CCA-744

547.466.1:541.183

Conference Paper

Surfaces and the Folding of Polypeptide Chains*

A. G. Walton

Case Western Reserve University, Division of Macromolecular Science, Cleveland, Ohio 44106, U.S.A.

Received September 1, 1972

An approach to the folding of polypeptide chains is explored theoretically and compared with experimental data. It appears that in synthetic chain folded polypeptide crystals, the fold period is statistically determined by the energetics of nucleation.

The broader question of chain folding in globular proteins is presented and an equation derived which shows that the fold length is likely to be a function mainly of the particular fold energy and the driving free energy derived from long range stabilization forces.

INTRODUCTION

Of the surfaces and interfaces which exist in biological systems, virtually none has been characterized, mainly because of the lack of sophisticated tools for experimental evaluation. The magnitude of the problem can be judged from the fact that structures of molecular dimensions are often involved in determining specific chemical processes. One such problem is associated with the surface structure and properties of globular proteins, which in large measure affects their efficacy in enzymic and immunogenic reactions. The manner in which blood proteins and cellular matter interact with prosthetic implants and the nature of the interface is another area of current concern.

A new class of polymeric surfaces which is currently being investigated in our laboratories, with a view to elucidation of behavior of implants is that of biopolymers. Films of copolypeptides may be prepared which have several characteristics in common with commercial plastics. Similarly single crystals of polyamino acids and polypeptides may be prepared under special circumstances, which gives us a chance to examine, at least indirectly, the surface structure. Some twenty or more biopolymers have now been crystallized, a summary of these materials is given in references 1 and 2. Whereas there has been continuing discussion concerning the regularity of folds at the surfaces of commercial polymer crystals, there are two aspects of polypeptide chain folding which enable more complete evaluation to be carried through. Firstly steric requirements restrict possible bond rotation angles and the forces involved in stabilizing conformations are fairly well established. Secondly, single crystal data for globular proteins establish the structure of certain types of folds unambiguously.

The purpose of this paper is to correlate experimental and theoretical information concerned with the nucleation of synthetic polypeptide crystals

---

* Based on a lecture presented at the *III International Summer School on the Chemistry of Solid/Liquid Interfaces*, Rovinj, Yugoslavia, July 1—5, 1972.

and the development of their surfaces and to propose a similar mechanism for the folding of globular proteins.

Although polypeptides may be crystallized in single crystal form with at least four known conformations[3-6] (apha, X beta, polyproline II and polyglycine II) the cross beta form only will be dealt with here.

Nucleation Theory

The classical nucleation theory of polymers assumes that the chains fold to form a nucleus and that the process is promoted by intermolecular forces but opposed by interfacial forces and chain folding[7]. There are two separate cases, homogeneous nucleation, where the folding and nucleus formation steps are spontaneous and heterogeneous nucleation, where the crystal growth is catalyzed by, and begins on, the surface of some foreign entity, often high molecular weight material[8].

a) Homogeneous Nucleation. — This process is probably encountered in globular protein formation and in simultaneous polymerization/crystallization reactions in general. A model chain folded (beta) nucleus is shown in Fig. 1. The Gibbs free energy of formation of the nucleus may be written[9]

$$\Delta G = v \Delta G_v + \Sigma_i A_i \sigma_i \tag{1}$$

where v is the volume of the nucleus, ΔG_v is the crystallization free energy per unit volume and the interfacial energetics are summed over the various surfaces i.

For the nucleus shown in Fig. 1 we may write

$$\Delta G = abl \Delta G_v + 2 ab\sigma_e + 2 bl\sigma_1 + 2 al\sigma_2 \tag{2}$$

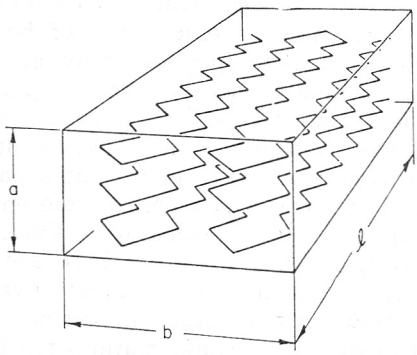

Fig. 1. Schematic diagram of a chain folded nucleus such as that occurring in a cross beta polypeptide chain. Interchain hydrogen bonds would lie perpendicular to the molecular axis and parallel to the fiber axis in the final crystallite.

