Hypothesis

Origins of globular structure in proteins

Vladimir Tolstoguzov*

Nestlé Research Centre, P.O. Box 44, Vers-chez-les-Blanc, CH-1000 Lausanne 26, Switzerland

Received 1 October 1998; received in revised form 5 January 1999

Abstract Thermodynamic incompatibility of polymers in a common solvent is possibly a driving force for formation and evolution of globular protein structures. Folding of polypeptide chains leads to a decrease in both excluded volume of molecules and chemical differences between surfaces of globular molecules with chemical information hidden in the hydrophobic interior. Folding of polypeptide chains results in 'molecular or thermodynamic mimicry' of globular proteins and in at least more than 10-fold higher phase separation threshold values of mixed protein solutions compared to those of classical polymers. Unusually high co-solubility might be necessary for efficient biological functioning of proteins, e.g. enzymes, enzyme inhibitors, etc.

© 1999 Federation of European Biochemical Societies.

Key words: Protein evolution; Thermodynamic incompatibility; Excluded volume; Mimicry; Transport in the cell; Protein folding

1. Introduction

One of the most fascinating problems of molecular evolution is why a compact globular conformation was chosen for proteins. A hypothesis related to the origins of globular structure in proteins has recently been proposed [1]. According to this hypothesis, membranes play a key role in the gradual evolution of primitive polypeptide chains from a random coil to a folded globular conformation. However, thermodynamic constraints in the bulk of mixed biopolymer solutions could also be a reason for the evolution of the primary structure of polypeptide chains towards globular conformations of proteins. This assumption is based on the phenomenon of limited thermodynamic compatibility (or briefly, incompatibility) of biopolymers in the common solvent water. This has been mainly studied in terms of miscibility of globular proteins with each other, with proteins of unfolded unordered structures (such as gelatin, casein and denatured proteins) and with various polysaccharides [2–5]. The study of phase behaviour of mixed biopolymer solutions is of great importance to composition-property relationships in food systems and could also be regarded as thermodynamic modelling of primitive biological systems. For instance, mixed solutions of proteins with anionic polysaccharides, such as linear polyacids, could be useful for modelling primitive mixtures of proteins with nucleic acids. So far, however, the importance of the thermodynamic incompatibility of biopolymers has not featured in discussions on the structure and biological activity of proteins in spite of the fact that this phenomenon was discovered more than 100 years ago [6] and has been intensively studied in the last two decades [7–10]. The objective of this paper is to consider a hypothesis that the incompatibility of biopolymers is a tool for forming biological preferences in globular protein structure. The main idea is that folded protein structures increase the miscibility (or co-solubility) of globular proteins with one another and with other biopolymers at the molecular level.

2. Incompatibility of biopolymers. Excluded volume. Phase diagrams

During the last 20 years experimental studies have shown that incompatibility of chemically and/or structurally different biopolymers is the rule rather than the exception [2–4,7–10]. The phase behaviour of biopolymers is quantitatively illustrated in Fig. 1. This is a typical phase diagram for a protein-I–protein-II–water system. The bold curve EFGD is a binodal separating the single- and two-phase states of the mixed solutions. In the region below the binodal, biopolymers are miscible at a molecular level and form single-phase mixed solutions. In the region above the binodal, biopolymers have limited co-solubility. For instance, upon mixing aqueous solutions of a protein A and a protein B, a mixture of composition C is obtained. This mixed solution C spontaneously breaks down into two liquid phases, D and E. The line DE is a tie-line. It connects the points representing the compositions of the co-existing phases. Phase D is rich in protein-I while phase E is rich in protein-II. The critical point G represents the system's composition where the two co-existing phases are of the same composition and volume. The rectilinear diameter passes from the critical point through the mid-tie-lines and gives the composition of systems breaking down

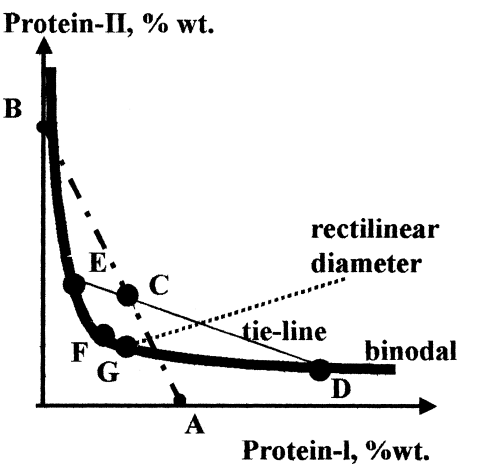

Fig. 1. Typical phase diagram for protein-I–protein-II–water and protein–polysaccharide–water systems.

*Fax: (41) (21) 785-8554.
E-mail: vladimir.tolstoguzov@rdls.nestle.com

0014-5793/99/$19.00 © 1999 Federation of European Biochemical Societies. All rights reserved.
PII: S0014-5793(99)00040-X

into phases of equal volume. The phase separation threshold F (the sum of point F co-ordinates) is the minimal bulk concentration of biopolymers at which phase separation occurs.

