Differences between the Sequential Collapse Folding Pathways of Apoleghemoglobin and Apomyoglobin

Fernando Bergasa-Caceres*,† and Herschel A. Rabitz‡

Department of Chemistry, Princeton University, Princeton, New Jersey 08544

Received: November 8, 2001; In Final Form: February 16, 2002

The sequential collapse model (SCM) is applied to reveal the folding pathway of apoleghemoglobin, and the results are compared to the folding pathway of apomyoglobin, a structurally similar protein. The folding pathway of apoleghemoglobin is found to differ significantly from that calculated for apomyoglobin by the SCM in previous work. There are two energetically equivalent initial contacts leading to two distinct folding pathways. One of the possible initial contacts is predicted to form between segments in helices E and H. The other is predicted to form between a segment in the region connecting helices C and E and helix G. In apomyoglobin the dominant initial contact was predicted to form between segments in helices B and G. The predicted differences are shown to be in general agreement with experimental results. The observed differences between the folding pathways of both proteins confirm experimental and theoretical observations that folding pathways are not evolutionarily conserved and that a significant degree of structural homology does not necessarily imply similar folding pathways for different proteins. This result suggests that several folding pathways are available for a given fold and that the dominant one for each protein is a function of the sequence and the experimental conditions.

1. Introduction

The sequential collapse model (SCM) is a recent theoretical proposal that strives to explain how the primary sequence of a protein determines its folding pathway. In the SCM the folding pathway is governed by the hydrophobic effect and by the topology of the intermediate states along the folding pathway.[1] The SCM has been able to successfully reproduce the folding pathway for a number of proteins at low resolution, including apomyoglobin, barnase and cytochrome *c*,[1] hen lysozyme, and α-lactalbumin.[2]

An interesting question for any theory attempting to predict protein folding pathways is whether a high degree of structural homology implies similar folding pathways. Several authors on the basis of theoretical analyses[3,4] have suggested that this is the case. Experimental results[5−10] have also confirmed that there is often a link between structural homology and similar folding pathways, although this might not be a general rule.[11] Recently, experimental results comparing the folding pathways of apoleghemoglobin and apomyoglobin, two structurally very similar proteins, have become available.[12] The results strongly question the validity of a direct correspondence between structural homology and a similar folding pathway, as the observed folding pathways for both proteins are broadly distinct.[12] Moreover, because both proteins appear to have diverged from a common ancestral gene,[13] this result puts into question the seemingly natural assumption that evolution conserves both the structure and the folding pathway of globular proteins.

The purpose of this paper is to show that the SCM is able to naturally account for most of the observed differences in the folding pathways of apomyoglobin and apoleghemoglobin. To fulfill this purpose, a brief review of the SCM is presented below. Then, the SCM folding pathway of apoleghemoglobin is calculated and compared with that of apomyoglobin. It is found that there are two possible first folding events leading to two distinct dominant SCM folding pathways for apoleghemoglobin. One of them implies the formation of a contact between helices E and H, the other implies the formation of a contact between amino acids 43−47 located between helices C and E in the native structure and helix G. This result is consistent with the experimental results.[12] These first intermediates along both SCM folding pathways of apoleghemoglobin differ from that formed along the SCM pathway of apomyoglobin, defined by a dominant initial contact formed between helices B and G[1] and a MGLIS formed by helices A, B, G, and H, consistent also with experimental results.[14−17] This difference in the first folding events leads to broadly distinct multistate folding pathways for both proteins, as discussed in the paper. Finally, conclusions are presented regarding the significance of the result for broader issues of protein folding.

2. Model

The SCM has been outlined in full detail elsewhere.[1] Here, only a brief review appropriate for the specific goals of this paper is presented. In the SCM the free energy of formation of a successful contact is written as

$$\Delta G_{\mathrm{cont}} = \Delta G_{\mathrm{loop}} + \Delta G_{\mathrm{int}} \qquad (1)$$

where ΔG_{loop} is the free energy change associated with loop closure and is expected to be positive because loop formation defines a state that has fewer conformational possibilities than an open protein chain, thus inducing a large entropic loss ΔS_{loop}. ΔG_{int} represents all the interactions that help stabilize the contact,

* To whom correspondence should be addressed.
† Current address: Endesa, C/Príncipe de Vergara 187, 28002 Madrid, Spain. E-mail: Fbergasa@Endesa.Es. Tel: 34 91 566 88 00. Fax: 34 91 213 95 56.
‡ E-mail: Hrabitz@Chemvax.Princeton.edu. Tel: 609 258 39 17. Fax: 609 258 67 46.

