Pacific Symposium on Biocomputing 4:590-600 (1999)

RECOGNITION BETWEEN DISORDERED POLYPEPTIDE CHAINS FROM CLEAVAGE OF AN α/β DOMAIN: SELF- VERSUS NON-SELF-ASSOCIATION

X.-M. YANG, R. E. GEORGESCU, J.-H. LI, W.-F. YU, HAIERHAN, M. L. TASAYCO*

Department of Chemistry
City College of the City University of New York
New York, NY 10031

Advances in structural biology have provoked a re-evaluation of the biological significance of the disordered state of proteins. We believe that the rules that govern structure, stability and kinetics in the molecular recognition between disordered polypeptide chains can be elucidated by studying processes that couple association with folding. The reassembly of single domain proteins by fragment complementation provides an excellent opportunity to study them. Since almost the complete sequence is available, although not on a single chain, most of the complementary fragments are expected to reassemble. However, that happens not to be the case. We have chosen *E. coli* thioredoxin (Trx), a small, single α/β-domain protein, as a model system to study the effect of the site and number of cleavages on the reassembly of complementary fragments. We have shown at atomic detail the reassembly after cleavage of a loop (1–73, 74–108)[1] and after cleavage of an α-helix (1–37, 38–108).[2] Although both sets of fragments produce native-like complexes, there are clear differences in the interface geometry, apparent stability of the folded state and mechanism of association/folding: (i) the apparent equilibrium dissociation constant for 1–37/38–108 complex (4 μM) is higher than the one for 1–73/74–108 complex (49 nM), (ii) the apparent rate constants of non-self-association are similar (about 10^3 $M^{-1}s^{-1}$), and (iii) only the 1–37 fragment self-associates under these experimental conditions. Here the competition between self- and non-self-association leads to an apparently less stable 1–37/38–108 complex.

1 Introduction

Advances in structural biology have provoked a re-evaluation of the biological significance of the disordered state of proteins. The finding of a protein from bacterial flagella[3,4] that might have a biological role in the disordered state is only one of the most striking examples.[5] The development of structure in a disordered protein region upon binding another molecule is not a new observation in Biology.[6] Neither is the fact that protein-protein interactions play a central role in the regulation of many biological phenomena. However, our understanding of these processes is still in its infancy, as attested by the difficulty in finding a protein complementary partner for a protein sequence taken at random.[7,8] Interestingly, recent developments in Bioinformatics[9] predict that long disordered regions in proteins might be more frequent than expected.[10]

Inspection of the structural databases demonstrates the presence of: (i) an apparently finite number of hydrophobic folding units;[11,13] (ii) monomeric proteins and protein interfaces that share the same folding unit;[12,14] and (iii) units that are assembled by pieces of various shapes and sizes.[15] A revision of the literature on single domain homo- and heterodimers with up to 126 residues indicates that (i) the stability is between 8 to 11 kcal/mol.[16,17,18,19,20] and (ii) the rate of association spans from 10^3 to 10^7 $M^{-1}s^{-1}$.[16,17,18,19,20,21] Thus the changes in the rates of association seem more dramatic than the ones for the stability. No consensus has been attained about the relationship between structure, stability and folding. Many aspects of molecular recognition are still obscure. To mention a few: the relation between stability and dimeric interface; or that among the rates of association, the conformational preference of the disordered state, and the geometry of the folded state. In this regard, perhaps, one of the most challenging enterprises is still to characterize the structure of a disordered monomer in equilibrium with its dimer.

We believe that the rules that govern structure, stability and kinetics in the molecular recognition between disordered polypeptide chains can be obtained by studying processes that couple association with folding. The reassembly of single domain proteins by fragment complementation provides an excellent opportunity to study them.[1,20,23,24,25,26] These studies, pioneered by Taniuchi,[22] are based on the fact that the same principles that govern protein folding are involved in the reassembly of two or more chains. Since almost the complete sequence is available, although not on a single chain, most of the complementary fragments are expected to reassemble. However, that happens not to be always the case.[25,27]

Based on Holmgreen's studies[28,29] on the cleavage of *E. coli* thioredoxin (Trx) we have chosen this small single α/β-domain protein, as a model system to study the effect of the site and number of cleavages on the reassembly of complementary fragments. We have shown at atomic detail the reassembly after cleavage of a loop (1–73, 74–108)[1] and after cleavage of an α-helix (1–37, 38–108).[2] Although both sets of fragments produce native-like complexes, there are clear differences in the interface geometry (see fig. 1), apparent stability of the folded state and mechanism of association/folding. Here we report results on the stability and kinetics involved in the association/folding of the 1–37 and 38–108 fragments and the analysis on two sets of complementary Trx fragments: One after cleavage of a loop, and another after cleavage of an α-helix.

