ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/7170150

Diffusion-collision of foldons elucidates the kinetic effects of point mutations and suggests control strategies of the folding process of helical proteins

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · JULY 2006

Impact Factor: 2.63 · DOI: 10.1002/prot.20980 · Source: PubMed

CITATIONS
2

READS
21

2 AUTHORS:

Emidio Capriotti
Heinrich-Heine-Universität Düsseldorf
59 PUBLICATIONS **1,891** CITATIONS

SEE PROFILE

Mario Compiani
University of Camerino
35 PUBLICATIONS **341** CITATIONS

SEE PROFILE

Available from: Mario Compiani
Retrieved on: 11 January 2016

PROTEINS: Structure, Function, and Bioinformatics 64:198–209 (2006)

Diffusion-Collision of Foldons Elucidates the Kinetic Effects of Point Mutations and Suggests Control Strategies of the Folding Process of Helical Proteins

Emidio Capriotti[1*] and Mario Compiani[1,2]
[1]*Laboratory of Biocomputing, University of Bologna, Bologna, Italy*
[2]*Department of Chemical Sciences, University of Camerino, Camerino, Italy*

ABSTRACT **In this article we use mutation studies as a benchmark for a minimal model of the folding process of helical proteins. The model ascribes a pivotal role to the collisional dynamics of a few crucial residues (foldons) and predicts the folding rates by exploiting information drawn from the protein sequence. We show that our model rationalizes the effects of point mutations on the kinetics of folding. The folding times of two proteins and their mutants are predicted. Stability and location of foldons have a critical role as the determinants of protein folding. This allows us to elucidate two main mechanisms for the kinetic effects of mutations. First, it turns out that the mutations eliciting the most notable effects alter protein stability through stabilization or destabilization of the foldons. Secondly, the folding rate is affected via a modification of the foldon topology by those mutations that lead to the birth or death of foldons. The few mispredicted folding rates of some mutants hint at the limits of the current version of the folding model proposed in the present article. The performance of our folding model declines in case the mutated residues are subject to strong long-range forces. That foldons are the critical targets of mutation studies has notable implications for design strategies and is of particular interest to address the issue of the kinetic regulation of single proteins in the general context of the overall dynamics of the interactome. Proteins 2006;64:198–209.**
© 2006 Wiley-Liss, Inc.

Key words: determinants of folding; protein folding mechanisms; folding kinetics; mutants; neural networks; prediction of folding rates; long-range intramolecular interactions

INTRODUCTION

Dealing with the inherent complexity of protein folding calls for suitable simplification strategies. The unexpected simplicity of folding[1] has led to minimalist models relying on a supposedly small set of determinants of the process.[1–5] Current views maintain that the folding kinetics of single domain two-state and three-state proteins depend on few gross-grained descriptors of the native state or the sequence. The primary determinant is the topology of the native state, described by the contact order (CO)[2] and related concepts.[6–11] The secondary factor is protein stability.[12,13] Within a different approach, dynamical studies also hinted at the possibility of a substantial reduction of the host of degrees of freedom to a limited set of critical variables that are linked to collective nonlinear excitations driving the slow dynamics of the protein.[14–18] Alternative simplification strategies rely on the notion that a reduced set of fundamental residues can be identified to account for the general thermodynamic[3,19] or kinetic[20] features of the folded protein.

In the case of the folding process of helical proteins, classical theories of helix-coil transitions[21] are available to describe the elementary events of folding. This provides useful tools to dissect the overall process and to connect local properties with global behavior. Taking advantage of these features we suggested that, for helical proteins, essential pieces of information about the kinetics of the folding process can be drawn from a limited set of key regions (foldons), corresponding to the initiation sites (ISs) of folding.[22] Foldons are immersed in stretches of the sequence that exhibit native helical structures. On this basis we can build up an effective reduced description of the folding dynamics that is valid in the full nonequilibrium domain.[23] Foldons are minimally frustrated segments of the sequence where the global interactions and the local propensities for secondary structure minimally conflict.[23] Foldons are crucial for the whole process to the extent that folding can be boiled down to the dynamics of the sole helices containing a foldon (IS helices for brevity). A suitable framework to represent the dynamics of foldons is the foldon diffusion-collision model (henceforth FDC model) that allows reconstruction of the folding dynamics of helical proteins in terms of the collisional dynamics of the IS helices.[23]

Grant sponsors: European Union's VI Framework Programme for the BioSapiens Network of Excellence project, and Italian Ministry for Research through the PRIN 2005 Project

*Correspondence to: Mario Compiani, Department of Chemical Sciences, University of Camerino, Via S. Agostino 1, 62032 Camerino MC, Italy. E-mail: mario.compiani@unicam.it

Received 20 September 2005; Revised 5 January 2006; Accepted 20 January 2006

Published online 11 April 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20980

© 2006 WILEY-LISS, INC.

TABLE I. Structural, Thermodynamic, and Dynamical Data of the λ-Repressor and Its Mutants Used in the Simulations Performed According to the FDC Model

Mutant	Helices	β_1^f	β_3^f	β_4^f	β_5^f	ΔG_{UF}	τ_{exp}	τ_{comp}
Wild type	1, 4, 5 (none)	0.099	[a]	0.050	0.048	−0.86	204 ± 25	213
Basic	1, 3, 4, 5 (3)		0.092			−2.65	12 ± 2	13
M15	1, 3, 4, 5 (1)	0.006				−0.31	100 ± 13	81
M20	1, 3, 4, 5 (1)	0.045				−1.73	17 ± 2	18
M37	1, 3, 4, 5 (2)	0.099	0.106			−1.48	10 ± 3	12
M49	1, 3, 4, 5 (3)		0.018			−2.11	17 ± 1	29
M63	1, 3, 4, 5 (4)		0.092	0.008		−2.35	18 ± 2	38
M66*	1, 3, 4, 5 (4)			0.044		−0.93	190 ± 40	13
M81	1, 3, 4 (5)			0.050	[a]	−1.63	16 ± 4	14

The ends of the crystallographic and predicted helices are detailed in Figure 1. Eight mutations of the λ-repressor have been examined. The basic mutant has two point mutations (glycine to alanine, G46A/G48A) in position 46 and 48, with respect to the wild type. The seven mutants (M15 to M81) have undergone a further change (alanine to glycine) in position 15, 20, 37, 49, 63, 66, and 81, respectively. Boldface labels indicate the mutated residues that belong to a foldon of the basic mutant (Fig. 1). In the second column, we show the IS helices that have been used in our simulations of folding. Foldons of the basic mutant extend over the following regions: 13–22, 48–49, 63–66, and 82–84. The 48–49 foldon (foldon 3) is lacking in the wild type. The numbers in parentheses in the second column list the helices that have been involved in the mutational changes on passing from the wild type to the current protein. For easing readability, we have omitted listing helix 3 in parentheses for the Mx mutants. The missing values of β_3^f for the wild type and β_5^f for M81 are attributable to the absence of the corresponding foldon. In the wild type the superscript a indicates that native helix 3 is not an IS helix. A new foldon appears in native helix 3 only in the basic mutant and all the other Mx mutants. In the M81 mutant, the superscript a indicates that, upon mutation of the basic mutant, IS helix 5 (shared by the wild type, the basic mutant and the Mx mutants M15 to M66) is turned into a non-IS helix. Helix 2, in addition, is never counted as an IS helix. The β_i^f values for the IS helices have been calculated from the entropy profile associated with each protein[23]. For the coalescence probability of the individual microdomain, we have used $\beta_i = \beta_i^f \beta_i^g$ leaving $\beta_i^g = \beta_{WT}^g$, $\forall i$ (see Materials and Methods). This entails that changes of β_i are entirely attributed to variations of β_i^f. To ease readability, constant values are not repeated. Missing values of β_i^f are equal to the last preceding value within the same column. ΔG_{UF} is the stability expressed in kcal/mole. In the two rightmost columns, we compare the folding times (in microseconds) τ_{comp}, computed according to the FDC model, with the experimental rates τ_{exp} (drawn from Ref. 30). The asterisk indicates that M66 will be considered an outlier for the reasons explained in the Discussion section.

The FDC model provides a minimal but realistic model of protein folding in that it introduces a major simplification with respect to the current applications of the bare diffusion-collision model (DC model) (for review see Ref. 24). As a matter of fact, the FDC model maintains that the minimal set of dynamical determinants of protein folding corresponds to the set of the IS helices. The helices hosting no foldons and the nonhelical regions, enter the model as generic stretches that merely connect the foldons. This entails a clear reduction of the number of variables to be handled. The folding of such generic regions involves non-rate-limiting stages of the overall folding process.

