Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 10703–10706, October 1996
Biochemistry

Rapid refolding of a proline-rich all-β-sheet fibronectin type III module

(10th fibronectin type III domain protein folding kinetics)

KEVIN W. PLAXCO, CLAUS SPITZFADEN, IAIN D. CAMPBELL, AND CHRISTOPHER M. DOBSON

Oxford Centre for Molecular Sciences, University of Oxford, New Chemistry Laboratory, South Parks Road, Oxford, England OX1 3QT

Communicated by Harry B. Gray, California Institute of Technology, Pasadena, CA, June 24, 1996 (received for review April 22, 1996)

ABSTRACT **Fibronectin type III modules contain approximately 90 residues and are an extremely common building block of animal proteins. Despite containing a complex all-β-sheet topology and eight prolines, the refolding of the 10th type III module of human fibronectin has been found to be very rapid, with native core packing, amide hydrogen bonding, and backbone conformation all recovered within 1 s at 5°C. These observations indicate that this domain can overcome many structural characteristics often thought to slow the folding process.**

A significant number of animal proteins are composed of multiple copies of a small number of extremely common protein domain types, termed modules (1). The reasons for the prevalence of these domains are unclear, but it may reflect an inherent thermodynamic or kinetic propensity of polypeptide chains to achieve these three-dimensional folds. Calorimetric data have indicated that some modules are very stable (2, 3), but folding kinetics have not yet been determined for any of the common module types. The fibronectin type III domain (FNIII) (Fig. 1) occurs more than 400 times in the protein sequence data base (4, 5) and is one of the most common protein modules. We have characterized the refolding kinetics of one of them, the tenth FNIII module of human fibronectin (^{10}FNIII), a 94-residue protein that lacks disulfide bonds and contains a single highly conserved tryptophan residue and eight prolines (6). Refolding of the protein was monitored using a wide variety of biophysical techniques after rapid dilution from 7.0 M guanidine hydrochloride (Gdn·HCl) solutions at 5°C, conditions under which the protein appears to lack significant residual structure.

MATERIALS AND METHODS

Recombinant ^{10}FNIII. ^{10}FNIII was prepared from a glutathione-*S*-transferase fusion protein expressed in *Escherichia coli* [clone generously provided by H. J. Mardon and K. E. Grant (7)]. The fusion protein was purified by affinity chromatography, cleaved with thrombin, and then further purified by cation-exchange chromatography. ^{15}N-labeled protein was prepared from transformed *E. coli* grown on M-9 minimal medium containing a labeled nitrogen source and purified as above. Deuterated samples were prepared by dissolving protein lyophilized from $^{2}H_2O$ in 7.0 M Gdn·^{2}HCl prepared by $^{2}H_2O$ exchange of protonated material.

Fluorescence Kinetics. Kinetic fluorescence data were obtained using a SX17-MV stopped flow fluorimeter (Applied Photophysics, Surrey, U.K.) thermostated at 5.0 ± 0.1°C. Excitation was at 280 ± 3 nm and emission detection above 310 nm was by use of the appropriate filters. ^{10}FNIII (0.3 mM) in 7.0 M Gdn·HCl was diluted 11-fold with 20 mM sodium acetate (pH 5.2) (acetate). Fluorescence data were also collected at pH 7.2 (refolding in 20 mM tris-hydroxymethyl hydrochloride), which were not significantly different (data not shown). Native fluorescence was determined by diluting native protein 11-fold with 0.7 M Gdn·HCl/acetate. Iodide exclusion was determined in an identical fashion, with the inclusion of 300 mM potassium chloride or iodide in the acetate refolding buffer. Reported kinetic values are an average of 25 traces fit using nonlinear regression (KALEIDAGRAPH, Abelbeck Software, Reading, PA). Confidence levels represent the standard deviation of the fitted parameters. Native values represent an average and standard deviation of 10 measurements.

