Biophysical Journal
Article

Biophysical Society

Quantifying Nonnative Interactions in the Protein-Folding Free-Energy Landscape

Paulo Ricardo Mouro,[1] Vinícius de Godoi Contessoto,[1] Jorge Chahine,[1] Ronaldo Junio de Oliveira,[2] and Vitor Barbanti Pereira Leite[1,*]

[1]Departamento de Física, Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, São Paulo, Brazil; and [2]Laboratório de Biofísica Teórica, Departamento de Física, Instituto de Ciências Exatas, Naturais e Educação, Universidade Federal do Triângulo Mineiro, Uberaba, Minas Gerais, Brazil

ABSTRACT Protein folding is a central problem in biological physics. Energetic roughness is an important aspect that controls protein-folding stability and kinetics. The roughness is associated with conflicting interactions in the protein and is also known as frustration. Recent studies indicate that an addition of a small amount of energetic frustration may enhance folding speed for certain proteins. In this study, we have investigated the conditions under which frustration increases the folding rate. We used a C_α structure-based model to simulate a group of proteins. We found that the free-energy barrier at the transition state (ΔF) correlates with nonnative-contact variation (ΔA), and the simulated proteins are clustered according to their fold motifs. These findings are corroborated by the Clementi-Plotkin analytical model. As a consequence, the optimum frustration regime for protein folding can be predicted analytically.

INTRODUCTION

Understanding the underlying folding mechanism of a protein to its functional compact three-dimensional structure is one of the great challenges of modern science. Failure in the process of a protein to achieve the correct folded native state can cause a series of pathological conditions, like neurodegenerative disorders, including Alzheimer's and Parkinson's diseases (1). Over the last decades, the energy-landscape theory has been a consistent framework in revealing protein-folding mechanisms (1–5). This theory states that the energy landscape of globular proteins resembles a funnel of structures progressively folded en route to the native state with its bottleneck narrowed at the transition state between unfolded and folded ensembles (6–9). The energy-landscape theory is successful in explaining, qualitatively and quantitatively, folding studies in theoretical (10–13) as well as experimental (14–17) investigations. Based on energy-surface theory, many computational models have been developed to predict folding mechanisms, rates, and stability parameters connected with experiments (9,18–22).

The protein-folding energy surface is multidimensional, and it has a funnelled topography as a function of the reaction coordinate, usually described by the fraction of native contacts, Q. The funnel theory describes protein folding as an ensemble of conformations gradually diffusing from the unfolded state (lower Q), high entropic and energetic conformations at the top of the surface funnel, to the native state (higher Q), the lowest entropic and energetic state at its bottom (23–28). Proteins are naturally designed through evolution, so that folding pathways toward the native state are not dominated by bumps due to local energetic traps (5,29–33). Nature-designed sequences have the ability to fold completely on biological timescales.

To fold, the funnel energy slope must be steep enough to overcome the roughness and minimize local energetic trapping (13), or energetic frustration (21). Energetic frustration occurs due to the impossibility of satisfying all favorable energetic interactions simultaneously during folding events (34–37), allowing the formation of nonnative contacts (pairs of residues not in contact at the native state). Kinetically foldable proteins are naturally selected throughout the evolution process so that the native state is minimally frustrated (34–36,38,39).

Clementi and Plotkin introduced a theoretical/analytical model that takes into account the effects of nonnative interactions on the folding rate and on the folding free-energy barrier (40). This model indicates that a nonzero amount of nonnative energy interaction may enhance protein folding

Submitted December 17, 2015, and accepted for publication May 17, 2016.

*Correspondence: vleite@sjrp.unesp.br

Editor: Daniel Raleigh.

http://dx.doi.org/10.1016/j.bpj.2016.05.041

rates. The theory for the folding speed limit upon the nonnative energetic frustration is well discussed in the literature (41,42), and studies have been done on specific proteins regarding nonnative interactions (26,43–45). In our previous article (46), using C_α structure-based models (SBMs) and homogeneous frustration, the maximum value of energetic frustration that can be added to the system to speed up folding rates was quantified. This maximum value was named the optimum energetic frustration (ϵ_f^{opt}) of each protein, and it was determined by analyzing the effect of the energetic frustration value (ϵ_{NN}) on free-energy barrier height and folding time. It was shown that not all proteins could have folding rates enhanced by including a small amount of frustration, i.e., ϵ_f^{opt} is zero. However, for all of the other proteins, ϵ_f^{opt} correlates with the free-energy barrier height (21) and the contact order parameter (47), clustering the proteins accordingly to their motif (48).

In this article, we seek the connection between the Clementi-Plotkin analytical model (40) and our previous computational results (46) through analysis of the underlying mechanism in nonnative-contact formation during folding. We show that the perturbed free-energy barrier variation ($\Delta\Delta F$) depends on the nonnative energetic contact parameter (ϵ_{NN}), and on the nonnative-contact variation (ΔA), in the folding transition state (Q^{ts}). These three quantities were calculated for a group of proteins using SBM simulations. The difference between the perturbed and nonperturbed SBM model, in both analytical and computational results, explains quantitatively the relevance of the nonnative-contact formation in the pretransition state for the maximal folding-speed-limit mechanisms. The former results of grouping the simulated proteins by their folding motif were also recovered.

MATERIALS AND METHODS

In this study, a group of 15 proteins (46) was used to investigate the effect of nonnative interactions in the transition state. Proteins were simulated using molecular dynamics and a coarse-grained C_α SBM (12,49). Despite the simplicity of this model, it can capture important characteristics of protein-folding mechanisms, as has been shown by computer simulations (9,49–51). The agreement between experimental and computational folding rates using the SBM is remarkable, as has been shown by Chavez et al. (21).The simulation details are explained in the Supporting Material.