The FDC model is in principle sensitive to the details of the residue sequence to the extent that it relies on pieces of thermodynamic and topological information deduced directly from the primary structure of the protein at hand. The crucial features include the location of the foldons within the protein sequence as well as estimates of the stability of the IS helices. The sensitivity of the model was confirmed in previous works[23,25] where we tested the FDC model on a set of nonhomologous two-state and three-state proteins, exploring folding rates spanning the microsecond to the millisecond range. Here we check the sensitivity of the FDC model to small changes in the sequence. More specifically, we show that the FDC model provides an effective sequence-specific tool that reproduces with satisfactory accuracy the variations of the folding rate ensuing from point mutations. We apply the FDC model to estimate the folding times of the mutants of two well-characterized two-state folders, the all-α proteins λ-repressor [Protein Data Bank (PDB) file, 1LMB4] and ACBP (PDB file, 2ABD). The folding rates of the wild types and mutants (displayed in Tables I and II) are computed with the FDC model and compared with the experimental values. In the Discussion, we show that the FDC model sheds light on the mechanistic reasons for the kinetic effects of point mutations in that we relate the alterations of the folding kinetics to the changes in the number and stability of the foldons. The FDC model is then used to address some key issues in the general theory of protein folding. In particular, we reconsider the effects on the folding rate of two features, CO and stability, that are currently viewed as the main determinants of the kinetics of the folding process.[2,6–11] Also, we comment on the limitations that are inherent in the FDC model. In particular, the decline of the performance of the FDC model is traced back to long-range interactions between the foldons that are poorly accounted for in the present version of the model. Finally, we discuss the role of foldons as critical residues for the regulation of the kinetic properties of the protein.

MATERIALS AND METHODS

Here we summarize the essential steps of the FDC method. In the FDC model,[23] we split the overall folding dynamics in local fast dynamics and global slow dynamics that are governed, respectively, by short-range (intrahelical) and long-range (interhelical) interactions. The fast

TABLE II. Structural, Kinetic, and Thermodynamic Data of the ACBP Protein

Mutant	Helix	Cons	Network	ΔG_{UF}	τ_{exp}	τ_{comp}	β^{f}_{A1}	β^{f}_{A2}	β^{f}_{A4}
Wild type				−8.08	4.5 ± 0.3	3.38	0.057	0.031	0.035
F5A*	*A1*	sf	1	−5.45	15.8 ± 0.2	2.98	0.075		
A9G	***A1***	s	1	−6.08	15.0 ± 1.2	19.49	0.002		
V12A*	*A1*	s	2	−6.58	13.7 ± 1.1	3.49	0.060		
L15A*	*A1*	s	2	−4.51	39.2 ± 4.3	3.57	0.058		
P19A				−7.17	3.1 ± 0.3	1.83	0.079	0.060	
D21A	*A2*			−7.68	3.1 ± 0.3	2.79	0.057	0.045	
L25A	***A2***	s		−6.06	3.7 ± 0.2	1.5		0.011	
F26A	***A2***			−6.55	4.8 ± 0.2	1.98		0.073	
Y28A	*A2*	sf	3	−5.56	1.3 ± 0.4	1.98		0.073	
Y28N	*A2*	s	3	−5.61	2.7 ± 0.3	3.74		0.029	
Y28F	*A2*	s	3	−6.82	2.7 ± 0.1	4.13		0.025	
Y31N	*A2*	f		−7.11	4.4 ± 0.2	3.77		0.029	
K32A	*A2*	sf	3	−6.47	1.9 ± 0.1	3.77		0.029	
K32E	*A2*	sf	3	−6.61	5.0 ± 0.2	2.98		0.041	
K32R*	*A2*	s	3	−5.86	7.8 ± 1.4	2.48		0.053	
Q33A*	*A2*	s	3	−5.05	0.6 ± 0.1	4.07		0.026	
A34G	*A2*		1	−7.00	3.1 ± 0.1	3.25		0.036	
T35A	*A2*			−6.80	2.0 ± 0.5	2.57		0.050	
I39A				−7.02	4.0 ± 0.3	4.00		0.026	
P44A				−6.85	4.6 ± 0.2	3.61		0.031	
K52M	A3			−8.56	2.7 ± 0.1	3.61			
K54A	A3	f		−6.67	7.2 ± 1.7	3.61			
K54M	A3	f		−7.86	6.3 ± 0.2	3.61			
E67A	*A4*			−7.28	1.8 ± 0.1	3.37			0.040
A69G	*A4*			−6.53	5.0 ± 0.5	6.68			0.008
Y73A*	***A4***	sf	1	−3.57	51.3 ± 11.8	3.17			0.045
Y73F	***A4***	sf	1	−8.11	2.2 ± 0.1	3.07			0.048
I74A*	***A4***			−6.74	15.0 ± 0.5	5.7			0.012
V77A*	***A4***	s	2	−6.54	33.4 ± 1.3	5.10			0.016
L80A*	*A4*	s	2	−3.51	85.4 ± 14.5	3.90			0.030

As shown in Figure 4, foldons are predicted in the 7–10, 25–26, and the 70–77 regions (boldface labels in the first column indicate mutations that affect the predicted foldons). The second column reports the labels of the native helices affected by the mutations studied in Ref. 29. Labels in italic designate the IS helices predicted by the neural network (*A*1, *A*2, and *A*4). Helix A3 is a normal helix that does not contain any foldons. The column with the heading "Cons" specifies the conserved residues that are important for stability (s) or function (f). The column with the heading "Network" highlights the three networks of strong interactions described in Refs. 28 and 29. ΔG_{UF} is the stability expressed in kcal/mole. τ_{exp} and τ_{comp} (in milliseconds) are the experimental and the computed folding times. Experimental stabilities and folding times are from Ref. 29. The folding probabilities β^{f} of IS helices *A*1, *A*2, and *A*4 are displayed in the last three columns. As in Table I, constant values are not repeated to ease readability. Missing values are equal to the last preceding value within the same column.

dynamics pertain to the helix-coil transitions resulting in the early formation of marginally stable protostructures (the IS helices) that undergo thermally activated diffusional motions and binary collisions. The slow dynamics describe the stochastic formation of the tertiary structure via progressive aggregation of the IS helices. The subsequent coagulation of the IS helices leads to formation of clusters (microdomains) of increasing rank (the rank equals the number of the IS helices composing the microdomain at hand). The birth of a new microdomain at the expense of the older ones with smaller rank hallmarks the transition to a new state along the folding pathway. Once all the microdomains participate in the globular cluster with the highest possible rank, the folding is complete.

In the context of the FDC dynamics, the crucial parameters are the probabilities β_{ij} that a successful collision takes place between microdomains *i* and *j*. Successful impacts result in the irreversible aggregation of the colliding microdomains. After an unsuccessful collision, the microdomains separate and start anew the diffusional search for their partner. The basic pieces of information necessary to implement the FDC model are collected by using a feed-forward neural network which is used to predict the secondary structure of helical proteins from the bare sequence. Because the FDC model describes the folding of helical proteins, we partition the space of structures predicted by the neural network into α and non-α structures. The specifics of the neural network are described in Ref. 23. The neural network is trained with the error backpropagation algorithm on a database comprising 822 proteins from the PDB.[26] The neural network used in this article was trained with single-sequence input so as to ensure maximal sensitivity to the details of the protein sequence. Structural predictions are exemplified in Figures 1 and 4.

Once we have located the native helices we are in a position to search for the foldons. To this aim, we process the outputs of the neural network to find out the position of the foldons and to estimate the probabilities of formation of the corresponding IS helices. This can be done by

