.

REFERENCES

1. Ando, N.; Barquera, B.; Bartlett, D. H.; Boyd, E.; Burnim, A. A.; Byer, A. S.; Colman, D.; Gillilan, R. E.; Gruebele, M.; Makhatadze, G. The Molecular Basis for Life in Extreme Environments. *Annual Review of Biophysics* **2021,** *50*, 343-372.
2. Bar-On, Y. M.; Phillips, R.; Milo, R. The biomass distribution on Earth. *Proceedings of the National Academy of Sciences* **2018,** *115*, 6506-6511.
3. McMahon, S.; Parnell, J. Weighing the deep continental biosphere. *FEMS Microbiology Ecology* **2014,** *87*, 113-120.
4. Royer, C. A. Application of pressure to biochemical equilibria: the other thermodynamic variable. *Methods Enzymol* **1995,** *259*, 357-377.
5. Pappenberger, G.; Saudan, C.; Becker, M.; Merbach, A. E.; Kiefhaber, T. Denaturant-induced movement of the transition state of protein folding revealed by high-pressure stopped-flow measurements. *Proceedings of the National Academy of Sciences* **2000,** *97*, 17-22.
6. Herberhold, H.; Winter, R. Temperature- and Pressure-Induced Unfolding and Refolding of Ubiquitin: A Static and Kinetic Fourier Transform Infrared Spectroscopy Study. *Biochemistry* **2002,** *41*, 2396-2401.
7. Kitahara, R.; Royer, C.; Yamada, H.; Boyer, M.; Saldana, J.-L.; Akasaka, K.; Roumestand, C. Equilibrium and Pressure-jump Relaxation Studies of the Conformational Transitions of P13MTCP1. *Journal of Molecular Biology* **2002,** *320*, 609-628.
8. Cioni, P.; Gabellieri, E.; Marchal, S.; Lange, R. Temperature and pressure effects on C112S azurin: volume, expansivity, and flexibility changes. *Proteins* **2014,** *82*, 1787-1798.
9. Tan, C. Y.; Xu, C. H.; Wong, J.; Shen, J. R.; Sakuma, S.; Yamamoto, Y.; Lange, R.; Balny, C.; Ruan, K. C. Pressure equilibrium and jump study on unfolding of 23-kDa protein from spinach photosystem II. *Biophys J* **2005,** *88*, 1264-1275.
10. Desai, G.; Panick, G.; Zein, M.; Winter, R.; Royer, C. A. Pressure-jump studies of the folding/unfolding of trp repressor. *J Mol Biol* **1999,** *288*, 461-475.
11. Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. Theory of protein folding: the energy landscape perspective. *Annual review of physical chemistry* **1997,** *48*, 545-600.
12. Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Structure, Function, and Bioinformatics* **1995,** *21*, 167-195.
13. Veitshans, T.; Klimov, D.; Thirumalai, D. Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Folding and Design* **1997,** *2*, 1-22.
14. Dill, K. A.; Chan, H. S. From Levinthal to pathways to funnels. *Nature structural biology* **1997,** *4*, 10-19.
15. Clementi, C.; Garcıa, A. E.; Onuchic, J. N. Interplay among tertiary contacts, secondary structure formation and side-chain packing in the protein folding mechanism: all-atom representation study of protein L. *Journal of molecular biology* **2003,** *326*, 933-954.
16. Clementi, C.; Nymeyer, H.; Onuchic, J. N. Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins. *Journal of molecular biology* **2000,** *298*, 937-953.
17. Nymeyer, H.; García, A. E.; Onuchic, J. N. Folding funnels and frustration in off-lattice minimalist protein landscapes. *Proceedings of the National Academy of Sciences* **1998,** *95*, 5921-5928.

18. Shimada, J.; Kussell, E. L.; Shakhnovich, E. I. The folding thermodynamics and kinetics of crambin using an all-atom Monte Carlo simulation. *Journal of molecular biology* **2001,** *308*, 79-95.

19. Shimada, J.; Shakhnovich, E. I. The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation. *Proceedings of the National Academy of Sciences* **2002,** *99*, 11175-11180.

20. Li, L.; Shakhnovich, E. I. Constructing, verifying, and dissecting the folding transition state of chymotrypsin inhibitor 2 with all-atom simulations. *Proceedings of the National Academy of Sciences* **2001,** *98*, 13014-13018.

