Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 11615–11617, October 1996
Biophysics

Diffusion-limited contact formation in unfolded cytochrome *c*: Estimating the maximum rate of protein folding

(laser photolysis/polymer dynamics)

STEPHEN J. HAGEN, JAMES HOFRICHTER, ATTILA SZABO, AND WILLIAM A. EATON*

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, Building 5, National Institutes of Health, Bethesda, MD 20892-0520

Communicated by Robin M. Hochstrasser, University of Pennsylvania, Philadelphia, PA, June 24, 1996 (received for review June 10, 1996)

ABSTRACT **How fast can a protein fold? The rate of polypeptide collapse to a compact state sets an upper limit to the rate of folding. Collapse may in turn be limited by the rate of intrachain diffusion. To address this question, we have determined the rate at which two regions of an unfolded protein are brought into contact by diffusion. Our nanosecond-resolved spectroscopy shows that under strongly denaturing conditions, regions of unfolded cytochrome *c* separated by ~50 residues diffuse together in 35–40 μs. This result leads to an estimate of ~(1 μs)$^{-1}$ as the upper limit for the rate of protein folding.**

Recent advances in the study of protein folding on the submillisecond time scale (1–7) and the observation of very rapid folding in many systems (8–17) raise an important question: how fast can a protein fold? The energy landscape theory of Wolynes and coworkers (18, 19) indicates that some proteins may fold without encountering a thermodynamic barrier, so that the folding rate is limited only by the speed of collapse to a compact structure. If regions of the polypeptide chain come into contact by diffusion (20–22), then the rate of intrachain diffusion provides the upper limit to the rate of collapse. We have studied this limit experimentally by determining the rate of diffusional contact formation between two regions of an unfolded protein. This rate may be seen as a unimolecular analogue to Smoluchowski's diffusion-limited rate for the bimolecular reaction of molecules free in solution, and it also allows us to estimate an upper limit for the rate of protein folding.

Jones *et al.* (1) measured the rate at which the heme of unfolded cytochrome *c* binds the methionine residues that are located 50–60 positions away on the 104-aa chain. To determine from this the rate of intrachain diffusion of the heme and ligand, we measured the rate of bimolecular binding of free methionine to the heme iron of a cytochrome *c* peptide. Our analysis is based on the two-step description represented by Fig. 1. In this model, the heme and ligand first diffuse together until they share a small reaction volume, forming an encounter complex. This encounter complex is identical to the geminate complex produced by photodissociation of a heme–ligand pair: it either dissociates (at a rate k_{D-}), or reacts (at the geminate rate k_{gem}) to form a covalent complex. In the steady-state approximation, the overall rate of binding is k_{on}, where

$$1/k_{on} = 1/k_{D+} + 1/(Kk_{gem}). \quad [1]$$

$K \equiv k_{D+}/k_{D-}$ is the equilibrium constant for forming the encounter complex. Eq. **1** can also be obtained from the theory of partially diffusion-controlled reactions (23, 24). We assume that the two-step description and Eq. **1** apply both to a unimolecular heme–ligand reaction—the reaction between two chemical groups residing on the same polypeptide chain—and to a bimolecular reaction—the binding of the heme to a ligand that is free in solution. Although K and k_{D+} are expected to be very different in the two cases, k_{gem} is presumed to be the same. Applying Eq. **1** to both the unimolecular and bimolecular reactions then allows us to calculate the intrachain diffusion rate k_{D+}^{uni} from the experimentally determined bimolecular and unimolecular rates, k_{on}^{bi} and k_{on}^{uni}, respectively. Wang and Davidson (25, 26) used a similar approach to show that ring closure of lambda DNA is reaction-limited (i.e., $k_{on} \approx Kk_{gem}$).

