1. Schroeder, R. & von Ahsen, U. *Nucleic Acids Mol. Biol.* **10**, 53–74 (1996).
2. Hermann, T. & Westhof, E. *Curr. Opin. Biotechnol.* **9**, 66–73 (1998).
3. Recht, M. I., Fourmy, D., Blanchard, S. C., Dahlquist, K. D. & Puglisi, J. D. *J. Mol. Biol.* **262**, 421–436 (1996).
4. Draper, D. E., Xing, Y. & Laing, L. G. *J. Mol. Biol.* **249**, 231–238 (1995).
5. Wank, H., Rogers, J., Davies, J. & Schroeder, R. *J. Mol. Biol.* **236**, 1001–1010 (1994).
6. Zapp, M. L., Stern, S. & Green, M. R. *Cell* **74**, 969–978 (1993).
7. Hermann, T. & Westhof, E. *J. Mol. Biol.* **276**, 903–912 (1998).
8. Rogers, J., Chang, A. H., von Ahsen, U., Schroeder, R. & Davies, J. *J. Mol. Biol.* **259**, 916–925 (1996).
9. Patel, D. J. *Curr. Opin. Chem. Biol.* **1**, 32–46 (1997).
10. Wallis, M. G. & Schroeder, R. *Prog. Biophys. Molec. Biol.* **67**, 141–154 (1997).
11. Jiang, L., Suri, A. K., Fiala, R. & Patel, D. *J. Chem. Biol.* **4**, 35–50 (1997).
12. Wang, Y. & Rando, R. R. *Chem. Biol.* **2**, 281–290 (1995).
13. Fourmy, D., Recht, M. I., Blanchard, S. C. & Puglisi, J. D. *Science* **274**, 1367–1371 (1996).
14. Kieft, J. S. & Tinoco Jr., I. *Structure* **5**, 713–721 (1997).
15. Correll, C. C., Freeborn, B., Moore, P. B. & Steitz, T. A. *Cell* **91**, 705–712 (1997).
16. Dallas, A. & Moore, P. B. *Structure* **5**, 1639–1653 (1997).
17. Wu, M. & Turner, D. H. *Biochemistry* **35**, 9677–9689 (1996).
18. Battiste, J. L., Tan, R., Frankel, A. D. & Williamson, J. R. *J. Biomol. NMR* **6**, 375–389 (1995).
19. Ye, X., Gorin, A., Ellington, A. D. & Patel, D. J. *Nature Struct. Biol.* **3**, 1026–1033 (1996).
20. Bayens, K. J., De Bondt, H. L. & Holbrook, S. R. *Nature Struct. Biol.* **2**, 56–62 (1995).
21. Wang, Y-X., Huang, S. & Draper, D. E. *Nucleic Acids Res.* **24**, 2666–2672 (1996).
22. Rich, A. & Kim. S. H. *Sci. Am.* **238**, 52–62 (1978).
23. Huang, S., Wang, Y-X. & Draper, D. E. *J. Mol. Biol.* **258**, 308–321 (1996).
24. Cai, Z. *et al. Nature Struct. Biol.* **5**, 203–212 (1998).
25. Legault, P., Li, J., Mogridge, J., Kay, L. E. & Greenblatt, J. *Cell*, **93**, 289–299 (1998).
26. Clouet-d'Orval, B., Stage, T. K. & Uhlenbeck, O. C. *Biochemistry* **34**, 11186–11190 (1995).
27. Weeks, K. M. & Crothers, D. M. *Science* **261** 1574–1577 (1993).
28. Pardi, A. *Nucleic Acids Res.* **261**, 350–380 (1995).
29. Jiang, F., Fiala, R., Live, D., Kumar, R. A. & Patel, D. J. (1996). *Biochemistry* **40**, 13250–13266.

Evidence for an unfolding and refolding pathway in cytochrome *c*

It has been suggested that three partially unfolded forms detected in a native state hydrogen exchange study of oxidized cytochrome *c* may represent sequential intermediates in an unfolding-refolding pathway. To better define the structure of each intermediate a 'stability labeling' method was used in which the stability of a given segment against unfolding was changed. The condition — folded or unfolded — of the stability-labeled segment in each intermediate could then be determined. The structures found are discrete and native-like and are just those necessary for a sequential pathway.

