Volume 334, number 3, 265–268 FEBS 13254 November 1993

Secondary structure of globular proteins at the early and the final stages in protein folding

Kunihiro Kuwajima[a,b,*], Gennady V. Semisotnov[c], Alexei V. Finkelstein[c], Shintaro Sugai[b,d], Oleg B. Ptitsyn[c]

[a]*Department of Physics, School of Science, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113, Japan*
[b]*Department of Polymer Science, Faculty of Science, Hokkaido University, Kita-ku, Sapporo, Hokkaido 060, Japan*
[c]*Institute of Protein Research, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russian Federation*
[d]*Department of Bioengineering, Faculty of Engineering, Soka University, Hachioji, Tokyo 192, Japan*

Received 16 August 1993; revised version received 22 September 1993

The ellipticities for an early transient intermediate in refolding observed by kinetic circular dichroism measurements at 220–225 nm for 14 different proteins are summarized, and the ellipticity values are compared with those for the final native proteins and also with the ellipticities expected from a physical theory of protein and polypeptide secondary structure. The results show that a substantial part of the protein secondary structure is in general formed in the earliest detectable intermediate in refolding and that the ellipticities in both the native and the intermediate states are consistent with the physical theory of protein secondary structure.

Protein folding; Folding intermediate; Secondary structure; Circular dichroism

1. INTRODUCTION

Protein folding is a sequential process with formation of an early transient structural intermediate that has an appreciable amount of the backbone secondary structure in the absence of the specific tertiary interactions [1–6]. The intermediate has been detected for a number of globular proteins by kinetic circular dichroism techniques including stopped-flow circular dichroism measurements [7–15], but the amount of the secondary structure formed in the intermediate strongly depends on the protein species [2,12]. In this paper, we summarize all available data (including unpublished ones) of the kinetic CD studies, and the ellipticities of the early intermediate and the final folded proteins are compared with each other. We show that the ellipticity of the intermediate as a rule is smaller than that of the final native protein, but the difference is usually not dramatic. This supports the idea that a substantial part of the protein secondary structure is formed in the earliest detectable intermediate and that the other part of the secondary structure is restored at subsequent stages of protein folding. The ellipticities in both the native and the intermediate states of the proteins studied are more or less consistent with the physical theory of protein and polypeptide secondary structure [16,17].

*Corresponding author. Fax: (81) (3) 3814 9717;
e-mail: kuwajima@tkyvax.phys.s.u-tokyo.ac.jp.

Abbreviations: CD, circular dichroism; UV, ultraviolet

2. MATERIALS AND METHODS

The proteins studied by the kinetic CD measurements include bovine and human α-lactalbumins, lysozyme, β-lactoglobulin A, ribonuclease A, chymotrypsinogen A, β-lactamase, carbonic anhydrase B, staphylococcal nuclease, dihydrofolate reductase, phosphoglycerate kinase, parvalbumin, tryptophan synthase β_2 subunit, and ferri-cytochrome *c*. Essentially for all of these, the transient intermediate (I) accumulates rapidly within the dead time of the measurement (about 10 ms in the stopped-flow measurement), and this first stage of refolding is followed by a slower process, further compaction of a protein molecule or formation of the specific tertiary structure.

Most of the CD data presented in this paper are those already published, and the sources of the proteins used are also shown in the previous reports; the appropriate references are given in Table I. The sources of human α-lactalbumin, staphylococcal β-lactamase and yeast 3-phosphoglycerate kinase are shown in [18,19], bovine ribonuclease A was from Sigma (Type XII-A) and further purified by ion-exchange chromatography on a CM Sephadex C-50 column with a salt gradient from 0.1 to 0.25 M NaCl at pH 8.0 (10 mM Tris-HCl), and bovine chymotrypsinogen A was from SIGMA (Type II, 6 × crystallized) and used without further purification. Refolding reactions of the proteins were induced by a concentration jump of the denaturant, urea or guanidine hydrochloride, in a stopped-flow mixing apparatus and monitored by rapid CD measurements as described previously. Details of the CD data will be presented elsewhere, and the ellipticity values at 220 (or 225) nm of the native state and the transient intermediate formed at the first stage in refolding are presented in this paper. The conditions, pH, temperature and ionic strength, used in the kinetic experiment for each protein are listed in Table I. The values of pH, temperature and ionic strengh are taken into account in calculations of the secondary structure contents based on the theory of protein secondary structure [16,17].