MATERIALS AND METHODS

Jones *et al.* (1) measured k_{on}^{uni} for the formation of heme–ligand intrachain complexes of cytochrome *c* unfolded in 5.6 M guanidine hydrochloride (GuHCl) at 40°C. They photodissociated the heme–carbon monoxide complex and used nanosecond-resolved spectroscopy to show that the heme, which is covalently attached to His-18, binds Met-65 and Met-80 (the native heme ligand) at a unimolecular rate $k_{on}^{uni} \approx (40\ \mu s)^{-1}$. To determine k_{D+}^{uni} from this result, we measured the bimolecular rate k_{on}^{bi} for the binding of free methionine to the heme of cytochrome *c*.

We eliminated competing intramolecular ligand-binding reactions by studying methionine binding to an 11-residue heme peptide obtained by enzymatic digestion of horse cytochrome *c*. The peptide, known as microperoxidase, consists of residues 11–21 of the intact cytochrome, including the heme group covalently attached to residues Cys-14, Cys-17, and His-18. Under pseudo first-order conditions (i.e., at ligand concentrations of ~5-100 mM and peptide concentrations of ~100 μM), bimolecular binding to the heme of microperoxidase occurs rapidly, at rates of ≈10^6 to 10^7 s^{-1}. At lower ligand concentrations the peptide aggregates, which rules out stopped-flow studies of ligand binding. To study the rapid rebinding, we used a nanosecond laser pulse to photodissociate the microperoxidase–methionine complex and then observed rebinding by collecting time-resolved absorption difference spectra in the heme Soret region (~390-450 nm). Under each set of solution conditions, we collected ~50 spectra at logarithmically spaced time intervals from $t = -10$ ns to $t = +30$ μs after the photolyzing laser pulse. We analyzed these spectra by singular value decomposition, which shows that the time-dependent spectral changes following photolysis essentially consist of a single, exponential relaxation to the equilibrium spectrum. The nanosecond-resolved spectrometer (27) and the singular value decomposition method of data analysis (28) have been described elsewhere. The microperoxidase samples contained 100 μM peptide and 12.5–100 mM *N*-acetyl-L-methionine in 5.6 M GuHCl, pH 6.5/0.1 M phosphate, sealed anaerobically within a quartz cuvette.

The photolysis yield was 6–7%. The bimolecular rebinding of methionine can be observed with greater signal-to-noise in

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*To whom reprint requests should be addressed.

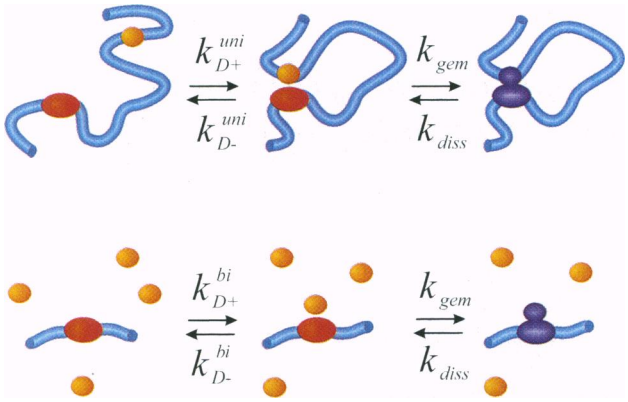

FIG. 1. The two-step model for ligand binding by a heme on a polypeptide. The heme and ligand diffuse together at a rate k_{D+} to form the "encounter complex," which then either separates (rate k_{D-}) or reacts ("geminate" rate k_{gem}) to give the covalent complex. (The covalent complex dissociates spontaneously at a rate $k_{diss} \ll k_{gem}$.) The model is shown for the unimolecular reaction (*Upper*) and the bimolecular reaction (*Lower*).

the presence of carbon monoxide (CO), although with more complicated kinetics. The CO binds tightly to the heme at equilibrium but is easily photolyzed, resulting in large spectral changes as free methionine binds to the heme after photolysis. The rate of methionine binding in such experiments is consistent with the data presented here, and the observed overall dissociation rate is consistent with the results of Jones *et al.* (1) and with our estimated chain relative diffusion constant (see below) $D \approx 4 \times 10^{-7}$ cm^2/s (unpublished results).