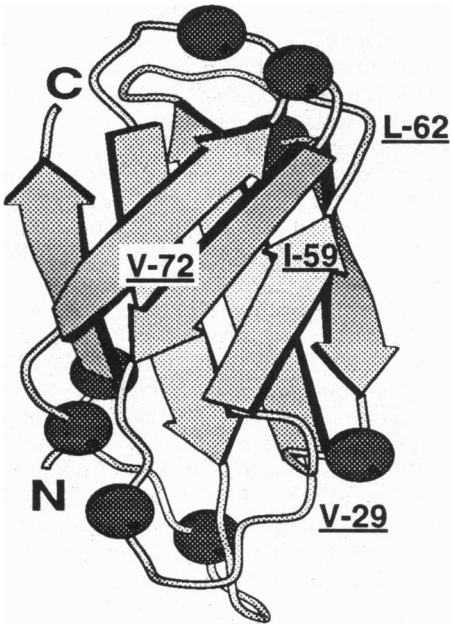

FIG. 1. Proline-rich all-β-sheet structure of ^{10}FNIII (4). Prolines are drawn as spheres. The residues observed during kinetic NMR experiments (see Fig. 4) are indicated. Val-72 is located on the rear sheet. Figure was produced using MOLSCRIPT (9).

Kinetic Circular Dichroism (CD). Far-UV CD results were obtained using a J720 spectropolarimeter (Jasco, Easton, MD) coupled to a SFM3 stopped-flow module (Biologic, Grenoble, France) thermostated to 5.0 ± 0.5°C. ^{10}FNIII (0.2 mM) in 7.0 M Gdn·HCl was diluted 11-fold with acetate and the ellipticity at 216 ± 2 nm was followed. Native ellipticity was determined by diluting native protein 11-fold with 0.7 M Gdn·HCl/acetate. A sum of 250 traces was fit using nonlinear regression (KALEIDAGRAPH). Confidence levels represent estimated fitting error (time constant) or the standard deviation of 10 repeated measurements (native ellipticity).

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.

Abbreviations: ^{10}FNIII, the 10th fibronectin type III domain of human fibronectin; Gdn·HCl, guanidine hydrochloride; ESI-MS, electrospray mass spectrometry; 1D, one-dimensional.

Amide Protection Recovery Kinetics. The recovery of amide hydrogen exchange protection was observed using pulse-label amide protection electrospray ionization mass spectrometry (ESI-MS) and NMR. Denatured perdeutero-^{10}FNIII solutions (5 or 400 μM, respectively) were diluted 11-fold with acetate, incubated for set time periods, pH-pulsed with 1 vol of 500 mM sodium borate (pH 10.0) for 10 ms, and pH-quenched with 1 vol of 500 mM acetic acid in a QFM5 module (Biologic) thermostated to 5.0 ± 0.5°C. Buffer exchanged samples [water or 20 mM sodium acetate (pH 4.8), respectively] were analyzed on a VGBioQ quadrapole ESI-MS (Fisons Instruments, Altringham, U.K.) or by ^{1}H-^{15}N-HSQC NMR spectroscopy on a Bruker Omega 600 MHz NMR spectrometer. Native protection levels were determined by pulse-labeling native (refolded for 2 h at 20°C) perdeutero-^{10}FNIII in refolding buffer containing 0.7 M Gdn·HCl. Control experiments with native protein suggest that the greater level of amide protection observed by NMR than by ESI-MS (Fig. 2*d*) is due to increased exchange during work-up of the latter (data not shown). Confidence levels represent estimated fitting error (time constant) or the estimated error in mass determinations.