3. RESULTS AND DISCUSSION

Table I shows the mean residue ellipticities of the 14

proteins for the early intermediate (θ_I) and the native (θ_N) states. The values for the proteins calculated from their X-ray structures (θ_N^X) and from their amino acid sequences using the theory of protein secondary structure (θ^T) are also included for comparison. The values of θ_N^X and θ^T were calculated by the equations:

$$\theta_N^X = \theta_\alpha f_\alpha^X + \theta_\beta f_\beta^X + \theta_r (1 - f_\alpha^X - f_\beta^X) \quad (1)$$

and $$\theta^T = \theta_\alpha f_\alpha^T + \theta_\beta f_\beta^T + \theta_r (1 - f_\alpha^T - f_\beta^T) \quad (2)$$

Here, f_α and f_β are the secondary structure contents of α-helix and β-sheet, respectively, calculated from the X-ray structures by the method of Kabsch and Sander (f_α^X and f_β^X) [20] or predicted by the theory of protein secondary structure (f_α^T and f_β^T) [16,17]. θ_α, θ_β and θ_r are ellipticities for the pure secondary-structure components, α-helix, β-sheet and irregular structure, respectively, taken from Bolotina et al. [21], and the values at three different wavelengths are given in the footnote of Table I. It should be noticed that the θ_α, θ_β and θ_r values have been obtained by using the experimental data on θ_N for 6 proteins and that the secondary structures of these proteins have been determined from their X-ray coordinates in a way similar to that of Kabsch and Sander [20,21]. Therefore, the θ_α, θ_β and θ_r values of Bolotina et al. used in this paper must be consistent with the f_α^X and f_β^X values defined by Kabsch and Sander.

Fig. 1 shows the relationship between the experimental ellipticities of the native proteins and the values calculated from their X-ray structures. It gives a very good correlation between these two quantities (the correlation coefficient $r = 0.82$), suggesting that the secondary-structure components dominate over the contributions

Table I

Observed ellipticities of 14 proteins compared with those calculated from their X-ray structures and predicted from the theory of protein secondary structure

	Conditions**				Observed			Calculated from X-ray structure				Predicted from the theory§§		
Protein*	pH	T (K)	I (M)	λ (nm)	θ_N $\times 10^{-3}$	θ_I $\times 10^{-3}$	References***	PDB§	f_α^X (%)	f_β^X (%)	θ_N^X $\times 10^{-3}$	f_α^T (%)	f_β^T (%)	θ^T $\times 10^{-3}$
RNS	4.0	283	1.5	220	−8.8	−4.1	SK	7RSA	13	29	−7.4	11	11	−3.0
BLG	3.0	278	0.5	220	−6.8	−10.8	7	–	–	–	–	27	19	−10.8
BLA	7.0	278	0.5	220	−10.2	−9.0	6,14	1ALC	25	9	−8.1	30	9	−9.9
HLA	8.2	296	0.03	225	−8.0	−8.0	SK	1ALC	25	9	−7.2	28	8	−8.2
LYZ	1.5	278	0.5	220	−8.7	−7.7	6	2LZT	26	10	−8.8	21	8	−6.3
CHT	3.0	278	0.1	220	−4.6	−4.6	SK	2CGA	6	31	−5.2	22	14	−7.9
SBL	7.2	296	0.4	225	−7.5	−5.3	SK	3BLM	29	15	−9.3	12	16	−3.0
CAB	8.0	296	0.03	220	−3.6	−3.6	13	2CAB	6	29	−4.7	16	13	−5.3
SNS	7.0	278	0.1	225	−9.7	−3.3	11	1SNC	19	26	−6.4	20	11	−5.5
DHF	7.8	288	0.03	220	−7.9	−4.4	12	5DFR	16	30	−8.6	28	10	−9.4
PGK	7.3	296	0.04	225	−6.0	−3.5	SK	3PGK	28	10	−8.4	31	14	−9.9
PAR	7.0	278	0.5	220	−13.4	−8.6	9	4CPV	41	2	−12.6	34	6	−10.9
WSY	7.8	285	0.3	220	−12.0	−6.1	10	1WSY	30	14	−11.0	37	9	−12.7
CYT	7.0	298	0.5	222.5	−11.1	−9.1	7	3CYT	33	2	−9.7	14	7	−3.1