10.1021/jp014075g CCC: $22.00
Published on Web 04/05/2002

including hydrophobic interactions, van der Waals interactions, hydrogen bonds, disulfide bonds, and salt bridges,[18] $\Delta G_{int} < 0$.

Upon formation of a contact in a protein of N amino acids, three regions are distinguished within the SCM: the contact region c of length n_c, the open connecting loop l of length n_l, and the free ends or tails of combined length $n_0 = N - (n_c + n_l)$. Up to a constant, ΔG_{loop} can be written as[1]

$$\Delta G_{loop} \approx -kT[n_l \ln(f_l/f_0) + n_c \ln(f_c/f_0) - {}^3/_2 \ln n_l] \quad (2)$$

where f_i represents the conformational freedom of the amino acids in a region i of the protein and the conformational freedom of the amino acids in the free ends of the protein f_0 is taken to be the same as that of the amino acids in the random coil. ΔG_{loop} can be shown to have two minima[1] as a function of loop length n_l: a deeper one at $65 < n_l < 85$ amino acids, called the optimal loop length, and a shallower one at $n_l \approx 3-4$ amino acids. The shallow minimum represents the shortest length over which the protein chain can reverse its direction. Because a few amino acids are required to form a stable contact, we define a minimal loop to be generated by a protein contact between segments of ~5 amino acids linked by a turn of 3−4 amino acids. Thus, the minimal loop size is $n_{min} \approx n_l + 10 = 13-14$ amino acids. In proteins consisting of ~100−150 amino acids, within the SCM, most of the contacts will form separated by distances of n_l, defining loops of length n_{min}, because no more than one or two primary contacts may form.[2] This behavior is consistent with experimental evidence showing that short-range contacts predominate over long-range ones in protein structures.[19]

Based on the formalism developed above, in the SCM, for proteins sufficiently long, the folding pathway is likely initiated by the formation of a contact between segments located at the optimal distance n_{op} of 65−85 amino acids. Formation of this initial contact, referred to as the primary contact, leads to a multistate folding pathway that includes an intermediate state with many of the properties of a molten globule,[20] and therefore referred to as the molten globule-like intermediate state (MGLIS).[1] Formation of the initial contact is followed by the folding of the residues located outside the primary loop, referred to as the tails of the protein (i.e., in the SCM this step is argued to occur, provided the tail segments are sufficiently long). This is followed by a hydrophobic collapse of the protein core in which the native topology of the protein is established.[2] The collapse is followed by an optimization subphase governed by the activation barriers generated by the need to exclude water from the interior of the protein core in order to fully establish the interactions that stabilize the native structure. The time lapse between the cooperative collapse and the optimization subphase cannot be determined in the abscence of a SCM theory of folding rates yet to be developed. Recent experimental evidence for protein L suggests that the acquisition of native structure might be simultaneous with chain collapse.[21] The consistency of the SCM with this result may only be determined when (a) a SCM theory of folding rates is available and (b) more data become available for more protein examples. Proteins that are shorter than 65−85 amino acids, henceforth referred to as short proteins in the SCM, must fold through the formation of loops close to minimal size because their short length precludes the existence of an initial long-range contact. There also may be larger proteins falling outside this model in which a stable primary contact does not form and for which the folding mechanism should be similar to that of short proteins. Only through the analysis of many proteins by SCM will the statistical pattern emerge about which cases are consistent. The previous work[1,2] and this paper aid in this process.

The SCM is operationally similar to recent theoretical efforts to develop simple models able to describe the intermediates along the folding pathway in the context of the so-called "funnel" view of the folding process.[22−25] Both in the SCM and in the funnel picture, one goal is to determine the intermediates along the folding pathway by minimizing relatively simple free energy functionals, at a much lower level of resolution than atomic. The SCM, however, shares much of the "old" view of the folding process in which the protein descends toward the free energy minimum through few intermediate steps.[26] The "new" view embodied in the funnel picture postulates instead a large number of intermediates, especially in the early folding stages. These two views might, however, be compatible, as recent theoretical results show that some folding pathways might be strongly statistically preferred in a free energy landscape that allows for a multiplicity of pathways.[27]