Kinetic NMR. Kinetic observation of one-dimensional (1D) NMR spectra was performed using a pneumatic mixer as described (8). Perdeutero Gdn·^{2}HCl-denatured ^{10}FNIII (10 mM) previously rapidly mixed with 11 vol of perdeutero-acetate, p^{2}H 5.2 (meter reading). Spectra were recorded at 5°C and a ^{1}H resonance frequency of 600 MHz. Native and denatured spectra were recorded in 0.64 and 7.0 M Gdn·HCl, respectively. When samples are refolded under the conditions required for this experiment (final protein concentration, 900 μM), a small proportion (estimated to be 10% of the total) remains unfolded and is seen as the large peak at 0.9 ppm in Fig. 4 *Bottom*. This material remains unfolded over the course of the experiment (1 h) and is not observed when refolding is initiated with high-efficiency mixers at 20-fold lower concentration. It should also be noted that fluorescence studies show no significant concentration dependence over the range 10–100 μM (time constants within the ranges 50 ± 3 and 195 ± 18 ms) and pulse-labeling ESI-MS, CD, and fluorescence experiments (Fig. 2 *a, c,* and *d*), conducted at a final concentrations of 0.5, 18 and 27 μM, respectively, display effectively identical kinetics and complete recovery of native signals.

RESULTS

Hydrophobic Core Formation. The formation of the hydrophobic core of ^{10}FNIII was detected by monitoring intrinsic tryptophan fluorescence. During folding this residue becomes sequestered from solvent and packed into the protein core, causing a reduction in fluorescence intensity. Two post-deadtime phases are observed in the fluorescence data, Fig. 2*a*, with time constants (τ) of 51 ± 2 ms and 184 ± 1 ms and relative amplitudes of 77 ± 3% and 23 ± 3%. A more direct measure of tryptophan solvent accessibility is the exclusion of charged fluorescence quenching agents from the hydrophobic core. The kinetics of iodide exclusion, Fig. 2*b*, exhibit two post-dead time phases with τ = 29 ± 2 ms and 179 ± 2 ms (compared with 41 ± 1 ms and 196 ± 2 ms for tryptophan fluorescence under conditions of identical ionic strength) and amplitudes of 70 ± 2% and 30 ± 2%, similar to the tryptophan fluorescence kinetics. The slightly more rapid kinetics may reflect the reduced sensitivity of this probe to fine details of packing geometry. Measurements of native ^{10}FNIII and fit residuals indicate that, within the experimental error of approximately 3%, native levels of fluorescence and iodide exclusion are recovered within 1 s and no other detectable fluorescence phases occur.

Secondary Structure Formation. The formation of native secondary structure and backbone hydrogen bonding were assessed using stopped-flow far-UV CD and pulsed-label amide protection in combination with ESI-MS and NMR. CD measurements at 216 nm, Fig. 2*c*, indicate the kinetics of β-sheet formation, τ = 55 ± 4 ms, and correspond closely to the kinetics of hydrophobic core formation. The recovery of

