



formed protein–DNA complex (3  $\mu$ l) was mixed with 3  $\mu$ l of reservoir solution containing 15% (w/v) PEG 4000, 100 mM 2-(N-Morpholino)ethanesulfonic acid (MES), pH 6.0, 100 mM  $\text{NH}_4\text{H}_2\text{PO}_4$  and 15% (v/v) ethyleneglycol. Crystals grew in 6–9 d as needles of  $0.05 \times 0.05 \times 1$  mm<sup>3</sup>. Crystal parameters are indicated in Table 1. Crystals were harvested and directly flash-frozen in liquid nitrogen. Data from two crystals were collected at 110 K at beamline X11 at DESY Hamburg on a CCD detector (MarResearch). Data were integrated with DENZO/SCALEPACK<sup>20</sup>.

**Structure determination and refinement.** The structure of the  $Z\alpha_{\text{DLM}}$ –Z-DNA complex was solved by molecular replacement using the program EPMR<sup>21</sup> with the  $Z\alpha_{\text{ADAR}}$ –Z-DNA complex (chains C and F; PDB code 1QBJ) as the search model (with Ala residues substituted for all solvent-exposed and additional nonidentical protein residues). Molecular replacement produced one clear solution per asymmetric unit. The correctness of this solution was judged by a quick drop of both R-factor and  $R_{\text{free}}$  after an initial round of rigid body and positional refinement, and by the formation of Watson–Crick base-paired dsDNA by crystallographic symmetry operations. Refinement was initially performed with CNS<sup>14</sup> using standard protocols. Model building and adjustments were done with vUSette zc (M.A. Rould, unpublished program). The model was inspected manually with  $\sigma_A$ -weighted  $2F_o - F_c$  and  $F_o - F_c$  maps, and progress in the model refinement was gauged by the decrease in the  $R_{\text{free}}$ . After CNS refinement converged at R-factor /  $R_{\text{free}}$  25.2 / 27.9%, further refinement was carried out with the program Refmac5 (ref. 22). TLS refinement<sup>23,24</sup> improved both R-factor and  $R_{\text{free}}$  significantly. The selection of an optimal set of TLS groups was crucial. Each base-paired nucleotide was divided in three segments – the ribose, the phosphorus atom plus both nonesterified oxygens and the base.  $Z\alpha_{\text{DLM}}$  was treated as one TLS group. Refinement statistics are in Table 1.

**Coordinates.** Atomic coordinates have been deposited in the Protein Data Bank (accession code 1J75)

#### Acknowledgments

We thank S. Rothenburg (Universitätsklinikum Hamburg-Eppendorf) for providing mouse cDNA and for carefully reading the manuscript. Helpful discussions with U. Müller, Y.A. Muller (MDC) and M.A. Rould (UVM) are gratefully acknowledged.

Correspondence should be addressed to T.S. *email:* [Thomas.Schwartz@mail.rockefeller.edu](mailto:Thomas.Schwartz@mail.rockefeller.edu)

Received 22 May, 2001; accepted 16 July, 2001.

1. Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L. & Sweet, R.M. *Nature* **362**, 219–223 (1993).
2. Clark, K.L., Halay, E.D., Lai, E. & Burley, S.K. *Nature* **364**, 412–420 (1993).
3. Gajiwala, K.S. & Burley, S.K. *Curr. Opin. Struct. Biol.* **10**, 110–116 (2000).
4. Schwartz, T., Rould, M.A., Lowenhaupt, K., Herbert, A. & Rich, A. *Science* **284**, 1841–1845 (1999).
5. Maas, S., Melcher, T. & Seeburg, P.H. *Curr. Opin. Cell Biol.* **9**, 343–349 (1997).
6. Herbert, A. & Rich, A. *J. Biol. Chem.* **271**, 11595–11598 (1996).
7. Fu, Y. *et al. Gene* **240**, 157–163 (1999).
8. Herbert, A. *et al. Proc. Natl. Acad. Sci. USA* **94**, 8421–8426 (1997).
9. Schwartz, T. *et al. J. Biol. Chem.* **274**, 2899–2906 (1999).
10. Schade, M. *et al. FEBS Lett.* **458**, 27–31 (1999).
11. Schade, M., Turner, C.J., Lowenhaupt, K., Rich, A. & Herbert, A. *EMBO J.* **18**, 470–479 (1999).
12. Takagi, T. *et al. J. Chem. Soc.* **2**, 1015–1018 (1987).
13. Brandl, M., Weiss, M.S., Jabs, A., Sühnel, J. & Hilgenfeld, R. *J. Mol. Biol.* **307**, 357–377 (2001).
14. Brunger, A.T. *et al. Acta Crystallogr. D* **54**, 905–921 (1998).
15. Wittig, B., Wolff, S., Dorbic, T., Vahrson, W. & Rich, A. *EMBO J.* **11**, 4653–4663 (1992).
16. Kawakubo, K. & Samuel, C.E. *Gene* **258**, 165–172 (2000).
17. Brandt, T.A. & Jacobs, B.L. *J. Virol.* **75**, 850–856 (2001).
18. Behlke, J., Ristau, O. & Schönfeld, H.-J. *Biochemistry* **36**, 5149–5156 (1997).
19. Schwartz, T. *et al. Acta Crystallogr. D* **55**, 1362–1364 (1999).
20. Otwinowski, Z. & Minor, W. *Methods Enzymol.* **276**, 307–326 (1997).
21. Kissinger, C.R., Gehlhaar, D.K. & Fogel, D.B. *Acta Crystallogr. D* **55**, 484–491 (1999).
22. Collaborative Computational Project, Number 4. *Acta Crystallogr. D* **50**, 760–763 (1994).
23. Winn, M.D., Isupov, M.N. & Murshudov, G.N. *Acta Crystallogr. D* **57**, 122–133 (2001).
24. Merritt, E.A. *Acta Crystallogr. D* **55**, 1997–2004 (1999).