Spontaneous nucleation occurs when

$$\frac{\delta \Delta G}{\delta a} = \frac{\delta \Delta G}{\delta b} = \frac{\delta \Delta G}{\delta l} = 0 \tag{3}$$

Assuming that a fold period l^* is »frozen in« during the nucleation process then

$$l^* = \frac{-4\sigma_e}{\Delta G_v} \tag{4}$$

where σ_e is the interfacial free energy of the fold surface. Now for a polypeptide the volume free energy (of crystallization) may be written

$$\Delta G_v = \Delta G_{inter} + \Delta G_{intra} - \Delta G_{solv} \tag{5}$$

where the free energy terms on the right are those referring to the change of interaction on a intermolecular and intramolecular basis as the chain passes from solvated to solid state.

In addition the major contribution to the end interfacial energy σ_e is the fold free energy ΔG_{fold}.

Thus the fold period for polypeptide chains nucleated homogeneously should be calculable on the basis of conformational theory provided that a reasonable model for the surface structure can be postulated, and is

$$l^* = \frac{-4\Delta G_{fold}}{\Delta G_{inter} + \Delta G_{intra} - \Delta G_{solv}} \tag{6}$$

b) *Heterogeneous Nucleation.* — If the nucleus of Fig. 1 were formed on a solid substrate, then equation (2) becomes

$$\Delta G = abl\Delta G_v + 2ab\sigma_e + 2al\sigma_2 + bl\sigma_1 + bl\sigma_{1_s} - bl\sigma_s \tag{7}$$

If the substrate is high molecular weight material of the same type as the crystallizing phase, then $\sigma_s = \sigma_1$ and $\sigma_{1s} = 0$.
Thus equation (7) becomes

$$\Delta G = abl\Delta G_v + 2ab\sigma_e + 2al\sigma_2 \tag{8}$$

and application of (3) and (5) gives

$$l^* = \frac{-2\Delta G_{fold}}{\Delta G_{inter} + \Delta G_{intra} - \Delta G_{solv}} \tag{9}$$

i. e. the fold period is halved.

In the case where the substrate is not high molecular weight material of the crystallizing phase, as would probably be the case for film cast material, then l^* may be expected to lie between these two extreme values.

It is noteworthy that equation (6) and (9) represent minimum values of l^* for polypetide crystals and that thermal aging would probably involve thickening and increase of l^* with time, as with commercial polymers. However, at this time there is no experimental evidence for such a phenomenon.

A second feature of equation (6) and (9) is that for alpha helices crystallizing from solution there is generally no conformational change ($\Delta G_{intra} = 0$) and ΔG_{inter} is small, ΔG_{fold} probably large, resulting in large l. Conversely there is a large change of conformational free energy for beta structures on crystallization (ΔG_{inter} large), ΔG_{fold} is fairly small and thus l will be small. Both of these features are found experimentally[1,2]. Also in strong solvents, where ΔG_{solv} is large the fold period is expected to increase, another observed

effect. For beta structures, the minimum observed fold periods is ~ 25 Å, most are approximately 60 Å but polyglutamic acid[2] is 120 Å.

Conformational Analysis

The intra chain energetics of polypeptide chains have been explored fairly extensively[10]. More recent developments have included inter chain energetics[11], inclusion of entropy and thus free energy[12] and solvent interactions[13]. Although it is not appropriate to include here the details of such calculations, the methodology employed in these previous references has been used.