Effect of nonnative interactions on the free-energy barrier

Two order parameters are used by Clementi and Plotkin to map the folding process in their analytical model: the fraction of native contacts (Q) and the fraction of nonnative contacts (A) (40). These two order parameters are normalized between 0 and 1. The fraction of nonnative contacts depends on Q: the more native contacts are formed, the fewer nonnative interactions are allowed. The theoretical model does not allow nonnative contacts for $Q = 1$. The native attraction is characterized by the mean attraction energy, ε ($\varepsilon < 0$), and two energy scales are used to analyze the nonnative contribution: the mean energy of a nonnative interaction (ϵ_{NN}) and the energetic variance of nonnative interaction (b^2). Nonnative interactions are designed to be weaker when compared to native interactions in a perturbative way. Expressions for thermal energy (E), free energy (F), and entropy (S) as a function of Q, A, and temperature (T) were normalized by the maximum number of contacts, M (Eqs. 7(a)–7(c) in (40)). The folding temperature of the model (T_F^0) was defined as the temperature at which folded and unfolded states have the same probability. Thus, at T_F^0, $F(0,A) \approx F(1,0)$ for $\epsilon_{NN} = b^2 = 0$. If A^* (Q) is the most probable value of A in a certain Q, the difference between the most probable value in the transition state, $A^*(Q^{ts})$, and in the unfolded state, A^* (0), is given by

$$\Delta A^*(Q^{\neq}) \equiv A^*(Q^{ts}) - A^*(0). \quad (1)$$

The variation in free energy between the unfolded and the transition states at T_F^0 is calculated. It is shown that the free-energy barrier height (ΔF) corresponds to the sum of the free-energy barrier between these states in the absence of nonnative forces (ΔF^0) plus a term referring to energetic frustration:

$$\frac{\Delta F}{T_F^0} = \frac{\Delta F^0}{T_F^0} + M\left(\frac{\epsilon_{NN}}{T_F^0} - \frac{b^2}{2T_F^{0\,2}}\right)\Delta A^*(Q^{\neq}). \quad (2)$$

Since ϵ_{NN} always has negative values, Eq. 2 can be analyzed in two conditions. If $\Delta A^*(Q^{\neq}) > 0$, more nonnative contacts are formed in the transition state than in the unfolded state, and the third term of Eq. 2 is negative, leading to a decrease in the free-energy barrier when compared with the barrier without nonnative interactions ($\Delta F < \Delta F^0$). This condition implies an increase in the folding rate. On the other hand, for $\Delta A^*(Q^{\neq}) < 0$, the opposite happens: nonnative interactions lead to an increase in the barrier, decreasing protein folding rates.

Nonnative-contact calculation

Nonnative contacts for the C_α SBM simulations were defined according to the following criteria:

1) Nonnative contacts cannot be listed in the first native-contact map generated from the protein native configuration (see the Supporting Material);
2) Nonnative contacts consist of any two nonbonded amino acid residues separated by at least four residues in the main chain and at a distance of up to 6 Å from each other;
3) Nonnative contacts cannot have a >30% probability of being found on folded structures ($Q > 0.9$), and if that is the case, these new contacts are included in the final native-contact list.

The difference in the number of nonnative contacts (ΔA) between the transition and the unfolded states is calculated by

$$\Delta A = \overline{A}(\mathrm{TS}) - \overline{A}(\mathrm{Unf}), \quad (3)$$

with $\overline{A}$ being the average of nonnative contacts formed in the transition state (*TS*) and in the unfolded state (*Unf*). The TS and the unfolded regions are defined using the free-energy profile as a function of the native contacts ($F(Q)$) for the respective protein (Fig. 1). The unfolded region is defined as the region that involves all states from $Q = 0$ to the configuration in which the protein reaches 15% of barrier height (ΔF) after the unfolded minimum. On the other hand, the transition region is defined as the region that involves all states located inside the free-energy barrier and having a free energy $F(Q) \geq 85\% \Delta F$. ΔF is calculated by the difference between the peak of the free-energy barrier ($F^{\neq} = F(Q^{ts})$) and its unfolded minimum ($F_{unf} = F(Q^u)$), i.e., $\Delta F = F^{\neq} - F_{unf}$, as shown in Fig. 1. The values of 15% and 85%, described above, were chosen because they delimit the unfolded and folded states and allow the presence of a pretransition region for all proteins studied in this work, including those with low free-energy barriers.

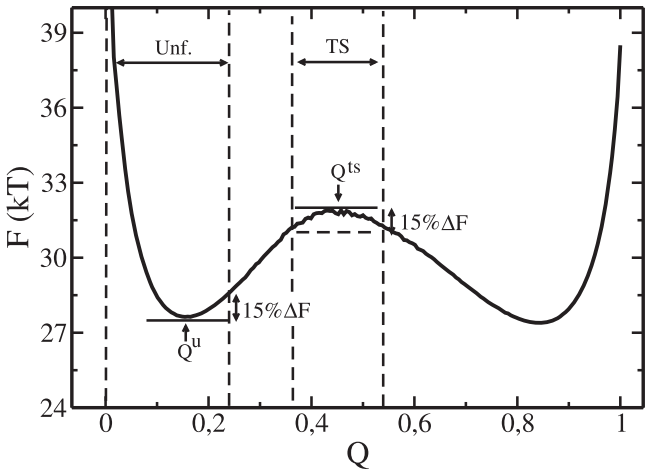

FIGURE 1 Free-energy profile (F) as a function of the native-contact fraction (Q) at the folding temperature (T_F) for a typical protein (Ubiquitin, PDB: 1UBQ) . Q^u is the native-contact fraction that corresponds to the first minimum in the unfolded state and Q^{ts} corresponds to the transition state at the free energy barrier peak. The region between the first two vertical dashed lines is defined as the unfolded region (*Unf*). The region delimited by the third and fourth vertical dashed lines, which involves the free-energy barrier peak, is defined as the transition region (*TS*). ΔF is the variation in free energy between the unfolded and transition states ($\Delta F = F(Q^{ts}) - F(Q^u)$).

RESULTS AND DISCUSSION

Correlation between nonnative-contact variation and the free-energy barrier without energetic frustration

In accordance with the Clementi-Plotkin model (40) in Eq. 2, the change in folding rates is directly related to the difference in the number of nonnative contacts that are formed in the transition and unfolded states (ΔA). As described in the previous section, if ΔA is positive, the addition of an attractive interaction between nonnative contacts decreases the free-energy barrier and increases the folding rate. To test the correlation between the number of nonnative contacts and the height of the free-energy barrier without frustration (ΔF^0), the number of nonnative contacts formed during the folding process for 15 proteins was calculated (see Table 1).