## Transition states and the meaning of $\Phi$ -values in protein folding kinetics

S. Banu Ozkan<sup>1,2</sup>, Ivet Bahar<sup>3</sup> and Ken A. Dill<sup>2</sup>

<sup>1</sup>Department of Chemical Engineering, Polymer Research Center, Bogazici University, Bebek 80815, Istanbul, Turkey. <sup>2</sup>Department of Pharmaceutical Chemistry, University of California San Francisco, California 94143-1204, USA. <sup>3</sup>Center for Computational Biology & Bioinformatics, and Department of Molecular Genetics & Biochemistry, School of Medicine, University of Pittsburgh, Pennsylvania 15213, USA.

**What is the mechanism of two-state protein folding? The rate-limiting step is typically explored through a  $\Phi$ -value, which is the mutation-induced change in the transition state free energy divided by the change in the equilibrium free energy of folding.  $\Phi$ -values ranging from 0 to 1 have been interpreted as meaning the transition state is denatured-like (0), native-like (1) or in-between. But there is no classical interpretation for the experimental  $\Phi$ -values that are negative or  $>1$ . Using a rigorous method to identify transition states via an exact lattice model, we find that nonclassical  $\Phi$ -values can arise from parallel microscopic flow processes, such as those in funnel-shaped energy landscapes.  $\Phi < 0$  results when a mutation destabilizes a slow flow channel,**

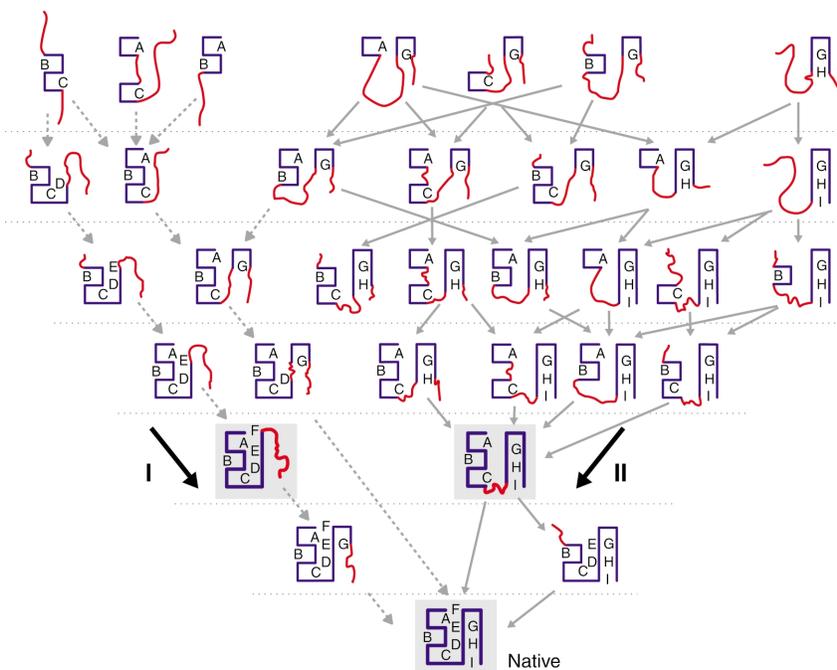
**causing a backflow into a faster flow channel.  $\Phi > 1$  implies the reverse: a backflow from a fast channel into a slow one. Using a 'landscape mapping' method, we find that  $\Phi$  correlates with the acceleration/deceleration of folding induced by mutations, rather than with the degree of nativeness of the transition state.**