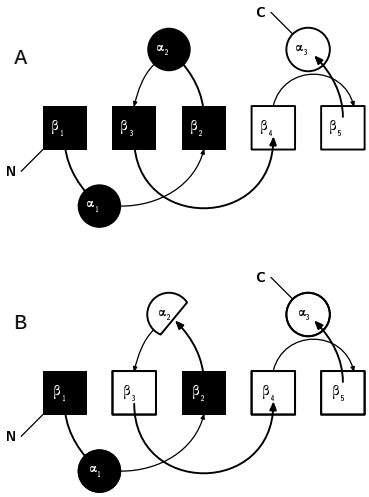

Figure 1: (A) Scheme of the topology of the 1–73/74–108 complex (B) Scheme of the topology of the 1–37/38–108 complex.[2] The regions corresponding to the N- and the C-fragment are depicted with solid and empty symbols, respectively.

2 Materials and Methods

The fragments were prepared in 10mM potassium phosphate (KP_i) at pH 5.7 as previously described.[30,2] Molecular Sieve chromatography was carried out with a Superdex Peptide 10/30 column equilibrated with 0.1 M KP_i at pH 7.0, using a Pharmacia LCC 500 with a detector at 214 nm. Far-UV CD measurements were obtained with samples in KP_i at pH 5.7 and 20°, using an AVIV-60DS instrument equipped with a thermostatic cell holder and cells with a pathlength of 1 or 2 mm. Five scans at 20 nm/min were averaged after substracting the signal corresponding to the buffer or excess of fragment. The NMR spectra were recorded using samples in KP_i at pH 6.5 and 90% H_20/10% D_20 and analyzed as previously decribed.[2] The N-fragment (0.9 μM for 1–73-fragment and 2 μM for the 1–37-fragment) was titrated with increasing amounts of C-fragment (0.1 to 6 μM) and incubated for 3 h in KP_i at pH 5.7 and 20 degrees. Afterwards, the fluorescence of samples and controls were measured using a PTI fluorimeter and 1 to 5 nm band pass. The normalized plot of the observed fluorescence change was fitted to the following equation to obtain the dissociation constant:

$$\Delta F_{\mathrm{m}} = \frac{1}{2}\Delta F_{\mathrm{int}} \left([N]_0 + [C]_0 + K_d - \sqrt{([N]_0 + [C]_0 + K_d)^2 - 4[N]_0[C]_0} \right). \quad (1)$$

Manual kinetic measurements using fluorescence spectroscopy were carried out and analized as previously described.[30]

3 Results and Discussion

3.1 Disordered State

Analysis of the isolated fragments by molecular sieve chromatography (see fig. 2) indicates that the 1–37 and 74–108 fragments or the 1–73 and 38–108 fragments behave like a 7–kD or a 19–kD globular protein, respectively.[32] Previous sedimentation equilibrium analysis of 1–73 and 74–108 fragments in KPi at pH 7.0[30] indicated the presence of only 4– and 8–kD monomeric species, respectively. Thus, the 1–37 and 38–108 fragments behave as a disordered monomer although the participation of rather compact dimers can not be ruled out.

The Far-UV-CD spectra of the isolated N- and C-terminal fragments (see fig. 3) show the typical traces of disordered peptides with a minimum at 198 nm. Moreover, the spectra of the N-fragments, which contain the disulfide bond (C32-C35), show a more pronounced bulge around 215 nm. In all cases, this bulge decreases in the presence of 4M GuHCl (data not shown), suggesting the presence of non random conformations.