Thermodynamic principles require that protein molecules populate not only their lowest free energy state but also all possible higher energy states according to their Boltzmann factors and that all molecules continually cycle through all of these states. Thus protein molecules must unfold and refold even under native conditions. A native state hydrogen exchange (HX) method can depict the structure and measure the free energy of the unfolded forms through which proteins cycle. The method has been used to study the globally unfolded state of oxidized equine cyt *c*[1] and three large scale partially unfolded intermediates[2]. The intermediates found are produced by the cooperative unfolding of one or more of the structural units shown in Fig. 1. These forms can be referred to operationally as Red open (R → r), Yellow open (Y → y), Green open (G → g), and Blue open (B → b), where upper case indicates the folded form of each unit and lower case the unfolded form. The units are color coded in order of increasing free energy for unfolding (red to blue; Fig 1).

It has been suggested that the partially unfolded forms detected in cyt *c* represent the intermediates rYGB, ryGB and rygB and that these are produced in a stepwise sequential manner according to reaction scheme 1[2]:

$$N(RYGB) \rightarrow I_1(rYGB) \rightarrow I_2(ryGB) \rightarrow I_3(rygB) \rightarrow U(rygb) \quad (1)$$

Because the HX experiments that detected these intermediates were performed under equilibrium native conditions, it follows that each individual unfolding reaction must be matched by an equivalent refolding reaction (the principle of

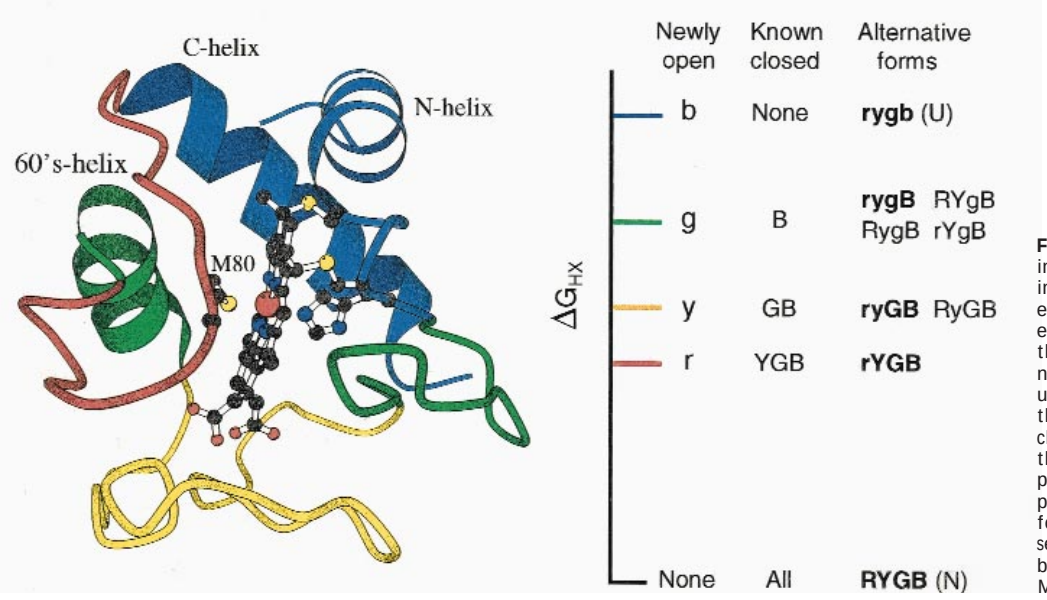

Fig. 1 The structural unfolding units in cyt *c*, color coded in order of increasing free energy for unfolding[2]. The energy level diagram shows the units measured to be newly opened in each unfolded state (lower case), the units known to be still closed (upper case), and therefore the alternative possible identities for each partially unfolded form. The forms specified by the present experiments are in bold. Figure produced using Molscript[35].