What is the folding mechanism of fast folding proteins? Many proteins fold and unfold rapidly with single-exponential (two-state) kinetics<sup>1–5</sup>. Rate processes that involve a single-exponential relaxation in both the forward and reverse directions have been traditionally interpreted in terms of rate limiting steps, called transition states (TS). This raises the question of what are the transition states for protein folding<sup>5–8</sup>. To determine them, classical rate theory suggests searching for structures associated with an energy or entropy barrier along a reaction coordinate. However, protein folding may be so fundamentally different from classical rate processes that there is no single microscopic reaction coordinate — that is, an ordered list of structures — that every chain follows<sup>9</sup>, and there may not be identifiable barriers of the traditional type, because these landscapes may be funnel-shaped. The questions of folding mechanism then become (i) what chain conformations are responsible for the observable relaxation rate, (ii) how do they cause this rate, and (iii) what is the experimental evidence?

The main experimental methodology for addressing this question has been mutational studies of folding rates and equilibrium constants<sup>10–13</sup>. This methodology, developed as  $\Phi$ -value analysis by Alan Fersht and his colleagues<sup>13–15</sup>, has been widely applied to many different proteins<sup>2,16–18</sup>. In  $\Phi$ -value analysis, a particular

## letters

**Fig. 1** Folding kinetics of a lattice model chain. The top row shows the high energy, high entropy conformations at the top of the funnel. Lower rows represent lower energy, lower entropy conformations deeper on the landscape. The bottom is the native (N) conformation. Channel I is the fast route, and channel II is slower. The boxed nonnative structures are the rate limiting steps for these channels, found from landscape mapping (see text).



amino acid is mutated. If the mutation changes the stability of the protein by an amount  $\Delta\Delta G$  (where  $\Delta G = G_N - G_D$ , N and D refer to native and denatured states, respectively, and the first  $\Delta$  refers to the change in stability that arises from the mutation), and if the mutation affects the folding barrier by an amount  $\Delta\Delta G^\ddagger$ , then the  $\Phi$ -value is:

$$\Phi = \Delta\Delta G^\ddagger / \Delta\Delta G \quad (1)$$

**$\Phi$  as a 'kinetic ruler' of nativeness of the transition state**  
Interpretations of  $\Phi$ -values are invariably based on Brønsted theory and the Hammond postulate of classical chemical reactions<sup>19</sup>. For protein folding,  $\Phi$  is usually regarded as a sort of 'kinetic ruler' indicating how native-like the transition state conformation of a monomer is; "a  $\Phi$ -value of 0 indicates that the structure is unfolded at the site of mutation as much as in the denatured state, and a  $\Phi$ -value of 1.0 means that the structure is folded at the site of mutation as much as in the native state"<sup>19</sup>. Fractional values of  $\Phi$  are taken to signify that structure is present but weakened<sup>19</sup>.

Typically, the nativeness of the transition state is defined by the coordinate  $\xi$ , where  $\xi = 0$  refers to denatured conformations, and  $\xi = 1$  to the native conformation. The kinetic ruler hypothesis means that  $\Phi$  is monotonically related to  $\xi$ ; however, this does not necessarily imply that the relationship is linear<sup>19</sup>. For example, if a kinetic ruler was linear, then  $\Phi_j = 0.75$  would mean that  $\xi = 0.75$  — that is, that amino acid *j* is 75% native when the protein is in its transition state.

The kinetic ruler hypothesis, which is deeply rooted in classical rate theories, has an important implication. For any one-dimensional quantity, such as a nativeness coordinate, there is the notion of 'betweenness'. That is, for some particular point  $\xi$  along the coordinate between D and N ( $\xi_D < \xi < \xi_N$ ), the conformations  $\xi$  are more native than D and more denatured than N. All protein conformations can be lined up along a single axis of nativeness, and somewhere along this axis will be transition state conformations  $\xi$  that are responsible for the single exponential time constant. It follows that the transition state must be between

the native and denatured states. That is, the kinetic ruler assumption implies that there cannot be nonclassical  $\Phi$  values:  $\Phi < 0$  would have the nonsensical meaning that the transition state is more denatured than the denatured state, and  $\Phi > 1$  would mean that the transition state is more native than the native state itself. Therefore, when nonclassical values are observed in experiments, they are typically dismissed as artifacts. However, 10–20% of the hundreds of measured  $\Phi$ -values for protein folding are outside this classical range<sup>20–23</sup>. In addition,  $\Phi$ -values as large as 8 have been measured (T. Logan, pers. comm.). Negative  $\Phi$ -values have also been observed in computer simulations<sup>24–27</sup>, where they have been interpreted in terms of nonnative contacts.