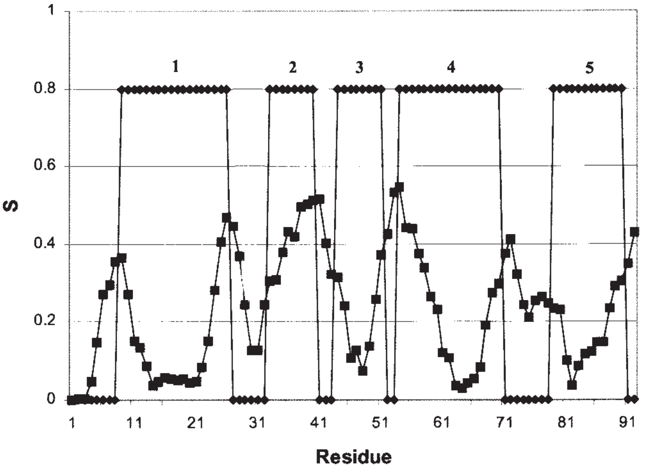

Fig. 1. Profile of the information entropy S versus residue number of the basic mutant of the λ-repressor calculated according to the neural network-based procedure of Ref. 22 (see text). The step function superimposed to the curve indicates the location of the five α-helices predicted by the neural network. Crystallographic boundaries of the helices (in parentheses), predicted helices (in parentheses), and distances from the preceding helix (in square brackets) in the basic mutant are: helix 1 (9–29) (predicted 9–26), helix 2 (33–40) (predicted 33–40) [3], helix 3 (44–51) (predicted 44–51) [3], helix 4 (59–69) (predicted 54–70) [7], helix 5 (78–88) (predicted 79–90) [8]. The entropy profile of the wild type differs from the present plot only in the region corresponding to foldon 3. Helix 3 is correctly predicted by the neural network also in the wild type but the entropy profile in that region, according to the criterion proposed below, does not reveal any foldons (see Fig. 3). To detect the protein's foldons, we look at the predicted helices with an entropy minimum that is lower than the threshold entropy $S = 0.416$ introduced in Ref. 31 (helices 1, 3, 4, and 5 in Fig. 1 fulfill this criterion). Because of the noise that affects the entropy signal[36] we associate foldons with minima whose depth is larger than 0.05.[22] Shallower minima are considered nonsignificant fluctuations of the signal. The predicted helices containing a foldon comprise the set of the IS helices. In the basic mutant (obtained performing mutations G46A/G48A on the wild type), the foldons span the 13–22, 48–49, 63–66, and 82–84 segments.

exploiting the profile of the information entropy S associated with each protein sequence (Figs. 1 and 4). S is the Shannon entropy of the discrete probability function defined by the outputs of the neural network.[22] The foldons are detected by applying the minimal entropy criterion[22] which is briefly stated in the legend to Figure 1.

The β's are factorized as $\beta = \beta^f\beta^g$, where β^f specifies the folding probability of the colliding IS helices or microdomains. β^g (orientational probability) accounts for the geometric factors relevant to the correct positioning of each helical microdomain within the resulting new microdomain. For any pair of interacting microdomains 1 and 2, β^f and β^g are factorized as $\beta_1^f\,\beta_2^f$ and $\beta_1^g\,\beta_2^g$ whereas $\beta_{12} = \beta_1^f\,\beta_2^f\,\beta_1^g\,\beta_2^g$. The evaluation of the β^fs is performed by taking into account the depth and the steepness of the entropy minimum. More details about the thermodynamic meaning of the β^f parameters and the procedure devised to compute them from the entropy profile are to be found in Ref. 23. In the FDC model, the values of β^f are biased in that we have chosen $\beta^f = 1$ for multihelical aggregates with rank >2.[23] A more refined procedure for estimating β^f of microdomains comprising two helices was presented in Ref. 25.

The β^g are usually computed from the three-dimensional (3D) structure of the protein. They are related to the loss of solvent accessible surface that the microdomains suffer as they take on their own native structure.[23,25] The program DSSP[27] provided the accessible surfaces of the various helices as well as the surface that is lost upon formation of multihelical microdomains. Because the structures of the mutants are not available, we assigned the same β^g derived from the wild type to all the mutants examined here. This is a reasonable approximation for the alanine-glycine mutants of 1LMB4 that are usually assumed to minimize variations in the network of interactions.[28] The several mutations performed on ACBP are less homogeneous under this respect, so that the same assumption is less safe.

The set of the β_{ij} and the geometric information regarding the relative positions of the IS helices and their sizes allow calculation of the mean first passage time for the coalescence of any pair of microdomains. Such a time defines the rate constant to be used in a master equation which describes the probability flux among different states of the folding process. Any change of state corresponds to some variation of the number and rank of the microdomains at any instant of time. The state with the highest rank possible represents the folded protein. The time required for the probability of the final state to reach a threshold value (0.6 in the present work as in most of the applications of the DC model[24]) defines the folding time of the protein. The basic equations used to pass from the thermodynamics of the elementary events to the kinetics of the whole folding process are reported in Refs. 24 and 25.

RESULTS

The structural, kinetic, and thermodynamic properties of the λ-repressor and ACBP have been examined by means of extensive single-site mutation experiments.[28–30] These proteins lend themselves as challenging benchmarks for the FDC model. Following the procedure reported in the Materials and Methods section, we use a neural network to predict the secondary structure of the wild types and their mutants. The same neural network is then used to determine the location and thermodynamic properties of the IS helices. In stating the results, we emphasize the data concerning the β^f parameters, that estimate the thermodynamic stability of the IS helices, and determine the coalescence probability of the colliding IS helices during the diffusion-collision dynamics (see Materials and Methods).

Analysis of 1LMB4 and Its Mutants

1LMB4 is the λ_{6-85} monomeric version of the N-terminal domain of the λ-repressor (see Ref. 30 and references therein). The mutants examined in this article are described in Table I.

The analysis of 1LMB4 starts from the entropy plot of the wild type and all its mutants. In Figure 1 we display the entropy plot of the basic mutant because it possesses the maximum number of foldons. According to the FDC

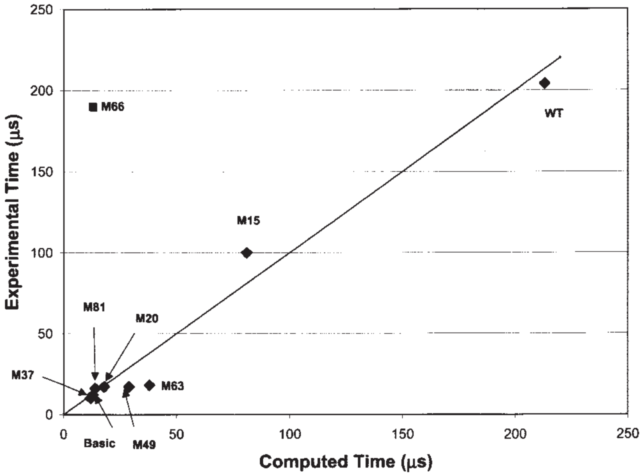

Fig. 2. Correlation of the computed and experimental folding times for 1LMB4. The plot visualizes the data of Table I. The correlation coefficient amounts to ρ = 0.67, but increases significantly (ρ = 0.98) if we exclude the unfavorable result obtained for M66. The outlier M66 is indicated by a square. The bisecting line (ideal case with ρ = 1) is shown to visually estimate the spread of the kinetic data. See Discussion for a more complete statistical analysis.

model, the critical regions for folding the basic mutant correspond to the four foldons indicated in Figure 1. The foldons and the IS helices are referred to according to the numbering of the native helices in which they are comprised (see Fig. 1). The mutations of 1LMB4 listed in Table I affect sequentially all of the five native helices of the wild type. As shown in Figure 1, helix 2 does not belong to the class of the IS helices. Accordingly, it does not contribute to the FDC dynamics. It should be noted that the wild type possesses only three foldons. In fact, native helix 2 is successfully predicted by the neural network but does not meet the requisites to qualify as an IS helix (see legend to Fig. 1).

Application of the FDC model results in estimates of the folding rate that are in good agreement with the experimental values (see Fig. 2 and Table I). The reason for the large deviation obtained for M66 is discussed below in the Discussion section. At this stage, the effects of the mutation are visible through the changes in the folding rates. However, we can improve our understanding of the modified folding dynamics of the mutants by tracing back the kinetic effects of the mutations to the properties of the foldons. We first discuss the mutations that affect the stabilities of the four foldons of the basic mutant. In M15, β_1^f is substantially reduced with respect to its original value and the rate is correspondingly lowered. Similarly, in M20 the mutation lowers β_1^f, albeit by a smaller amount than in M15, and the change in the folding rate is less dramatic. In M37 the mutation affects helix 2 that does not belong to the set of the IS helices. The entropy profile (not shown) makes it evident that mutation propagates only weakly to IS helix 3 so that the ensuing minimal change in β_3^f leaves the rate practically unaltered.

In M49 and M63 foldon 3 and, respectively, foldon 4 become less stable and the rate is lowered. The result obtained for M49 shows that a strong modulation of β_3^f entails only a very moderate variation of the folding rate. The rates of M49 and M63 also show that the same relative variation in different foldons may result in the same slight kinetic change. On the contrary, the folding times of the M63 and M66 mutants exhibit largely unequal susceptibilities to the same mutation performed in different positions within the same foldon (foldon 4).

Mutations like those resulting in the basic mutant or in the M81 mutant modify the folding kinetics also through a second mechanism involving the change of the number of the foldons. Actually, the double mutation glycine to alanine in positions 46 and 48 (G46A/G48A), performed on the wild type to get the basic mutant, results in the birth of the strong foldon 3. The subsequent mutation leading to the M81 mutant, in turn, involves the death of foldon 5. Figure 3 visualizes the modifications of the entropy profile associated with these two mutations.