TABLE 1 Data Obtained for the 15 Proteins Studied and Sorted by ΔF^0

Protein	PDB	Amino Acids	M	RCO	ΔF^0 (k_BT)	ΔA ($\times 10^{-3}$)
α_3D	2A3D	73	136	0.095	0.23	−5.44
PtABD	1BDC	60	102	0.086	0.36	−0.98
EnHD	1ENH	54	111	0.13	0.71	−9.81
IM9	1IMP	86	178	0.11	1.56	−4.44
HHCC	1HRC	104	246	0.11	1.79	−7.10
PtL	2PTL	60	136	0.18	2.13	2.05
ADA2h	1PBA	81	175	0.14	2.38	3.60
PtG	2K0P	56	139	0.17	2.95	4.17
CI2	1CIS	66	152	0.16	3.00	−1.25
SH3	1FMK	61	152	0.19	3.89	3.94
Ubiquitin	1UBQ	76	188	0.15	4.34	1.27
HPr	1HDN	85	222	0.18	5.06	1.30
CSPTm	1G6P	66	180	0.17	5.69	5.00
TWIg	1WIU	93	253	0.20	5.86	10.19
αAIT	2AIT	74	196	0.19	6.21	11.68

Data included in the table for the proteins shown are the Protein Data Bank (PDB) code, the number of amino acids, the number of native contacts (M), the relative contact order (RCO), the free-energy barrier without energetic frustration (ΔF^0), and the variation in the fraction of nonnative contacts (ΔA).

The average number of nonnative contacts formed during the folding process varies for each protein. For example, Fig. 2 shows the average number of nonnative contacts formed during the folding process for two proteins, EnHD (Fig. 2 *a*) and SH3 (Fig. 2 *b*). Fig. 2 shows that for both proteins, there is an increase in the formation of nonnative contacts up to the maximum value at the pretransition region (*P-TS*). After the maximum value, the number of nonnative contacts has a monotonic decrease. Beyond the small difference in the average number of nonnative contacts formed, these proteins can be differentiated by their values and signals of ΔA. The protein EnHD, and also the proteins α_3 D, PtABD, EnHD, HHCC, IM9, and CI2, formed more nonnative contacts on average in the unfolded than in the transition region, i.e., $\Delta A < 0$. In this case, according to Eq. 2, it is expected that the addition of an energetic frustration in these proteins will be responsible for reducing the folding rates. On the other hand, for proteins PtL, ADA2h, PtG, SH3, Ubiquitin, HPr, TWIg, CSPTm, and αAIT, the formation of nonnative contacts in the transition state exceeds the

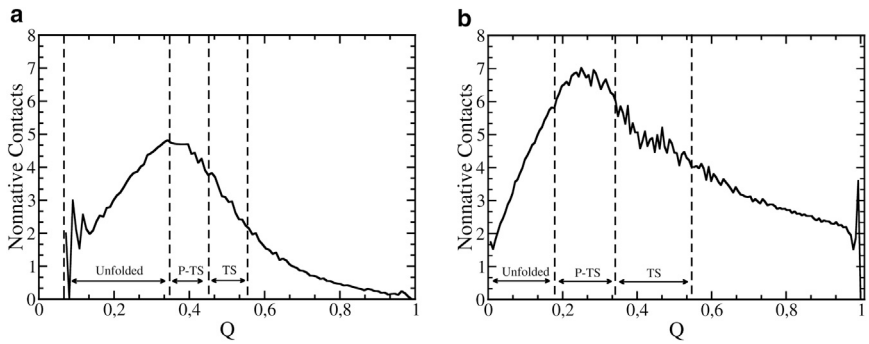

FIGURE 2 Average number of nonnative contacts ($\overline{A}$) as a function of the native-contact fraction (Q) at the folding transition temperature (T_F) for the proteins (*a*) EnHD and (*b*) SH3. Vertical dashed lines delimit the defined regions for the unfolded (*Unf*), transition (*TS*), and pretransition (*P-TS*) states. $\overline{A}$ reaches its maximum at the pretransition state for both cases.

number formed in the unfolded state, and therefore, $\Delta A > 0$. ΔA values calculated for all 15 proteins are shown in Table 1.

The correlation between the free-energy barrier height (ΔF^0, in kT units), without the energetic frustration term, and the variation in the number of nonnative contacts (ΔA) for the 15 proteins studied are shown in Fig. 3. The free-energy barrier correlates with the nonnative-contact variation by a factor of 0.81 (linear correlation) and it clusters the simulated proteins by their fold motifs. This result shows a strong correlation between these two variables and indicates that α-helix proteins make, on average, fewer nonnative contacts in the transition state than in the unfolded state. The opposite effect happens with proteins with β-sheets. This result appears not to be dependent on the chain length. The positive linear correlation coefficient indicates that proteins with a low variation in nonnative contacts (ΔA) have a low free-energy barrier and that the free-energy barrier increases with the increase in variation of nonnative contacts.

The product of ΔF^0 by native contact order was found to be the best parameter to correlate with optimum frustration values, according to the principal-component analysis shown by Contessoto et al. (46). It will be seen that the ΔA values for each protein are also strongly correlated with relative contact order (RCO) $\times$ ΔF^0, as can be seen in Fig. S1. The linear correlation of 0.84 between these parameters indicates that free-energy barrier height multiplied by the contact order is a good parameter for predicting the effect of the energetic frustration on these proteins, as was suggested in the earlier study (46). The ΔF^0 and RCO values are shown in Table 1.