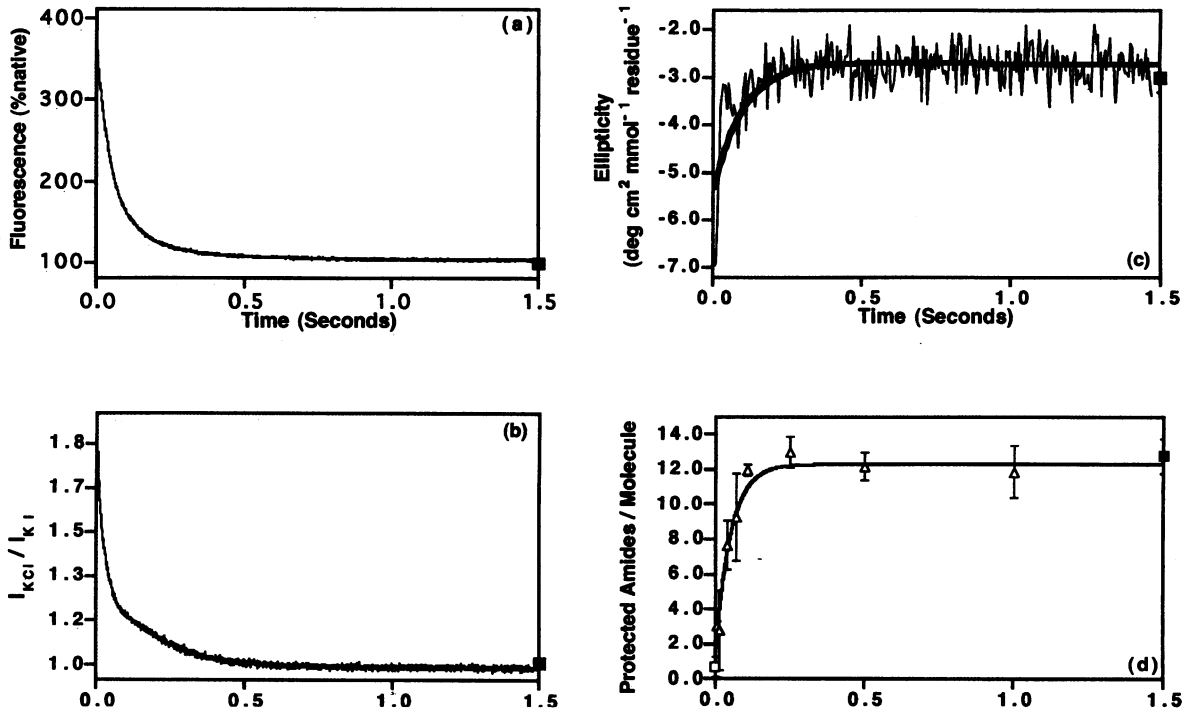

FIG. 2. ^{10}FNIII refolding at 5°C. (*a*) Intrinsic fluorescence kinetics. (*b*) Iodide exclusion kinetics (expressed as the ratio of fluorescence in 300 nM KCl and KI). (*c*) Far-UV CD kinetics (216 nm). (*d*) Pulse-label amide protection recovery kinetics. The heavy lines in *c* and *d* represent exponential fits of τ = 55 ms and 51 ms, respectively. Traces represent averages of 25 (*a* and *b*) or 250 (*c*) experiments, and error bars denote the standard deviation of repeated measurements. Estimated accuracy of fluorescence (*a* and *b*) amplitudes is ±3%. Native ^{10}FNIII values are indicated by filled squares, and the amide protection of the denatured protein (*d*) is indicated by an open square.

amide exchange protection, a sensitive measure of the formation of stable backbone hydrogen bonding, is shown in Fig. 2*d*. The data fit a single exponential with $\tau = 51 \pm 6$ ms, also corresponding to the faster intrinsic fluorescence phase. Because of the limited signal-to-noise ratio of CD and ESI-MS we are not able to ascertain the existence of any small amplitude slow phases in these processes. Control experiments demonstrate that the ellipticity and number of protected amides observed after 1 s are those of native ^{10}FNIII and NMR spectra of pulse-labeled protein demonstrate that the protection pattern of the refolded protein is highly native (Fig. 3).

Formation of Tertiary Contacts. Fluorescence, CD, and amide protection data, therefore, suggest that native core packing and backbone conformation are rapidly formed during the refolding of ^{10}FNIII. These techniques are, however, insufficient to prove the identity of this folding product with the native structure at the level of the tertiary contacts of individual atoms. A more rigorous, if less temporally resolved, assay for the occurrence of native protein is based on the emergence of its characteristic ^{1}H-NMR chemical shift pattern during refolding: the high-field methyl region in the 1D ^{1}H-NMR spectrum of native ^{10}FNIII has a characteristic fingerprint of well-dispersed methyl resonances (Fig. 4*a*). 1D spectra recorded after the initiation of refolding demonstrate that the methyl fingerprint of the native protein is reproduced within the 20 s mixing and acquisition of these stopped flow NMR experiments (Fig. 4*c*).