There are three new aspects to the calculations reported here.

a) Solvation Energies. — Since parameters are currently available only for interaction between polypeptide chains and water, the calculations are limited to this solvent. Hopfinger[13] has presented a complete table of solvation energies for poly amino acids in which the conformational free energy has been optimized. However, to solve equation (6) or (9), $\Delta G_{solv} - \Delta G_{intra}$ is required. The most strainghtforward method of calculation is achieved if the beta form is assumed in solution and $\Delta G_{intra} = 0$. Solvation energies are presented in Table I and differ from those quoted by Hopfinger by ΔG_{intra}.

TABLE I

Hydration Energies of Polyamino Acids in the β Conformation*

Polymer	Energy	Entropic Contrib. ($-T\Delta S$)	Free Energy
Poly-L-glutamic Acid	— 14.0	— 1.7	— 15.7
Polyglycine	— 4.4	—	
Poly-L-alanine	— 10.0	—	

b) Fold Energies. — The fold free energy is the crux of the problem in that the surface structure of the crystal defines ΔG_{fold} and also the fold period. Whereas there may be many different variations of beta folds involved in the surface structure of polypeptide crystals, there are two simple forms which commonly occur in globular proteins involving four peptide units[14]. The structure of these folds is shown in Fig. 2. In the type II fold there is a close contact which allows only glycine in the third position; this will be referred to as the glycine fold. The other form (fold I) can accommodate any residue and will be called the amino fold.

The fold energies have been calculated here for polyglycine in fold types I and II and for poly-L-alanine and poly-L-glumatic acid in the amino type I fold. Results are presented in Table II.

As can be seen, three contributions have been considered, changes in intramolecular energy, solvation of the fold and changes in intermolecular energy arising from broken hydrogen bonds.

Two contributions which have not been taken into account are: *a)* interfold energies, *b)* entropy of residues in the fold. Both of these parameters are extremely difficult to obtain unless specific geometry is known. For purposes of convenience, contributions by both are taken as zero.

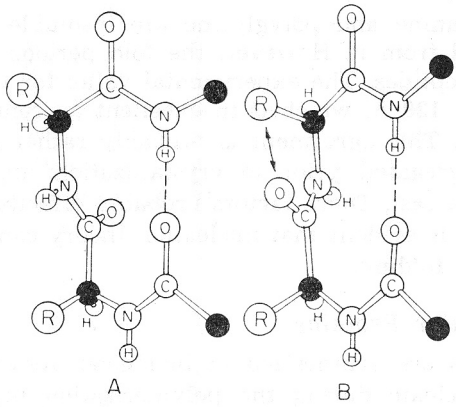

Fig. 2. Two possible structures for the beta fold. Fold II can only accommodate glycine in the position indicated by short >C = 0, R distance.

TABLE II

Fold Energies for Polyamino Acids in the β Fold*

Polymer	Intramol.	Solvation	Intermol.	Total
Glycine Fold (II)				
Polyglycine	— 0.8	— 7.9	+ 50.0	+ 29.1
Amino Fold (I)				
Polyglycine	— 4.5	— 7.9	+ 50.0	+ 37.6
Poly-L-alanine	+ 2.0	— 4.0	+ 65.0	+ 63.0
Poly-L-glutamic Acid	— 45.0	— 22.0	+ 80.0	+ 63.0

* kcals/total fold (five residues)

It is evident that since equation (6) and (9) involve ratios of free energies, errors in the absolute magnitude will tend to be minimized. There are though evident approximations which make accurate prediction of fold periods difficult.

c) Calculated Fold Periods. — With the preceding limitations, we have calculated the fold period of three polyamino acids for crystallization from water; namely, polyglycine with folds I and II and poly-L-alanine and poly-L-glutamic acid using the amino fold.

The data in Table III are based on the homogeneous nucleation equation (6). It can be seen that the calculated fold periods fall in the range 18—145 Å which is virtually identical with the fold periods observed for beta polypeptides.