The birth of a foldon in IS helix 3 on passing from the wild type to the basic mutant is responsible for the nearly 20-fold decrease of the folding time. This effect is partially reversed in the M15 mutant through the stability change mechanism. In M15, foldon 1 is dramatically destabilized to the extent that its β^f is nearly nullified, resulting in a change of the rate that is, however, half that observed in the wild type. A similar though less neat effect is visible in the M81 mutant. In this case, although foldon 5 disappears (β_5^f value lacking in Table I), we get a minor change in the folding time to the extent that the kinetic properties of M81 are nearly the same as those of the basic mutant. The quite different kinetic susceptibility to foldon subtraction is probably attributable to the different location of the foldons involved (foldon 1 or 3). Actually, we expect that the entropic effect elicited by a mutation is larger when the target foldon has an internal location (foldon 3) rather than when it is part of the N-terminus or C-terminus helices (foldon 5). Inspection of the 3D structure of 1LMB4 shows that there is also a structural ground for the relatively minor importance of foldon 5. As a matter of fact, IS helix 5 hardly contacts the body of the protein. The small contact area (see also Table 4 in Ref. 25) suggests that it can be viewed as a nearly autonomous folding unit. Therefore, it is likely that the rate-limiting step of the folding process is driven essentially by foldons 1, 3, and 4.

The different response to the strong destabilization of foldon 1 in M15 and foldon 4 in M63 is also quite striking. This effect suggests that the folding process is less sensitive to the destabilization of a weak foldon than to a similar modulation performed on a strong foldon. This finding and the failure to reproduce correctly the rate of the M66 mutant deserve a more detailed discussion that is deferred to the Discussion section.

Analysis of ACBP and Its Mutants

Experimental data on ACBP are taken from two detailed mutational analyses.[28,29] The entropy profile of ACBP is displayed in Figure 4. The folding kinetics of ACBP is quite faithfully reproduced by the three IS helices (*A*1, *A*2, and *A*4) out of the four native helical segments, predicted by

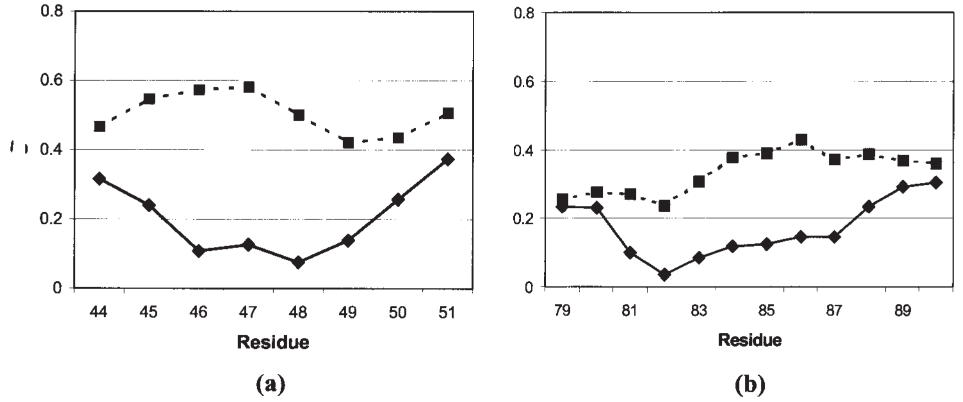

Fig. 3. Details of the information entropy profile of 1LMB4 versus residue number (see Fig. 1) which illustrate the birth and death of a foldon attributable to point mutations. **a:** Performing the basic mutation G46A/G48A on the wild type, changes locally the entropy profile of helix 3 (squares) into an entropy minimum that is eligible as a foldon (diamonds). A new foldon appears and helix 3 is turned into an IS helix. This corresponds to adding a microdomain to the FDC model. **b:** IS helix 5 of the basic mutant (diamonds) turns to a normal helix (squares) upon mutating residue 81. The new shallow minimum of helix 5 of M81 has a depth smaller than the threshold value 0.05 (see legend to Fig. 1). The minimum of the new entropy profile does not comply with the criterion presented in Ref. 23. This signals that foldon 5 disappears. The folding dynamics of M81 depends on foldons 1, 3, and 4.

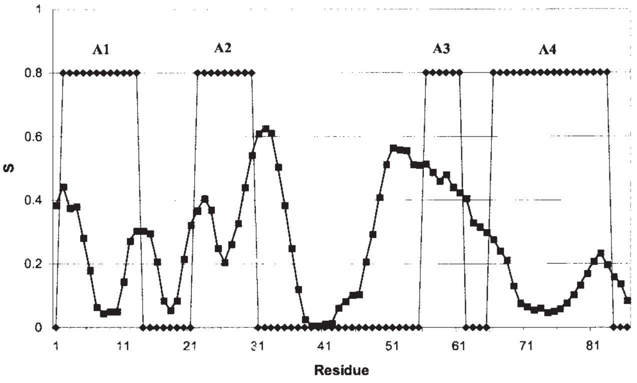

Fig. 4. Entropy profile of ACBP, drawn following the procedure reported in Ref. 22. Native helical traits are marked by the step function superimposed on the entropy plot. Crystallographic boundaries of the helices (in parentheses), predicted helices (in parentheses), and distances from the preceding helix (in square brackets) in ACBP are: helix *A*1 (3–15) (predicted 2–13), helix *A*2 (21–36) (predicted 22–30) [5], helix A3 (52–62) (predicted 56–61) [15], helix *A*4 (65–84) (predicted 66–83) [2].[29] According to the defining criterion for foldons (see legend to Fig. 1), only helices *A*1, *A*2, and *A*4 are counted as IS helices.

the network and shown in Figure 4. Helix A3 does not belong to the set of the IS helices.

The folding times resulting from the application of the FDC model are summarized in Table II, where they are supplemented with details concerning the foldons and their stabilities. The folding times of many mutants are predicted with moderate deviations from the experimental values. To ease further analysis, the most significant deviations (exceeding by a factor 2.5 the experimental value) are signaled by an asterisk.

Generally, the least perturbing Ala-Gly mutants are well predicted by the FDC model even when they undergo strong intramolecular interactions (for example, A9G and A34G). A69G is not at all involved in any intramolecular interaction network and is a fortiori correctly predicted.

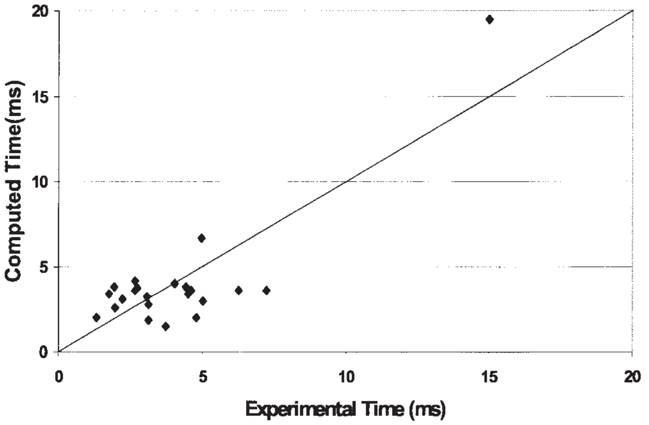

Fig. 5. Correlation of the computed and experimental folding times for ACBP. The outliers (asterisks in Table II) are not included in the plot for the sake of the readability of the diagram. The bisecting line is shown to visually estimate the spread of the kinetic data.

Mutations L25A and F26A modify the stability of foldon *A*2 but are not involved in any interaction network. On the whole, the changes in local stability are therefore sufficient to describe the ensuing change of the folding rate within the assigned tolerance range. Mutations affecting the residues participating in the interaction networks 1, 2, and 3 (Table II) are discussed in more detail in the Discussion section.

The results obtained for ACBP are epitomized in Figure 5. The data reported in Figure 5 (correlation with experimental rates $\rho = 0.85$) do not include the outliers of Table II (marked with an asterisk) to which we devote a more thorough analysis in the Discussion section.

DISCUSSION

The present application of the FDC model to the mutants of 1LMB4 and ACBP is aimed primarily at ascertaining the sequence sensitivity of the model to point mutations. The general conclusion suggested by the results of Tables I and II and Figures 2 and 5 is that the dynamics of foldons capture the essentials of the folding process and the FDC model exhibits a resolution at the residue level that enables it to perceive the perturbations of the folding kinetics induced by point mutations. The factors that limit the sequence sensitivity of the FDC model are thoroughly discussed below.

To comment on the implications of our results, we remind that the basic tenet of the FDC model is the existence of a simplified description of folding in terms of a reduced set of essential residues (foldons). Accordingly, the FDC model suggests that one of the native helices in each of 1LMB4 and ACBP wild types is not counted among the IS helices (compare Fig. 1 with Table I and Fig. 4 with Table II). The folding dynamics of 1LMB4 and ACBP (wild types) are then ruled by four and, respectively, three foldons. Intriguingly, the number of foldons is not invariant upon mutation. As shown in Table I, in the basic mutant of 1LMB4, an additional foldon is created as a consequence of the G46A/G48A double mutation, whereas foldon 5 is destroyed in mutant M81. We come back later to these specific cases.