The Denatured State. Experiments demonstrate that the rapid refolding of the ^{10}FNIII domain is not due to significant residual structure in the denatured protein. NMR spectra recorded in 7.0 M Gdn·HCl exhibit the "collapsed" appearance typical of unstructured proteins (Fig. 4*b*). ^{1}H-^{15}N and ^{1}H-^{13}C-HSQC spectra lack the shift dispersion indicative of structured proteins and are consistent with significant populations of both *cis* and *trans* proline isomers (data not shown). All of the labile protons of ^{10}FNIII exchange in 7.0 M Gdn·HCl within 10 ms at pH 9.0 (Fig. 2*d*). Refolding kinetics remain unchanged ($\tau = 57 \pm 2$ and 303 ± 20 ms) from 6.0, 7.0, or 8.0 M Gdn·HCl (final Gdn·HCl concentration held fixed at 0.727 M) and for samples denatured for 6 months at 4°C or 30 min at 80°C. Refolding remains rapid ($\tau = 10$ and 319 ms) from 7.0 M Gdn·HCl at 65°C and from 7.0 M Gdn·HCl in 70% dimethyl sulfoxide ($\tau = 58$ and 455 ms), conditions under which significant residual structure is unlikely to persist.

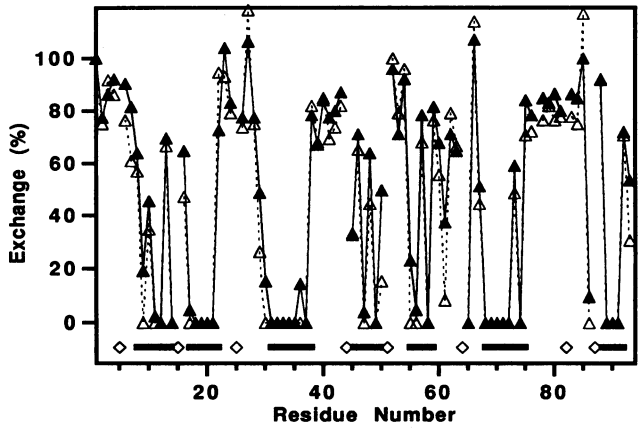

FIG. 3. (*Upper*) Similarity between the pulse-labeled amide exchange protection pattern of ^{10}FNIII 2 s after the initiation of refolding (open triangles) and native material (solid triangles) indicates that the formation of native β-sheet hydrogen bonding occurs within that time. (*Lower*) Proline residues are indicated by diamonds and β-sheet residues are indicated by solid lines. Small differences in exchange protection at positions near proline residues 5, 15, 51, and 64 may reflect proline isomerization driven reorganization in an otherwise fully native structure.

DISCUSSION

Rapid Refolding of an ^{10}FNIII Domain. The refolding kinetics of the topologically complex proline-rich all-β-sheet ^{10}FNIII domain are counter to expectations drawn from the refolding kinetics of a number of previously characterized proteins. Early studies of the refolding of β-sheet proteins (10–13) suggested that, perhaps due to the nonlocal contact requirements of β-sheets, β-sheet formation is slow ($\tau =$ 3–1200 s), an effect that could potentially have been exacerbated by the complex folding topology of the ^{10}FNIII domain (6, 14). The rapid refolding of ^{10}FNIII and the recently reported refolding kinetics of the *Bacillus* cold shock protein (15) suggests, however, that β-sheet formation is not intrinsically slow. The rapid folding of the eight-proline ^{10}FNIII domain is also unexpected because proline *cis–trans* isomerization is known to produce a population of slowly folding molecules in the refolding kinetics of many proteins (16–18). Although the proline residues in native ^{10}FNIII are all in the favored *trans* conformation (6), theory suggests that 70–90% of denatured ^{10}FNIII molecules will contain at least one incorrect proline isomer and, because *cis–trans* isomerization is slow (10–100 s), at 5°C the large majority of incorrectly configured molecules should require hundreds to thousands of seconds to fold (19, 20).