The ellipticity values are shown as mean residue ellipticities in deg cm^2/dmol. The θ_α, θ_β and θ_r values to calculate θ_N^X and θ^T are: $\theta_\alpha = -35.2 \times 10^3$, $\theta_\beta = -17.0 \times 10^3$, $\theta_r = 3.5 \times 10^3$ at 220 nm; $\theta_\alpha = -35.6 \times 10^3$, $\theta_\beta = -12.9 \times 10^3$, $\theta_r = 3.5 \times 10^3$ at 222.5 nm; and $\theta_\alpha = -34.3 \times 10^3$, $\theta_\beta = -6.8 \times 10^3$, $\theta_r = 3.1 \times 10^3$ at 225 nm [21].

*RNS, bovine ribonuclease A; BLG, bovine β-lactoglobulin A; BLA, bovine α-lactalbumin; HLA, human α-lactalbumin; LYZ, hen lysozyme; CHT, bovine chymotrypsinogen A; SBL, staphylococcal β-lactamase; CAB, bovine carbonic anhydrase B; SNS, staphylococcal nuclease A; DHF, *E. coli* dihydrofolate reductase; PGK, yeast phosphoglycerate kinase; PAR, carp parvalbumin III; WSY, *E. coli* tryptophan synthase β_2 subunit; CYT, horse ferri-cytochrome *c*.

**Experimental conditions, pH, temperature (T), ionic strengh (I) and wavelength (λ), used for observation of θ_N and θ_I and also for calculation of θ^T.

***SK means unpublished data of G.V. Semisotnov and K. Kuwajima.

§File names of the X-ray data from the Brookhaven Protein Data Bank (PDB) [29]. For BLG, the X-ray coordinates [30] have not yet been deposited in the PDB. For BLA and HLA, the X-ray data of baboon α-lactalbumin are used; the X-ray structure of the human protein has recently been reported [31] but has not yet been deposited in PDB (the July issue of 1992). For CAB, the X-ray data of human carbonic anhydrase B whose sequence shows 53 replacements and 1 deletion out of 261 residues when compared with the sequence of the bovine protein [32] are used. For WSY, the f_α^X and f_β^X values were calculated for the β_2 subunit from the X-ray data of the $\alpha_2\beta_2$ complex of the *Salmonella typhimurium* protein. For CYT, the X-ray data of tuna cytochrome *c* (oxidized) are used.

§§The values predicted from the theory of secondary structure in a globular state of proteins [16,17] and calculated by Equation **(2)** (see text). Amino acid sequences for all proteins except CAB are from the sequence data base of Micro Genic of Beckman Instruments Inc. The sequence for CAB is from [32].