In the next section, the SCM folding pathway of apoleghemoglobin is calculated and compared with that of apomyoglobin. Both pathways are found to be different, and the differences are then shown to be generally consistent with existing experimental data. This result differs from previous theoretical analyses that hypothethize that the folding pathways of the proteins in the globin family should be similar.[28]

3. Results

3.1. Folding Pathway of Apoleghemoglobin. As members the globin family, soybean Apoleghemoglobin has 143 residues, while sperm whale apomyoglobin has 153 residues. Their structures are very similar,[29,30] allowing for a close alignment of their amino acid sequences.[12] The native structure of sperm whale apomyoglobin shows eight α-helical segments labeled A−H; the native structure of soybean apoleghemoglobin is very similar, but for the absence of the short D helical segment replaced by a short loop. In this section, the SCM folding pathway of apoleghemoglobin is calculated and compared to the previously determined SCM folding pathway of apomyoglobin[1] and to experimental data.[12]

3.2. Calculational Method. We follow closely the method introduced previously[1] to determine the SCM folding pathway of apomyoglobin, cytochrome *c*, barnase, and ribonuclease A. A summary of the procedure is presented below. The primary contact is determined by the minimum value of ΔG_{hyd} for segments of five amino acids located 65−85 residues apart along the sequence. Since the identification of the primary contact is determined by the hydrophobicity of the segments forming the contact, polarity values obtained from the Fauchere−Pliska scale[31] were assigned to each residue. To summarize the procedure, the hydrophobicity P_k of each residue is added over a contact window of 5 amino acids, resulting in a polarity P_i of a potential contact segment centered at residue i. To determine the best contact, the P_i value of a segment centered at residue i is added to the P_j value of a segment centered at residue j located 65−85 residues away from i, to give a contact propensity $P_{ij} = P_i + P_j$. The ij pair along the sequence separated by 65−85 residues that produces the highest value of P_{ij} is selected as the primary contact. Differences in P_{ij} larger than ~0.4 reflect differences in ΔG_{hyd} larger than kT.[31]

For the purpose of determining the activation barriers E^i governing the formation of native contacts in the cooperative collapse phase, the amino acids L, W, F, V, M, and I are assigned hydrophobicity values from the Fauchere−Pliska scale. All other amino acids are considered to be non-hydrophobic and assigned a hydrophobicity of zero. A segment size of 15

amino acids is chosen, and the results are robust for windows between 13 and 17 amino acids. This length is sufficient for even the largest possible secondary structure elements (i.e., the ω loops[32]) in the cooperative collapse sequence.

The hydrophobicities P_k of 15 consecutive residues centered at residue j are summed, resulting in a hydrophobicity value H_j. The H_j's are calculated for all possible segments of 15 amino acids along the protein sequence. To determine the sequence of folding events, the 15 amino acids with the lowest H_j which do not overlap with each other are sequentially chosen. These segments are assumed to reach their native structure in increasing order of H_j, because the activation energy E^j for each protein segment is assumed to be directly proportional to its hydrophobicity represented by H_j (i.e., a large H_j value means that more water needs to be excluded and more hydrophobic side chains fixed in the protein core upon structure optimization).

3.3. Primary Contacts of Apoleghemoglobin. The best predicted primary contacts in apoleghemoglobin are established between (a) residues 43−47 and 109−113 with a $P_{ij} = 11.4$ and (b) residues 65−69 and 136−140 with a $P_{ij} = 11.2$, the difference in formation propensity between primary contacts a and b is smaller than kT, $\Delta P_{ij}(a,b) = 0.2 < kT$, so the protein will be assumed to fold through the two pathways defined by the two primary contacts in parallel. It is possible that there are constraints in the subsequent folding phases for any of the two dominant folding pathways that prevent the attainment of the native structure. The next best primary contact is established between residues 27−31 and 101−105 with $P_{ij} = 10.7$. The predicted primary contacts occur (a) between a segment comprising the C-terminal amino acid of helix C and four adjacent residues in the native structure and helix G in the native structure and (b) between segments located in helices E and H in the native structure. This is distinct from the predicted dominant primary contact of apomyoglobin established between residues 28−32 in helix B and 111−115 in helix G.[1] The predicted primary contact a for apoleghemoglobin correlates well with experimental observations. Residues L109, L110, T112, I113, and K114 in the primary contact region are among the first to enter the folding process in good correlation with the SCM prediction. No residues have been experimentally probed in the 43−47 segment,[12] so no conclusions can be drawn at this stage regarding its involvement in the earliest folding stages. The location of primary contact b also correlates well with the experimental result that helix H and a segment of helix E including the hydrophobic residues L65, F66, L68, and V69 included in primary contact b are involved in the earliest detectable folding events along the folding pathway within ∼6 ms of refolding initiation.[12] The hydrophobic residue L136 included in primary contact b is also observed to be involved in the earliest events along the folding pathway, while I140 is also observed to fold somewhat slower.[12] Residues Y133, D134, E135, and A137 are also involved in the first detectable intermediate, consistent with the predicted location of primary contact b. Finally, it must be noted that the SCM prediction that there are two possible folding pathways involving segments included in helices G, E, and H in the native structure in the formation of the first stable contacts along two parallel folding pathways is not the only possible interpretation of the proton exchange results. It is also possible on the basis of the experimental results that helices H and G form a first stable contact on which helix E would dock afterward in a single pathway.[12] This issue might be resolved when experimental time resolution is significantly improved and would constitute a significant test for the SCM.[16] The existence of two parallel pathways correlates well also with the observation that the refolding kinetics of apoleghemoglobin can be well approximated by a double exponential.[12]