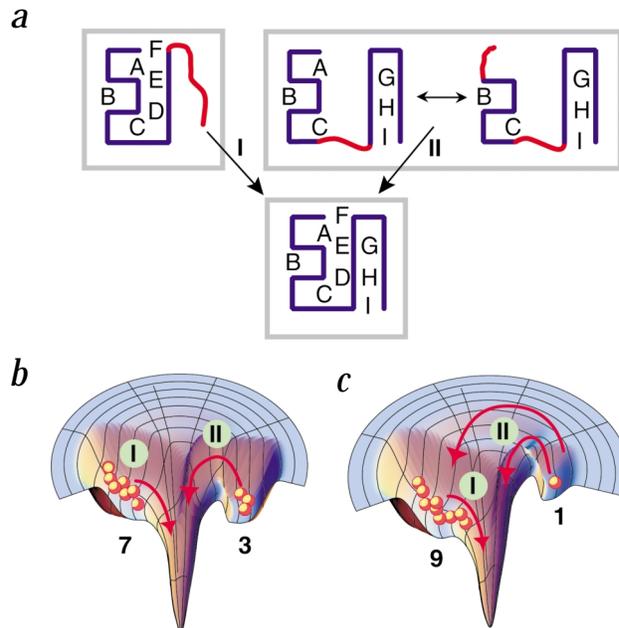
Here we show, using an exact lattice model, that there is a broader physical basis for nonclassical values,  $\Phi < 0$  and  $\Phi > 1$ . We find that although classical  $\Phi$ -values are restricted to systems having a single reaction coordinate, nonclassical  $\Phi$ -values can arise from parallel, coupled flows — for example, in funnel-shaped energy landscapes<sup>9,28</sup>.

### Model

Our aim is to explore a model of folding that meets certain minimal requirements. First, having a model that has single exponential folding is essential for studying two-state kinetics. Second, we wanted a model that, unlike classical mass-action models, specifically examines individual chain conformations and sequence-structure relationships and whose folding is directed toward a unique native state by an energy function — that is, a statistical mechanical model. Only in this manner can we draw conclusions about folding at the level of individual chain conformations. Third, we wanted a model of sufficient simplicity to be able to obtain the kinetics completely and rigorously, without *ad hoc* assumptions about the transition state.

Currently, the only models that meet these requirements are short chain lattice Go models. Go models are the principal theoretical models for exploring microscopic steps in two-state protein folding kinetics<sup>29</sup>. In this type of model, each native contact has an attractive potential  $\epsilon$ . We studied the 16-mer on two-dimensional square lattices (Fig. 1). The denatured state is all

**Fig. 2** A negative  $\Phi$ -value is defined as an acceleration of folding following a destabilizing mutation. Destabilizing (increasing the energy) of a step in the slow flow channel (II) rechannels the flow into the faster channel (I). **a**, Rate limiting conformational steps along channels I and II. **b**, For wild type protein, greater population and faster flow is along route I. **c**, Mutant protein has a destabilized energy well (right side), forcing more population of channel I. Hypothetical populations of 7:3 shifted to 9:1 illustrate a  $\Phi$ -value of approximately  $-0.30$ .



possible self-avoiding walks except for the native conformation. Self-avoiding walks are configurations in which no two monomers occupy the same lattice site. The native state has the lowest possible energy, with the nine native contacts A through I (Fig. 1). Each rate constant for the transition from the  $j^{\text{th}}$  conformation to the  $i^{\text{th}}$  is:

$$A_{ij} = \exp(-\Delta G_{ij}/RT) \quad (2)$$

where  $\Delta G_{ij} = \exp(-\nu <(\Delta r_{ij})^2>^{1/2}) \exp(-(q_i - q_j) \varepsilon H(q_i, q_j) / RT)$ ,  $q_i$  is the number of native contacts in conformation  $i$ ,  $<(\Delta r_{ij})^2>^{1/2}$  is the root mean square (r.m.s.) deviation in the residue coordinates of the two conformations evaluated after their optimal superimposition,  $\nu$  is a scaling parameter penalizing/favoring the transitions between dissimilar/similar conformations, and  $H(q_i, q_j)$  is the Heavyside step function, equal to 1 for  $q_i > q_j$  and zero otherwise. Here, we use  $\varepsilon = -2.5 RT$  and  $\nu = 1$ . Bonds have unit length, and we use  $\varepsilon' = -1.75 RT$  for mutations of native contacts.