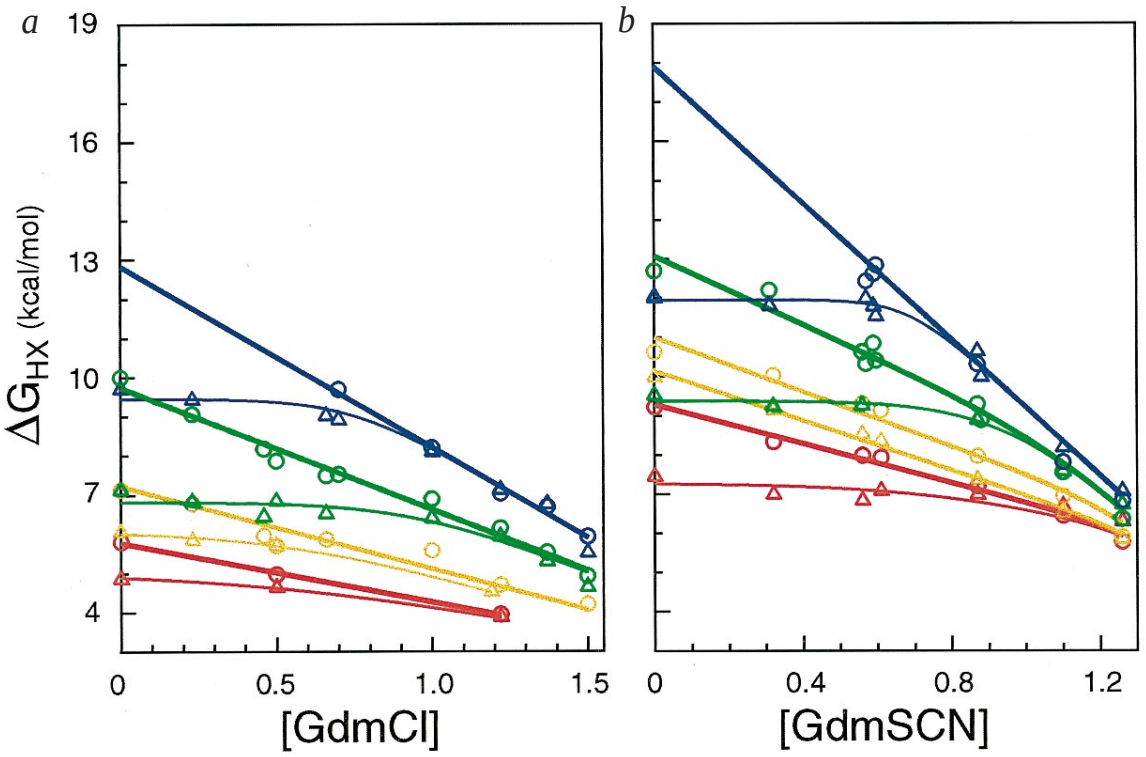

Fig. 2 ΔG_{HX} against denaturant concentration for the H-D exchange of various NHs in ***a***, oxidized (from ref. 2) and ***b***, reduced equine cyt *c* (30 ºC). For each cooperative unfolding unit, one marker residue (large m) and one joining residue (m ~0) are shown (except Yellow in reduced). The color code is the same as in Fig. 1. The marker residues are (top to bottom) Leu 98 (Blue), Leu 68 (Green), Leu 64 and Lys 60 (Yellow), and Ile 75 (Red). The joining residues are (top to bottom; for oxidized and reduced respectively): Phe 10 and Tyr 97 (Blue), Tyr 67 and Met 65 (Green), Gly 37 (Yellow, oxidized only), Ile 85 and Tyr 74 (Red). Different joining residues are shown for reduced cyt *c* because the increase in the ΔG of the large openings unmasks new local exchange pathways for some residues that are markers in oxidized cyt *c* while obscuring the merging behavior of others. Replicated data points represent independent runs at pD 7, 8 and 9 to test for EX2 behavior and for the validity of using higher pD to measure the slower NHs.

microscopic reversibility). Therefore the unfolding sequence, if correct, should reveal a sequential refolding pathway.

The problem addressed here arises because the HX data so far available identify only the newly unfolded segment(s) in each large unfolding but are silent concerning the role of segments that have already exchanged their hydrogens in lower free energy openings. This produces ambiguity in the identity of the partially unfolded forms, as indicated in Fig. 1. A method for distinguishing these alternatives is implemented here.

The approach is to stabilize one cooperative unfolding unit and see how this 'stability labeling' appears in the other unfoldings. Initial experiments presented here stabilize the red unit by changing the redox state of the heme group. Reduced and oxidized cyt *c* have essentially identical structures[3–8], but the reduced form is far more stable. Much of the additional stability represents the increased strength of the Met 80-S ligation to the reduced heme iron. The affinity of methionine for the reduced heme iron, measured in direct binding experiments, is greater than for the oxidized iron by 3.2 kcal mol^{-1} (Table 1). One expects that the additional ligation energy will stabilize the red loop against unfolding because Met 80 is in the red loop (Fig. 1) and its unfolding breaks the bond to the heme iron. This is indeed the case. We then study the higher energy segmental unfoldings. If these unfoldings occur independently of the red loop, then their energy levels will be unaffected by the redox state. However, if they include the red unfolding then they should experience the same ~3.2 kcal mol^{-1} increment in unfolding free energy.

Native state HX

The results for reduced and oxidized cyt *c* obtained by the native state HX method using denaturant concentration as a variable are compared in Fig. 2. The method is designed to measure otherwise invisible large unfoldings by enhancing them to the point that they come to dominate measurable HX behavior. Hydrogens that exchange through small local fluctuations are insensitive to denaturant (slope = m ~0). Hydrogens that exchange through large unfoldings, referred to as marker protons, are sharply accelerated (large m), reflecting the denaturant-dependent stabilization of the open states that they represent. When a large unfolding is promoted in this way, it can come to dominate the exchange of all the hydrogens that it exposes. This is seen as the merging of many NHs into a common HX isotherm. Two representative NHs, a marker NH and a joining NH, for each of the four HX isotherms in cyt *c* are shown in Fig. 2*a*. More complete data measure many residues that join the different isotherms[2], indicating the segments that are newly unfolded in each large unfolding. In oxidized cyt *c* these are the segments color coded in Fig. 1.