The ^{1}H-^{15}N-HSQC spectrum of the isolated fragments shows a narrow dispersion of amide protons typical of disordered peptides (see fig. 4). Standard 3D-NMR analysis of 1–37 fragment indicates sequential NOE connectivities (data not shown) between protons on adjacent residues consistent with the dihedral angles of an average backbone conformation. The lack of non-sequential NOEs and the small deviation of chemical shifts from the random coil values (data not shown) indicate mainly a random coil peptide. This observation makes less likely, but does not eliminate, the possible contribution of a small population of compact dimeric species in rapid exchange with disordered monomeric species.

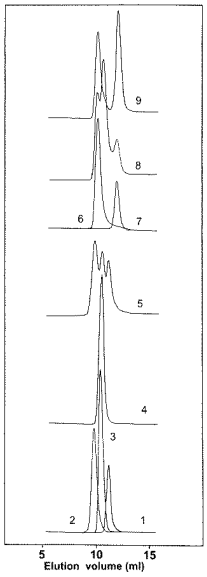

Figure 2: Overlay of chromatograms obtained after injection of 33 μM samples in KP_i (unless otherwise indicated) into a Superdex Peptide 10/30 column under folding conditions: (1) 74–108 fragment, (2) 1–73 fragment, (3) uncleaved Trx, (4) 1:1 stoichiometric mixture (1–73/74–108), (5) 1:1 stoichiometric mixture (1–73/74–108) in 4M GuHCl/KP_i, (6) 38–108 fragment, (7) 1–37 fragment, (8) 1:1 stoichiometric mixture (1–37/38–108), and (9) 1:1 stoichiometric mixture (1–37/38–108) in 4M GuHCl/KP_i.

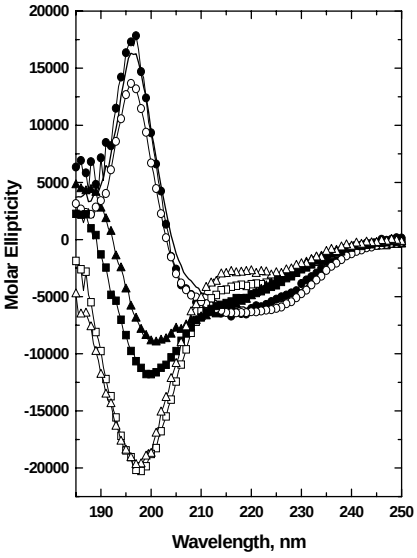

Figure 3: (A) Overlay of Far-UV-CD spectrum of fragments: 1-37 (solid triangle), 74–108 (empty triangle), 38–108 (empty square),1–73 (solid square); of complexes: 1–37/38–108 (solid sphere), 1–73/74–108 (empty sphere); of uncleaved Trx (solid line). The protein concentration in all cases was 20μM with exception of a ratio 40μM/10μM for 1–37/38–108. All spectra were corrected for the contribution of buffer or excess of isolated fragments.

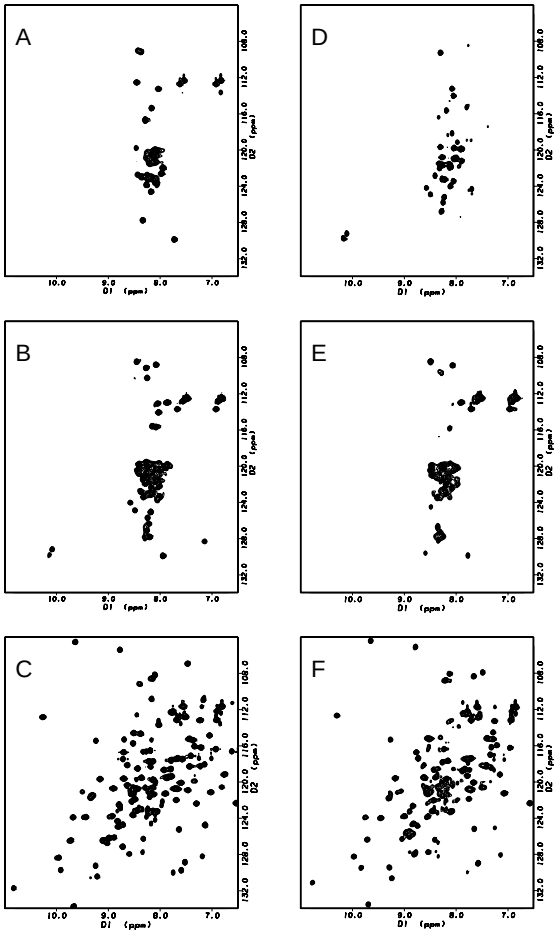

Figure 4: ^{1}H-^{15}N-HSQC spectra of ^{15}N-labeled fragments. (A) 74–108, (B) 1–73, (C) a 1:1 stoichiometric mixture of 1–73 and 74–108, (D) 1–37, (E) 38–108, (F) a 1:1 stoichiometric mixture of 1–37 and 38–108. Spectra D and E were taken from a previous report.[2]