#### Treatment of kinetics

The time evolution of the conformational ensemble is controlled by the master equation:

$$d\mathbf{P}(t) / dt = \mathbf{A} \mathbf{P}(t) \quad (3)$$

where  $\mathbf{P}(t)$  is the  $N$ -dimensional vector of the instantaneous probabilities of the  $N$  conformations, and  $\mathbf{A}$  is the  $N \times N$  transition matrix. We find the instantaneous probabilities of the individual conformations by simultaneous solution of these  $N$  equations, which gives:

$$\mathbf{P}(t) = \exp(-\mathbf{A}t) \mathbf{P}(0) = \exp(-\mathbf{\Lambda}t) \mathbf{B}^{-1} \mathbf{P}(0) \quad (4)$$

where  $\mathbf{P}(0)$  is the vector of initial probabilities,  $\mathbf{\Lambda}$  is the diagonal matrix of the eigenvalues of  $\mathbf{A}$  and  $\mathbf{B}$  is the matrix of eigenvectors. We find the exact kinetics of the model without resorting to assumptions or to limited sampling, such as Monte Carlo. Crucial to our present study is the application of a general, rigorous and unambiguous method to identify the rate limiting steps<sup>30</sup>. The lowest nonzero eigenvalue represents the frequency of the slowest transition mode that contributes to the folding process. The eigenvector that corresponds to this eigenvalue (relaxation time) describes the populations responsible for the single exponential, and hence these populations are rigorously what should be called transition state conformations.

#### Folding routes of model proteins.

In the resultant folding routes for a model protein, the native structure emerges systematically, with a single relaxation time, through multiple coupled microscopic routes (Fig. 1). In terms of the individual chain conformations, folding does not follow a single sequential pathway. The native structure of the protein has the two-dimensional lattice model equivalent of a two-stranded antiparallel  $\beta$ -sheet adjacent to a helix. The folding process

involves two main routes. Along the dominant folding channel I, the fastest folding with greatest flux, the rate limiting step has the native helix and the first but not the second strand. Along channel II, which is slower than I and has smaller flux, the rate limiting structure has a partial or complete helix and the full sheet, but the two secondary structures are not yet assembled together. Channel II resembles the diffusion-collision model of Karplus and Weaver<sup>31</sup>. An experimental example of parallel flow channels that involves a rapid helix formation and a slower  $\beta$ -sheet formation is lysozyme<sup>32</sup>.

#### Physical basis for nonclassical $\Phi$ -values

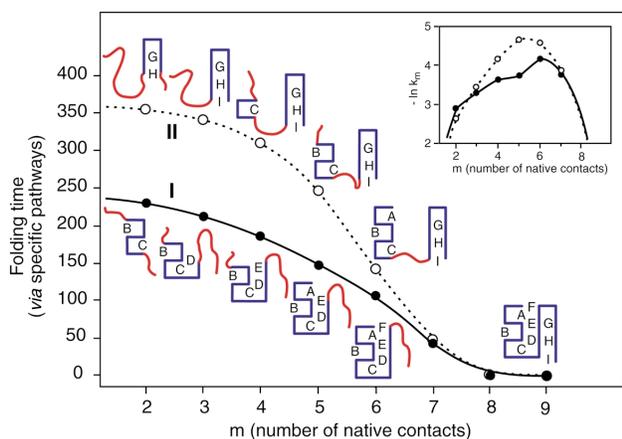
By definition, a negative  $\Phi$ -value would mean that a mutation that destabilizes the native state speeds up folding. In the model, destabilizing contact G (reducing the attractive potential by 30%) gives  $\Phi = -0.36$ . The mutation destabilizes the  $\beta$ -sheet, slows the flow through channel II and causes increased helical populations in channel I; the overall folding flow rate *via* channel I increases (Fig 2). All the  $\beta$ -sheet contacts have negative  $\Phi$ -values (Table 1). Synergistic effects such as this have been observed in other simulations<sup>27,33</sup>.

The explanation for  $\Phi > 1$  is the reverse. Destabilizing a contact in the fast channel redirects the flow through the slow channel. When the contact D is weakened by 30%,  $\Phi = 0.99$ . When D is weakened by 50%,  $\Phi = 1.38$ . We conclude that nonclassical  $\Phi$ -val-

**Table 1**  $\Phi$ -values resulting from 30% destabilization of native contacts

| Type of native contact | $\Phi$ -value |
|------------------------|---------------|
| A                      | 0.012         |
| B                      | 0.096         |
| C                      | 0.251         |
| D                      | 0.990         |
| E                      | 0.093         |
| F                      | 0.035         |
| G                      | -0.357        |
| H                      | -0.296        |
| I                      | -0.085        |

# letters



**Fig. 3** Landscape mapping. A chain is put into a specific conformation with  $m$  native contacts before time  $t = 0$ . Folding is initiated at  $t = 0$ . The passage time required for each single step transition between macroconformations is computed. The y-axis shows the cumulative time to reach the native state, which is the sum of times for the individual steps. Preforming the helix (I) speeds up folding more than preforming the  $\beta$ -sheet (II). Inset: The logarithms of the individual step rates  $k_m$  map the relative energies on the landscape. For example, the barrier along the  $\beta$ -sheet route is higher than for the helix route.

ues can identify parallel routes in folding pathways and are not experimental artifacts. Because models that are based on assuming single reaction coordinates and sequential paths cannot explain nonclassical  $\Phi$ -values, we believe that experimental observations of nonclassical  $\Phi$ -values give evidence of protein folding steps that are not sequential.

