

Energetic heterogeneity selects the folding nucleus of PDZ protein

Alexander Kluber^{1,1}, Cecilia Clementi^{1,1}

Abstract

(250 word max) Energetic heterogeneity in structure-based models.

Keywords: structure-based model, Protein folding, folding nucleus

1. Introduction

Protein folding has come to serve as an important touchstone for understanding biomolecular organization. One where simplified theories and models have been able to make important connections with experiment. Folding is possible on biological timescales because evolution has crafted the interactions in the folded structure to be in harmony when compared to the average misfolded alternative. The

Simplified protein models serve as an important testbed for connecting energy landscape theory to experimental observations. In particular, structure-based models have demonstrated that the topology of the folded state plays an important role in shaping the energy landscape [1].

Since native contacts are the central component to structure-based models and

For simplicity the strengths of the native contact interactions in structure-based models are commonly taken to be uniform across the protein. Since we do not know a priori the appropriate values to assign.

Theoretical work by Plotkin and Onuchic showed that both energetic and structural heterogeneity can affect the folding mechanism and folding free energy barrier.

Work by Cho by has shown that structure-based models of β proteins are more robust with respect to energetic perturbation than α -helical proteins. Experimental evidence allows suggests that homologous α -helical proteins can exhibit a wider spectrum of folding mechanisms (Fersht Homeodomain, Clarke spectrins, Radford Im7). It is

By modeling energetic heterogeneity .

Understanding the sensitivity of model predictions with respect to parameterization is fundamental computation. For example, even amongst. We parameterize our model to be consistent with . choice of parameterization. how sensitive a model parameterization is ways.

put native contacts in .have a . Theoretical work has shown that . The original off-lattice variation [1].

Despite having the potential .

Single folding trajectories perform diffusion that is bias towards the native state.

The probability fluxes of folding trajectories are guided by the contours of the energy landscape.

The ensemble of folding routes can be imagined as following riverbeds in the free energy funneled landscapes. Upon close inspection these folding routes are braided amongst the microcorrugations .

Structure-based models have now become common practice in modeling of large biomolecules as they give access to longer timescale dynamics of biomolecules that fold to reasonably well-defined structures. Structure-based models are supported theoretically by the energy landscape theory of protein folding and in particular the principle of minimal frustration. The principle of minimal frustration states that heteropolymers searching for particle folded structure.

[1] [2] [3] [4]

Heuristically speaking .

Structure-based models started by using uniform contact strengths for all interactions (the homogeneous model). This would accurately describe cases where the average contact energies is representative of

[4]

Heterogeneous contact energies have been implemented in the literature by: Matysiak, Clementi Karanicolas, Brooks Cho, Wolynes

There are several simplified models of proteins that use funneling in their construction. The model of Karanicolas, Brooks incorporates energetic and backbone heterogeneity into .

The folding mechanisms of all β proteins are robustly captured by structure-based models because heterogeneity in their native and non-natives contacts is self-averaging.

The quest for self-averaging properties.

A self-averaging properties.

Physical descriptors that characterize an esemble. For example, the radius of gyration and average collapse time of heteropolymer is a self-averaging property of sequences with the same hydrophobic content (average intrachain attraction).

Considering all the contact energy distributions with similar characteristics (mean and variance), how do we show that ours leads to different physics?

Hypothesis: Arbitrary heterogeneity will lead to 1) broadening of the transition and 2) lower free energy barrier. [5]

Oztop and coworkers [6] have previously suggested that real proteins have significant energetic heterogeneity by comparing the dispersion of ϕ -values to dispersion in contact loop lengths: $\overline{\delta\phi^2}$ vs. $\frac{\overline{\delta l^2}}{l^2}$

2. Results

2.1. Folding kinetics and thermodynamics

Energetic heterogeneity alters the folded and unfolded states: the unfolded state has more residual structure and the folded state frays. This results in smaller smaller energetic and entropic changes upon folding ΔE_{N-U} and ΔS_{N-U} . The folding temperature usually lowers because the difference in energy decreases more than the difference in entropy $T_f \approx \frac{\Delta E_{N-U}}{k_B \Delta S_{N-U}}$.