In the case of 1LMB4 wild type, the reduction of variables accomplished by the foldon dynamics highlights the notable selectivity of the FDC model. Actually, that the maximal set of irrelevant variables is neglected and, correspondingly, the minimal set of critical residues for reconstructing the folding process is taken into account can be appreciated if we compare the FDC model with other reduced pictures of the folding dynamics. A cogent example is reported in Ref. 30. In that work, the same set of mutants of 1LMB4 as in Table I was studied, by assuming that all the five native helices contribute to the folding dynamics. Also, they were taken to obey a slow diffusional process depicted in terms of the DC model[24] as in the FDC model. The two approaches differ as to the determination of the β^f parameters which, in Ref. 30, is based on the AGADIR algorithm. Despite the major simplification achieved, the FDC model is more accurate because less satisfactory correlations with the experimental rates are obtained in Ref. 30. A quantitative statistical analysis of our results is performed below after detailed discussion of the sources of error inherent in the FDC model.

However, it should be noted that the simplification of the folding dynamics inherent in the FDC model does not imply neglect of the critical stages of the process. On the contrary, the simplified picture makes it easier to investigate the kinetics and the sequence of the critical folding events. For example, the FDC model was useful to elucidate the essential steps of folding of three-state proteins and to reproduce the switch from a two-state to a three-state folding mechanism in proteins belonging to the same family.[25] Further support to the completeness of the FDC description is provided by evidence regarding the involvement of the foldons in the transition state of the folding process of 1LMB4 and ACBP (M. Compiani, E. Capriotti, and M. Vendruscolo, unpublished results). This is to be expected because of the remarkable effectivity of the FDC model emerging from the present article and previous works.[23,25] Finally, the key role assigned by the FDC model to the residues in foldons $A1$ and $A4$ of ACBP and their bordering regions (residues 5, 9, 12, 15, and 73, 74, 77, and 80) is confirmed by experimental data showing the participation of the same residues in the stabilization of the rate-limiting native-like structure (RLNLS) via tertiary contacts occurring between the IS helices $A1$ and $A4$.[29] Notably, existence of such an RLNLS is in keeping with the predictions of the FDC model. Actually, our calculations show that only two intermediate steps are slightly populated during the folding process. They correspond to coalescence of helices $A1$ with $A4$, and $A1$ with $A2$, with aggregation $A1$–$A4$ occurring before aggregation $A1$–$A2$. In general, we expect that the foldons mediate the essential interactions, both local and long range.

These findings make us confident about relying on the foldons to investigate the mechanisms that underpin the effects of mutations. The first general conclusion is that the "accelerator pedals" of 1LMB4 and ACBP lie in the foldons. The clearest examples are the basic mutant, M15 and M66 of 1LMB4, as well as the A9G, Y73A, I74A, and V77A mutants of ACBP. It should be noted that mutations performed in the residues not included in any foldons but lying in their immediate neighborhood may affect the folding kinetics to the extent that the change of the entropy profile extends to the entropy minimum that defines the foldon proper. These effects, no matter how small, are visible in the M20, M37, and M81 mutants of 1LMB4 and the F5A, V12A, L15A, P19A, D21A, and L80A mutants of ACBP, as well as in most mutants corresponding to mutations performed in IS helix $A2$.

A basic consequence is that foldons are the critical targets of mutations that are intended to induce large kinetic effects. The corollary of this statement is that mutations are expected to be kinetically neutral in the case they affect coil regions or the non-IS helices. To be sure, this does not rule out that other mutations affecting the foldons elicit very modest effects. This is the case of M20, M49, and M63 of 1LMB4, or L25A, F26A, and Y73F of ACBP.

That the foldon dynamics turn out to be sensitive to sequence-specific features (within the limits discussed below) is especially evident from the study of M15 and M20 and also M48 and M49 of 1LMB4, where our calculations correctly predict that two mutations falling in the same foldon induce remarkably different modulations of the folding rate. The FDC model performs worse on other mutants in which the mutations affect alternative positions within the same foldon (e.g., mutant M63 as compared with mutant M66 of 1LMB4). These cases are discussed in more detail in the sequel of this section.

The second general lesson we learn from the 1LMB4 and the ACBP case studies is that two main mechanisms are at work in controlling the dynamics of folding. The first kind of control is through the modulation of the β^f of the available IS helices, whereas a further kind of regulation is made possible through the mutation-induced change of the number of foldons and IS helices. These two mechanisms suggest that the FDC model is in principle consistent with the notion proposed in the current literature that CO and stability are the two major determinants of folding.[2,6–13] This poses the intriguing question of how these two key features of folding are related to the stability and distribution of foldons. To discuss this topic, let us introduce the essential stability β_{ess} as an estimate of the contribution of local interactions to the overall stability of the protein under study. Essential stability is defined as $\beta_{ess} = \Sigma\beta_i^f$, with index i running over the set of the IS helices. The stability factor is clearly related to β_{ess} that, in turn, reflects the intrinsic helical propensity of the IS helices (see the Introduction).[23] This implies that our results are consistent with but also more specific than the general conclusion that reinforcing the propensities of the native helical structures accelerates the folding process.[13,32–35] Instead, CO is to be related to the average separation of foldons in sequence (see below for further comments on CO and foldon topology).

Clearly, by insisting on β_{ess}, we are stressing that the FDC model depends critically on local interactions. However, the visible failure of the FDC model applied to some particular mutations [notably M66 of 1LMB4, and L15A as well as the Y73A to the L80A mutants of ACBP (except Y73F)] calls for a thorough reconsideration of the basic assumptions and limitations of the FDC model. In this connection, it is convenient to mention the principal sources of error of the FDC model. We start by noting that prediction of the α-helices is affected by some noise. Sometimes, the neural network mispredicts some native helices. In the proteins of interest here, the effects of noise are signaled by the presence of underpredicted or overpredicted residues at the boundaries of the helices (compare the crystallographic and predicted helices in the legends to Figs. 1 and 4). This side effect is, however, ineluctable because some level of noise in the output signal is necessary to ensure the generalization capability of the neural network.[36] Interestingly, the poor performance of the FDC model in predicting the folding time of M66 seems to be amenable to factors having different origin, otherwise we could hardly successfully predict the moderate effect of the remarkable destabilization of the same foldon 4, performed in mutant M63. In addition, that the generalization capability of the neural network is not in question is also hinted at by the finding that in Ref. 30, the reconstruction of the folding dynamics of the same set of proteins without exploiting neural network-based methods, similarly fails to reproduce the experimental folding rate of M66.

Having ruled out the noise affecting the entropy signal as a performance-limiting factor, we are brought back to consider global interactions. In general, the delicate balance between local and global interactions is a matter of debate in the literature.[37–41] Precise assessment of the relative strength of these interactions is normally made even more elusive as they are susceptible to large variations along the protein's sequence.[42] If we neglect the contribution of the random coil regions, we can surmise that the changes of total stability of the protein can be approximately decomposed into a variation of local (intrinsic) stability of the helices plus a contribution from the interhelical forces (long-range or packing interactions).[43] Evidently, variations of β_{ess} cannot account neither for changes in the long-range interactions among the colliding IS helices nor the stability change of the non-IS helices. If we neglect the non-IS helices, only in the case the helix packing interactions are roughly fixed on mutating the wild-type protein, the change of β_{ess} is expected to reflect the overall change of stability. Nonetheless, these approximations inherent in the estimation of the stability factor through β_{ess} are somewhat reduced because stability changes attributed to long-range interactions are implicitly, albeit only partially, taken into account in the FDC model. Actually, a moment's reflection shows that the geometrical factors β^g capture, at least to some extent, the influence of global forces because these factors depend on the mutual orientation and position assumed by any couple of IS helices and microdomains within the native structure. This seems to be the case for the wild types and the mutants for which we get good estimates of the folding times (Tables I and II). Conversely, in the less favorable cases extrapolating the β^g from the wild type to any mutants (section Materials and Methods) may be conducive to poorly estimate the changes of tertiary interactions. Such an approximation can be mitigated only to the extent that the actual variation of β^g is negligible with respect to the change of β_{ess}.

The preceding arguments suggest that the performance of the FDC model is critically dependent on the engagement of the mutated residues in long-range interactions. More precisely, the more biased toward local forces the balance of the interactions the mutated residue is involved in, the more effective the FDC model in describing the folding dynamics.