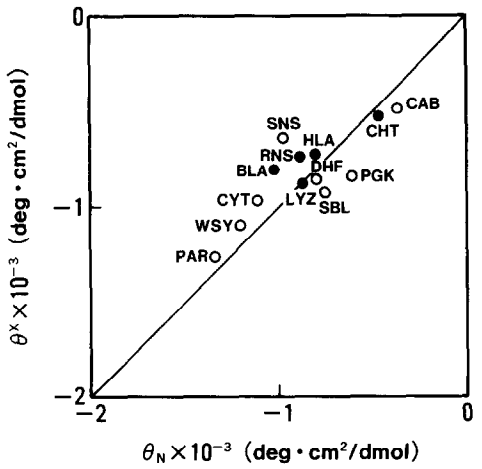

Fig. 1. The ellipticity values expected from the X-ray structures versus the experimentally observed ellipticities for the native proteins at 220, 222.5 or 225 nm (θ_N). The expected ellipticities were calculated from their X-ray structures and reference ellipticities for the pure secondary structure components, α-helix, β-sheet and irregular structure (see Equation (1)). The thin solid line is a diagonal. Filled and open circles refer to the proteins with and without disulfide bonds, respectively. The correlation coefficient $r = 0.82$ between θ_N^X and θ_N.

of aromatic groups and disulfide bonds in the far UV CD spectra in these proteins, although the ellipticities of some other proteins are known to be more strongly influenced by the aromatic groups and/or disulfide bonds [22–24]. It also shows that the ellipticity at 220–225 nm is a reasonably good empirical measure of the secondary structure.

Fig. 2 presents the correlation between the experimental values of ellipticity for 14 proteins in their early intermediates and the native states. We can see that for 6 proteins (BLA, HLA, LYZ, CHT, CAB, CYT) $\theta_I \cong \theta_N$, for 7 proteins (RNS, SBL, DHF, PGK, PAR, SNS, WSY)$|\theta_I| < |\theta_N|$, and for one (BLG) $|\theta_I| > |\theta_N|$. Thus, the general result of the comparison of the far UV CD between the early intermediate and the native state supports the idea that the main part of secondary structure is formed very early in protein folding and that the secondary structure is then stabilized further at the subsequent stages of folding [1–6]. As the ellipticity of the early intermediate could be influenced by a small amount of urea or guanidine present in the solutions of refolding experiments, the similarity between θ_I and θ_N may be even greater if the traces of the denaturants were absent.

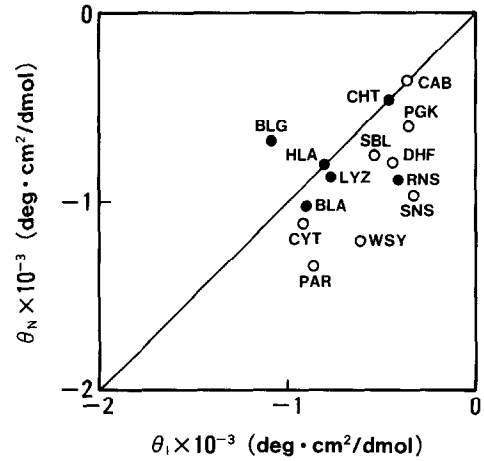

Fig. 2. The observed ellipticities in the native state (θ_N) versus those in the early kinetic intermediate (θ_I) of the 14 proteins studied ($r = 0.44$).

Fig. 3 shows the comparison of both θ_N and θ_I with the values calculated from the theory of protein secondary structure [16,17]. Cytochrome *c* was omitted from this comparison because the known contribution of a covalently attached heme to stabilization of its α-helix [25] is beyond the scope of this theory. We have used the version of the theory that takes into account, in some crude way, long-range interactions and therefore can be