The same general pattern of small and large opening behavior seen in oxidized cyt *c* (Fig. 2*a*) is observed in the reduced protein (Fig. 2*b*). The same marker protons show large m values (large unfolding reactions) and these increase in the same sequence and by about the same factors. Evidently, very similar large unfoldings are preserved. We focus on the behavior of the marker protons as probes for the large unfoldings. The free energy of each of the large unfoldings is greatly increased in reduced cyt *c*, while local

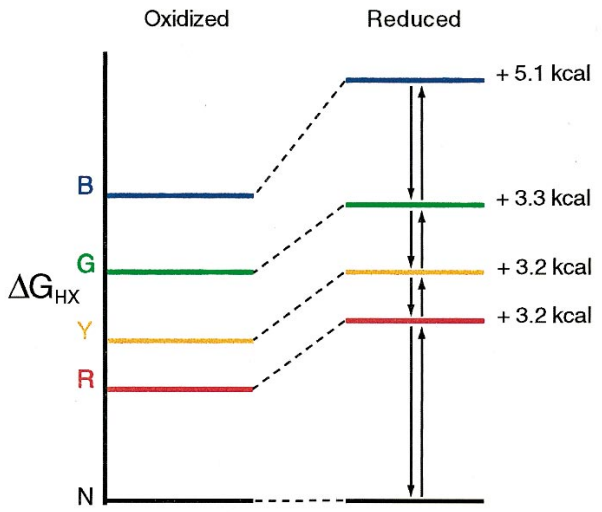

Fig. 3 The effect of redox state on stability of the partial and global unfoldings. The red open, yellow open, and green open states all show the 3.2 kcal mol^{-1} increment associated with the increase in Met 80-S to heme iron affinity upon reduction, indicating that the red unit is unfolded in the yellow open state and the green open state. The putative unfolding-refolding pathway is shown by the arrows. The additional 1.9 kcal mol^{-1} increment in the final unfolding step (blue unfolding) may reflect the exposure of the unfavorably buried net charge (+1) on the oxidized heme iron[36], consistent with the demonstration that the first step in refolding involves a large scale molecular collapse[37] that forms the blue intermediate[25].

fluctuations that control the HX of the non-marker protons are much less affected.

Met 80 is in the red loop of cyt *c* and its affinity for the heme iron in reduced cyt *c* is increased by 3.2 kcal mol^{-1} (Table 1). The unfavorable free energy of the red unfolding is increased in reduced cyt *c* by 3.2 kcal mol^{-1}, equal to the heme ligation increment, in keeping with the expectation that Met 80 should separate from the heme when the red loop unfolds (K_D ~300 mM in oxidized and ~1 mM in reduced cyt *c*; Table 1). This behavior is indicated by the Ile 75-NH which provides a marker for the red unfolding in both oxidized and reduced cyt *c* (Fig. 2). It is clear that the higher energy unfolding units do not unfold with the red loop since their hydrogens exchange much more slowly. These observations demonstrate, as expected (Fig. 1), that the identity of the red open form is rYGB.

In oxidized cyt *c* Lys 60 and Leu 64 provide markers for unfolding of the yellow loop. In reduced cyt *c* they both exhibit the same large m value but separate somewhat in ΔG since one or the other has some residual protection or acceleration in the partially unfolded condition (Fig. 2*b*). Their averaged increase in unfolding free energy is 3.2 kcal mol^{-1} (2.7 and 3.6 kcal mol^{-1} measured). Leu 68 marks the green unfolding in both reduced and oxidized cyt *c*. The Leu 68-NH is stablized by 3.3 kcal mol^{-1} in the reduced protein. Thus the three partially unfolded intermediates in reduced cyt *c* show equally large additional stability, measured at 3.2, 3.2 and 3.3 kcal mol^{-1} (Table 2, Fig. 3).

The difference in free energy of 3.2 kcal mol^{-1} between the oxidized and reduced forms found for the yellow open and the green open states indicates that the red loop is unfolded also in these states. Thus the yellow open form is ryGB and not RyGB, eliminating the two-fold ambiguity. The green open form is either rygB or rYgB. The form rYgB, with only Y and B folded, is not excluded by the data but seems less likely on structural grounds since the folded Y segment would then have no docking surface to stabilize it.