### Is $\Phi$ a kinetic ruler for folding events?

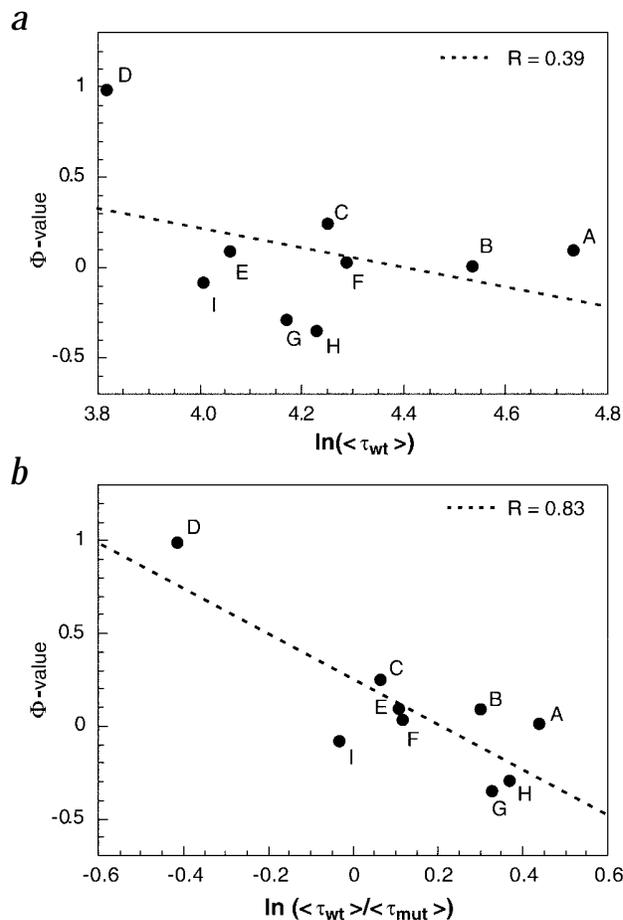
What is the physical meaning of a  $\Phi$ -value? For folding landscapes that have multiple microscopic reaction coordinates, such as funnels, there is no single quantity  $\xi$  that applies at the same time to all microscopic trajectories. Thus, there is no kinetic ruler, and no obvious structural property that will be related monotonically to the measurable property  $\Phi$ . Therefore, we devised a general computational experiment, called landscape mapping, to find a physical explanation of the  $\Phi$ -values on our energy landscapes. If a classical chemical reaction could be initiated from any specific point along its reaction coordinate, and the time required to reach the product from that point is measured, an unambiguous measure of reaction progress would be obtained. For example, if molecules are forced to have reactant-like structures (before the transition state) and are then allowed to proceed to react from that point, the time to product will be long. However, if molecules are forced into conformations that are highly product-like (after the transition state), the time to product will be short. By fixing molecular structures into specific conformations, starting the reaction and then measuring the time to product, you could map out the kinetic distances between conformations. We do this here for our folding model.

Terms that are pertinent to our model are defined below. A 'conformation' (or 'microconformation') is a single arrangement of chain monomers in space. For the 16-mer on a square lattice, the total number of conformations is 802,075. A 'micro-route' is one particular trajectory between two conformations,  $i$  and  $j$ , with single microscopic rate constant  $k_{ij}$ . A 'macroconformation' is the ensemble of all conformations having a particular specified set of contacts. We have a total of 267 macroconformations, including the native state. A 'macroroute' between two macroconformations,  $m$  and  $n$ , is the collection of all micro-

routes from the one macroconformation to the other. The 'passage time' ( $\tau_{mn}$ ) between macroconformations, found from the best fit single exponential, is faster than  $1/k_{ij}$  because the macroroute includes all direct and indirect microroutes between the two macroconformations. Two macroconformations are 'successive' if they differ by one additional contact. For example (GH) and (GHI) along channel II (Fig. 1) are two successive macroconformations.

In landscape mapping, we fix  $m$  native contacts, giving an initial macroconformation. At time  $t = 0$ , folding is initiated from that ensemble. We measure the passage time  $\tau_{mn}$  required to reach the successive macroconformation  $n$  along a particular macroroute. Not surprisingly, the more native-like the starting ensemble, the shorter is the time required to reach the native state (Fig. 3). Because the energy difference between conformations  $m$  and  $n$  on the landscape is the logarithm of passage time  $\tau_{mn}$ , we have  $E_{act, mn} = -RT \ln k_{mn}$ , where the stepwise macroroute rate constant is  $k_{mn} = 1/\tau_{mn}$  (inset of Fig. 3).