The theory of Plotkin and Onuchic [5] treats energetic and structural heterogeneity explicitly. According to the Plotkin theory a random perturbation to the contact energies will tend to lower the free energy barrier and this is seen in simulations of random perturbations to the homogeneous model.

The configurational diffusion coefficient is not discussed in the theory of Plotkin. Energetic heterogeneity decreases the configurational diffusion coefficient. The theory of Portman and coworkers [7] supplies a method for calculating the rate law prefactor in terms of the dynamics of crossing the high-dimensional free energy saddle-point.

2.2. Folding mechanism

For arbitray heterogeneity the folding nucleus shifts from globular (located in one place) to more ramified (broken in disconnected). The overall locus of structure formation . Mechanism

Energetic heterogeneity

The folded state becomes frayed with increasing heterogeneity. Folding temperatures usually go down because native state begins to fray

We have found that structure-based models

Adding energetic heterogeneity to the structure-based model for PDZ shifts the center of the folding nucleus. The folding nucleus shifts from the beta1-beta2 hairpin in the homogeneous model to the C-terminus beta strands in the heterogeneous model.

We conclude that topology and heterogeneity are important for shaping the folding of PDZ.

Our procedure optimizes the model parameters in order to reproduce a given experimental observable. We chose to optimize our model using $\Delta\Delta G$'s from ϕ -value analysis (as in Matysiak 2004), however the procedure used is general.

We claim that

Comparisons with the Wolynes group's frustratometer suggest.

Uniform contact strengths result in fluctuations that correlate with the growth of the folding nucleus. Contacts are collectively pulled into the nucleus.

Heterogeneity blurs the surface of the nucleus. The folding nucleus of the .

If the prefactors are

Heterogeneity has been addressed in theoretical models of protein folding [5].

Non-Markovian memory effects alter the rate prefactor and local barrier crossing coordinate [8, 7, 9].

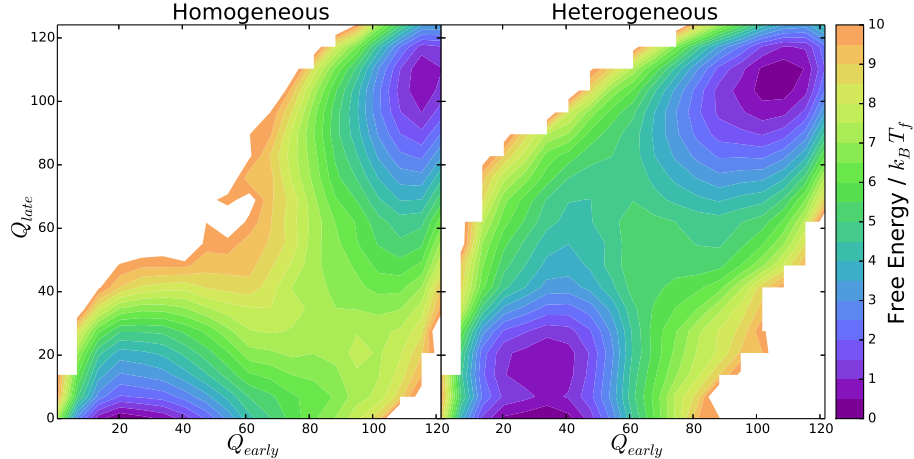


Figure 1: Optimizing contact energies of the PDZ domain reveal folding through an alternative nucleus due to the coupling of energetic and entropic fluctuations. Reaction coordinates Q_{early} (Q_{late}) are the number of contacts that form with more (less) than 0.5 probability in homogeneous SBM transition state. Random perturbations of the same magnitude do not exhibit the same concerted shift in mechanism.

The effect of non-native interactions on folding rates [10, 11].

Simulations of replicas with randomly distributed native interactions indicate what features are shared amongst all parameter sets with the same heterogeneity. Random heterogeneity with the same variance as our optimized interactions represent the noise level.

Comparing replicas of the same random native heterogeneity to our optimized energies indicate what properties are generally shared between heterogeneous parameter sets versus what properties are specific to our results.

We discuss our results in the context of a Random Energy Model of folding that treats native heterogeneity specifically [5]. Mean-field theories of folding.