A careful analysis of each isolated fragment indicates that they behave like disordered expanded monomeric species with exception of the 1–37 fragment which shows an interesting behaviour: (i) The far-UV CD spectrum of a freshly prepared solution of this fragment shows concentration dependence (data not shown), although only the apparent 7–kD globular species are present in so-

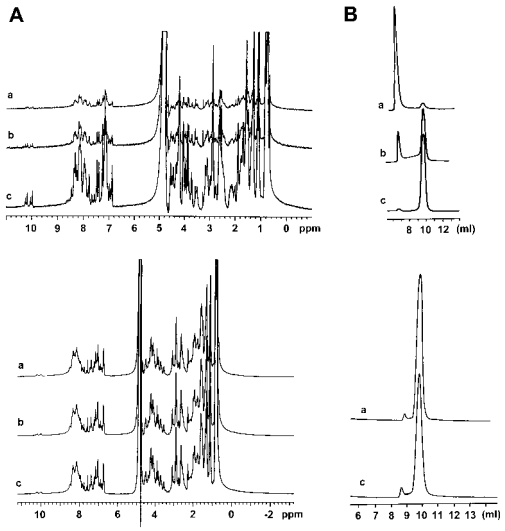

Figure 5: (A) Overlay of 1D-NMR spectrum at 20°, (B) Overlay of molecular sieve chromatograms. In all cases, the top and bottom panel correspond to the 1–37 and 1–73 fragments, respectively, after 8 days (a), 4 days (b) and at the beginning (c) of the NMR data acquisition.

lution (see fig. 2). (ii) The 1D-NMR spectrum of freshly prepared solutions within the range of 50 to 300 μM (data not shown) shows no concentration dependence of the linewidth of the resonances. (iii) Both, the intensity of the resonances in the 1D-NMR spectrum of this fragment in solution (1 mM) and the peak of molecular sieve chromatography that correspond to the apparent 7–kD globular specie decreases slowly in concert with the appearance of a peak at the void volume (see fig. 5). (iv) The large oligomers, wich elute in the void volume of the column, do not dissociate upon dilution or precipitate. (v) Finally, the presence of large oligomers do not broaden the peaks of the 1D-NMR spectrum of the apparent 7–kD globular species. This behaviour is consistent with the presence of disordered expanded monomers in rapid equilibrium with compact dimers, which themselves slowly self-associate into large stable oligomers. Evidence to support this interpretation awaits further studies of the 1–37 fragment by sedimentation analysis.

3.2 Folded State

Analysis of the 1:1 stoichiometric mixture of complementary fragments and Trx by molecular sieve chromatography (see fig. 2) indicates the presence of

species with the same Stokes radii that behave as a 12–kD globular protein.[32] The far-UV CD spectra (see fig. 3) and the pattern of NOE connectivities of the 1:1 stoichiometric mixtures (1–73/74–108 and 1–37/38–108) are in complete agreement with previous NMR analysis[1,2] that show striking similarities between these complexes and the uncleaved Trx.[33] These results unequivocally demonstrate the successful reassembly by fragment complementation of the α/β domain of Trx after cleavage of the loop before β_5 or the α_2-helix (see fig. 1).

3.3 Stability

Titration of each fluorescent N-fragment with an excess of its complementary C-fragment indicates that the value of the apparent K_D for the 1–37/38–108 complex is higher (4μM) than the one for the 1–73/74–108 complex (49nM). Comparison of the molecular sieve chromatography under folding conditions of the 1:1 stoichiometric mixtures of the complementary fragments at chemical denaturing and folding conditions (see fig. 2) indicates that each mixture undergoes a different degree of association/folding and dissociation/unfolding within the column, and that the 1–37/38–108 complex is the weakest complex. Moreover, the 1D-NMR spectrum of the 1–37/38–108 complex shows broader resonances (data not shown) compared to the other complex, suggesting a rapid equilibration of the fragments with their complex. In summary, our results indicate that the 1–37/38–108 complex is apparently less stable than the 1–73/74–108.