If we turn to a more thorough examination of Table I, it is quite clear that this interpretation applies to our results for the two critical residues 15 and 66 of 1LMB4. Actually, our explanation is consistent with a recent investigation stressing that Ala 15 and Ala 66 are linked by a very strong long-range mutual interaction and correspond to the most stable residues of 1LMB4.[42] A semiquantitative demonstration of this basic idea is provided by the histograms of Figure 6, showing that the largest discrepancy between total stability change and essential stability change is maximal for the worst predicted mutant M66. This confirms that long-range interactions are the main causes for the decline of performance of the FDC model. A look at the native structure offers additional insight on the different roles of residues 20 and 63 as compared with residues 15 and 66. The latter amino acids are seen to have a larger number of spatial neighbors (within 4 Å) that are somewhat distant in sequence. Residue 15 contacts with the 50's residues whereas residue 66 contacts with the 70's residues.

The asymmetric effect of the same Ala-Gly mutation in positions 15 and 66 and the different precision attained by the FDC predictions for the two mutants can be rational-

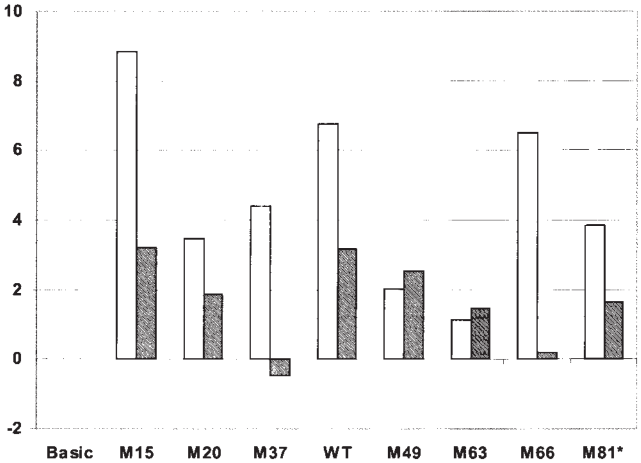

Fig. 6. Histograms displaying thermodynamic parameters of 1LMB4. White bars denote $-10(\Delta G_{UF} - \Delta G_{UF}^{bas})/\Delta G_{UF}^{bas}$ (superscript refers to the basic mutant). Gray bars show $-10(\beta_{ess} - \beta_{ess}^{bas})/\beta_{ess}^{bas}$. The difference between any two coupled bars tells us how much the essential stability lags behind the real stability change. The largest discrepancy between the two indicators occurs for M66, for which the FDC calculation visibly fails (see Table I). Misprediction of the folding rate is thus traced back to the maximally defective estimate of the mutant's stability provided by β_{ess}.

ized by means of a similar argument. Residue 15 lies in the most stable foldon, as evident from the corresponding β_1^f value (Table I). Unlikely, residue 66 belongs to the second least stable foldon. For the latter mutant, the variation in long-range interactions has a major relative weight with respect to the change of short-range interactions (i.e., β_{ess}). Conversely, the local contribution to stability reflected by β_{ess} is likely to be more significant for M15 than for M66, so that the change in essential stability of M15 describes more accurately the total stability change.

Also, in the case of ACBP, we maintain that marked discrepancies between experimental and computed folding rates are likely to occur whenever significant native tertiary interactions involve the mutated residue. The proposed explanation is consistent with the finding that most of mutations of Table II for which we get the worst results (marked with an asterisk) are directed onto residues that are conserved for stability and belong to networks of interhelical interactions.[28] More precisely, as shown in Table II, the most relevant interactions take place among Phe5, Ala9, and Ala34 with Tyr73 (cluster 1); Val12, Leu15, and Val77 with Leu80 (cluster 2); Tyr28 and Lys32 with Gly33 (cluster 3). Remarkably, most of the interhelical interactions associated with the mispredicted mutants of Table II involve the foldon-spanning regions of IS helices *A*1 and *A*4, or their immediate neighborhood (with the exception of K32R and Q33A). This confirms the fact that the foldons are involved in the key long-range interactions. Instead, the A9G and A34G mutants are relatively well predicted although the mutated residues are interacting within cluster 1, probably because Ala to Gly mutations are known to minimally interfere with the preexisting network of interactions.[28] The shift of the force balance helps to also explain the data for K54A and A69G. These two mutations involve approximately the same destabilization, but the FDC result for the former is worse than for the latter because K54 has more van der Waals interactions.[28] The subset of minimally perturbing mutations of 2ABD comprises A9G, Y28F, A34G, A69G, and Y73F that are well predicted. The combined effect of long-range interactions and the defective extrapolation of β^g are visible in the unsuccessful prediction of Y73A as compared with the good prediction of Y73F. Mispredictions of I74A and V77A are presumably attributable to the wrong extrapolation of β^g. The unique feature of the A9G mutant that justifies its exclusion from the set of the outliers (despite its participation in interaction cluster 1) is the same we have invoked to explain the asymmetric effects of the mutations in M15 and M66 of 1LMB4. Accordingly, the satisfactory prediction of A9G is attributed to the local interactions dominating the force balance in the first (most stable) foldon of 2ABD.

Properties such as conservatism, stability, and engagement in transition state-like states are often associated with the putative determinants of folding. The detailed analysis summarized in Table II shows that these requisites apply to the foldons or their neighboring residues. This reinforces our claim that foldons comprise the fundamental regions for the folding dynamics. As far as stability as one of the determinants of folding is concerned, Ref. 28 indicates that many of the residues that substantially contribute to the stability of ACBP are conserved [marked with an (s) in Table II]. Careful scrutiny of Table II shows that these residues are found in the foldons or close to their ends. The paramount importance of foldons is also confirmed by the finding that the four foldon residues A9, Y73, I74, and V77 are involved in the formation of the RLNLS as stressed in Ref. 29. The remaining RLNLS residues F5, V12, L15, and L80 are also quite close to the ends of the two external foldons *A*1 and *A*4 (see legend to Table II). In addition, Table II shows that the eight RLNLS residues (F5, A9, V12, L15, Y73, I74, V77, and L80) are conserved. This is also in keeping with the conclusions of our previous preliminary work on the conservatism of the foldons of aligned proteins.[44] Finally, independent preliminary evidence about the participation of foldons in the transition state (M. Compiani, E. Capriotti, and M. Vendruscolo, unpublished results) is also consistent with the conservation of the residues involved in the transition states.[13,45,46]

The same reasoning illustrated in Figure 6 can be conducted also for ACBP. Figure 7(a and b) helps to trace back the failure of the FDC model on some mutants of ACBP to the fact that the modulations of the essential stability β_{ess} are no longer dominating over the interactions with distant amino acids. In Figure 7(a) we compare, for all the mutants of Table II, the relative deviations in stability (with respect to the wild type) and the relative errors incurred by the FDC estimates with respect to the experimental values of the folding time. The most remarkable peaks are coincident in both histograms showing that the most defective predictions of the folding kinetics are linked essentially to substantial alterations of stability. The true source of error emerges clearly from Figure 7(b) which visualizes the divergence of the relative change of