The importance of energetic heterogeneity could be considered.

If heterogeneity is self-averaging.

3. Discussion

Proteins are minimally frustrated heteropolymers that can fold consistently to well defined structures. Since their biological function requires interacting with other molecules, residual frustration occurs in their folded structures. It can be important to consider.

with respect to the function that have evolved to perform.

The model assumes the native state to be uniformly minimally frustrated. This leads to a natural preference for contacts that are close in sequence, such as helices, turns, and hairpins, due to their small entropic cost. However, real

proteins may have localized frustration due to functional constraints during evolution (e.g. on binding faces).

The interplay of the backbone bias and existence of residual frustration can explain the three studied cases. In the cases of

PDZ and S6 have backbone and/or contact frustration in their folded structures in regions that are predicted to be too structured in the transition state in comparison with experiment. On the other hand SH3 has a minimally frustrated turn where it compares favorably with experiment.

How can we effectively capture local interactions that are not really minimally frustrated using structure-based models? - Optimize contact interactions - Optimize backbone interactions - Optimize both contact and backbone interactions - Use contact strengths derived from transferable potential - Use transferable backbone potential

Hypothesis: If I simultaneously optimize the contact and dihedral strengths will increased flexibility compensate for contact heterogeneity?

Raises question, is the nonlocal contact heterogeneity/frustration also important or the corrections only important for local interactions?

Sam Cho (2009) suggests that predictions will be more robust for proteins with a high nonlocal/local ratio. Our work indicates that the of .

Flexibility modulates the impact of heterogeneity/frustration.

However, real proteins are may have localized frustration in their folded states from functional considerations (Ferriero), which may be alleviated when they e.g. bind their intended partner.

The discrepancies of the homogeneous structure-based model with experiment observations can be understood when considering localized frustration explains.

This is why we observe repulsive native interactions.

have localized frustration in their native structures indicates that this hypothesis .

real proteins

The homogeneous structure-based model assumes that protein interactions are uniformly unfrustrated throughout. However that may not be the case given that proteins have evolved their sequences to perform particular functions as well as fold. The folding nucleus 1 of PDZ includes residues frustration in the turn region.

Future work to design the beta hairpin of PDZ to favor nucleus 1.

Gianni 2007 has previously suggested that PDZ captures the essential features of nucleation-condensation folding mechanisms whereby the transition state ensemble is a diffuse nucleus center around the termini. Frustration.

All the parameters are not free in the sense that they are correlated. The Jacobian does not have full rank.

Protein design

See [12] for discussion of capillarity model of folding. Issues raised are: the scaling of the effective size of the folding nucleus; the breadth of the interface separating folding and unfolded phases; the size scaling of the folding barrier;

applicability of mean-field versus capillarity models of folding; what are the size of the heterogeneity effects? ;

According to the capillarity model of folding [12] energetic heterogeneity should roughen the interface bringing capillarity model.

See [?] for discussion of “multiple folding nuclei” (MFN) perspective of Thirumalai. See [13]: discusses “multiple folding nuclei” versus “diffuse nucleus model” versus BW capillarity picture.

Shakhnovich has discussed a “specific folding nucleus” [14] where specific interactions must be formed.

4. Materials and Methods

4.1. Structure-based model

We use a “C_α” structure-based model derived from [1], where the model Hamiltonian $H = H_{bonded} + H_{nonbonded}$ has a term that applies a local bias to the backbone H_{bonded} and a term for the long-range interactions between residue beads $H_{nonbonded}$. The functional forms of these terms are,

$$H_{bonded} = \sum_{bonds} k_b(r_{ij} - r_{ij}^0)^2 + \sum_{angles} k_\theta(\theta_{ijk} - \theta_{ijk}^0)^2 + \quad (1)$$

$$\sum_{dihedrals} k_\phi[\cos(\phi_{ijkl} - \phi_{ijkl}^0) + \frac{1}{2} \cos(3(\phi_{ijkl} - \phi_{ijkl}^0))]$$

$$(2)$$

$$(3)$$

$$H_{non-bonded} = \sum_{native} \epsilon_{ij} V_{ij}^{cont}(r_{ij}) + V_{ij}^{contex}(r_{ij}) \sum_{non-native} \epsilon_{ij}^{ex} \left(\frac{r_{ex}}{r_{ij}} \right)^{12} \quad (4)$$

Non-native contacts are given a purely repulsive potential of fixed strength $\epsilon_{ij}^{ex} = 1$ while native contacts are allowed to be attractive or repulsive with heterogeneous strengths. When two beads have an attractive interactions

Residue pairs that are not in contact in the native structure are given a purely repulsive

Contact maps were created using Shadow Map[15] via the SMOG webserver[16].