21. García, A. E.; Onuchic, J. N. Folding a protein in a computer: an atomic description of the folding/unfolding of protein A. *Proceedings of the National Academy of Sciences* **2003,** *100*, 13898-13903.

22. Tzul, F. O.; Schweiker, K. L.; Makhatadze, G. I. Modulation of folding energy landscape by charge–charge interactions: Linking experiments with computational modeling. *Proceedings of the National Academy of Sciences* **2015,** *112*, E259-E266.

23. Tripathi, S.; Garcia, A. E.; Makhatadze, G. I. Alterations of Nonconserved Residues Affect Protein Stability and Folding Dynamics through Charge-Charge Interactions. *J Phys Chem B* **2015,** *119*, 13103-13112.

24. Tripathi, S.; Makhatadze, G. I.; Garcia, A. E. Backtracking due to residual structure in the unfolded state changes the folding of the third fibronectin type III domain from tenascin-C. *J Phys Chem B* **2013,** *117*, 800-810.

25. Chen, C. R.; Makhatadze, G. I. ProteinVolume: calculating molecular van der Waals and void volumes in proteins. *BMC Bioinformatics* **2015,** *16*, 101-107.

26. Chen, C. R.; Makhatadze, G. I. Molecular determinant of the effects of hydrostatic pressure on protein folding stability. *Nature Communications* **2017,** *8*, 14561.

27. Avagyan, S.; Vasilchuk, D.; Makhatadze, G. I. Protein adaptation to high hydrostatic pressure: Computational analysis of the structural proteome. *Proteins: Structure, Function, and Bioinformatics* **2019,** *88*, 584-592.

28. Chen, C. R.; Makhatadze, G. I. Molecular Determinants of Temperature Dependence of Protein Volume Change upon Unfolding. *Journal of Physical Chemistry. B* **2017,** *121*, 8300-8310.

29. Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation* **2008,** *4*, 435-447.

30. Ferrenberg, A. M.; Swendsen, R. H. Optimized monte carlo data analysis. *Computers in Physics* **1989,** *3*, 101-104.

31. Mitra, L.; Hata, K.; Kono, R.; Maeno, A.; Isom, D.; Rouget, J.-B.; Winter, R.; Akasaka, K.; García-Moreno, B.; Royer, C. A. Vi -Value Analysis: A Pressure-Based Method for Mapping the Folding Transition State Ensemble of Proteins. *Journal of the American Chemical Society* **2007,** *129*, 14108-14109.

32. Maxwell, K. L.; Wildes, D.; Zarrine-Afsar, A.; De Los Rios, M. A.; Brown, A. G.; Friel, C. T.; Hedberg, L.; Horng, J. C.; Bona, D.; Miller, E. J. Protein folding: defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Science* **2005,** *14*, 602-616.

Figures and Legends

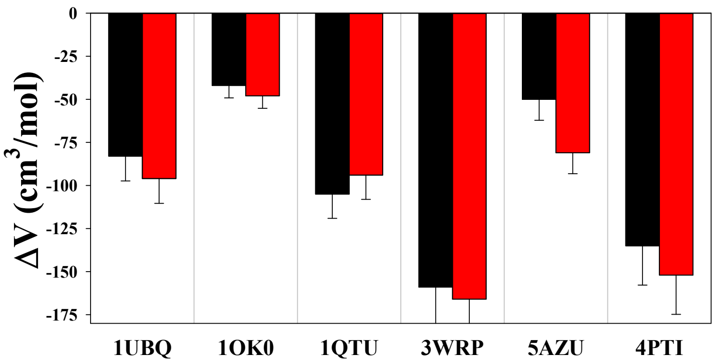

Figure 1. Comparison of the results of calculations, ΔV_{Tot} (red), with the experimentally measured total volume changes upon unfolding of six proteins studied here, ΔV_{Exp} (black).