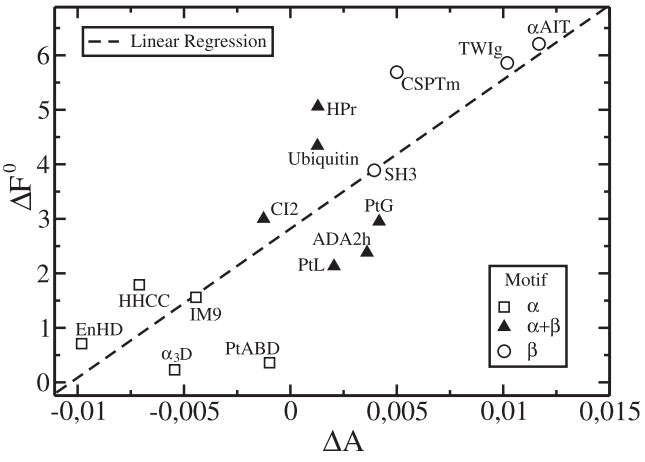

FIGURE 3 The free-energy barrier (ΔF^0) as a function of nonnative-contact-fraction variation (ΔA) for all proteins studied. Proteins are represented by their fold motif according to the SCOP database criterion (48): α (*squares*), β (*circles*), and $\alpha + \beta$ (*triangles*). The linear-fit correlation to the data is 0.81. ΔF^0 is strongly correlated to ΔA, and it clusters the simulated proteins by their motif in three distinct groups. The data were extracted from Table 1 using ΔF^0 (without energetic frustration).

Comparing analytical and computational results

In this section, the ΔA values and the variation in the free-energy barrier ($\Delta\Delta F = \Delta F^{\text{frus}} - \Delta F^0$) when the energetic frustration (ϵ_{NN}) is added to the C_α model are presented. The results were compared with those predicted by the Clementi-Plotkin model (Eq. 2) and the optimal energetic frustration from our previous study (46).

This study uses homogeneous energetic frustration with no sequence dependence. In our previous article (46), it was shown through principal-component analysis and partial least-squares analysis that the results seem to depend weakly on the frustration potential. In another study, a heterogeneous frustration potential was employed (9) in which energetic frustration was achieved by the addition of a random energy term characterized by a distribution width, b. An increase in folding temperature was observed for some values of b^2 and a decrease for high values. These results are analogous to the case of homogeneous frustration. More details regarding the potential used to add frustration to the C_α model can be seen in the Supporting Material. The ϵ_{NN} values used were 0.05, 0.10, and 0.20.

Nonnative-contact formation with energetic frustration

For all proteins studied here, an increase in the nonnative interaction leads to an increase in the formation of nonnative contacts during the folding process. This increase does not occur homogeneously in all ensembles; it becomes more evident in the unfolded and pretransition states. Therefore, depending on the value of ϵ_{NN}, a gradual increase in the amount of nonnative contacts is formed before the pretransition state and it can be higher than the average amount formed in the transition region. Thus, the gradual increase in the energetic frustration accounts for the reduction of ΔA values until it becomes negative, which, according to the Clementi-Plotkin model, makes the increase in frustration unfavorable to the folding process.

Fig. 4 shows the nonnative-contact variation (ΔA) as a function of the energetic frustration term (ϵ_{NN}) included in the simulation model. Each protein responds differently when a small quantity of frustration is added. Proteins formed exclusively by α-helixes, such as EnHD and HHCC, shown in Fig. 4, already had a negative value of ΔA, even without the energetic frustration term, and continue to have more negative values of ΔA when ϵ_{NN} is incremented. For $\alpha + \beta$ proteins, which had, initially, a positive value of ΔA, an increase in the ϵ_{NN} values causes a gradual change of their ΔA values and signals. Ubiquitin and ADA2h had their signals of ΔA inverted when the frustration term was 0.05 and 0.1, respectively. These values of ϵ_{NN}, which inverted the signal of ΔA, correspond exactly to the optimal energetic frustration determined for these proteins by Contessoto et al. (46). Very similar behavior occurs with CSPTm and αAIT proteins formed mainly by β-sheets. Such proteins had their ΔA values decreased gradually, but only reached a

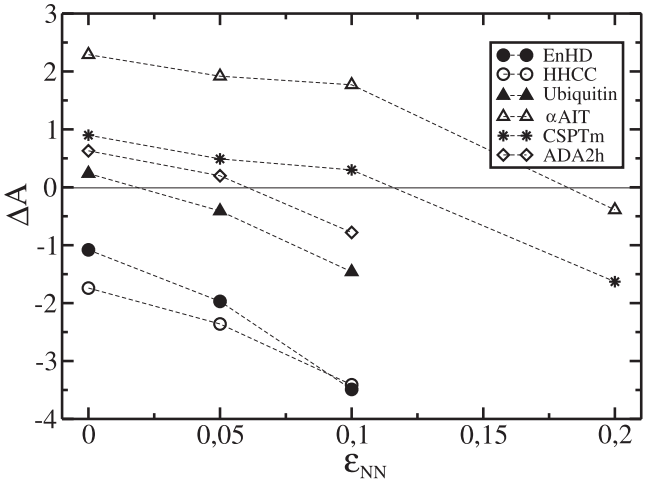

FIGURE 4 Variation in the formation of nonnative contacts (ΔA) according to the energetic frustration term (ϵ_{NN}) added in the simulation for six of our simulated proteins. Both axes are calculated at the respective folding temperature. The black line divides the positive and negative values of ΔA. The dotted lines connecting the points are a visual guide. ΔA is the difference between the numbers of nonnative contacts formed in the transition region and the unfolded region.