Normally, chemically and/or structurally different biopolymers are compatible if they form soluble complexes. When there is no attraction between unlike macromolecules, each of them shows a preference to be surrounded by macromolecules of the same type. Macromolecules compete for space in the mixed solution, where phase behaviour becomes sensitive to entropic factors (given by the excluded volume). Phase separation results in separation of the unlike macromolecules into co-existing phases.

The phase behaviour of biopolymer mixtures can be predicted from the excluded volume of the macromolecules [10–12]. Because molecules are not penetrable by each other, a minimal distance between two spherical molecules of a globular protein equals the sum of their radii or the diameter of one of them. Accordingly, the excluded volume around each protein molecule from which the centres of other protein molecules are expelled is eight-fold greater than that of the protein molecule itself. The excluded volume is significantly greater for non-spherical macromolecules [13]. An increase in excluded volume results in a decrease in compatibility. Normally, classical flexible chain polymers are poorly co-soluble in a common solvent. Their mixed solutions tend to be completely separated into phases containing nearly pure individual polymers [13–15].

Table 1 gives examples of phase separation threshold values for mixtures of some proteins and proteins with polysaccharides. It illustrates the effect of protein conformation. Generally, phase separation occurs at about 1–4% for mixtures of linear polysaccharides with gelatin or casein, which can be regarded as polypeptides differing in molecular weight, chain flexibility, heterogeneity and degree of association. The phase separation threshold is about 4% or higher for globular protein–polysaccharide mixtures and more than 12% for mixtures of globular proteins [2–4,7–10]. The competition between macromolecules for space in solution also determines partitioning of the solvent water between the co-existing phases, i.e. the phase diagram asymmetry. The greater the difference in biopolymer molecular weight, the greater the shift of the binodal towards the concentration axis of the biopolymer with the lower excluded volume.

3. Origins of globular structure. Molecular mimicry

The first remarkable feature of globular proteins is the more than 10-fold higher phase separation threshold values of their mixed solutions compared to those of classical polymers and unfolded polypeptide chains, e.g. proteins such as gelatin and casein. The second feature is the compatibility of many of them, in spite of great differences in amino acid composition and structure. These features might be of key importance for biological functions of proteins, especially enzymes and enzyme inhibitors. The conversion of a polypeptide chain into a folded compact molecule results in a decrease in its volume. The compactness, rounded shape, rigidity and limited number of accessible hydrophobic side groups means that the phase behaviour of proteins differs greatly from that of common polymers. Another result of the conversion of the polypeptide chain into the folded compact molecule with chemical information hidden inside is a decrease in chemical differences between surfaces of non-identical macromolecules and in their affinity to each other and to the solvent.

Formation of globular protein can be regarded as a molecular mimicry resulting in the surprising features of globular proteins. The thermodynamic consequences of mimicry of globular proteins are a marked similarity in phase behaviour of protein mixtures. Thermodynamic incompatibility only occurs between proteins belonging to different classes (albumins, globulins, glutelins and prolamines) within the Osborne classification based on protein solubility [16]. Thus, molecular mimicry could govern collective functioning of proteins with one another and with other biopolymers in biological systems. Phase separation threshold values seem to be a useful measure of the quality of protein mimicry. The idea of protein molecular mimicry implies an analogy with other types of biological mimicry.

4. Compatibility control in biological systems

Unfolding of globular proteins usually decreases their co-solubility (Table 1) with other proteins, including the same native protein [7–10]. By contrast, partial proteolysis of proteins increases their co-solubility with other biopolymers [10]. This was shown by limited trypsin proteolysis of the 11S broad bean globulin in concentrated biphasic mixtures with Ficoll. After a short (10-min) incubation the biphasic mixed solution is converted into a single-phase state due to a modification of the protein, where the molecule loses the least ordered segment, preserves conformational stability and becomes more compact and hydrophilic [10,17]. This could contribute to activation of protoenzymes by cleavage, like that of inactive forms of many enzymes, e.g. conversion of trypsinogen into trypsin, pepsinogen into pepsin, etc.

Proteins were possibly the first natural and versatile catalytic agents. Their evolution can be regarded as a process of functional adaptation to the surroundings. A model of molecular evolution has to involve a stringent rejection of macromolecules that do not meet some functional criteria in interactions with the surroundings. Phase separation could be a mechanism for selection and rejection of molecules with in-

Table 1
Phase separation thresholds of some protein and protein–polysaccharide mixtures [7,8]

Protein-I+protein-II (or +polysaccharide)	Phase separation threshold (%)	Conditions
Gelatin+legumin	8.4	pH 7.0; 40°C
Casein+soybean globulins	12	pH 6.9; 20°C
Ovalbumins native+thermodenatured	13.3	pH 6.7; 20°C
Ovalbumin+soybean globulins	19.7	pH 6.6; 40°C
Casein+Na alginate	3.0	pH 7.2
Gelatin (MW = 170 kDa)+Na alginate (MG = 50%)	1.67	pH 6.0; 40°C
Gelatin (MW = 170 kDa)+Na alginate (MG = 50%)	1.36	pH 6.0; 40°C

sufficient mimicry, e.g. a newly separated phase, rich in unnecessary proteins, to be degraded and reused, should be a better solvent for hydrolases. The principle of phase separation and precipitation of macromolecules with insufficient mimicry could then be used for immune defence. The principle of denaturation, phase separation, precipitation and hydrolysis of macromolecules could then be used for nutrition.