The starting point of our investigation is the homogeneous structure-based model derived from [?] which places one bead per residue at the C_α positions of the backbone and creates attractive interactions between beads that are .

The energy scale of the model ϵ is constrained to its average value of the optimization.

places one bead per residues at the alpha-carbon positions of the backbones. Beads that are in contact in the folded structure are given attractive (or repulsive) native contact interactions. structure-based model is based off of

Our simplified model places one bead at the alpha-carbon of each residue. Beads have an excluded volume radius of 4 angstroms. Residues that are in contact in the folded structure, “native contacts”, are given an attractive gaussian interaction. The strengths of the native contacts is varied by the optimization algorithm in order to reproduce the .

A structure-based model is a simplified

$$H_{bonded} = \sum_{bonds} k_b (r_{ij} - r_{ij}^0)^2 + \sum_{angles} k_\theta (\theta_{ijk} - \theta_{ijk}^0)^2 + \quad (5)$$

$$\sum_{dihedrals} k_\phi [\cos(\phi_{ijkl} - \phi_{ijkl}^0) + \frac{1}{2} \cos(3(\phi_{ijkl} - \phi_{ijkl}^0))] \quad (6)$$

$$(7)$$

The bonded constants used in this work are taken from Clementi et.al[1] and are (all in (kJ/mol)): $k_b = 20000$, $k_\theta = 40$, $k_\phi = 1$.

The non-bonded Hamiltonian contains long-range interactions between beads.

$$H_{non-bonded} = \sum_{native} \epsilon_{ij} V_{ij}^{cont}(r_{ij}) + \sum_{non-native} \epsilon_{ij}^{ex} \left(\frac{r_{ex}}{r_{ij}} \right)^{12} \quad (8)$$

The Gaussian potential class takes the following forms,

$$V_{ij}^G(r_{ij}) = -e^{-\frac{(r_{ij} - r_{ij}^0)^2}{2\sigma_{ij}}} \quad (9)$$

$$V_{ij}^{Grep}(r_{ij}) = \frac{1}{2} \left[\tanh \left(- \left(\frac{r_{ij} - r_{ij}^0 - \sigma_{ij}}{\sigma_{ij}} \right) \right) + 1 \right] \quad (10)$$

and requires adding an additional excluded volume to the corresponding pair in order to ensure that the contact potential is not perturbed at its equilibrium distance r_{ij}^0 ,

$$V_{ij}^{Gexc}(r_{ij}) = \left(\frac{r_{ex}^0}{r_{ij}} \right)^{12} \left(1 - e^{-\frac{(r_{ij} - r_{ij}^0)^2}{2\sigma_{ij}}} \right) \quad (11)$$

Homogeneous structure-based model developed by Clementi Implementation in gromacs using Shadow map, Gaussian contacts. SMOG. Noel, Lammert

4.2. Parameter learning

Parameter fitting algorithm developed by Matysiak, Clementi

$$\vec{f}^{sim}(\vec{\epsilon}) \approx \vec{f}^{sim}(\vec{\epsilon}^{(0)}) + \mathbf{J} \cdot \delta \vec{\epsilon} \quad (12)$$

Where (dropping the vector notation) $\delta \epsilon$ is some change in the model parameters and \mathbf{J} encodes how that parameter change affects our simulation observable,

$$(\mathbf{J})_{ij} = \frac{\partial f_i^{sim}}{\partial \epsilon_j} \quad (13)$$