Stability and Folding. The relatively high stability of ^{10}FNIII ($\Delta G_f = -24$ kJ/mol; C.S. and I.D.C., unpublished data) suggests a mechanism for rapid refolding. If the overall stability of the domain correlates to stability in regions of native conformation in transient folding nucleation sites (with an average $G =$ -0.25 kJ/mol per residue relatively few native contacts would

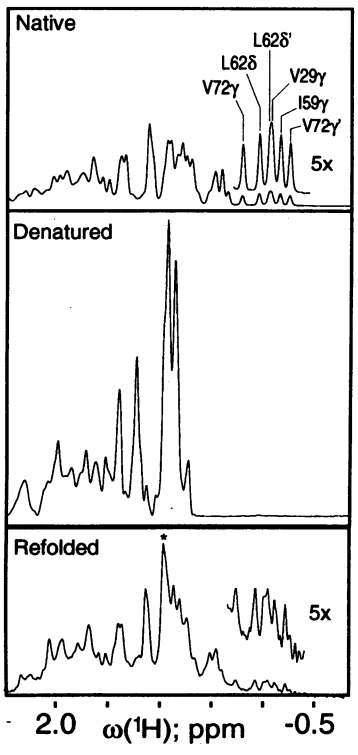

FIG. 4. Aliphatic region of 1D ^{1}H-NMR spectra spectra of native protein (*Top*), denatured protein (*Middle*), and a sum of 4 spectra recorded 16–20 s after the initiation of refolding (*Bottom*). The peak indicated by an asterisk (*Bottom*) arises from a small amount (10%) of protein that remains permanently unfolded under the conditions employed in this experiment (final protein concentration, 900 μM). It is not observed in pulse-labeling samples refolded at lower (40 μM) protein concentrations and no concentration dependence is observed in refolding experiments conducted with final concentrations ranging from 0.5 to 100 μM (see text).

be required to form a stable nucleus), then the unproductive re-unfolding of such intermediates will not compete with the folding process and folding would be accelerated (21). Specifically, the highly stable core of ^{10}FNIII might, despite the potential destabilization of several kJ/mol inherent in such a conformation (22, 23), accommodate incorrectly configured *cis* prolines, all of which occur in interstrand loops, in a rapidly formed and extremely native-like intermediate (24–27). The presumed stability of such a native-like folding intermediate could, potentially, also accelerate the rate of proline isomerization by torsional destabilization of *cis* prolines in near native conformations (28). Moreover, irrespective of which mechanisms are responsible, the domain appears to generate native core packing, secondary structure, backbone hydrogen bonding, and tertiary contacts very rapidly despite the potential difficulties of incorrectly configured proline containing loops and a complex all-β-sheet topology.

The Putative Advantages of Rapid Refolding. Limited predominantly to structural roles (3), FNIII modules presumably lack some of the selective constraints of enzymes. This might allow evolution to act at the level of protein folding to ensure that the module folds quickly to prevent misfolding or aggregation. This may be of particular importance for modular protein domains as the occurrence of multiple sequential copies of a single domain type may cause the incorrect association of corresponding structural elements from neighboring domains; such a misfolding/dimerization event has been observed for the all-β-sheet CD2 protein (29). A putative selective advantage for the rapid refolding of FNIII modules also occurs in the muscle protein titin, which contains more than one hundred FNIII modules and is thought to act as a highly elastic linker to maintain myosin filament centering in the sarcomere. The elastic rebound of titin, which is known to elongate reversibly by 60–400%, is postulated to occur via the rapid and spontaneous refolding of tension-unfolded FNIII modules (30, 31). Our results demonstrate that it is possible for FNIII domains to refold at a rate consistent with this hypothesis. Similar elongation has been noted in fibronectin itself (32), and FNIII modules occur in several collagens (4, 5), where they may also play a role in macromolecular elasticity.

We acknowledge the kind help of Dr. Carol Robinson and Ms. Evonne Chung in conducting the ESI-MS experiments. We are also indebted to Dr. Jochen Balbach for aid in the kinetic NMR experiments and Drs. Jane Clarke and Sheena Radford for helpful comments. The Oxford Centre for Molecular Sciences is supported by the U.K. Biotechnology and Biological Sciences Research Council, the Engineering and Physical Sciences Research Council, and the Medical Research Council. This work was also supported in part by the Wellcome Trust (I.D.C.) and an International Research Scholars award from the Howard Hughes Medical Institute (C.M.D.).