**Fig. 4** Correlation of  $\Phi$  with  $\tau_{mut}$  and  $\tau_{wt}$ . **a**,  $\Phi$  correlates only weakly with a kinetic ruler property,  $\langle \tau_{wt} \rangle$ , the average folding time required to reach the native state. **b**,  $\Phi$  correlates more strongly with gatekeeping,  $\langle \tau_{mut} \rangle / \langle \tau_{wt} \rangle$ , the change in folding time due to the mutation.



**Macroroutes have barriers; microroutes do not**

These simulations lead to the following conclusions. First, there is a fundamental difference between microroutes and macroroutes, and a fundamental difference between conformations and macroconformations. Even though the microroutes have no barrier and follow funnel-like landscapes, the macroroutes do have kinetic barriers. This results from a balance of two effects: (i) the velocity of conformational changes (along microroutes) increases monotonically down the landscape, but (ii) the number of microroutes diminishes down the landscape. The product of these two factors leads to bottleneck steps along the macroroutes. The maximum time per step (Fig. 3) identifies the slowest macrosteps, which are those having  $m = 6$  and 5 native contacts along channels I and II, respectively. This barrier results from a property of the landscape, and not from a property of a microroute. The barrier is due to a reduction in the numbers of routes at the bottom of the energy funnel, and not to an energetic problem along any one microroute, in these simulations.

**Folding involves nucleated zipping**

Second, we see a broad heterogeneity of rates if the chain is started from a partially formed conformation. Often the native state is reached faster by starting from a particular set of two native contacts than when started from some other set of six native contacts. The nonuniqueness of the folding nucleus was first established in a model system that shows kinetic partitioning similar to the present model<sup>34</sup>. The heterogeneity results because partially formed conformations are typically committed to a given folding route and must find uphill routes on the energy landscape in order to reach fast downhill folding routes. In contrast, open conformations almost always flow rapidly downhill.

We mapped the full landscape; we tested every possible starting set of native contacts. Our results can be interpreted using two ideas. (i) As noted by many previous investigators<sup>35–39</sup>, there are folding nuclei — that is, certain sets of contacts that provide much greater kinetic accessibility to the native state than others. (ii) The sequences of folding events are zippers<sup>40</sup>: on average, the most local contacts form first, secondary structures form sequentially and nonlocal contacts form later, in a series of minimum conformational entropy-loss steps.

Finally, we conclude that  $\Phi$ -values are not kinetic rulers of the progress toward the native state — for example, the  $\Phi$ -value for contact  $k$  shows little correlation with the average time  $\langle \tau_{wt} \rangle_k$  required for stabilizing the contact  $k$  in the folding of the wild type protein. The correlation coefficient is 0.39 (Fig. 4a). In contrast,  $\Phi$ -values correlate with the change in rates caused by mutations (Fig. 4b). Sites having positive  $\Phi$ -values indicate where destabilizing mutations decelerate folding. Negative  $\Phi$ -values indicate where destabilizing mutations accelerate folding. The absolute value,  $|\Phi|$ , defines 'gatekeeper' contacts — that is, the degree to which a contact controls the flow process

towards the native state. Sites having  $|\Phi| \gg 0$  are gatekeepers; sites with  $\Phi \sim 0$  have little flow control. Classical  $\Phi$ -values result when opening a gate controls a forward flow. Nonclassical  $\Phi$ -values result when opening a gate controls a backflow, resulting in a redirection of the folding flow through some alternative macroroutes. In principle, the landscape mapping strategy used here is also feasible for experimental determinations of the shapes of folding energy landscapes.

**Acknowledgments**

We thank J. Schonbrun, D. Thirumalai, A. Fersht, D. Goldenberg and A. Robertson for helpful comments and the NIH for grant support. We would also like to acknowledge a TUBITAK fellowship to S.B.O. We are grateful to J. Schreurs for the preparation of the figures.

Correspondence should be addressed to K.A.D. *email: dill@maxwell.ucsf.edu*

Received 7 March, 2001; accepted 18 July, 2001.