Setting equation 12 equal to the vector of experimental observables \bar{f} and collecting the error $\delta f = f^{sim} - \bar{f}$ to the left-hand side yields,

$$-\delta f = \mathbf{J} \delta \epsilon \quad (14)$$

Which can be solved for the update to the model parameters $\delta \epsilon$ that will bring the simulated observables closer to the experimental values using standard tools for ill-posed problems (e.g. damped least-squares[17]; Singular Value Decomposition).¹ Overall this yields an iterative protocol for updating the model parameters which takes the form (on iteration n),

$$\epsilon^{(n+1)} = \epsilon^{(n)} + \delta \epsilon^{(n)} \quad (15)$$

The general procedure outlined above is iterated until satisfactory convergence. The specific form of the Jacobian matrix \mathbf{J} depends on the specific observable being reproduced. Recall that the Jacobian is the partial derivative of the observable with respect to one of the model parameters,

$$(\mathbf{J})_{ij} = \frac{\partial f_i}{\partial \epsilon_j} \quad (16)$$

In most cases, the Jacobian \mathbf{J} takes the simple form of a correlation function. For example, if \bar{f} is any mechanical observable (i.e. anything computed solely from the coordinates), then,

$$\frac{\partial f_i}{\partial \epsilon_j} = -\beta \left[\left\langle f_i \frac{\partial H}{\partial \epsilon_j} \right\rangle - \langle f_i \rangle \left\langle \frac{\partial H}{\partial \epsilon_j} \right\rangle \right] \quad (17)$$

This further simplifies for the linear Hamiltonian given by equation ??,

$$\frac{\partial f_i}{\partial \epsilon_j} = -\beta [\langle f_i V_j \rangle - \langle f_i \rangle \langle V_j \rangle] \quad (18)$$

In fact, the higher derivatives can also be obtained because they will always be joint cumulants of f_i and the potential energies conjugate to the model parameters, V_k (e.g. $\frac{\partial f_i}{\partial \epsilon_j \partial \epsilon_k}$ will be a second-order joint cumulant between f_i , V_j , and V_k). If desired a higher order method could be constructed using the second derivative in the Taylor expansion. However since the linear method convergences rather quickly and higher order methods are much more computationally demanding this direction hasn't been pursued. The Jacobian used to

¹Note that even though there may be many more model parameters than observables, correlations between the corresponding potential energies mean that these parameters are not all truly independent. One way to quantify the lack of independence between parameters is by looking at the rank of the correlation matrix $c_{ij} = \langle V_i V_j \rangle - \langle V_i \rangle \langle V_j \rangle$, which has been found to have very low rank in this work (data available upon request).

reproduce experimental $\Delta\Delta G$'s is slightly more involved and so is addressed in the next section.

In order to make the simulation and experimental energy scales comparable we put them in terms of $k_B T$ at their respective temperatures, then we multiply the experimental $\Delta\Delta G$'s by the following ratio $r = \frac{\Delta\Delta G_{sim}}{\Delta\Delta G_{exp}}$ in order to make the averages of the two equal (r is only calculated from the initial homogeneous simulations and fixed thereafter). This can be understood as removing the systematic error and is justified because the coarse-grain model is defined on an arbitrary energy scale. Consequently we reproduce the true heterogeneity in the data, which is the deviation from the mean.

Modeling hydrophobic truncations as deletion of native contacts.

In simulation we model a mutation by perturbing the potential energy by ΔH_i for the i -th mutation,

$$H'_i = H + \Delta H_i \quad (19)$$

Naturally ΔH_i is calculated by subtracting some fraction of the contact energy from the mutated residue.

$$\Delta H_i = - \sum_j w_j^i \epsilon_j V_j \quad (20)$$

where the weight w_j^i is calculated as the average fraction of heavy atom contacts lost when a mutation in the experimental structure. The purpose of the weight w_j^i is to allow for different mutations at the same site. Mutations should perturb the energy proportional to the fraction of contacts they delete (e.g. L30V versus L30A). Then the free energy change resulting from the perturbation ΔH_i can be calculated using Zwanzig's relation [18],

$$\Delta G_i^X = -k_B T \ln \langle e^{-\beta \Delta H_i} \rangle_X \quad (21)$$

Where $X \in U, TS, N$ indicates the state we are perturbing, the unfolded (U), transition state (TS), or native state (N), respectively. The free energy profile along a reaction coordinate Q , e.g. the number of native contacts, is used to classify structures into states U, TS, N . The boundaries of the states are taken as $\frac{1}{3}k_B T$ from the corresponding minimum (for U,N states) or maximum (for TS). Q has been shown to be an adequate reaction coordinate for folding in structure-based models [19].