The next best predicted contact for apoleghemoglobin is established between segments located in helices B and G in the native structure. The location of this contact in the native structure is roughly similar to that of the best predicted contact for apomyoglobin and its $P_{ij} = 10.7$ is only $\sim kT$ less stable than the primary contacts for the dominant folding pathways a and b. This small difference could potentially be overcome by a single mutation given that a single hydrophobic amino acid contributes $\sim 3-4\ kT$ to P_{ij}.[31] This result suggests that several folding pathways may be available within a protein family to reach the native fold and that the dominant pathway for a given protein is a function of its amino acid sequence and of the experimental conditions.[33] The availability of more than one folding pathway for a given fold is possibly a mechanism aimed to provide robustness.

3.4. Folding of the Apoleghemoglobin Tail in Pathway a. Formation in apoleghemoglobin of primary contact a between residues 43−47 and the C-terminal region of helix G leaves two unfolded tails including the residues that form helices A, B, and most of C in the native structure in the N-terminal tail, and helix H in the C-terminal tail. There is also an open primary loop formed between residue 48 and helix G. The residues in the N-terminal tail are experimentally observed to fold slowly.[12] The MGLIS state of folding pathway a for apoleghemoglobin as defined in the SCM (i.e., the primary contact plus folded tails) should include at least most of helices A, B, C, G, and H. The SCM predicts in general that the tail will fold faster than the open primary loop. Residues in helix F included in the primary loop were not probed in the experiment,[12] so no conclusions can be drawn about the stage at which they become involved in the folding process.

The N-terminal tail in folding pathway a is 42 residues long, significantly longer than the tails defined by the primary contact in the proteins studied previously.[1] In fact, the length of the N-terminal apoleghemoglobin tail a is comparable to the length of existing proteins, shorter than n_{op}, that fold spontaneously and whose folding mechanism has not yet been studied in detail within the SCM context. The folding of the N-terminal apoleghemoglobin tail a may follow a mechanism similar to that of proteins of comparable length, possibly different from the simple nucleation process around the primary contact assumed to allow for the folding of very short tails in the SCM.[1]

3.5. Folding of the Apoleghemoglobin Tail in Pathway b. Formation in apoleghemoglobin of primary contact b between helices E and H leaves an unfolded tail including the residues that form helices A, B, and C in the native structure and an open primary loop formed mainly by the residues between helices E and H in the native structure. The residues in the tail are experimentally observed to fold slowly.[12] The MGLIS state of apoleghemoglobin for folding pathway b as defined in the SCM (i.e., the primary contact plus folded tails) should include at least most of helices A, B, C, E, and H. The SCM predicts in general that the tail will fold faster than the open primary loop. As explained for pathway a, residues in helix F included in the primary loop were not probed in the experiment,[12] so no conclusions can be drawn about the stage at which they become involved in the folding process.