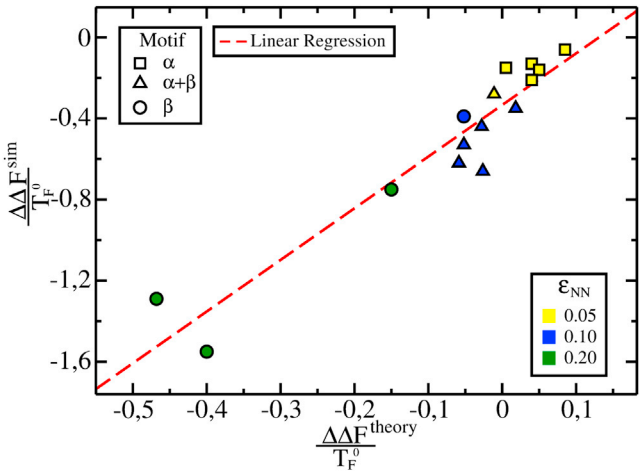

FIGURE 5 Analytical and simulated free-energy barrier variations when energetic frustration is added ($\Delta\Delta F = \Delta F^{frus} - \Delta F^0$). Each protein is represented by its respective motif according to the SCOP database criterion (48)—α (*squares*), β (*circles*), and $\alpha + \beta$ (*triangles*)—and colored by the energetic frustration (ϵ_{NN}) added to the simulation: 0.05, 0.1, and 0.2. Both axes are normalized by the respective folding temperature in the absence of ϵ_{NN} (T_F^0). The dotted line is the linear regression of the two data sets. The values of $\Delta\Delta F^{sim}$ are obtained by simulation using the weighted histogram analysis method with Q (the fraction of native contacts) as an order parameter. The values of $\Delta\Delta F^{theory}$ are obtained using Eq. 2 as reference with $b = 0$ and ΔA calculated by the simulation of each protein. The linear correlation to the data is 0.95. To see this figure in color, go online.

negative value at larger degrees of frustration ($\epsilon_{NN} = 0.2$). In all cases, when proteins formed more nonnative contacts in the unfolded state than in the transition state, ϵ_{NN} was close to the optimal energetic frustration value determined previously (46). In this case, when proteins are in their optimum frustration degree, they are at the same regime as those proteins in which the increase of frustration slows folding.

The vast majority of proteins have an unfolded dimension that scales with the protein chain length. An increase in the formation of nonnative contacts in the unfolded region can occur when an attractive energetic frustration is added to the simulation, and more unfolded compact structures can be formed, as shown in Fig. 4. Compaction of unfolded structures has been identified and studied computationally and experimentally in some proteins (52–56). This compaction is associated with the formation of hydropobic cores or with the formation of medium/long-range contacts. These unfolded structural aspects seem to play an important role in the thermodynamics and kinetics of the folding process (52,57).

Changes in the free-energy barrier height

To compare the variation of the simulated energetic barrier with and without the frustration term, a value of energetic frustration close to the optimal values already determined (46) was used; $\epsilon_{NN} = 0.05$ for α-helix proteins and Ubiquitin, $\epsilon_{NN} = 0.1$ for PtL, PtG, ADA2h, CI2, SH3, and HPr; and $\epsilon_{NN} = 0.2$ for CSPTm, αAIT, and TWIg.

The comparison between the variation in the free-energy barrier height, $\Delta\Delta F^{sim}$, obtained by the simulations, and $\Delta\Delta F^{theory}$, given by Eq. 2, is shown in Fig. 5. $\Delta\Delta F^{theory}$ is calculated using the M, T_f^0, and ΔA parameters calculated by simulation. Fig. 5 shows that the simulated results and the analytical model predictions are strongly correlated, with linear fit correlation 0.95. Thus, when the proteins are folded with a small degree of frustration close to their optimal value, the computational and the analytical result have the same behavior. Also, Fig. 5 shows the results clustered according to their folding motifs; β proteins have lower $\Delta\Delta F$ and high ϵ_{NN}, $\alpha + \beta$ proteins have intermediate $\Delta\Delta F$ and ϵ_{NN}, whereas α-proteins have high $\Delta\Delta F$ and low ϵ_{NN}, which is in agreement with the analytical model (Eq. 2) and the results from Fig. 3.

CONCLUSIONS

The difference in the number of nonnative contacts formed between the unfolded and the transition states showed a strong correlation with the height of the free-energy barrier, since proteins with a low free-energy barrier were those with fewer nonnative contacts formed in the transition state than in the unfolded state. In addition, the results suggest that proteins formed mainly by α-helix structures form more nonnative contacts in the unfolded state, in such a way that the ratio between the numbers of nonnative contacts formed in the transition state makes the energetic frustration unfavorable for them.

Energetic frustration, as it has been modeled in this study, proved to be responsible for increasing the formation of nonnative contacts, especially in the unfolded state. Frustration has also proved to reach a maximum threshold as the

ΔA values is decreased, leading proteins to form fewer nonnative contacts in the transition state than in the unfolded state. The energetic frustration value, which changes the sign of ΔA for almost all proteins tested, was close to the optimal frustration determined previously by Contessoto et al. (46). This corroborates the results expected by the theory developed by Clementi and Plotkin and makes a connection with the former computational analysis. Despite all the approximations that were made for the simulation and the analytical models, both resulted in agreement in 95% of the cases, as can be seen in Fig. 5.

Although it appears to be controversial that the presence of nonnative contacts may favor the folding process, this can be explained mainly by the fact that when inserting an energetic frustration, we add an attractive interaction energy in the midst of the contacts, which according to the results obtained, leads to an increase in the nonnative collapse. The increase in the nonnative collapse may favor access to other structures, which may be responsible for bringing amino acids close to each other, which in turn favors the formation of native contacts, accelerating the kinetics of the process. One question that arises from these studies is about the nature of the changes and stabilizations observed when a nonnative potential is introduced. Such a stabilization may originate with entropic or enthalpic factors. In preliminary studies, we observed variations dependent on the protein motifs. Further studies should not only investigate different motifs, but also analyze such effects in terms of their dependence on the type of frustration used.

SUPPORTING MATERIAL

Supporting Materials and Methods and one figure are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(16)30388-5.

AUTHOR CONTRIBUTIONS

P.R.M., J.C., and V.B.P.L. designed research. P.R.M. performed research. P.R.M. and V.G.C. contributed analytical tools. Data were analyzed by P.R.M., V.G.C., R.J.O., J.C., and V.B.P.L. All authors contributed to the manuscript writing.

ACKNOWLEDGMENTS

P.R.M. and V.G.C. were supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil. R.J.O. was funded by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG). V.B.P.L. was funded by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 2014/06862-7). R.J.O., J.C., and V.B.P.L. were supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). This research was supported by resources supplied by the Center for Scientific Computing (NCC/GridUNESP) of the São Paulo State University (UNESP).