RESULTS AND DISCUSSION

Fig. 2 shows the spectrum of the methionine complex (at 100 mM methionine) together with that of the photoproduct at 10 ns. The photoproduct spectrum is identical to the spectrum obtained by Jones *et al.* (1) for the photodissociated CO complex of unfolded cytochrome *c*. It is the spectrum of a five-coordinate deoxyheme like deoxymyoglobin, indicating that His-18 remains bound to the proximal side of the heme iron after photolysis. The subsequent rebinding of methionine, observed as a decrease in the amplitude of the difference

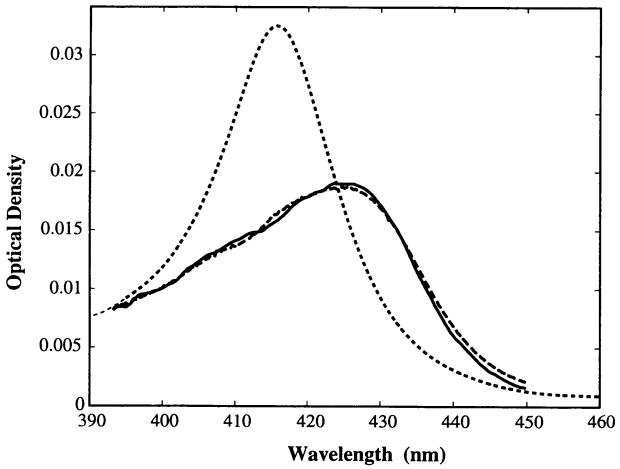

FIG. 2. Equilibrium spectrum (dotted line) and spectrum of the 10-ns photoproduct (solid line) of the *N*-acetyl-L-methionine complex of microperoxidase. Also shown is the deoxyheme spectrum observed by Jones *et al.* (1) after photolysis of the CO complex of cytochrome *c* (dashed line). The optical density scale is that of the microperoxidase photoproduct, with the other two spectra scaled to the correct relative amplitude.

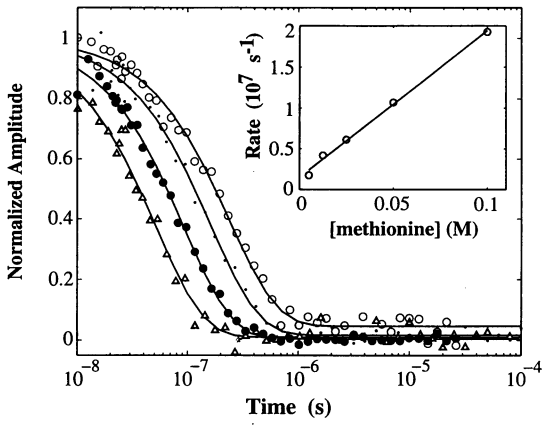

FIG. 3. Time-dependent amplitude (normalized to unity for time $t \rightarrow 0$) of the absorption difference spectrum after photodissociation of the heme–methionine complex by a 10-ns laser pulse at 533 nm. Data shown are for methionine concentrations of 12.5 mM (○), 25 mM (·), 50 mM (●), and 100 mM (△), in 5.6 M GuHCl, pH 6.5/0.1 M phosphate at T = 40°C. The amplitudes of the difference spectra can be fit to a single exponential decay (solid lines). (*Inset*) The rate obtained from the exponential fit is proportional to the free methionine concentration, as expected for bimolecular, pseudo first-order rebinding.

spectrum, consists of a single exponential phase (Fig. 3). The rate of the exponential is proportional to the free methionine concentration (Fig. 3 *Inset*), with a slope equal to the bimolecular binding rate, $k_{on}^{bi} \approx 1.8 \times 10^8$ M^{-1}·s^{-1} at 40°C. We combine this rate with Eq. **1** [and the data of Jones *et al.* (1)] to calculate the intrachain diffusion rate of unfolded cytochrome *c*.