3.4 Kinetics

The kinetics of the association/folding process between the fluorescent 1–37 fragment and its non-fluorescent partner was monitored by fluorescence spectroscopy under pseudo first order conditions using an excess of the non-fluorescent fragment (data not shown). More than one concentration dependent observed rate constant was obtained by fitting each kinetic trace (see table 1). This is to be contrasted with the presence of only one concentration dependent observed rate constant for the other set of complementary fragments.[30] The unexpected concentration dependent kinetics of the association/folding between the 1–37 and 38–108 fragments and the concentration dependence of the molecular sieve chromatogram, the far-UV CD spectrum and the 1D-NMR spectrum of the 1–37 fragment, lead us to believe that the fastest kinetic phase reflects a rapid equilibrium between disordered expanded monomers and rather compact homodimers of this fragment. In fact, previously reported studies of bacterial luciferase[34] indicate that the same

Table 1: Kinetic parameters for the association/folding process between 1–37 and 38–108 fragments

$[N]^a$ (μM)	$[C]^a$ (μM)	k_1^b ($s^{-1} \times 10^{-3}$)	k_2^b ($s^{-1} \times 10^{-3}$)	k_3^c ($s^{-1} \times 10^{-4}$)	A_1^e (%)	A_2^e (%)	A_3^e (%)
1	30	286 ± 4	43 ± 1	33 ± 0.1	-75	15	85
1	20	146 ± 3	29 ± 2	27 ± 0.1	-76	14	86
1	10	108 ± 1	13 ± 1	19 ± 0.4	-76	11	89

[a]Final concentration of the 1–37 (N) and 38–108 (C) fragments. [b]Bimolecular phase. [c]Apparent unimolecular phase. [e]Relative amplitudes. *During the fitting A_1 was fixed to a value derived from controls.

monomeric specie might undergo self- and non-self-association. Assuming that the first phase corresponds to self-association and the second one to non-self-association, analysis of the concentration dependence of the second observed rate constant (k_2) yields a value of 1500 $M^{-1}s^{-1}$ for the apparent rate constant of association (k_{on}) between the 1–37 and 38–108 fragments. Comparison of this rate with the one (1300 $M^{-1}s^{-1}$) for the association between the 1–73 and 74–108 fragments,[30] indicates similar rates of association for both sets of complementary fragments. In fact, the second and third kinetic phases are observed in both association/folding processes and might be indicative of a similar mechanism of non-self-association between these complementary Trx fragments.

4 Conclusions

Cleavage of the α/β-domain of Trx at either the loop preceding β_5 or near the N-terminus of the α_2-helix yields complementary fragments able to reassemble with a native-like backbone topology (see fig. 1) and defined side chain packing.[1,2] Various lines of evidence indicate that the 1–37/38–108 complex is apparently less stable than the 1–73/74–108 complex despite the larger complementary interface (hydrophobic and electrostatic patches). This might reflect a decrease in stability due to shortening of α_2-helix by the appearance of new termini.[2] However, there is evidence that the competition between self- and non-self-association of the 1–37 fragment favors its thermodynamically more stable large oligomer and leads to an apparently less stable 1–37/38–108

complex.

Acknowledgments

We thank Francisco Figueirido and the NMR facility from Hunter college for their invaluable support. M. L. T. is an NSF Career Awardee (MCB-9507255), recipient of financial support from NIGMS (5R29GM53808-02) and the RCMI grant from NIH to CCNY.