SUPPORTING CITATIONS

References (58–63) appear in the Supporting Material.

REFERENCES

1. Dobson, C. M. 2003. Protein folding and misfolding. *Nature.* 426:884–890.
2. Baldwin, R. L. 1995. The nature of protein folding pathways: the classical versus the new view. *J. Biomol. NMR.* 5:103–109.
3. Dill, K. A., and H. S. Chan. 1997. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* 4:10–19.
4. Pande, V. S., AYu Grosberg, and T. Tanaka. 1997. On the theory of folding kinetics for short proteins. *Fold. Des.* 2:109–114.
5. Onuchic, J. N., H. Nymeyer, …, N. D. Socci. 2000. The energy landscape theory of protein folding: insights into folding mechanisms and scenarios. *Adv. Protein Chem.* 53:87–152.
6. Leopold, P. E., M. Montal, and J. N. Onuchic. 1992. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. USA.* 89:8721–8725.
7. Frauenfelder, H., S. G. Sligar, and P. G. Wolynes. 1991. The energy landscapes and motions of proteins. *Science.* 254:1598–1603.
8. Wolynes, P. G., J. N. Onuchic, and D. Thirumalai. 1995. Navigating the folding routes. *Science.* 267:1619–1620.
9. Wang, J., R. J. Oliveira, …, V. B. P. Leite. 2012. Topography of funneled landscapes determines the thermodynamics and kinetics of protein folding. *Proc. Natl. Acad. Sci. USA.* 109:15763–15768.
10. Nymeyer, H., A. E. García, and J. N. Onuchic. 1998. Folding funnels and frustration in off-lattice minimalist protein landscapes. *Proc. Natl. Acad. Sci. USA.* 95:5921–5928.
11. Shoemaker, B. A., J. Wang, and P. G. Wolynes. 1997. Structural correlations in protein folding funnels. *Proc. Natl. Acad. Sci. USA.* 94:777–782.
12. Whitford, P. C., J. K. Noel, …, J. N. Onuchic. 2009. An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins.* 75:430–441.
13. Oliveira, R. J., P. C. Whitford, …, J. Wang. 2010. Coordinate and time-dependent diffusion dynamics in protein folding. *Methods.* 52:91–98.
14. Fersht, A. R. 1995. Characterizing transition states in protein folding: an essential step in the puzzle. *Curr. Opin. Struct. Biol.* 5:79–84.
15. Garcia-Mira, M. M., M. Sadqi, …, V. Muñoz. 2002. Experimental identification of downhill protein folding. *Science.* 298:2191–2195.
16. Nettels, D., I. V. Gopich, …, B. Schuler. 2007. Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc. Natl. Acad. Sci. USA.* 104:2655–2660.
17. Chung, H. S., J. M. Louis, and W. A. Eaton. 2009. Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. *Proc. Natl. Acad. Sci. USA.* 106:11837–11844.
18. Koga, N., and S. Takada. 2001. Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model. *J. Mol. Biol.* 313:171–180.
19. Chavez, L. L., J. N. Onuchic, and C. Clementi. 2004. Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. *J. Am. Chem. Soc.* 126:8426–8432.
20. Snow, C. D., E. J. Sorin, …, V. S. Pande. 2005. How well can simulation predict protein folding kinetics and thermodynamics? *Annu. Rev. Biophys. Biomol. Struct.* 34:43–69.
21. Gosavi, S., L. L. Chavez, …, J. N. Onuchic. 2006. Topological frustration and the folding of interleukin-1β. *J. Mol. Biol.* 357:986–996.
22. Chu, X., L. Gan, …, J. Wang. 2013. Quantifying the topography of the intrinsic energy landscape of flexible biomolecular recognition. *Proc. Natl. Acad. Sci. USA.* 110:E2342–E2351.
23. Bryngelson, J. D., and P. G. Wolynes. 1987. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA.* 84:7524–7528.
24. Fersht, A. R. 1997. Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.* 7:3–9.