In summary, the stability labeling results suggest that the three partially unfolded forms in cyt *c* are rYGB, ryGB, rygB. These are exactly the forms necessary for a linear stepwise unfolding sequence, as considered in scheme 1. A role for rYgB would allow a bifurcated pathway.

An alternative possibility, in principle, is that the unfolding of the yellow unit and the green unit may break the Met 80-S to heme iron linkage even though the red unit does not unfold. This seems unlikely. The yellow unit and almost all of the green unit are remote from the heme iron. An exception is the Tyr 67 phenolic OH in the 60's helix (see Fig. 1), the removal of which would not sever the S–Fe bond. The possible role of a remote effect can be judged from behavior in the destabilized acid molten globule form of oxidized cyt *c* where tertiary interactions are globally relaxed. In this case, about two-thirds of the amplitude of the 695 nm absorbance characteristic for the S–Fe ligation is retained (at the high salt concentration, 0.5 M or more, used here). In reduced cyt *c* the ligation strength is further increased by 300-fold (Table 1). It therefore appears that in order to separate the ligation the Met 80-S must be actively removed, for example by moving away the segment of main chain that holds it (the red loop). This view is supported by the quantitative agreement of the measured unfolding increment in all cases with the full 3.2 kcal mol^{-1} increment in ligation strength. Independently, the measured denaturant-dependent m values are inconsistent with the unfolding of the yellow or the green unit alone and agree with the concerted unfolding pictured in scheme 1 (see Fig. 7 of ref. 2).

Folding models and observed intermediates

Different protein folding models lead one to expect very different intermediate forms. The classical model for protein folding[9,10] envisions pathways composed of a small number of discrete cooperative metastable forms. At the other extreme, some theoretical folding models picture essentially a continuum of very many non-cooperative intermediates with no definite order of folding events[11–13].

The conformations that proteins experience on the way from U to N must exist as high energy forms even under native conditions and may be detected by the native state HX method. Therefore, in favorable cases the method can be used to test the different models by comparing their predictions with the character of the high energy forms actually found. A number of different proteins have now been studied. Discrete intermediates have been found in cyt *c*[2], ribonuclease H[14], a thermophilic rubredoxin[15], and apo cyt b_{562}[16,17]. The intermediates found are few, discrete and cooperative and are composed of native-like secondary structural elements.

In two other proteins, barnase and CI2, large scale subglobal unfoldings were not initially observed[18,19]. A possible conclusion is that defined intermediates do not exist, although a continuum of high energy unfolded forms, which should be identifiable as such, was also not observed. However, early results for barnase were inconclusive because the protein adopts EX1 behavior[19,20], which was thought to preclude the search for large scale openings[21]. More recent work[22] attributes the EX1 behavior in barnase to promotion of the distinct intermediate expected from earlier studies. The CI2 protein[21,23] is small (64 residues) and appears to provide a poor model for native state HX studies, as described[24].

This record suggests that the free energy landscape of proteins in the space between U and N may be commonly dominated by a

Table 1 Methionine to heme iron binding affinity[1]

	K_D (mM) reduced	K_D (mM) oxidized	ΔΔG kcal mol^{-1}
MP11	1.55	300	3.2
+0.5 M NaCl	0.91	200	3.2
Cyt *c*[2]	0.38	48	2.9
+5.6 M GdmCl	9.2		
+5.6 M GdmCl[3]	10.5		

[1]The binding affinity of methionine for the oxidized and reduced heme iron of microperoxidase 11 (MP11) was measured as described in the methods section. The results are in good agreement with the prior result of Schejter and Plotkin[38], measured in a modified cyt *c* (dicarboxymethyl Met 80 cyt *c*) with dimethyl sulfide as an extrinsic ligand, and also replicate the lower affinity found in 5.6 M GdmCl for reduced MP11 by Hagen *et al.*[39].
[2]From ref. 38, at 25 ºC.
[3]From ref. 39 at 22 ºC.

limited manifold of discrete metastable, cooperative forms, consistent with expectations from classical folding models. In no case has a continuum of non-cooperative large scale openings been detected.

Interest then centers on the specific structure of the intermediates and their possible role in kinetic folding. A classical pathway model views intermediates as native-like, productive and sequential. Other models view intermediates as misfolded, non-native and obstructive. One wants to compare these model predictions with the intermediate unfolded forms found in proteins. The present experiments move toward this goal by making it possible to more clearly define the identity of intermediates detected by native state HX.