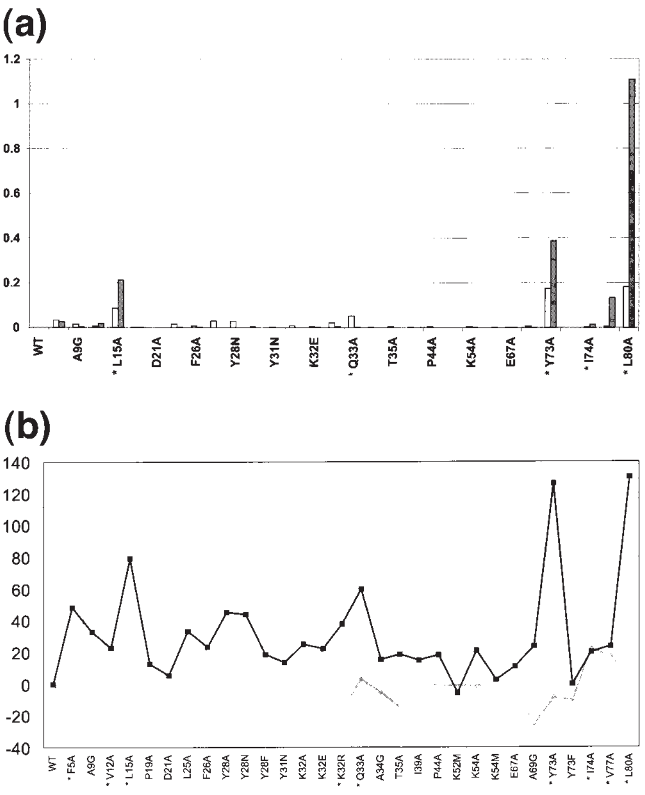

Fig. 7. Histograms displaying thermodynamic and kinetic parameters of ACBP. **a:** White bars denote $-10(\Delta\Delta G_{UF}/\Delta G_{UF})$. $\Delta\Delta G_{UF}$ is the variation of ΔG_{UF} (free energy change of the folding process) on passing from the wild type to the current mutant. Gray bars visualize the deviation of the computed folding rate $(\tau_{comp} - \tau_{exp})^2/6{,}000$ for the current mutant. The numerical proportionality factors were introduced to bring both histograms to the same scale. Interestingly, the most remarkable deviations occur for the same mutants that essentially correspond to the outliers of Table II (marked with an asterisk). This indicates that large stability changes are correlated with large mispredictions of the FDC model. **b:** Trends of the relative deviations of the free-energy change accompanying the folding process ΔG_{UF} and the essential stability β_{ess} (see text) of the mutants of ACBP. The plotted parameters are $-100(\Delta\Delta G_{UF}/\Delta G_{UF})$ (diamonds) and $-100(\Delta\beta_{ess}/\beta_{ess})$ (squares). Δf indicates the difference of the generic parameter f on changing from the wild type to the mutated protein. Positive variations of these parameters indicate loss of stability. Superscript WT refers to the wild type. The two plots exhibit the most sensible divergence for the outliers of Table II (labels with asterisk). For the same mutants, (a) shows that the FDC model leads to sensible underestimation of the folding rate. Comparison of (a) and (b) hints at the fact that mispredicted rates are attributable to the defective estimation of the current ΔG_{UF} through β_{ess}. The exceptional cases of I74A and V77A are commented on in the text.

β_{ess} from the relative change of total thermodynamic stability of the protein (changes are with respect to the wild type). Comparison of the plot of Figure 7(b) with the histograms of Figure 7(a) indicates that the mutants for which the change of β_{ess} is maximally deficient in mirroring the stability change coincide with the mispredicted mutants [peaks of Fig. 7(a)]. This provides compelling evidence that the precision of the FDC model is reduced so long as nonlocal interactions (not accounted for by β_{ess}) become too strong as compared with the local interactions.

After stressing the factors limiting the validity of our prediction method, a more complete correlation analysis can be performed to quantify the reliability of the FDC model. The remarks made in this Discussion have clarified that it makes sense to restrict statistical analysis to the set of the minimally perturbing mutations that allow safe extrapolation of the β^g factors from the wild type and for which local energetic factors dominate over long-range interactions. Therefore, we consider all the mutants of 1LMB4 except M66 and the set of the A9G, Y28F, A34G, A69G, and Y73F mutants of 2ABD. For the sake of brevity, we denote collectively these mutants the AGYF set. To perform the statistical check under the most stringent conditions, we have compared two statistical indicators (correlation coefficient ρ and Spearman correlation ρ_S[47]) with and without the two data that presumably have overwhelming weight within the set (1LMB4 wild type and mutant A9G of 2ABD). The relevant statistics for the AGYF set are: $\rho_S = 0.\ 98$ and $\rho = 0.99$ (A9G included) and $\rho_S = 0.98$ and $\rho = 0.95$ (A9G excluded). The corresponding estimates of ρ and ρ_S for the Mx mutants of 1LMB4 are $\rho_S = 0.99$, $\rho = 0.99$ (including the wild type), and $\rho_S = 0.99$, $\rho = 0.94$ (excluding the wild type). Interestingly, the calculations performed in Ref. 30 lead to $\rho = 0.56$ and $\rho = 0.82$ (with and without M66, respectively) to be compared with our data in Figure 2, $\rho = 0.67$, and $\rho = 0.99$ (with and without M66, respectively).

Correlations for the AGYF set of 2ABD are: $\rho_S = 0.70$, $\rho = 0.67$ (excluding A9G) and $\rho_S = 0.83$, $\rho = 0.98$ (including A9G). Expectedly, correlations decline when estimates are computed over the whole set of 2ABD mutants (outliers in Table II excluded). Including A9G we get $\rho_S = 0.28$, $\rho = 0.85$. This is to be ascribed to the fact that, in this case, most of the mutations do not belong to the AGYF set. The remarkable reduction of ρ_S is mainly attributable to the incapability of the FDC model to reproduce the tiny fluctuations (with respect to the wild-type rate) of the folding rate of those mutants in which the mutations affect portions of the sequence that are distant from the foldons. In these cases, the mutations in question hardly affect the entropy profile of the nearest foldon resulting in a nearly constant folding time. Excluding A9G, both figures decrease dramatically indicating that the good correlation depends heavily on the individual A9G mutant. To be sure, this statistical analysis has only provisional character because of the paucity of the available data. Nonetheless, we believe that the importance of foldons is quite safely established also by the qualitative result that, irrespective of the accuracy of the FDC predictions, the foldons comprise the kinetically hot residues of the proteins studied, i.e., those residues that upon mutation may be conducive to dramatic changes of the folding time. The explanatory value of the FDC model resides mainly in the identification of the critical residues for the kinetic control of the folding process with the foldons. The foldons are also useful to establish a link between folding properties and specific intramolecular interactions. In this respect, we have ascertained that the most frequent regulation mechanism relies on the modulation of the foldon stability.

A second effective mechanism to control protein folding kinetics emerges from the case study of 1LMB4, where inducing the birth or death of foldons in the wild type gives rise to substantial changes of the folding kinetics. The death or birth of foldons (Fig. 2 and Table I) changes the number of the microdomains taking part in the folding dynamics and modifies the distances in sequence of the extant foldons. To address the issue of how the separation in sequence of the foldons is quantitatively related to the CO, one needs to resort to the 3D structures of the mutants which, however, are not available. Nonetheless, a qualitative relationship is clearly present as, seemingly, both variables share a common physical origin. Actually, in both cases, we are confronted with entropic effects on the folding dynamics because the separation in sequence either of the foldons or the contacting residues (within the native structure) determines the volume of the configuration space to be explored before effective collisions promote formation of native contacts. That this may be the relevant factor is suggested by the arguments invoked in recent discussions, where the good correlation between CO and folding rates is interpreted in terms of entropy effects.[48,49]

The FDC model holds promise to be useful in a more general sense. For example, detection of critical residues is the main goal of protein engineering methods that investigate the transition state and the folding pathway at the residue level.[50,51] In this sense, the FDC model is a promising substitute that, because of its sequence sensitivity, can be used to perform a preliminary screening of the putative kinetically hot residues. Much in the same spirit as the single-site thermodynamic mutation method,[42] the FDC model lends itself to conduct simulated mutation experiments that might be useful to direct mutation studies onto the minimal set of putative controllers of the folding process, thus avoiding blind and extensive experimental mutant analysis. In this respect, the FDC model exemplifies how the study of the folding mechanism may be instrumental to the rational design of proteins with specified kinetic properties.[52]

At a more general level of organization, the current focus on the integration of proteins within complex interaction networks brings to the foreground the issue of control. In this framework, one can take advantage of the FDC model as an effective predictor of the kinetic effects of mutations, to link kinetic control of single protein folding with the temporal dynamic changes at network scale.[53]

Viewed in these terms, the FDC model can be fruitfully used as a powerful tool to connect the molecular level to the higher levels of analysis proper to functional modules, as required by the ongoing transformation of cell biology into a modular cell biology.[54]

Our results emphasize the role of foldons as cooperative semi-independent units. In this respect, a quite meaningful finding regards the striking correlation between the helicities of the IS helices, estimated in the framework of the native structure from the β^f parameters, and the experimental helical content of the same IS helices isolated from the remainder of the protein.[23] The central role of foldons is also in accord with recent speculations about the modularity of protein folding mechanisms and the importance of preorganized elements of secondary structure.[55,56]

In Ref. 56, the relevant role played by cooperative units was anticipated to have "important implications for a variety of protein properties including cooperativity, stability, design, evolution and function." The FDC model seems to offer new avenues on most of these items. That the FDC model sheds light on stability, evolution, and design is quite evident from the arguments reported in this discussion. It should be added that the results of the present article and previous works on two- and three-state folders[23,25] confirm that the sequential stabilization of foldons and their aggregates provide a quite general key for also understanding the folding mechanisms of helical proteins. In particular, the use of the FDC model as a unifying mechanism of folding and a tool to quantitate the cooperative character of folding mechanisms is illustrated in a forthcoming article (M. Compiani, submitted).

ACKNOWLEDGMENTS

E.C. acknowledges financial support from a grant delivered by the European Union's VI Framework Programme for the BioSapiens Network of Excellence project. M.C. has been funded by the Italian Ministry for Research also through the PRIN 2005 Project. The authors also acknowledge the Laboratory of Biocomputing of the University of Bologna for the access to computational facilities.