25. Chahine, J., R. J. Oliveira, …, J. Wang. 2007. Configuration-dependent diffusion can shift the kinetic transition state and barrier height of protein folding. *Proc. Natl. Acad. Sci. USA.* 104:14646–14651.
26. Oliveira, R. J., P. C. Whitford, …, V. B. Leite. 2010. The origin of non-monotonic complex behavior and the effects of nonnative interactions on the diffusive properties of protein folding. *Biophys. J.* 99:600–608.
27. Xu, W., Z. Lai, …, J. Wang. 2012. Configuration-dependent diffusion dynamics of downhill and two-state protein folding. *J. Phys. Chem. B.* 116:5152–5159.
28. Oliveira, A. B., Jr., F. M. Fatore, …, V. B. P. Leite. 2014. Visualization of protein folding funnels in lattice models. *PLoS One.* 9:e100861.
29. Onuchic, J. N., P. G. Wolynes, …, N. D. Socci. 1995. Toward an outline of the topography of a realistic protein-folding funnel. *Proc. Natl. Acad. Sci. USA.* 92:3626–3630.
30. Onuchic, J. N., Z. Luthey-Schulten, and P. G. Wolynes. 1997. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* 48:545–600.
31. Eaton, W. A., P. A. Thompson, …, J. Hofrichter. 1996. Fast events in protein folding. *Structure.* 4:1133–1139.
32. Eaton, W. A., V. Muñoz, …, J. Hofrichter. 1997. Submillisecond kinetics of protein folding. *Curr. Opin. Struct. Biol.* 7:10–14.
33. Ozkan, S. B., I. Bahar, and K. A. Dill. 2001. Transition states and the meaning of Φ-values in protein folding kinetics. *Nat. Struct. Biol.* 8:765–769.
34. Shakhnovich, E. I., and A. M. Gutin. 1989. Formation of unique structure in polypeptide chains. Theoretical investigation with the aid of a replica approach. *Biophys. Chem.* 34:187–199.
35. Bryngelson, J. D., and P. G. Wolynes. 1989. Intermediates and barrier crossing in a random energy-model (with applications to protein folding). *J. Phys. Chem.* 93:6902–6915.
36. Goldstein, R. A., Z. A. Luthey-Schulten, and P. G. Wolynes. 1992. Optimal protein-folding codes from spin-glass theory. *Proc. Natl. Acad. Sci. USA.* 89:4918–4922.
37. Abkevich, V. I., A. M. Gutin, and E. I. Shakhnovich. 1994. Free energy landscape for protein folding kinetics: Intermediates, traps, and multiple pathways in theory and lattice model simulations. *J. Chem. Phys.* 101:6052–6062.
38. Bryngelson, J. D., J. N. Onuchic, …, P. G. Wolynes. 1995. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins.* 21:167–195.
39. Oliveira, L. C., R. T. H. Silva, …, J. Chahine. 2006. Frustration and hydrophobicity interplay in protein folding and protein evolution. *J. Chem. Phys.* 125:084904.
40. Clementi, C., and S. S. Plotkin. 2004. The effects of nonnative interactions on protein folding rates: theory and simulation. *Protein Sci.* 13:1750–1766.
41. Plotkin, S. S. 2001. Speeding protein folding beyond the G(o) model: how a little frustration sometimes helps. *Proteins.* 45:337–345.
42. Plotkin, S. S., and J. N. Onuchic. 2002. Understanding protein folding with energy landscape theory. Part I: Basic concepts. *Q. Rev. Biophys.* 35:111–167.
43. Zarrine-Afsar, A., S. Wallin, …, H. S. Chan. 2008. Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding. *Proc. Natl. Acad. Sci. USA.* 105:9999–10004.
44. Zarrine-Afsar, A., Z. Zhang, …, H. S. Chan. 2012. Kinetic consequences of native state optimization of surface-exposed electrostatic interactions in the Fyn SH3 domain. *Proteins.* 80:858–870.
45. Zhang, Z., and H. S. Chan. 2010. Competition between native topology and nonnative interactions in simple and complex folding kinetics of natural and designed proteins. *Proc. Natl. Acad. Sci. USA.* 107:2920–2925.
46. Contessoto, V. G., D. T. Lima, …, V. B. P. Leite. 2013. Analyzing the effect of homogeneous frustration in protein folding. *Proteins.* 81:1727–1737.
47. Plaxco, K. W., K. T. Simons, and D. Baker. 1998. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* 277:985–994.
48. Murzin, A. G., S. E. Brenner, …, C. Chothia. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247:536–540.
49. Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. 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. *J. Mol. Biol.* 298:937–953.
50. Clementi, C., A. E. García, and J. N. Onuchic. 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.
51. de Mendonça, M. R., L. G. Rizzi, …, N. A. Alves. 2014. Inferring a weighted elastic network from partial unfolding with coarse-grained simulations. *Proteins.* 82:119–129.
52. Kohn, J. E., I. S. Millett, …, K. W. Plaxco. 2004. Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. USA.* 101:12491–12496.
53. Kazmirski, S. L., K.-B. Wong, …, V. Daggett. 2001. Protein folding from a highly disordered denatured state: the folding pathway of chymotrypsin inhibitor 2 at atomic resolution. *Proc. Natl. Acad. Sci. USA.* 98:4349–4354.
54. Garcia, P., L. Serrano, …, M. Bruix. 2001. NMR and SAXS characterization of the denatured state of the chemotactic protein CheY: implications for protein folding initiation. *Protein Sci.* 10:1100–1112.
55. Hodsdon, M. E., and C. Frieden. 2001. Intestinal fatty acid binding protein: the folding mechanism as determined by NMR studies. *Biochemistry.* 40:732–742.
56. Meng, W., N. Lyle, …, R. V. Pappu. 2013. Experiments and simulations show how long-range contacts can form in expanded unfolded proteins with negligible secondary structure. *Proc. Natl. Acad. Sci. USA.* 110:2123–2128.
57. Shortle, D. 1996. The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J.* 10:27–34.
58. Berman, H. M., J. Westbrook, …, P. E. Bourne. 2000. The protein data bank. *Nucleic Acids Res.* 28:235–242.
59. Sobolev, V., R. C. Wade, …, M. Edelman. 1996. Molecular docking using surface complementarity. *Proteins.* 25:120–129.
60. Noel, J. K., P. C. Whitford, …, J. N. Onuchic. 2010. SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res.* 38:W657–W661.
61. Van Der Spoel, D., E. Lindahl, …, H. J. C. Berendsen. 2005. GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26:1701–1718.
62. Ferrenberg, A. M., and R. H. Swendsen. 1988. New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.* 61:2635–2638.
63. Ferrenberg, A. M., and R. H. Swendsen. 1989. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* 63:1195–1198.

Biophysical Journal, Volume *111*

Supplemental Information

Quantifying Nonnative Interactions in the Protein-Folding Free-Energy Landscape

Paulo Ricardo Mouro, Vinícius de Godoi Contessoto, Jorge Chahine, Ronaldo Junio de Oliveira, and Vitor Barbanti Pereira Leite

Supporting Material

Structure-based C_α model

The Cα model is a structure-based model which considers the amino acid sequence as simple spheres centralized at the position of the alpha carbon, in accordance with the structure data deposited in the Protein Data Bank (PDB) (1). During the dynamic to which the model is submitted, the potential that defines the energy of the conformations is of G$\bar{o}$-type, in which the main idea is to give importance to the interactions between amino acids residing in native contacts, and then choose the energy of contacts that minimize the total energy of the native state (1, 2).