- Jackson S.E. & Fersht A.R. *Biochemistry* **30**, 10428–10435 (1991).
- Millia, M. & Sauer, R.T. *Biochemistry* **33**, 1125–1133 (1994).
- Huang, G.S. & Oas, T.S. *Biochemistry* **34**, 3884–3892 (1995).
- Schindler, T., Herrler, M., Marahiel, M.A. & Schmid, F.X. *Nature Struct. Biol.* **2**, 663–673 (1995).
- Gujarro, J.I., Morton, C.J., Plaxco, K.W., Campbell, I.D. & Dobson, C.M. *J. Mol. Biol.* **276**, 657–667 (1998).
- Matthews, C.R. *Annu. Rev. Biochem.* **62**, 653–683 (1993).
- Laurents, D.V. & Baldwin, R.L. *Biophys. J.* **75**, 428–434 (1998).
- Englander, S.W. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 213–238 (2000).
- Dill, K.A. & Chan, H.S. *Nature Struct. Biol.* **4**, 10–19 (1997).
- Fersht, A.R., Leatherbarrow, R.J. & Wells T.N.C. *Nature* **322**, 284–286 (1986).
- Beasty, A.M. *et al. Biochemistry* **25**, 2965–2974 (1986).
- Goldenberg, D.P., Frieden, R.W., Haack, J.A. & Morrison, T.B. *Nature* **338**, 127–132 (1989).
- Fersht, A.R., Leatherbarrow, R.J. & Wells T.N.C. *Biochemistry* **26**, 6030–6038 (1987).
- Matouschek, A., Kellis, J.T., Serrano, L. & Fersht, A.R. *Nature* **340**, 122–126 (1989).
- Matouschek, A. & Fersht, A.R. *Methods Enzymol.* **202**, 81–112 (1991).
- Alexander, P., Orban, J. & Bryan, P. *Biochemistry* **31**, 7243–7248 (1992).
- Grantcharova, V.P., Riddle, D.S., Santiago, J.V. & Baker, D. *Nature Struct. Biol.* **5**, 714–720 (1998).
- Martinez, J.C., Pisabarro, M.T. & Serrano, L. *Nature Struct. Biol.* **5**, 721–729 (1998).
- Fersht, A.R. *Structure and mechanism in protein science* (W. H. Freeman, New York: 1999).
- Matouschek, A., Serrano, L. & Fersht, A.R. *J. Mol. Biol.* **224**, 819–835 (1992).
- Gay, G., Ruiz-Sans, J., Davis, B. & Fersht, A.R. *Proc. Natl. Acad. Sci. USA* **91**, 10943–10946 (1994).
- Nolting, B. & Andert, K. *Proteins* **41**, 288–298 (2000).
- Goldenberg, D.P. *Nature Struct. Biol.* **6**, 987–990 (1999).
- Daggett, V., Li, A., Itzhaki, S.L., Otzen, D.E. & Fersht, A.R. *J. Mol. Biol.* **257**, 430–440 (1996).
- Lazaridis, T. & Karplus, M. *Science* **278**, 1928–1931 (1997).
- Li L., Mirny A.L. & Shakhonivch, E.I. *Nature Struct. Biol.* **7**, 336–342 (2000).
- Shea, J., Onuchic, J.N. & Brooks, C.L. *J. Chem. Phys.* **113**, 7663–7671 (2000).
- Bryngelson, J.D., Onuchic, J.N., Socci, N.D. & Wolynes, P.G. *Proteins* **21**, 167–195 (1995).
- Ueda, Y., Taketomi, H. & Go, N. *Int. J. Peptide. Res.* **7**, 445–459 (1975).
- Hoang, T.X. & Cieplak, M. *J. Chem. Phys.* **112**, 6851–6862 (2000).
- Karplus, M. & Weaver, D.L. *Nature* **260**, 404–406 (1976).
- Matagne, A., Radford, S.E. & Dobson, C.M. *J. Mol. Biol.* **267**, 1068–1074 (1997).
- Ladurner, A.G., Itzhaki, S.L., Daggett, V. & Fersht A.R. *Proc. Natl. Acad. Sci. USA* **95**, 8473–8478 (1998).
- Guo, Z. & Thirumalai, D. *Folding Des.* **2**, 377–391 (1997).
- Pande, V.S., Grosberg, A.Y., Rokhsar, D. & Tanaka, T. *Curr. Opin. Struct. Biol.* **8**, 68–79 (1998).
- Klimov, D.K. & Thirumalai, D. *J. Mol. Biol.* **282**, 471–492 (1998).
- Dokholayan, N.V., Buldrey, S.V., Stanley, H.E. & Shakhonivch, E.I. *J. Mol. Biol.* **296**, 1183–1187 (2000).
- Fersht, A.R. *Curr. Opin. Struct. Biol.* **7**, 3–9 (1997).
- Galzitskaya, O.V. & Finkelstein, A.V. *Proc. Natl. Acad. Sci. USA* **99**, 11299–11304 (1999).
- Dill, K.A., Fiebig, K.M. & Chan H.S. *Proc. Natl. Acad. Sci. USA* **90**, 1942–1946 (1993).