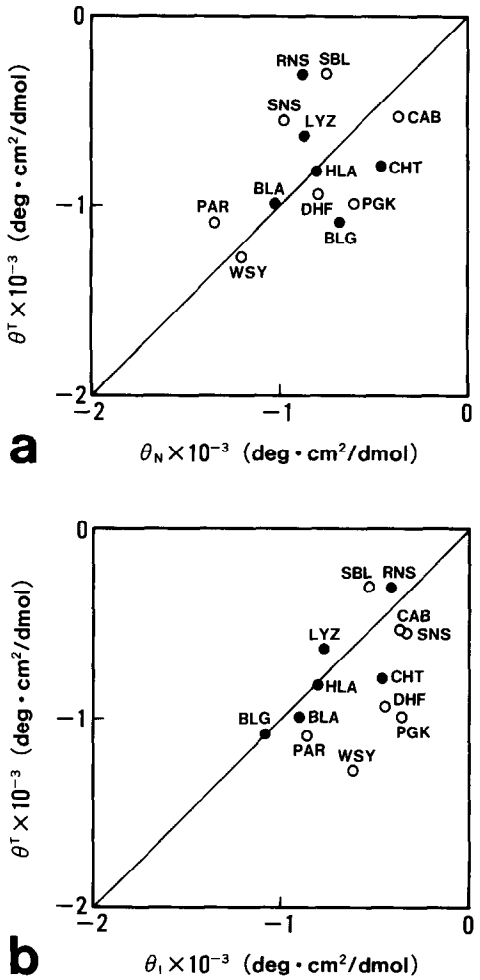

Fig. 3. The mean residue ellipticities of the 13 proteins, predicted from the theory of protein secondary structure and reference ellipticities for the pure secondary structure components, α-helix, β-sheet and irregular structure (see Equation (2)) versus the observed ellipticities in the native state (θ_N) ($r = 0.43$) (a), and those in the early intermediate (θ_I) of the proteins ($r = 0.41$) (b).

applied to compact molecules, e.g. the molten globule state. The similar results have been, however, obtained using the other version of the theory [17] that takes into account only short and middle range interactions, hence describing the non-compact forms of proteins [26,27]. Therefore, although we do not know to what extent the proteins are compact at the early stages of their folding, practically this uncertainty does not influence our results. The figure shows that the theory does not contradict the experimental data and predicts the ellipticities that are comparable with both θ_N and θ_I, and the r is 0.43 between θ^T and θ_N and 0.41 between θ^T and θ_I. One could not expect a much better correlation, as θ_N may be influenced by aromatic groups and/or disulfide bonds [22,28], the secondary structures of the refolding intermediate may be destabilized by a residual concentration (~ 0.5 M) of urea or guanidine in our experiment, and also the theory includes approximations, especially with respect to long range interactions.

Acknowledgements: We thank Professor C. Sander (EMBL, Germany) for providing us with the DSSP program to calculate the structural fractions from the X-ray coordinates and Professor I. Tanaka (Hokkaido Univ., Japan) for the use of a MicroVAX/VAXserver. This study was made when K.K. visited the Institute of Protein Research (USSR) and also when G.V.S. visited Hokkaido University and the University of Tokyo. It was supported by the Japan Society for the Promotion of Science and by Academy of Sciences of the USSR.