TABLE III

Calculated Fold Periods for Some Polyamino Acids Nucleated from Aqueous Solution

Polymer	l*
Polyglycine [Glycine fold (II)]	18 Å
Polyglycine [Amino fold (I)]	20 Å
Poly-L-alanine [Amino fold (I)]	23 Å
Poly-L-glutamic acid [Amino fold (I)]	143 Å

In fact poly-L-alanine and polyglycine are insoluble in water and thus cannot be crystallized from it. However, the fold periods are perhaps typical of the small apolar peptides. The experimental value for beta poly-L-glutamic acid is approximately 120 Å, which is in excellent agreement with the calculations reported here. This agreement is probably rather fortuitous since the acid is probably aggregated prior to crystallization and nucleates heterogeneously as it is film cast. There factors probably effectively counteract each other. Nevertheless it is evident that nucleation theory can effectively account for polypeptide chain folding.

Nucleation of Globular Proteins

Polypetide chains are synthesized in biological systems at the ribosome surface and may nucleate during the polymerization process or may form amorphous »oil drops« before nucleating into the conformation and structure of the globular protein. The folding process is undoubtedly driven by much the same forces as for polypeptide chain folding, namely improvement of intra and inter chain interactions and optimization of solvation and surface forces.

In principle, equation (1) might be applied but it does not lead to a useful formulation.

As an approach to this problem the model used in Fig. 3 will be explored. In this case we assume that a structure (not necessarily spherical) is formed

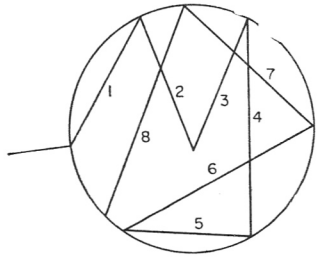

Fig. 3. Schematic representation of lengths l_i of conformation of the ith type in a nucleating globular protein.

in which the free energy is expressed as a combination of chain energetics as before

$$\Delta G = n\overline{\Delta G}_v + x\,\Sigma_i \Delta G_f \tag{10}$$

Here, $\overline{\Delta G}_v$ is the average free energy change/peptide between solvent and protein phase, n is the number of peptide units and x the number of folds. ΔG_f is the fold energy of the ith fold. In this case the surface energetics are accounted for in the partial solvation ($\overline{\Delta G}_v$ term) and solvation and chain folding once again inhibit the development of tertiary structure.

Rather than choosing to maximize equation (10) and assume that folding occurs as an energy barrier is overcome, we shall assume that the tertiary structure is initiated at some critical value of n^* such that $\Delta G \leq 0$.

If the structure and conformation are initiated at this point ($\Delta G = 0$) then the equation can be solved for lengths of various conformation l_i as specified in Fig.3.

$$e.\,g. \qquad n^* = n_1^* + n_2^* + n_3^* \ldots$$

where the n's are those corresponding to the various conformations and $l_1^* = c_1 n_1^* \quad l_2^* = c_2 n_2^* \quad l_3^* = c_3 n_3^* \quad etc.$

where c_1, c_2 *etc.* are the characteristic peptide repeat distances (1.5 Å for alpha helix, 3.5 Å extended beta, *etc.*).

Then

$$n^* = \frac{l_1^*}{c_1} + \frac{l_2^*}{c_2} + \ldots = -\left[\frac{\Delta G_f^*(1) + \Delta G_f^*(2) + \ldots}{\Delta G_v}\right] \tag{11}$$

If equation (11) were applied to a polyamino acid crystal as before, we would have

$$l^* = -c\,\frac{\Delta G_f^*}{\Delta G_v^*}$$

i. e. an equation identical in form to equations (6) and (9).

It seems intuitively likely therefore that equation (11) can be formulated as a family of equations of the form

$$l_i^* = \frac{-c_i \Delta G_f(i)}{\Delta G_v} \tag{12}$$

This equation indicates that the length of a particular conformation in the nucleus of a globular protein is dictated only by the fold energy — a rather significant fact.

A test of the above equation is possible once the conformational energetics of the various folds for proteins are known.