Because this bimolecular rate is nearly diffusion-limited,† one may expect that the unimolecular heme–methionine reaction is rate-limited by intrachain diffusion. In fact, an analysis of the observed rates shows that k_{on}^{bi} is large enough that, even if it is reaction-controlled, the unimolecular reaction must still be dominated by intrachain diffusion. Using Eq. **1**, we calculate the intrachain diffusion rate k_{D+} of the polypeptide. From the bimolecular form of Eq. **1**, $1/K^{bi}k_{gem} = 1/k_{on}^{bi} - 1/k_{D+}^{bi}$, whereas from the unimolecular form, $1/k_{D+}^{uni} = 1/k_{on}^{uni} - 1/K^{uni}k_{gem}$. Because k_{gem} is presumed to be the same for the unimolecular and bimolecular reactions, we can eliminate it from these two equations, giving

$$1/k_{D+}^{uni} = 1/k_{on}^{uni} - (K^{bi}/K^{uni})(1/k_{on}^{bi} - 1/k_{D+}^{bi}). \quad [2]$$

The terms on the right-hand side are either known from experiment or can be estimated. k_{D+}^{bi} lies within the range $0 \leq 1/k_{D+}^{bi} \leq 1/k_{on}^{bi}$. For small reaction volumes, $K^{bi}/K^{uni} = \int d\mathbf{r} \exp[-U(\mathbf{r})/k_BT]$, where $U(\mathbf{r})$ is the potential of mean force between the heme and its ligand. If the chain adopts a random coil configuration, the ratio of equilibrium constants is $K^{bi}/K^{uni} = (2\langle r^2\rangle\pi/3)^{3/2}$, which has been called the Jacobsen–Stockmayer factor (25, 26, 30). Here $\langle r^2\rangle$ is the mean-squared separation between the heme and ligand.‡ We consider the

†The bimolecular encounter complex forms at the Smoluchowski diffusional rate, $k_{D+}^{bi} = 2\pi D_0 a \approx 6 \times 10^8$ M^{-1}·s^{-1}. Here D_0 is essentially the free methionine diffusion constant ($\approx 10^{-5}$ cm^2/s) and $a \approx 1.5$ Å is the effective reaction radius. The value of a, which is smaller than typical for the sum of Smoluchowski reaction radii, was obtained by Miers *et al.* by modeling geminate binding of CO to heme (29). The above Smoluchowski rate has been reduced by a factor of two because methionine binds only to the distal side of the heme. Stereochemical factors that reduce the rate of diffusion-controlled reaction between the heme and ligand are incorporated by shrinking the value of a.

‡For a wide variety of proteins unfolded in 6 M GuHCl at 20°C, the dimension of the unfolded molecule scales with the number of residues

heme–Met-80 reaction, for which $\langle r^2 \rangle \approx (85\ \text{Å})^2$, although the final result is nearly identical for the heme–Met-65 reaction. Eq. 2 then gives $1/k_{D+}^{uni} = 34$–$40\ \mu s$, where the range in k_{D+}^{uni} reflects the range of possible values for k_{D+}^{bi}. Note that this argument does not require a microscopic model or estimate for k_{D+}^{bi}, although the range of values calculated for k_{D+}^{uni} does depend on the ratio of equilibrium constants K^{bi}/K^{uni}. Of the 40 μs required for the heme to bind to a residue located 62 positions away, we find that 35–40 μs represents the formation of the encounter complex.§ Thus we obtain the characteristic time scale, $1/k_{D+}^{uni} \approx 35$–$40\ \mu s$ for the diffusion-controlled formation of a ≈60-residue loop in an unfolded polypeptide.

This result represents the first experimental determination of the time scale for contact formation between two regions of an unfolded polypeptide. Can we relate this to the rate of collapse of a random coil to a compact structure? Although a number of theoretical models have been proposed (e.g., refs. 20–22 and 37), unfortunately the actual mechanism of polypeptide collapse remains unknown. Even for homopolymers, the process of collapse remains poorly understood. Ostrovsky and Bar-Yam (38) proposed that hydrophobic collapse of a homopolymer occurs essentially through intrachain diffusion, whereas de Gennes (39) and Grosberg *et al.* (40) described hydrophobic collapse as a hydrodynamically controlled crumpling.