Differentiating this expression with respect to our model parameter yields the difference of perturbed and unperturbed averages,

$$\frac{\partial \Delta G_i^X}{\partial \epsilon_j} = -\beta \left[\left\langle \frac{\partial H'_i}{\partial \epsilon_j} \right\rangle'_X - \left\langle \frac{\partial H}{\partial \epsilon_j} \right\rangle_X \right] \quad (22)$$

where the primed average is taken with respect to Boltzman weights of H'_i , and can also be written as,

$$\frac{\partial \Delta G_i^X}{\partial \epsilon_j} = -\beta \left[\frac{\left\langle e^{-\beta \Delta H_i} \frac{\partial (H + \Delta H_i)}{\partial \epsilon_j} \right\rangle_X}{\langle e^{-\beta \Delta H_i} \rangle_X} - \left\langle \frac{\partial H}{\partial \epsilon_j} \right\rangle_X \right] \quad (23)$$

Given our linear Hamiltonian from equation ?? and the perturbation in 20 this simplifies to,

$$\frac{\partial \Delta G_i^X}{\partial \epsilon_j} = -\beta \left[\frac{\langle e^{-\beta \Delta H_i} (1 - w_j^i) V_j \rangle_X}{\langle e^{-\beta \Delta H_i} \rangle_X} - \langle V_j \rangle_X \right] \quad (24)$$

Finally, the Jacobian for the $\Delta \Delta G$'s is found from combining these expressions for the individual states as follows,

$$\frac{\partial \Delta \Delta G_i^\ddagger}{\partial \epsilon_j} = \frac{\partial \Delta G_i^{TS}}{\partial \epsilon_j} - \frac{\partial \Delta G_i^U}{\partial \epsilon_j} \quad (25)$$

$$\frac{\partial \Delta \Delta G_i^o}{\partial \epsilon_j} = \frac{\partial \Delta G_i^N}{\partial \epsilon_j} - \frac{\partial \Delta G_i^U}{\partial \epsilon_j} \quad (26)$$

References

- [1] C. Clementi, P. a. Jennings, J. N. Onuchic, How native-state topology affects the folding of dihydrofolate reductase and interleukin-1beta., *Proc. Natl. Acad. Sci.* 97 (11) (2000) 5871–6. doi:10.1073/pnas.100547897.
URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=18526&tool=pmcentrez&rend>
- [2] S. Matysiak, C. Clementi, Optimal combination of theory and experiment for the characterization of the protein folding landscape of S6: how far can a minimalist model go?, *J. Mol. Biol.* 343 (1) (2004) 235–48. doi:10.1016/j.jmb.2004.08.006.
URL <http://www.ncbi.nlm.nih.gov/pubmed/15381433>
- [3] S. Matysiak, C. Clementi, Minimalist protein model as a diagnostic tool for misfolding and aggregation., *J. Mol. Biol.* 363 (1) (2006) 297–308. doi:10.1016/j.jmb.2006.07.088.
URL <http://www.ncbi.nlm.nih.gov/pubmed/16959265>
- [4] S. Cho, Y. Levy, P. Wolynes, Quantitative criteria for native energetic heterogeneity influences in the prediction of protein folding kinetics, *Proc. Natl. Acad. Sci.*
URL <http://www.pnas.org/content/106/2/434.short>
- [5] S. S. Plotkin, J. N. Onuchic, Structural and energetic heterogeneity in protein folding. I. Theory, *J. Chem. Phys.* 116 (12) (2002) 5263. doi:10.1063/1.1449866.
URL <http://link.aip.org/link/JCPSA6/v116/i12/p5263/s1&Agg=doi>