From equilibrium intermediates to a kinetic pathway

The present results indicate that the large scale unfoldings detected in the native state HX experiment produce the intermediates rYGB, ryGB, rygB, and the unfolded state rygb. It is striking that each of these forms could be produced from the next lower free energy form by the unfolding of one cooperative unit, as in reaction scheme 1. If unfolding does proceed mainly through this implied sequence, then the reverse sequence must dominate refolding. This is so because the HX experiments that allowed detection of these intermediates were performed under equilibrium native conditions. It follows that each individual unfolding reaction must be matched by an equivalent refolding reaction.

Clarke and Fersht[18,19] stress the principle that an equilibrium measurement like native state HX cannot prove a kinetic sequence all by itself. An analogous principle is that even kinetic information cannot prove a kinetic mechanism all by itself. What one does in practice is to pursue relevant information of all kinds — kinetic, equilibrium, and theoretical — until the weight of evidence makes some conclusion seem probable. Some relevant kinetic information for cyt *c* is already available. Reaction scheme 1 derived from the native state HX experiment postulates that the first intermediate in cyt *c* folding has only the blue unit folded (rygB). This very same form, with only the N- and C-terminal helices folded, has been independently described as the earliest folding intermediate in kinetic folding experiments[25]. Kinetic work further shows that the N/C helix intermediate accumulates in refolding due to trapping of the misorganized green loop[26]. This observation matches the native state HX results which identify the green segment as the next unit required to fold (rygB → ryGB). A similar equilibrium-kinetic concurrence has been found in ribonuclease H where a partially unfolded intermediate found by native state HX[14] was independently described in kinetic folding experiments[27].

Another kind of evidence is noteworthy. In the proposed refolding sequence, the initially formed structure, B in rygB, comprises the N-terminal and C-terminal helical segments. These segments when placed into solution are known to interact to form independently stable helical structure[28,29]. This provides an attractive rationale for the blue unit to fold first. The subsequent steps in scheme 1 extend this rationale; each consecutively added unit is stabilized by a docking surface provided largely by the previously folded unit (Fig. 1).

In summary, native state HX results for cyt *c* and other proteins reveal a small number of discrete cooperatively structured forms rather than a large number of indeterminate intermediates in the free energy landscape between U and N. The three partially unfolded intermediates found for cyt *c* are just the ones necessary to provide a sequential pathway. Independent kinetic experiments replicate some of the specific steps in the proposed pathway.

Methods

HX measurements were performed with equine cyt *c* (highest grade from Sigma) as a function of increasing concentrations of denaturant. GdmSCN rather than GdmCl was used with reduced cyt *c* to more strongly destabilize the protein (~1.5-fold) without excessively high salt concentrations which degrade the NMR signal. Additional KCl was added to samples with low denaturant concentrations (<0.5 M) to maintain constant high ionic strength. Buffers used were 0.1 M phosphate for pD 7.0 to 8.5 and 0.1 M pyrophosphate for pD 9.

HX was initiated by passing reduced cyt *c* in H_2O buffer through a Sephadex G25 column equilibrated with deoxygenated (argon bubbled) D_2O buffer containing 4.5 mM ascorbic acid. The protein solution was collected directly into an NMR tube under flowing argon. The tube was capped, flame sealed to prevent O_2 entry, and transferred to the NMR spectrometer. 2D NMR spectra were collected to monitor H to D exchange of the observable amides as a function of time of incubation at 30 ºC. To accelerate exchange and data collection for the slowest NHs, we found it possible to raise the pD of the protein solutions up to pD 9. In this range protein stability is independent of pH, HX rates respond ideally (10-fold rate change per pH unit), and no alkaline transition occurs.

Magnitude COSY spectra were collected (Bruker AMX 500 spectrometer) using pulsed field gradients and processed with the Felix software package on a Silicon Graphics workstation. Residual HOD signal was suppressed by low power presaturation during the recycle time. NH-CαH cross peak volumes were integrated, baseline subtracted, normalized against the non-exchanging heme bridge 4 peak, plotted against exchange time, and fit by standard methods. Measured rates were compared to the rates expected for each freely exposed NH at the same conditions[30,31] in order to obtain K_{op} and ΔG_{HX}, according to the standard equations[32,33]. EX2 conditions were assured by demonstration of ideal pH dependence.

The binding affinity of methionine to the reduced and the oxidized heme iron in cyt *c* was determined (30 ºC, pH 7) in binding experiments that used doubly blocked methionine (N-acetyl methionine methyl ester, to avoid charge effects) and an

Table 2 Unfolding free energy (kcal mol^{-1})[1]

Unfolding	(marker)	Reduced	Oxidized	ΔΔG
Blue	(Leu 98)	17.9	12.8	5.1
Green	(Leu 68)	13.1	9.8	3.3
Yellow	(Leu 64)	11.0	7.4	3.2
	(Lys 60)	10.1		
Red	(Ile 75)	9.2	6.0	3.2

[1]Free energies of unfolding at zero denaturant were determined from HX rates of the marker protons (Fig. 2; 30 ºC, pD 7.0–9.0) using the standard equations[32,33] and reference values[30,31].