The tail in folding pathway b of apoleghemoglobin is 64 residues long, even longer than the N-terminal tail generated in pathway a. Then, it is also likely that the folding of the apoleghemoglobin tail b may follow a mechanism similar to

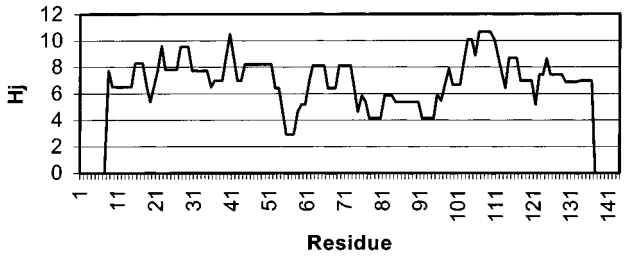

Figure 1. Relative hydrophobicity H_j versus the center of the 15 amino acid segments for apoleghemoglobin. The plot shows that inside the primary loop in folding pathway a, the native structure should be attained sequentially by segments 48−63, 71−85, and 85−99. In folding pathway b, the native structure should be attained sequentially by segments 71−85 and 85−99. Windows shorter than 15 amino acids at the ends of the plot were set to zero.

that of proteins of comparable length, possibly different from the simple nucleation process around the primary contact assumed to allow for the folding of very short tails in the SCM.[1]

3.6. Optimization Subphase of Apoleghemoglobin in Pathway a. The hydrophobicity values H_j for the cooperative collapse of the primary loop of apoleghemoglobin are shown in Figure 1. Within the primary loop defined by primary contact a, it is predicted that regions centered on residues 56, 78, and 92 should be the first to attain their native structure after the collapse of the primary loop. These regions form the C-terminal part of helix E, most of helix F and the N-terminal part of helix G. The H_j values are equal for regions 78 and 92, so they should attain their native structure more or less simultaneously. In apomyoglobin, the first regions to attain their native structure inside the primary loop were centered on residues 59 and 89, correlating well with the regions centered at 56 and 92 in folding pathway a of apoleghemoglobin.[1]

3.7. Optimization Subphase of Apoleghemoglobin in Pathway b. In folding pathway b it is predicted that regions centered on residues 78 and 92 inside the primary loop should be the first to attain their native structure after the collapse of the primary loop. The optimization subphase b of apoleghemoglobin is significantly different from that of apomyoglobin. In apomyoglobin, the first regions to attain their native structure inside the primary loop were centered on residues 59 and 89.[1] The difference is not surprising due to the large distinction in the location of the primary contact in both proteins. The minimum of H_j centered on residue 56 in apoleghemoglobin corresponding to the region centered around 59 in apomyoglobin is outside the primary loop in folding pathway b of apoleghemoglobin.

To summarize, the SCM folding pathways of apoleghemoglobin and apomyoglobin are distinct. Moreover, there are two distinct and coexisting dominant folding pathways for apoleghemoglobin. In terms of the helices observed in the native structure, the SCM predicts the following dominant folding pathways for both proteins:

Apomyoglobin: U → BG → ABGH → N

Apoleghemoglobin a: U → (43−47)G → ABCGH → N

Apoleghemoglobin b: U → EH → ABCEH → N

These differences correlate well with the experimental results.

4. Conclusions

In this paper, the SCM folding pathway of two proteins with similar structures, apoleghemoglobin and apomyoglobin, were compared. Their folding pathways were found to differ significantly, and the differences were shown to be generally consistent with existing experimental results. Moreover, two distinct and parallel folding pathways were found to be dominant in apoleghemoglobin as compared to a single one for apomyoglobin. This difference in folding pathway within the globin family is unexpected.

The results presented here are consistent with previous observations that folding pathways are not be evolutionarily conserved. The nonconservation of the folding pathway could be taken to imply that the main determinants of the native structure lie in the relative contributions of the amino acids of the primary sequence to the stability of the final state rather than in the sequence of events leading to the native structure.[12] This rather extreme hypothesis could possibly be tested through directed mutagenesis experiments, if, for example, several suitable mutants for the proteins studied could be found that are able to reach the native structure through formation of a different primary contact than the one defined by the native primary sequence. On the other hand, a small degree of degeneracy in the number of folding pathways associated with a single structural type could just represent an added degree of robustness to mutations in the folding code. These are important issues that will only be resolved when laboratory data for the folding pathways of a significant number of structurally related proteins becomes available.