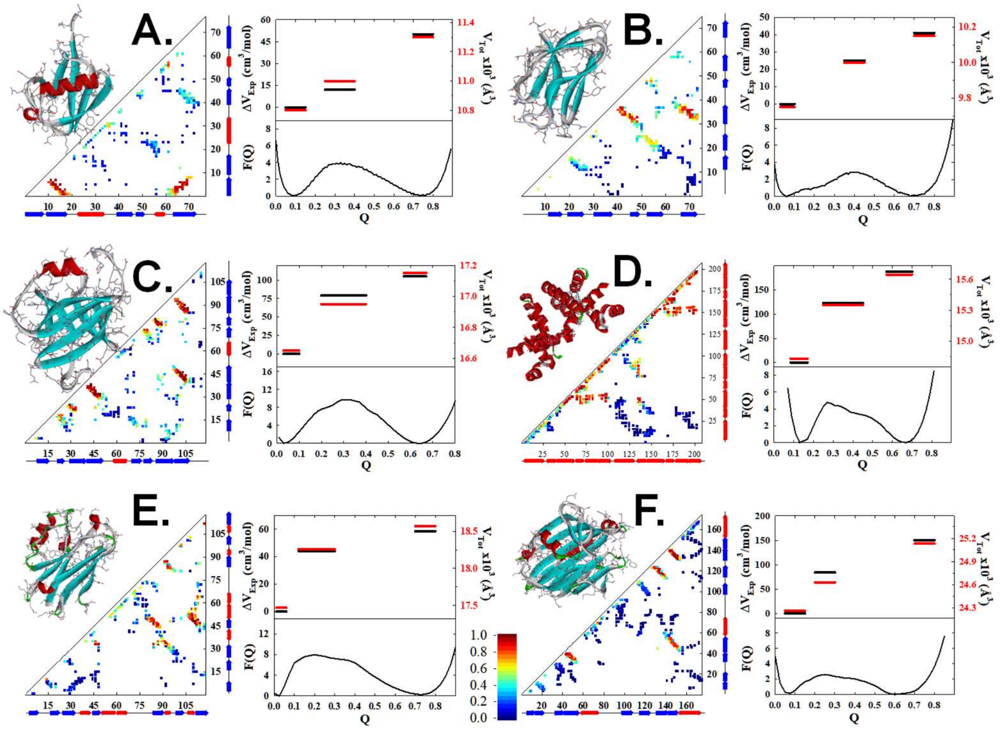

Figure 2. Comparison of the results of calculations from SBM and experiments on the activation volume of folding/unfolding reaction for six proteins studied here. **A.** Ubiquitin (1UBQ); **B.** Tendamistat (1OK0); **C.** P13-oncogene (1QTU); **D.** Trp-repressor (3WRP); **E.** Azurin (5AZU); **F.** PhotosystemII 23 kDa protein (4RTI). Each of the six panels shows (clock counter-wise starting in the upper left corner): the cartoon of the corresponding protein structure, the contact plot based on the x-ray structure, color-coded by fraction of contacts formed for the TSE, weighted probability of the potential energy as a function of Q (fraction of native contacts), and comparison of relative volumes of unfolded, TS and folded ensembles from the experiments (black) with computed, for each ensemble, values (red).

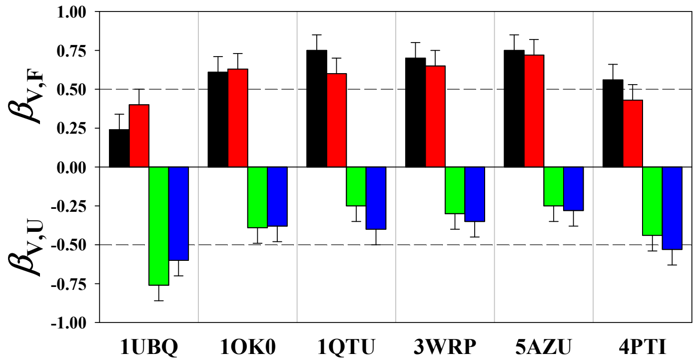

Figure 3. Comparison of the β_V values from experiment ($\beta_{V,F}$ - black, $\beta_{V,U}$ - green) with calculations ($\beta_{V,F}$ - red; $\beta_{V,U}$ - blue).

Supplementary Methods

All-atom structure based potentials were generated using SMOG (version 2.0.3) web server http://smog-server.org [1] with default parameter sets [2]. The following PDB entries were used: 1UBQ - ubiquitin; 1OK0 - tendamistat; 1QTU - P13-oncogene; 3WRP - Trp-repressor; 5AZU - azurin; 4RTI - PhotosystemII 23 kDa protein. The contacts were identified from PDB coordinates through use of the Shadow Contact Map algorithm [3] with a cutoff distance of 6 Å, shadowing radius of 1 Å and residue sequence separations of 3. Atom pairs that are not identified as contacts are assigned an excluded volume interaction. The bond lengths and angles, improper and planar dihedral angles of the protein are maintained by harmonic potentials. The potentials are assigned such that the native configuration of each bond and angle is considered the minimum. The final form of the potential energy function for AA-SBM model is:

$$
\begin{aligned}
V = {} & \sum_{bonds} \varepsilon_r \,(r - r_o)^2 + \sum_{angles} \varepsilon_\Theta \,(\Theta - \Theta_o)^2 + \sum_{\substack{improper\\ /planar}} \varepsilon_\chi \,(\chi - \chi_o)^2 + \sum_{backbone} \varepsilon_{BB}\, F_D(\phi) \\
& + \sum_{side-chai} \varepsilon_{SC}\, F_D(\phi) + \sum_{contact} \left\{ \varepsilon_C(i,j) \left[a\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - b\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] \right\} \\
& + \sum_{non-contact} \varepsilon_{NC}(i,j) \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}
\end{aligned}
$$

$$F_D(\phi) = [1 - \cos(\phi - \phi_o)] + \big[1 - \cos\big(3(\phi - \phi_o)\big)\big]/2$$

With all parameters having the default values as reported in [2].

Gromacs 4.6.7 was used as the computation engine to run the simulations [4]. To enhance sampling efficiency and accelerate equilibration, the replica exchange molecular dynamics (REMD) method [5] as implemented in Gromacs [4] was used. We used 20-24 replicas spaced by 0.5 K that were centered around the transition temperature for a given protein. Exchange was attempted every 5000 time steps, and coordinates were saved every 1000 integration steps. REMD was combined with Langevin dynamics (time step $\tau = 0.0005$ ps) for $5 \cdot 10^8$ time steps per replica. The fraction of number of native contacts (defined as any native pair within 1.5 times the native distance) formed as a function of time, Q, was used as a global reaction coordinate. Potential energy as a function of Q from all replicas was analyzed by Weighted Histogram Analysis Method (WHAM) to calculate the free energy profiles and the constant volume heat capacity [6].

A set of 200 random structures for each state (unfolded, folded and TS) identified from the analysis of the free energy profiles were extracted and energy minimized to adjust bond length and add hydrogens. Energy minimization was performed with Gromacs 4.6.7 for 1,000 steps using the Steepest Descent minimization algorithm with GBSA implicit solvent model and dielectric of 80. The volume for each structure, V_{SE}, was calculated using PV algorithm with starting volume probe radius of 0.08 Å, surface probe minimum distance of 0.1 Å. The volume of hydration was calculated from the polar and non-polar molecular surface areas as [7]:

$$V_{\mathrm{Hyd}} = (k_{\mathrm{NP}} \cdot \mathrm{MSA}_{\mathrm{NP}}) + (k_{\mathrm{P}} \cdot \mathrm{MSA}_{\mathrm{P}})$$

with k_{NP}=0.38 Å and k_P=0.03 Å. The final volume V_{Tot} is the sum V_{SE} and V_{Hyd} [7].

SI REFERENCES

1. J. K. Noel, M. Levi, M. Raghunathan, H. Lammert, R. L. Hayes, J. N. Onuchic and P. C. Whitford, SMOG 2: A Versatile Software Package for Generating Structure-Based Models, *PLoS Comput Biol*, 2016, **12**, e1004794.
2. P. C. Whitford, J. K. Noel, S. Gosavi, A. Schug, K. Y. Sanbonmatsu and J. N. Onuchic, An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields, *Proteins*, 2009, **75**, 430-441.
3. J. K. Noel, P. C. Whitford and J. N. Onuchic, The Shadow Map: A General Contact Definition for Capturing the Dynamics of Biomolecular Folding and Function, *Journal of Physical Chemistry. B*, 2012, **116**, 8692-8702.
4. B. Hess, C. Kutzner, D. Van Der Spoel and E. Lindahl, GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation, *Journal of chemical theory and computation*, 2008, **4**, 435-447.
5. Y. Sugita and Y. Okamoto, Replica-exchange molecular dynamics method for protein folding, *Chemical physics letters*, 1999, **314**, 141-151.
6. A. M. Ferrenberg and R. H. Swendsen, Optimized monte carlo data analysis, *Computers in Physics*, 1989, **3**, 101-104.
7. C. R. Chen and G. I. Makhatadze, Molecular determinant of the effects of hydrostatic pressure on protein folding stability, *Nature Communications*, 2017, **8**, 14561.