In spite of these uncertainties, it seems to us that a polypeptide cannot collapse to a compact structure faster than a small loop can form. To obtain the rate for such loops, we can scale our measured rate using the simplest theory for diffusion-controlled contact formation in a polymer (24). For a random coil polypeptide, this theory predicts that loops of n residues form at a rate proportional to $n^{-3/2}$. From a survey of protein structures, Leszczynski and Rose (41) have found that the shortest loops in proteins contain ~6-10 residues, which would then be expected to form in ~1-3 μs. The equilibrium theory of Thirumalai and coworkers (42, 43) predicts that the fastest loops (~10 residues, shorter loops forming more slowly because of chain stiffness) form 30–40 times faster than loops of 50–60 residues, again suggesting a minimum time of ~1 μs. Thus we conclude that $\sim 10^6\ s^{-1}$ should be an approximate upper limit on the rate of collapse of a random coil protein to a compact structure.

Finally, we suggest that the mechanism of protein collapse and an upper limit for its rate have biological significance. The hydrophobic residues of polypeptides synthesized *in vivo* must be concealed rapidly if aggregation is to be prevented in unchaperoned proteins. Once a molecule is compact, with no "sticky" hydrophobic patches on its surface, formation of the native structure may proceed more slowly. A successful sequence must not only form a functional folded structure but must also collapse rapidly.

as if the chain were a random coil (31). That is, the mean-squared separation between residues separated by n positions is given by $\langle r^2 \rangle \approx C_n n l^2$, where $l = 3.8$ Å is the distance between α-carbon atoms of adjacent residues and $C_n \approx 8$ is the observed value of Flory's characteristic ratio (32) in 6 M GuHCl. Therefore, we expect $\langle r^2 \rangle \approx (85\ \text{Å})^2$ for $n = 62$.

§With this value of k_{D+}^{uni}, we can use the Smoluchowski-equation model of Szabo *et al.* (24) to estimate the diffusion constant D for the relative motion of two positions on a polypeptide chain. The relation (24) $k_{D+}^{uni} \approx 3Da(3/2\pi)^{1/2}/\langle r^2 \rangle^{3/2}$ indicates $D \approx 4 \times 10^{-7}\ \text{cm}^2/\text{s}$, which is ~20 times smaller than the expected monomer diffusion rate $D_0 \approx 10^{-5}\ \text{cm}^2/\text{s}$. Our estimate for D is consistent with the only other experimental data on end-to-end diffusion of polypeptides. Haas and coworkers have used intrachain fluorescence energy transfer to study the equilibrium distribution of end-to-end distances r and the diffusion constant D. For a 20-residue segment of protein unfolded in 6 M GuHCl at 22°C, they find $D \approx 4.7 \times 10^{-7}\ \text{cm}^2/\text{s}$ (33–36).

We thank Abel Schejter, Peter Wolynes, and Robert Zwanzig for helpful discussion.