Acknowledgment. H. R. acknowledges support from the Petroleum Research Fund of the American Chemical Society

References and Notes

(1) Bergasa-Caceres, F.; Ronneberg, T. A.; Rabitz, H. A. *J. Phys. Chem. B* **1999**, *103*, 9749.
(2) Bergasa-Caceres, F.; Rabitz, H. A. *J. Phys. Chem. B* **2001**, *105*, 2874.
(3) Hollecker, M.; Creighton, T. E. *J. Mol. Biol.* **1983**, *168*, 409.
(4) Krebs, H.; Schmid, F. X.; Jaenicke, R. *J. Mol. Biol.* **1983**, *169*, 619.
(5) Stackhouse, T. M.; Onuffer, J. J.; Matthews, C. R.; Ahmed, S. A.; Miles, E. W. *Biochemistry* **1988**, *27*, 824.
(6) Jaenicke, R. *Prog. Biophys. Mol. Biol.* **1987**, *49*, 117.
(7) Plaxco, K. W.; Spitzfaden, C.; Campbell, I. D.; Dobson, C. M. *J. Mol. Biol.* **1997**, *270*, 763.
(8) Chiti, F.; Taddei, N.; White, P. M.; Bucciantini, M.; Magherini, F.; Stefani, M.; Dobson, C. M. *Nat. Struct. Biol.* **1999**, *6*, 1005.
(9) Martínez, J. C.; Serrano, L. *Nat. Struct. Biol.* **1999**, *6*, 1010.
(10) Riddle, D. S.; Grantcharoav, V. P.; Santiago, J. V.; Alm, E.; Ruczinski, I.; Baker, D. *Nat. Struct. Biol.* **1999**, *6*, 1016.
(11) Ternstrom, T.; Mayor, U.; Akke, M.; Oliveberg, M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14854.
(12) Nishimura, C.; Prytulla, S.; Dyson, H. J.; Wright, P. E. *Nat. Struct. Biol.* **2000**, *7*, 679.
(13) Landsmann, J.; Dennis, E. S.; Higgings, T. J. V.; Appleby, C. A.; Kortt, A. A.; Peacock, W. J. *Nature* **1986**, *324*, 166.
(14) Jennings P. A.; Wright P. E. *Science* **1993**, *262*, 892.
(15) Loh S. N.; Kay M. S.; Baldwin R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 5446.
(16) Ballew, R. M.; Sabelko, J.; Gruebele, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5759.
(17) Eliezer, D.; Yao, J.; Dyson, H. J.; Wright, P. E. *Nat. Struct. Biol.* **1998**, *5*, 148.
(18) Dill, K. A. *Biochemistry* **1990**, *29*, 7133.
(19) Schulz, G. E.; Schirmer, R. H. *Principles of Protein Structure*; Springer-Verlag: New York, 1979.
(20) Kuwajima, K. *Proteins: Struct. Funct. Genet.* **1989**, *6*, 87.
(21) Plaxco, K. W.; Millett, I. S.; Segel, D. J.; Doniach, S.; Baker D. *Nature Struct. Biol.* **1999**, *6*, 554.
(22) Sali, A.; Shakhnovich, E.; Karplus, M. *Nature* **1994**, *369*, 248.
(23) Bryngelson, J. D.; Onuchic J. N.; Socci, N. D.; Wolynes, P. G. *Proteins: Struct. Funct. Genet*. **1995**, *21*, 167.
(24) Shoemaker, B. A.; Wolynes, P. G. *J. Mol. Biol.* **1999**, *287*, 657.
(25) Shoemaker, B. A.; Wang, J.; Wolynes, P. G. *J. Mol. Biol.* **1999**, *287*, 675.
(26) Levinthal, C. J. *J. Chim. Phys.* **1968**, *65*, 44.

(27) Lazaridis, T.; Karplus, M. *Science* **1999**, *278*, 1928.
(28) Ptitsyn, O. B.; Ting, K. L. *J. Mol. Biol.* **1999**, *291*, 671.
(29) Ellis, P. J.; Appleby, C. A.; Guss, J. M.; Hunter, W. N.; Ollis, D. L.; Freeman, H. C. *Acta Crystallogr.* **1997**, *D53*, 302.
(30) Kuriyan, J.; Wilz, S.; Karplus, M.; Petsko, G. A. *J. Mol. Biol.* **1986**, *192*, 133.
(31) Fauchère, J. L.; Pliska, V. *Eur. J. Med. Chem.* **1983**, *18*, 369.
(32) Leszcynski, J.; Rose, G. D. *Science* **1986**, *234*, 849.
(33) Damberger, F.; Nikonova, L.; Horst, G.; Peng, G.; Leal, W. S.; Wüthrich, K. *Protein Science* **2000**, *9*, 1038.