1. Bork, P. & Bairoch, A. (1995) *Trends Biochem. Sci.* **20** (Suppl.).
2. Litvinovich, S. V. & Ingham, K. C. (1995) *J. Mol. Biol.* **248,** 611–626.
3. Politou, A. S., Gautel, M., Pfuhl, M., Labeit, S. & Pastore, A. (1994) *Biochemistry* **33,** 4730–4737.
4. Bork, P. & Doolittle, R. F. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 8990–8994.
5. Bork, P., Downing, A. K., Kieffer, B. & Campbell, I. D. C. *Q. Rev. Biophys.*, in press.
6. Main, A. L., Harvey, T. S., Baron, M., Boyd, J. & Campbell, I. D. (1992) *Cell* **71,** 671–678.
7. Mardon, H. J. & Grant, K. E. (1994) *FEBS Lett.* **340,** 197–201.
8. Balbach, J., Forge, V., van Nuland, N., Winder, S. L., Hore, P. J. & Dobson, C. M. (1995) *Nat. Struct. Biol.* **2,** 866–870.
9. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–955.
10. Goto, Y. & Hamaguchi, K. (1982) *J. Mol. Biol.* **156,** 911–928.
11. Ropson, I. J., Gordon, J. I. & Frieden, C. (1990) *Biochemistry* **29,** 9591–9599.
12. Rudolph, R., Siebendritt, R., Nesslauer, G., Sharma, A. K. & Jaenicke, R. (1990) *Proc. Natl. Acad. Sci.* **87,** 4625–4629.
13. Varley, P., Groneborn, A. M., Christensen, H., Wingfield, P. T., Pain, R. H. & Clore, M. G. (1993) *Science* **260,** 1110–1113.
14. Orengo, C. A., Jones, D. T. & Thornton, J. M. (1994) *Nature (London)* **372,** 631–634.
15. Schindler, T., Herrler, M., Marahiel, M. A. & Schmid, F. X. (1995) *Nat. Struct. Biol.* **2,** 663–668.
16. Schmid, F. X. & Baldwin, R. L. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 4764–4768;.
17. Kiefhaber, T., Quaas, R., Hahn, U. & Schmid, F. X. (1990) *Biochemistry* **29,** 3061–3070.
18. Schmid, F. X. (1993) *Adv. Prot. Chem.* **44,** 25–66.
19. Creighton, T. E. (1978) *J. Mol. Biol.* **125,** 401–406.
20. Kiefhaber, T., Kohler, H. & Schmid, F. X. (1992) *J. Mol. Biol.* **224,** 217–220.
21. Finkelstein, A. V. (1991) *Proteins Struct. Funct. Genet.* **9,** 23–27.
22. Levitt, M. (1981) *J. Mol. Biol.* **145,** 251–263.
23. Schreiber, G. & Fersht, A. R. (1993) *Biochemistry* **32,** 11195–11203.
24. Schmid, F. X. (1981) *Eur. J. Biochem.* **114,** 105–109.
25. Schmid, F. X. & Blaschek, H. (1981) *Eur. J. Biochem.* **114,** 111–117.
26. Evans, P. A., Kautz, R. A., Fox, R. O. & Dobson, C. M. (1989) *Biochemistry* **28,** 362–370.
27. Chazin, W. J., Kordel, J., Drakenberg, T., Thulin, E., Brodin, P., Grundstrom, T. & Forsén, S. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 2195–2198.
28. Cook, K. H., Schmid, F. X. & Baldwin, R. L. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 6157–6161.
29. Murray, A. J., Lewis, S. J., Barclay, A. N. & Brady, R. L. (1995) *Proc. Natl. Acad. Sci. USA* **76,** 7337–7341.
30. Soteriou, A., Clarke, S., Marin, S. & Trinick, J. (1993) *Proc. R. Soc. London B* **254,** 83–86.
31. Erickson, H. P. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 10114–10118.
32. Erickson, H. P., Carrell, N. A. & McDonagh, J. (1981) *J. Cell Biol.* **91,** 673–678.