1. Jones, C. M., Henry, E. R., Hu, Y., Chan, C. K., Luck, S. D., Bhuyan, A., Roder, H. Hofrichter, J. & Eaton, W. A. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 11860–11864.
2. Huang, G. S. & Oas, T. G. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 6878–6882.
3. Phillips, C. M., Mizutani, Y. & Hochstrasser, R. M. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 7292–7296.
4. Williams, S., Causgrove, T. P., Gilmanshin, R., Fang, K. S., Callender, R. H., Woodruff, W. H. & Dyer, R. B. (1996) *Biochemistry* **35,** 691–697.
5. Nölting, B., Golbik, R. & Fersht, A. R. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 10668–10672.
6. Ballew, R. M., Sabelko, J. & Gruebele, M. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 5759–5764.
7. Pascher, T., Chesick, J. P., Winkler, J. R. & Gray, H. B. (1996) *Science* **271,** 1558–1560.
8. Alexander, P., Orban, J. & Bryan, P. (1992) *Biochemistry* **31,** 7243–7248.
9. Elöve, G. A., Bhuyan, A. K. & Roder, H. (1994) *Biochemistry* **33,** 6925–6935.
10. Khorasanizadeh, S., Peters, I. D. & Roder, H. (1996) *Nat. Struct. Biol.* **3,** 193–205.
11. Kragelund, B. B., Robinson, C. V., Knudsen, J., Dobson, C. M. & Poulsen, F. M. (1995) *Biochemistry* **34,** 7217–7224.
12. Kuszewski, J., Clore, G. M. & Gronenborn, A. M. (1994) *Protein Sci.* **3,** 1945–1952.
13. Ptitsyn, O. B. (1995) *Adv. Protein Chem.* **47,** 83–229.
14. Schindler, T., Herrler, M., Marahiel, M. A. & Schmid F. X. (1995) *Nat. Struct. Biol.* **2,** 663–673.
15. Sosnick, T. R., Mayne, L., Hiller, R. & Englander, S. W. (1994) *Nat. Struct. Biol.* **1,** 149–156.
16. Baldwin, R. L. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 2627–2628.
17. Waldburger, C. D., Jonsson, T. & Sauer, R. T. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 2629–2634.
18. Bryngelson, J. D. & Wolynes, P. G. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7524–7528.
19. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995) *Proteins Struct. Funct. Genet.* **21,** 167–195.
20. Karplus, M. & Weaver, D. L. (1976) *Nature (London)* **260,** 404–406.
21. Karplus, M. & Weaver, D. L. (1979) *Biopolymers* **18,** 1421–1437.
22. Karplus, M. & Weaver, D. L. (1994) *Protein Sci.* **3,** 650–668.
23. Shoup, D. & Szabo, A. (1982) *Biophys. J.* **40,** 33–39.
24. Szabo, A., Schulten, K. & Schulten, Z. (1980) *J. Chem. Phys.* **72,** 4350–4357.
25. Wang, J. C. & Davidson, N. (1966) *J. Mol. Biol.* **15,** 111–123.
26. Wang, J. C. & Davidson, N. (1966) *J. Mol. Biol.* **19,** 469–482.
27. Hofrichter, J., Henry, E. R., Ansari, A., Jones, C. M., Deutsch, R. M. & Sommer, J. H. (1994) *Methods Enzymol.* **232,** 387–415.
28. Henry, E. R. & Hofrichter, J. (1992) *Methods Enzymol.* **210,** 129–192.
29. Miers, J. B., Postlewaite, J. C., Zyung, T., Chen, S., Roemig, G. R., Wen, X., Dlott, D. D. & Szabo, A. (1990) *J. Chem. Phys.* **93,** 8771–8776.
30. Jacobson, H. & Stockmayer, W. H. (1950) *J. Chem. Phys.* **18,** 1600–1606.
31. Damaschun, G., Gast, K., Müller-Frohne, M. & Zirwer, D. (1996) *Biopolymers*, in press.
32. Flory, P. J. (1969) *Statistical Mechanics of Chain Molecules* (Wiley, New York).
33. Haas, E., Katchalski-Katzir, E. & Steinberg, I. Z. (1978) *Biopolymers* **17,** 11–31.
34. Beechem, J. M. & Haas, E. (1989) *Biophys. J.* **55,** 245–269.
35. Gottfried, D. S. & Haas, E. (1992) *Biochemistry* **31,** 12353–12362.
36. Buckler, D. R., Haas, E. & Scheraga, H. A. (1995) *Biochemistry* **34,** 15965–15978.
37. Dill, K. A., Fiebig, K. M. & Chan, H. S. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 1942–1946.
38. Ostrovsky, B. & Bar-Yam, Y. (1994) *Europhys. Lett.* **25,** 409–414.
39. de Gennes, P. G. (1985) *J. Phys. Lett.* **46,** L639–L642.
40. Grosberg, A. Y., Nechaev, S. K., Shakhnovich, E. I. (1988) *J. Phys.* **49** 2095–2100.
41. Leszczynski, J. F. & Rose, G. D. (1986) *Science* **234,** 849–855.
42. Guo, Z. & Thirumalai, D. (1995) *Biopolymers* **36,** 83–102.
43. Camacho, J. & Thirumalai, D. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 1277–1281.