Table IV contains a few data which indicate that the fold lengths (coherence lengths) are of the correct order.

TABLE IV

Fold Periods and Coherence Lengths for β Folded Chains

Protein	Form	l
Chrysopa flava	X β fiber	25—30 Å
Carboxypeptidase	Globular (8 chains)	21—31 Å
Ribonuclease	Globular (3 chains)*	25—28 Å
Lysozyme	Globular (2 chains)	20—25 Å

* Other longer β segments are bent.

Evidently if all folds in globular proteins were of the same structure and the general chain conformation were known or could be predicted, then the points of folding could be enumerated and the tertiary structure of new proteins could be predicted.

CONCLUSIONS

The development of tertiary structure and chain folding in synthetic polypeptides can be described by a classical nucleation formulation combined with theoretical conformational analysis. The three major factors involved with controlling the fold periods and surface structure of polypeptide crystals are fold energy, bulk of the peptide and side chains and interactions with the surrounding solvent.

Chain folding in globular proteins is a considerably more complicated process but can be approached in a similar manner. The approach indicates that long range microthermodynamic requirements must be met. It seems probable that in contrast to the crystallization of many synthetic polypeptides there is a concommitant development of conformation and tertiary structure in globular proteins. This new formulation does not preclude the role of specific residues in directing folds (glycine is likely to lower fold energy) particularly if unusual folds are formed, but it does indicate that the role of specific peptides in directing tertiary and surface structure is probably not as great as had been formerly assumed.

REFERENCES

1. A. G. Walton and J. Blackwell, *Biopolymers*, Academic Press, New York, 1973.
2. A. G. Walton in *Fifth Jerusalem Symposium on Proteins and Polymers*, Academic Press, New York, in press.
3. F. J. Padden and H. D. Keith, *J. Appl. Phys.* **36** (1964) 2987.
4. J. C. Andries and A. G. Walton, *J. Mol. Biol* **56** (1971) 515.
5. J. C. Andries and A. G. Walton, *Biopolymers* **9** (1969) 523.
6. J. C. Andries, J. M. Anderson, and A. G. Walton, *Biopolymers* **10** (1971) 1049.
7. F. P. Price, Chapter 8 in *Nucleation*, A. Zettlemoyer (Editor), Dekker, New York, 1969.
8. D. J. Blundell and A. Keller, *J. Macromol. Sci.* (Phys.) **B 2** (1968) 301.
9. A. G. Walton, Chapter 5 in reference 7.
10. H. A. Scheraga, *Advan. in Phys. Org. Chem.* **6** (1968) 103.
11. A. J. Hopfinger and A. G. Walton, *J. Macromol. Sci.* (Phys.) **63** (1972) 101.
12. F. R. Brown, A. J. Hopfinger, and E. R. Blout, *J. Mol. Biol.* **4** (1971) 731.
13. A. J. Hopfinger, *Macromolecules* **4** (1971) 731.
14. P. N. Lewis, F. A. Momany, and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **68** (1971) 2293.

IZVOD

Površine i presavijanje polipeptidskih lanaca

A. G. Walton

Opisana je korelacija između eksperimentalnih podataka i teorijske informacije o nukleaciji sintetskih polipeptida, kao i o stvaranju oblika njihove površine. Prikazani su osnovi teorije nukleacije, homogene i heterogene, kao i specijalni slučaj nukleacije globularnih proteina. Konformacijska analiza proteina zasniva se na procjeni pojedinačnih doprinosa ukupnoj volumnoj slobodnoj energiji polimera. U tablicama su navedeni podaci za energije hidratacije i presavijanja, kao i veličine konformacijskih perioda za poliglicin, poli-L-alanin i poli-L-glutaminsku kiselinu. Navedeni su i podaci za periodu i koherentne duljine nekih β-lanaca.

CASE WESTERN UNIVERSITY
CLEVELAND, OHIO, USA

Primljeno 1. rujna 1972.