1. M. L. Tasayco and K. Chao, *Proteins: Structure, Function and Genetics*, 22:41–44, 1995.
2. X.-M. Yang, W.-F. Yu, J.-H. Li, J. Fuchs, J. Rizo, and M. L. Tasayco, *J. Amer. Chem. Soc.*, 128:7185, 1998.
3. G. W. Daughdrill, M. S. Chadsey, J. E. Karlinsey, K. T. Hughes, and F. W Dahlquist, *Nature Structural Biology*, 4:285–291, 1997.
4. G. W. Daughdrill, L. J. Hanely, and F. W. Dahlquist, *Biochemistry*, 37:1076–1082, 1998.
5. K. W. Plaxco and M. Gross, *Nature*, 386:657–658, 1997.
6. A. K. Dunker, E. Garner, S. Guilliot, P. Romero, K. Albrecht, J. Hart, Z. Obradovic, C. Kissinger, and J. Villafranca, in *Pacific Symposium on Biocomputing '98*, pages 473–484, 1998.
7. F. Pazos, M. Helmer-Clitterich, G. Aurelio, and A. Valencia, *J. Mol. Biol.*, 271:511–523, 1997.
8. S. Jones and J. M. Thornton, *J. Mol. Biol.*, 272:133–143, 1997.
9. P. Romero, Z. Obradovic, C. Kissinger, J. E. Villafranca, and A. K. Dunker, in *Proceedings of the IEEE Conference on Neural Networks*, volume 1, pages 90–95, 1997.
10. P. Romero, Z. Obradovic, C. R. Kissinger, J. E. Villafranca, E. Garner, S. Guilliot, and A. K. Dunker, in *Pacific Symposium on Biocomputing '98*, pages 437–448, 1998.
11. C.-J. Tsai and R. Nussinov, *Protein Science*, 6:24–42, 1997.
12. C.-J. Tsai and R. Nussinov, *Protein Science*, 6:1426–1437, 1997.
13. I. Holm and C. Sander, *Nucleic Acids Res.*, 26:316–319, 1998.
14. R. Preisner, A. Goede, and C. Frommel, *J. Mol. Biol.*, 280:535–550, 1998.
15. M. H. Zehfus, *Protein Science*, 6:1210–1219, 1997.
16. M. Munson, K. S. Anderson and L. Regan, *Folding and Design*, 2:77–87, 1997.
17. M. E. Milla and R. T. Sauer, *Biochemistry*, 34:3344–3351, 1995.
18. H. Wendt, L. Leder, H. Harma, I. Jelesarov, A. Baici and H. R. Bosshard, *Biochemistry*, 36:204–213, 1997.

19. A. D. Kippen and A. R. Fersht, *Biochemistry*, 34:1464–1468, 1995.
20. G. de Prat Gay and A. R. Fersht, *Biochemistry*, 33:7957–7963, 1994.
21. L. M. Gloss and C. R. Matthews, *Biochemistry*, 36:5612–5623, 1997.
22. H. Taniuchi, G. R. Parr, and M. A. Jullierat, *Methods of Enzymology*, 131:185–217, 1986.
23. J. Sancho and A. R. Fersht, *J. Mol. Biol.*, 224:741–747, 1992.
24. M. L. Tasayco and J. Carey, *Science*, 255:594–597, 1992.
25. N. Kobayashi, S. Honda, H. Yoshi, H. Uedaira, and E. Munekata, *FEBS Letters*, 366:99–103, 1995.
26. J. L. Neira, B. Davis, A. G. Ladurner, A. M. Buckle, G. de Prat Gay, and A. R. Fersht, *Folding & Design*, 1:189–208, 1996.
27. A. G. Ladurner, L. S. Itzhaki, G. de Prat Gay, and A. R Fersht, *J. Mol. Biol.*, 273:317–329, 1997.
28. I. Slabý and A. Holmgren, *J. Biol. Chem.*, 250:1340–1347, 1975.
29. H. Reutimann, B. Straub, and P. L. Luisi, *J. Biol. Chem.*, 256:6796–6803, 1981.
30. A. Chaffotte, J.-H. Li, M. Goldberg, and M. L. Tasayco, *Biochemistry*, 36:16040–16048, 1997.
31. O. Zhang, L. Kay, J. Olivier, and J. D. Forman-Kay, *J. Biomol. NMR*, 4:845–858, 1994.
32. N. Sevieux and J.-H. Li, *J. Student Res.*, 1:25–28, 1997.
33. H. J. Dyson, A. Holmgren, and P. E. Wright, *Biochemistry*, 28:7074–7087, 1989.
34. A. C. Clark, S.W. Raso, J.F. Sinclair, M.M. Ziegler, A.F. Chaffotte, and T.O. Baldwin, *Biochemistry*, 36:1891–1899, 1997.