11-residue peptide, microperoxidase 11 (MP11), derived from cyt *c* (residues 11–21). MP11 contains the heme covalently bonded to the chain by the normal thioether bridges at Cys 14 and 17 and the His 18 ligand. To avoid N-terminus to heme iron interaction that might compete with methionine binding and promote cross chain aggregation, the N-terminus of MP11 was acylated by reaction with acetic anhydride[34]. Binding curves were measured by mixing variable amounts of the methionine derivative with constant concentrations of MP11 in a stopped-flow spectrometer. Binding was determined by absorbance changes. A three syringe stopped-flow instrument (Biologic SFM3) was used to provide variable mixing ratios, isolate the reduced MP11 from contact with air, and allow fast measurement of the binding reaction before dithionite-dependent side reactions intervened in the experiments with reduced heme. For reduced MP11, the first two syringes contained oxidized MP11 (56 μM), one also held acetyl methionine methyl ester (~5 mM), and a small volume of sodium dithionite was added from the third syringe in a second mixing to reduce the MP11. These solutions were deoxygenated by argon bubbling before loading.

Acknowledgments
This work was supported by the NIH. We thank H. Luo who participated in the early HX experiments.

Yujia Xu, Leland Mayne and S. Walter Englander

Johnson Research Foundation, Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

Correspondence should be addressed to S.W.E. *email: walter@hx2.med.upenn.edu*

Received 21 April, 1998; accepted 30 July, 1998.

1. Bai, Y., Milne, J.S., Mayne, L. & Englander, S.W. *Proteins* **20**, 4–14 (1994).
2. Bai, Y., Sosnick, T.R., Mayne, L. & Englander, S.W. *Science* **269**, 192–197 (1995).
3. Takano, T. & Dickerson, R.E. *J. Mol. Biol.* **153**, 79–94 (1981).
4. Takano, T. & Dickerson, R.E. *J. Mol. Biol.* **153**, 95–115 (1981).
5. Berghuis, A.M. & Brayer, G.D. *J. Mol. Biol.* **223**, 959–976 (1992).
6. Qi, P.X., Di Stefano, D.L. & Wand, A.J. *Biochemistry* **33**, 6408–6417 (1994).
7. Qi, P.X., Beckman, R.A. & Wand, A.J. *Biochemistry* **35**, 12275–12286 (1996).
8. Banci, L. *et al. Biochemistry* **36**, 9867–9877 (1997).
9. Kim, P.S. & Baldwin, R.L. *Annu. Rev. Biochem.* **51**, 459–489 (1982).
10. Kim, P.S. & Baldwin, R.L. *Annu. Rev. Biochem.* **59**, 631–660 (1990).
11. Leopold, P.E., Montal, M. & Onuchic, J.N. *Proc. Natl. Acad. Sci. USA* **89**, 8721–8725 (1992).
12. Bryngelson, J.D., Onuchic, J.N., Socci, N.D. & Wolynes, P.G. *Proteins* **21**, 167–195 (1995).
13. Dill, K.A. & Chan, H.S. *Nature Struct. Biol.* **4**, 10–19 (1997).
14. Chamberlain, A.K., Handel, T.M. & Marqusee, S. *Nature Struct. Biol.* **3**, 782–787 (1996).
15. Hiller, R., Zhou, Z.H., Adams, M.W.W. & Englander, S.W. *Proc. Natl. Acad. Sci. USA* **94**, 11329–11332 (1997).
16. Fuentes, E.J. & Wand, A.J. *Biochemistry* **37**, 3687–3698 (1998).
17. Fuentes, E.J. & Wand, A.J. *Biochemistry*, **37** , 9877–9883(1998).
18. Clarke, J., Itzhaki, L.S. & Fersht, A.R. *Trends Biochem. Sci.* **22**, 284–287 (1997).
19. Clarke, J. & Fersht, A.R. *Folding & Design* **1**, 243–254 (1996).
20. Clarke, J., Hounslow, A.M., Bycroft, M. & Fersht, A.R. *Proc. Natl. Acad. Sci. USA* **90**, 9837–9841 (1993).
21. Neira, J.L., Itzhaki, L.S., Otzen, D.E., Davis, B. & Fersht, A.R. *J. Mol. Biol.* **270**, 99–110 (1997).
22. Dalby, P.A., Clarke, J., Johnson, C.M. & Fersht, A.R. *J. Mol. Biol.* **276**, 647–656 (1998).
23. Itzhaki, L.S., Neira, J.L. & Fersht, A.R. *J. Mol. Biol.* **270**, 89–98 (1997).
24. Bai, Y. & Englander, S.W. *Proteins* **24**, 145–151 (1996).
25. Roder, H., Elöve, G.A. & Englander, S.W. *Nature* **335**, 700–704 (1988).
26. Sosnick, T.R., Mayne, L., Hiller, R. & Englander, S.W. *Nature Struct. Biol.* **1**, 149–156 (1994).
27. Raschke, T.M. & Marqusee, S. *Nature Struct. Biol.* **4**, 298–304 (1997).
28. Wu, L.C., Laub, P.B., Elöve, G.A., Carey, J. & Roder, H. *Biochemistry* **32**, 10271–10276 (1993).
29. Kuroda, Y. *Biochemistry* **32**, 1219–1224 (1993).
30. Connelly, G.P., Bai, Y., Jeng, M.-F., Mayne, L. & Englander, S.W. *Proteins* **17**, 87–92 (1993).
31. Bai, Y., Milne, J.S., Mayne, L. & Englander, S.W. *Proteins* **17**, 75–86 (1993).
32. Hvidt, A. & Nielsen, S.O. *Adv. Protein Chem.* **21**, 287–386 (1966).
33. Englander, S.W. & Kallenbach, N.R. *Q. Rev. Biophys.* **16**, 521–655 (1984).
34. Munro, O.Q. & Marques, H.M. *Inorg. Chem.* **35**, 3752–3767 (1996).
35. Kraulis, P.J. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
36. Cohen, D.S. & Pielak, G.J. *J. Am. Chem. Soc.* **117**, 1675–1677 (1995).
37. Sosnick, T.R., Mayne, L. & Englander, S.W. *Proteins* **24**, 413–426 (1996).
38. Schejter, A. & Plotkin, B. The binding characteristics of the cytochrome *c* heme iron. *Biochem. J.* **255**, 353–356 (1988).
39. Hagen, S. J., Hofrichter, J. & Eaton, W. E. Rate of intrachain diffusion of unfolded cytochrome *c*. *J. Phys. Chem.* B **101**, 2352–2365 (1997).