- [6] B. Öztóp, M. Ejtehadi, S. Plotkin, Protein Folding Rates Correlate with Heterogeneity of Folding Mechanism, *Phys. Rev. Lett.* 93 (20) (2004) 208105. doi:10.1103/PhysRevLett.93.208105.
URL <http://link.aps.org/doi/10.1103/PhysRevLett.93.208105>
- [7] J. J. Portman, S. Takada, P. G. Wolynes, Microscopic theory of protein folding rates. I. Fine structure of the free energy profile and folding routes from a variational approach, *J. Chem. Phys.* 114 (11) (2001) 5069. doi:10.1063/1.1334662.
URL <http://link.aip.org/link/JCPSA6/v114/i11/p5069/s1&Agg=doi>
- [8] J. J. Portman, S. Takada, P. G. Wolynes, Microscopic theory of protein folding rates. II. Local reaction coordinates and chain dynamics, *J. Chem. Phys.* 114 (11) (2001) 5082. doi:10.1063/1.1334663.
URL <http://link.aip.org/link/JCPSA6/v114/i11/p5082/s1&Agg=doi>
- [9] S. Plotkin, P. Wolynes, Non-Markovian Configurational Diffusion and Reaction Coordinates for Protein Folding, *Phys. Rev. Lett.* 80 (22) (1998) 5015–5018. doi:10.1103/PhysRevLett.80.5015.
URL <http://link.aps.org/doi/10.1103/PhysRevLett.80.5015>
- [10] S. Plotkin, Speeding protein folding beyond the G model: How a little frustration sometimes helps, *Proteins Struct. Funct. Bioinforma.* 345 (2001) 337–345. doi:10.1002/prot.0000.
URL <http://onlinelibrary.wiley.com/doi/10.1002/prot.1154/full>
- [11] C. Clementi, S. Plotkin, The effects of nonnative interactions on protein folding rates: theory and simulation, *Protein Sci.* (2004) 1750–1766doi:10.1110/ps.03580104.Theoretical.
URL <http://onlinelibrary.wiley.com/doi/10.1110/ps.03580104/full>
- [12] P. Wolynes, As simple as can be?, *Nat. Struct. Mol. Biol.*
URL <http://www.nature.com/nsmb/journal/v4/n11/abs/nsb1197-871.html>
- [13] J. Chen, J. D. Bryngelson, D. Thirumalai, Estimations of the size of nucleation regions in globular proteins., *J. Phys. Chem. B* 112 (50) (2008) 16115–20. doi:10.1021/jp806161k.
URL <http://www.ncbi.nlm.nih.gov/pubmed/19367923>
- [14] E. I. Shakhnovich, Folding nucleus : specific or multiple ? Insights from lattice models and experiments In this commentary , I compare and discuss different (1998) 108–111.
- [15] J. K. Noel, P. C. Whitford, J. N. Onuchic, The shadow map: a general contact definition for capturing the dynamics of biomolecular folding and function., *J. Phys. Chem. B* 116 (29) (2012) 8692–702. doi:10.1021/jp300852d.
URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3406251&tool=pmcentrez&rend=full>

- [16] J. K. Noel, P. C. Whitford, K. Y. Sanbonmatsu, J. N. Onuchic, SMOG@ctbp: simplified deployment of structure-based models in GRO-MACS., *Nucleic Acids Res.* 38 (Web Server issue) (2010) W657–61. doi:10.1093/nar/gkq498.
URL [http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2896113&tool=pmcentrez&re](http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2896113&tool=pmcentrez&rendition=full)
- [17] D. Marquardt, An algorithm for least-squares estimation of nonlinear parameters, *J. Soc. Ind. Appl. ...* 11 (2).
URL <http://epubs.siam.org/doi/pdf/10.1137/0111030>
- [18] R. W. Zwanzig, High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases, *J. Chem. Phys.* 22 (8) (1954) 1420. doi:10.1063/1.1740409.
URL <http://link.aip.org/link/JCPSA6/v22/i8/p1420/s1&Agg=doi>
- [19] S. S. Cho, Y. Levy, P. G. Wolynes, P versus Q: structural reaction coordinates capture protein folding on smooth landscapes., *Proc. Natl. Acad. Sci.* 103 (3) (2006) 586–91. doi:10.1073/pnas.0509768103.
URL [http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1334664&tool=pmcentrez&re](http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1334664&tool=pmcentrez&rendition=full)