REFERENCES

1. Baker D. A surprising simplicity to protein folding. Nature 2000;405:39–42.
2. Plaxco KW, Riddle DS, Grantcharova V, Baker D. Simplified proteins: minimalist solutions to the protein folding problem. Curr Opin Struct Biol 1998;8:80–85.
3. Scala A, Dokholyan NV, Buldyrev SV, Stanley HE. Thermodynamically important contacts in folding of model proteins. Phys Rev E 2001;63:032901(1–4).
4. Vendruscolo M, Paci E, Dobson CM, Karplus M. Three key residues form a critical contact network in a protein folding transition state. Nature 2001;409:641–645.
5. Vendruscolo M, Dokholyan NV, Paci E, Karplus M. Small-world view of the amino acids that play a key role in protein folding. Phys Rev E 2002;65:061910(1–4).
6. Dill KA, Fiebig KM, Chan HS. Cooperativity in protein-folding kinetics. Proc Natl Acad Sci USA 1993;90:1942–1946.
7. Weikl TR, Palassini M, Dill KA. Cooperativity in two-state protein folding kinetics. Protein Sci 2004;13:822–829.
8. Fiebig KM, Dill KA. Protein core assembly processes. J Chem Phys 1993;98:3475–3487.
9. Zhou H, Zhou Y. Folding rate prediction using total contact distance. Biophys J 2002;82:458–463.
10. Galzitskaya OV, Garbuzynskiy SO, Ivankov DN, Finkelstein AV. Chain length is the main determinant of the folding rate for proteins with three-state folding kinetics. Proteins 2003;51:162–166.
11. Ivankov DN, Finkelstein AV. Prediction of protein folding rates from the amino acid sequence-predicted secondary structure. Proc Natl Acad Sci USA 2004;101:8942–8944.
12. Dinner A, Karplus M. The roles of stability and contact order in determining protein folding rates. Nat Struct Biol 2001;8:21–22.
13. Plaxco KW, Simons KT, Ruczinski I, Baker D. Topology, stability, sequence and length: defining the determinants of two-state protein folding kinetics. Biochemistry 2000;39:11177–11183.
14. Garcia AE. Large-amplitude nonlinear motions of proteins. Phys Rev Lett 1992;68:2696–2699.
15. Tirion MM. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys Rev Lett 1996;77:1905–1908.
16. Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. Proteins 1993;17:412–425.

17. Roccatano D, Daidone I, Ceruso M, Bossa C, Di Nola A. Selective excitation of native fluctuations during thermal unfolding simulations: horse heart cytochrome c, a case study. Biophys J 2003:84: 1876–1883.
18. Daidone I, Amadei A, Roccatano D, Di Nola A. Molecular dynamics simulation of protein folding by essential dynamics sampling: folding landscape of horse heart cytochrome c. Biophys J 2003;85: 2865–2871.
19. Freire E, Murphy KP. Molecular basis of cooperativity in protein folding. J Mol Biol 1991;222:687–698.
20. Demirel MC, Atilgan A, Jernigan RL, Burak E, Bahar I. Identification of kinetically hot residues in proteins. Protein Sci 1998;7:2522–2532.
21. Zimm B, Bragg JK. Theory of phase transition between helix and random coil in polypeptide chains. J Chem Phys 1959;31:526–535.
22. Compiani M, Fariselli P, Martelli P-L, Casadio R. An entropy criterion to detect minimally frustrated intermediates in native proteins. Proc Natl Acad Sci USA 1998;95:9290–9294.
23. Compiani M, Capriotti E, Casadio R. Minimally frustrated helices determine the folding mechanism of small helical proteins. Phys Rev E 2004;69:051905(1–8).
24. Karplus M, Weaver DL. Protein folding dynamics: the diffusion-collision model and experimental data. Protein Sci 1994;3:650–668.
25. Stizza A, Capriotti E, Compiani M. A minimal model of three-state folding dynamics of helical proteins. J Phys Chem B 2005;109: 4227–4233.
26. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
27. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
28. Kragelund BB, Poulsen K, Andersen VK, et al. Conserved residues and their role in the structure, function and stability of acyl-coenzyme A binding protein. Biochemistry 1999;38:2386–2394.
29. Kragelund BB, Osmark P, Neergaard TB, et al. The formation of a native-like structure containing eight conserved hydrophobic residues is rate limiting in two-state protein folding of ACBP. Nat Struct Biol 1999;6:594–600.
30. Burton RE, Myers JK, Oas TG. Protein folding dynamics: quantitative comparison between theory and experiment. Biochemistry 1998;37:5337–5343.
31. Casadio R, Compiani M, Fariselli P, Martelli P-L. A data base of minimally frustrated α-helical segments extracted from proteins according to an entropy criterion. Proc Int Conf Intell Syst Mol Biol 1999;7:66–76.
32. Viguera AR, Villegas V, Aviles FX, Serrano L. Favourable native-like helical local interactions can accelerate protein folding. Fold Des 1996;2:23–33.
33. Chiti F, Taddei N, Webster P, et al. Acceleration of the folding of acylphosphatase by stabilization of local secondary structure. Nat Struct Biol 1999;6:380–387.
34. Cavagnero S, Dyson HJ, Wright PE. Effect of H helix destabilizing mutations on the kinetic and equilibrium folding of apomyoglobin. J Mol Biol 1999;285:269–282.
35. Garcia C, Nishimura C, Cavagnero S, Dyson HJ, Wright PE. Changes in the apomyoglobin folding pathway caused by mutation of the distal histidine residue. Biochemistry 2000;39:11227–11237.
36. Compiani M, Fariselli P, Casadio R. Noise and randomlike behavior of perceptrons: theory and application to protein structure prediction. Phys Rev E 1997;55:7334–7343.
37. Avbely F, Moult J. Role of electrostatic screening in determining protein main conformational preferences. Biochemistry 1995;34: 755–764.
38. Abkevich VI, Gutin AM, Shakhnovich EI. Impact of local and non-local interactions on thermodynamics and kinetics of protein folding. J Mol Biol 1995;252:460–471.
39. Unger R, Moult J. Local interactions dominate folding in a simple protein model. J Mol Biol 1996;259:988–994.
40. Saven JG, Wolynes PG. Local conformational signals and the statistical thermodynamics of collapsed helical proteins. J Mol Biol 1996;257:199–216.
41. Hardin C, Luthey-Schulten Z, Wolynes PG. Backbone dynamics, fast folding, and secondary structure formation in helical proteins and peptides. Proteins 1999;34:281–294.
42. Hilser VJ, Dowdy D, Oas TG, Freire E. The structural distribution of cooperative interactions in proteins: analysis of the native state ensemble. Proc Natl Acad Sci USA 1998;95:9903–9908.
43. Gunasekaran K, Eyles SJ, Hagler AT, Gierasch LM. Keeping it in the family: folding studies of related proteins. Curr Opin Struct Biol 2001;11:83–93.
44. Compiani M, Fariselli P, Martelli P-L, Casadio R. Neural networks to study invariant features of protein folding. Theor Chem Acc 1999;101:21–26.
45. Ptitsyn OB, Ting K-LH. Non-functional conserved residues in globins and their possible role as folding nucleus. J Mol Biol 1999;291:671–682.
46. Mirny LA, Shakhnovich EI. Universally conserved positions in protein folds: reading evolutionary signals about stability, folding kinetics and function. J Mol Biol 1999;291:177–196.
47. Mood AM, Graybill FA, Boes DC. Introduction to the theory of statistics. Tokyo: McGraw-Hill Kogakusha; 1974.
48. Fersht AR. Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism. Proc Natl Acad Sci USA 2000;97:1525–1529.
49. Makarov DE, Plaxco KW. The topomer search model: a simple, quantitative theory of two-state protein folding kinetics. Protein Sci 2003;12:17–26.
50. Fersht AR. Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. New York: W.H. Freeman; 1999.
51. Matouschek A, Kellis JT Jr, Serrano L, Fersht AR. Mapping the transition state and pathway of protein folding by protein engineering. Nature 1989;340:122–126.
52. Guerois R, Serrano L. Protein design based on folding models. Curr Opin Struct Biol 2001;11:101–106.
53. Bork P, Jensen LJ, von Mering C, Ramani AK, Lee I, Marcotte EM. Protein interaction networks from yeast to human. Curr Opin Struct Biol 2004;14:292–2995.
54. Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. Nature 1999;402(Suppl):C47–C52.
55. Myers JK, Oas TG. Preorganized secondary structure as an important determinant of fast protein folding. Nat Struct Biol 2001;8:552–558.
56. Maity H, Maity M, Englander SW. How cytochrome c folds and why: submolecular foldon units and their stepwise sequential stabilization. J Mol Biol 2004;343:223–233.