The expression defining the energy of a configuration Γ based on the native conformation Γ^0 for the model Cα-model is given by,

$$\begin{aligned} V(\Gamma,\Gamma_o) = & \sum_{bonds} \epsilon_r (r-r_o)^2 + \sum_{angles} \epsilon_\theta (\theta-\theta_o)^2 \\ & + \sum_{dihedrals} \epsilon_\phi \left\{ [1-cos(\phi-\phi_o)] + \frac{1}{2}[1-cos(3(\phi-\phi_o))] \right\} \\ & + \sum_{contacts} \epsilon_C \left[5\left(\frac{d_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{d_{ij}}{r_{ij}}\right)^{10} \right] \\ & + \sum_{non-contacts} \epsilon_{NC} \left(\frac{\sigma_{NC}}{r_{ij}}\right)^{12} \end{aligned} \tag{S1}$$

In Equation S1, the first term refers to a harmonic potential representing the bond between two adjacent α-carbons in which r_0 is the distance between the two carbons connected in the native structure. The second sum also forms a harmonic potential, but this time, an angular harmonic potential formed by three α-carbons in sequence in the polypeptide chain. θ_0 is the angle formed by the three residues in the native conformation. The third term of the expression takes into consideration the torsion carried by the chain

in which, ϕ_0 is the dihedral angle formed by four α-carbons in sequence. The fourth term accounts for the interaction between non-bonded α-carbons i and j , but which make contact in the native structure. For this is used a potential 10-12, in which d_{ij} is the value of the distance between the carbons which make a native contact. A native contact is defined by a map created by CSU software (3). Finally, the last term refers to all α-carbons that do not form a native contact. This term is used to keep the maximum approach distance between α-carbons. $\sigma_{NC} = 4\mathring{A}$ (volume of the carbon atom in the model). The constants ($\epsilon_r = 100$, $\epsilon_\theta = 20$, $\epsilon_\phi = 1$ and $\epsilon_{NC} = 1$) are given in ϵ_c units (2, 4).

The potential used to add frustration to the Cα-model is shown in the equation S2.

$$V_f(r) = -\epsilon_{NN} \exp\left[-\frac{(r_{ij} - \bar{d})^2}{{\sigma_f}^2}\right] \tag{S2}$$

$\bar{d}$ is the average of the distances of the native contacts, $\sigma_f = 1\ \mathring{A}$ and ϵ_{NN} is the frustration parameter in ϵ_c units. In this study $\epsilon_{NN} = 0.00$; $\epsilon_{NN} = 0.05$; $\epsilon_{NN} = 0.1$ and $\epsilon_{NN} = 0.2$ are used.

Simulation details

The topology and initial coordinates files were generated by Structure-based Models in Gromacs (SMOG) available online (5). The dynamics was performed using molecular dynamics package *GROMACS version 4.5-5* (6) using a stochastic integrator and the Berendsen thermal coupling. The simulations without frustration were made with 10^9 steps (having the information stored every 5000 steps with 0.5 fs of time step) and equilibrated after 10^7 steps. The simulations using $\epsilon_{NN} \neq 0$ were made with 5×10^8 steps (having the information stored every 5000 steps with 0.5 fs of time step) and equilibrated

after 10^6 steps. The native contact fraction (Q) was the order parameter used to accompany the folding process, and a native contact was accepted in a Γ conformation if the distance between the amino acid in this conformation was less than $1.2d_{ij}$, in which d_{ij} is the distance between the residues in the native structure. The dynamic was realized with the computational resources of GridUnesp.

The thermodynamic profiles, such as thermal energy, free energy, entropy and specific heat were calculated by the method of multiple histograms (WHAM - *Weighted Histograms Analysis Method (7, 8)* in which the folding temperature was defined as the temperature at which F(0) ≈ F(1).

Figure

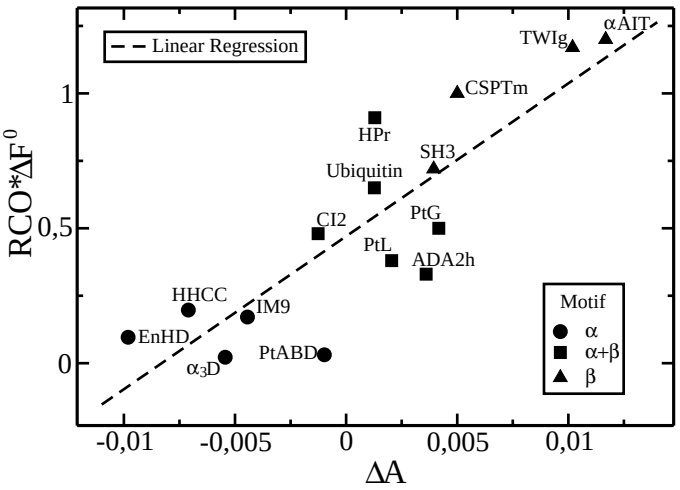

Figure S1

Figure S1. Free energy barrier and relative contact order ($\Delta F^0 \times RCO$) as a function of nonnative contact fraction variation (ΔA) for all proteins studied. Proteins are represented by their fold motif according to the SCOP database criterion (9): α (circles), β (triangles) and $\alpha + \beta$ (squares). The linear fit correlation to the data is 0.84. The data were extracted from Table I by using (ΔF) without energetic frustration.

Supporting References

1. Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. 2000. The protein data bank. Nucleic. Acids. Res. 28:235–242.

2. Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. 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. J. Mol. Biol. 298:937–953.

3. Sobolev, V., R. Wade, G. Vried, and M. Edelman. 1996. Molecular docking using surface complementarity. Proteins: Struct. Funct. Genet. 25:120–129.

4. Contessoto, V. G., D. T. Lima, R. J. Oliveira, A. T. Bruni, J. Chahine, and V. B. P. Leite. 2013. Analyzing the effect of homogeneous frustration in protein folding. Proteins: Struct. Funct. Bioinf. 81:1727–1737.

5. Noel, J. K., P. C. Whitford, K. Y. Sanbonmatsu, and J. N. Onuchic. 2010. SMOG@ctbp: simplified deployment of structure-based models in GROMACS. Nucleic Acids Research. .

6. Van Der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. 2005. GROMACS: fast, flexible, and free. J. Comp. Chem. 26:1701–1718.

7. Ferrenberg, A. M., and R. H. Swendsen. 1988. New monte carlo technique for studying phase transitions. Phys. Rev. Lett. 61:2635–2638.

8. Ferrenberg, A. M., and R. H. Swendsen. 1989. Optimized monte-carlo data analysis. Phys. Rev. Lett. 63:1195–1198.

9. Murzin, A. G., S. E. Brenner, T. Hubbard, and C. Chothia. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247:536–540.