REFERENCES

[1] Ptitsyn, O.B. (1987) J. Protein Chem. 6, 273–293.
[2] Kuwajima, K. (1989) Proteins 6, 87–103.
[3] Kim, P.S. and Baldwin, R.L. (1990) Annu. Rev. Biochem. 59, 631–660.
[4] Ptitsyn, O.B. (1992) in: Protein folding (Creighton, T.E., Ed.) pp. 243–300, Freeman, New York.
[5] Kuwajima, K. (1992) Curr. Opin. Biotechnol. 3, 462–467.
[6] Matouschek, A., Serrano, L. and Fersht, A.R. (1992) J. Mol. Biol. 224, 819–835.
[7] Kuwajima, K., Hiraoka, Y., Ikeguchi, M. and Sugai, S. (1985) Biochemistry 24, 874–881.
[8] Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S. and Nagamura, T. (1987) FEBS Lett. 221, 115–118.
[9] Gilmanshin, R.I. and Ptitsyn, O.B. (1987) FEBS Lett. 223, 327–329.
[10] Kuwajima, K., Sakuraoka, A., Fueki, S., Yoneyama, M. and Sugai, S. (1988) Biochemistry 27, 7419–7428.
[11] Goldberg, M.E., Semisotnov, G.V., Friguet, B., Kuwajima, K., Ptitsyn, O.B. and Sugai, S. (1990) FEBS Lett. 263, 51–56.
[12] Sugawara, T., Kuwajima, K. and Sugai, S. (1991) Biochemistry 30, 2698–2706.
[13] Kuwajima, K., Garvey, E.P., Finn, B.E., Matthews, C.R. and Sugai, S. (1991) Biochemistry 30, 7693–7703.
[14] Ptitsyn, O.B. and Semisotnov, G.V. (1991) in: Conformations and Forces in Protein Folding (Nall, B.T. and Dill, K.A., Eds.) pp. 155–168, American Association for the Advancement of Science, Washington, DC.
[15] Ikeguchi, M., Sugai, S., Fujino, M., Sugawara, T. and Kuwajima, K. (1992) Biochemistry 31, 12695–12700.
[16] Ptitsyn, O.B. and Finkelstein, A.V. (1983) Biopolymers 22, 15–20.
[17] Finkelstein, A.V. Program 'ALB' for protein and polypeptide secondary structure calculation and prediction (1983). Deposited in the Brookhaven Protein Data Bank, Upton, New York, and (a more user-friendly version) in the EMBL, Heidelberg, Germany; distributed by request.
[18] Semisotnov, G.V., Rodionova, N.A., Razgulyaev, O.I., Uversky, V.N., Gripas, A.F. and Gilmanshin, R.I. (1991) Biopolymers 31, 119–128.
[19] Semisotnov, G.V., Vas, M., Chemeris, V.V., Kashparova, N.J., Kotova, N.V., Razgulyaev, O.I. and Sinev, M.A. (1991) Eur. J. Biochem. 202, 1083–1089.
[20] Kabsch, W. and Sander, C. (1983) Biopolymers 22, 2577–2637.
[21] Bolotina, I.A., Chekhov, V.O., Lugauskas, V.Y., Finkelstein, A.V. and Ptitsyn, O.B. (1980) Mol. Biol. (USSR) 14, 891–901.
[22] Woody, R.W. (1985) in: The Peptides, Vol. 7 (Udenfried, S., Meienhofer, J. and Hruby, V.J., Eds.) pp. 15–113, Academic Press, San Diego.
[23] Manning, M.C. and Woody, R.W. (1989) Biochemistry 28, 8609–8613.
[24] Rodriguez-Romero, A., Arreguin, B. and Hernandez-Arana, A. (1989) Biochim. Biophys. Acta 998, 21–24.
[25] Fisher, W.R., Taniuchi, H. and Anfinsen, C.B. (1973) J. Biol. Chem. 248, 3188–3195.
[26] Ptitsyn, O.B. and Finkelstein, A.V. (1980) Q. Rev. Biophys. 13, 339–386.
[27] Finkelstein, A.V., Badretdinov, A.Y. and Ptitsyn, O.B. (1991) Proteins 10, 287–299.
[28] Bolotina, I.A. (1987) Mol. Biol. (USSR) 21, 1625–1635.
[29] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) Eur. J. Biochem. 80, 319–324.
[30] Papiz, M.Z., Sawyer, L., Eliopoulos, E.E., North, A.C.T., Findlay, J.B.C., Sivaprasadarao, R., Jones, T.A., Newcomer, M.E. and Kraulis, P.J. (1986) Nature 324, 383–385.
[31] Acharya, K.R., Ren, J., Stuart, D.I., Phillips, D.C. and Fenna, R.E. (1991) J. Mol. Biol. 221, 571–581.
[32] Tashian, R.E., Hewett-Emmett, D., Stroup, S.K. and Goodman, M. (1980) in: Biophysics and Physiology of Carbon Dioxide (Bauer, C., Gros, G. and Barkl, H., Eds.) pp. 165–176, Academic Press, New York.