Engineering an intertwined form of CD2 for stability and assembly

The amino-terminal domain of CD2 has the remarkable ability to fold in two ways: either as a monomer or as an intertwined, metastable dimer. Here we show that it is possible to differentially stabilize either fold by engineering the CD2 sequence, mimicking random mutagenesis events that could occur during molecular evolution. Crystal structures of a hinge-deletion mutant, which is stable as an intertwined dimer, reveal domain rotations that enable the protein to further assemble to a tetramer. These results demonstrate that a variety of folds can be adopted by a single polypeptide sequence, and provide guidance for the design of proteins capable of further assembly.

We have previously described the crystal structure of a metastable, intertwined form of the N-terminal domain of CD2 (Fig. 1) in which two polypeptide chains are entangled head-to-tail during folding[1]. The intertwined form of a β-barrel domain of this cell adhesion molecule from the surface of T-lympkocytes provides a rare opportunity to study molecular changes regulating partitioning between two alternate folds for a protein. A small fraction (15 %) of recombinant protein was found to fold in this alternative manner when expressed as a glutathione S-transferase (GST) fusion protein; the GST was believed to act only to tether two polypeptide chains in close proximity during folding hence increasing the effective local concentration of protein. The intertwined dimeric fold is unusual in having two lobes exhibiting immunoglobulin superfamily (IgSF) folds — to which both polypeptide chains contribute strands — separated by a substantial, buried hydrophilic core. The residues buried in this hydrophilic core region are surface exposed in the monomer where they form the cell adhesion surface[2–4]. Buried hydrophilic cores are rare in proteins, and in this case this region is likely to contribute to the observed metastability of the dimer. This intertwined form of CD2 is believed to represent an extreme example of a '3D domain-swapped' protein[5]. Here we demonstrate that partitioning of protein between the two folded states can be regulated through mutagenesis and show that it is possible to design stable oligomeric forms.

Simple mutations regulate the monomer-dimer transitions

Single amino acid mutations in the buried, hydrophilic core of the dimer (Fig. 1) lead to drastic changes in the relative proportion of monomer and dimer produced (Fig. 2). Within the crystal structure of the intertwined dimer virtually every possible hydrogen bond or ionic interaction across the buried interface that can be made is formed. This intricate network of electrostatic bonds enhances interactions between the two lobes of the interdigitated folded form of the protein and ensures balancing of all charged groups in this buried region. One set of mutants was designed to probe the role of this interface in dimer formation. Two mutants (E29R, K43E) which disrupted salt bridges formed between the domains, significantly reduced the amount