A globular, weakly functioning protein could be initially selected and then evolutionarily improved. Unlike natural selection for the two other operations of molecular evolution, reproduction, i.e. making macromolecular copies and their modification (mutation), the inherited genes are responsible. The evolution of globular structures presumably also included controlling thermodynamic properties of the proteins. For instance, the rigidity of a compact globule prevents swelling, and hence an increase in excluded volume and viscosity.

Contributory factors to rigidity are (i) formation of stiff ordered α-helix and pleated-sheet structural clusters; and (ii) shortening and an increase in the stiffness (vitrification) of the polypeptide chain segments connecting the structural clusters, i.e. formation of ridged glassy clips drawing ordered structural clusters together to form a densely packed globule. The melted glassy segments connecting one- and two-dimensional secondary structural segments (α-helices and β-sheets) seem to lead to the liquid globule (the molten globule, i.e. a reversible intermediary state between the native and fully denatured forms [18,19]). It has been shown that the group additivity approach gives a good estimation of glass transition temperatures of globular proteins [20]. A denaturational change of partial heat capacity of proteins also might reflect melting of the glassy state globule, making its interior accessible to the solvent water [21]. Protein folding might be under thermodynamic control, i.e. attempts to achieve a state of lower free energy. Presumably for this reason an interesting feature of oligomeric proteins (e.g. the 11S storage globulins from oilseeds and leguminous seeds and the enzyme ribulosebisphosphate carboxylase/oxygenase from green leaves) is that their polypeptide chains, differing greatly in molecular weight and amino acid composition, form structural domains thermodynamically equivalent to each other [9].

The evolution in thermodynamic properties of proteins could account for the formation of new hybrid macromolecules (e.g. proteoglycans, glycoproteins, protein–polysaccharide conjugates) comprising two or several incompatible biopolymers bound covalently [22,23]. A transition from compatibility to incompatibility conditions of unlike polymer parts could greatly change the shape and size (excluded volume) of combined macromolecules and the solubility and viscosity of their solutions. This transition could also be regarded as a model of the folding–unfolding behaviour of globular proteins.

5. Collective transport of proteins in the cell

Thermodynamic incompatibility of proteins and other biopolymers is possibly important in many processes occurring in the cell. One of them is the transport of protein molecules between the organelles of the cell. Phase separation of (newly synthesised or modified) protein molecules from the surrounding biopolymer solution (cytosol) may lead to liquid dispersed particles (e.g. phase D in Fig. 1) becoming a vesicle for the collective transport of protein molecules in the cell. This phase separation could be responsible for budding transport vesicles from the organelles and their migration throughout the unwettable cytosol until meeting and coalescing with the wettable surface of a target organelle. This hypothesis [21] does not contradict the idea about phase separation in the cytoplasm of the cell formulated by Walter and Brooks [24] and the mechanisms of intracellular protein transport proposed by Rothman [25]. Protein biosynthesis leads to an increase in excluded volume. According to Le Chatelier's principle the reaction of a highly volume-occupied system (e.g. cytosol) upon an increase in concentration of a biopolymer is to minimise the excluded volume effects. This can be achieved by processes such as phase separation of the cytosol with an increase in concentration of dispersed particles (e.g. phase D in Fig. 1), aggregation and crystallisation of biopolymers. Presumably, for this reason the assistance of molecular chaperons (thermodynamically compatible functional proteins) is absolutely necessary for controlling the association of macromolecules, a correct assembly of oligomeric proteins and prevention of incorrect folding of proteins [26].

6. Concluding remarks

The two principal features of globular proteins compared to classical flexible chain polymers are: (i) compatibility of most of them belonging to the same Osborne class and (ii) higher co-solubility of globular proteins belonging to different Osborne classes. These features are due to: (i) a low excluded volume of the macromolecules; (ii) smaller chemical differences between surfaces of the macromolecules, their interactions with each other and the solvent water; and (iii) the polyelectrolyte nature of biopolymers which enhances co-solubility due to the contribution of low-molecular-weight counterions to an increased mixing entropy.

Thermodynamic incompatibility of biopolymers differing in structure and chemical composition could draw the evolutionary trend towards preferable amino acid compositions providing more compact macromolecules with more chemically similar surfaces. The hypothesis of thermodynamic constraints as a tool for formation and evolutionary improvement of thermodynamic mimicry of protein structures is based on a number of experimental observations. It should also be noted that thermodynamic mimicry and its evolution might be used by other biopolymer constructions, e.g. viruses and the regulation of genes.

Both thermodynamic compatibility of proteins and their incompatibility (including phase separation) in mixed biopolymer solutions seem to be biologically significant. Co-solubility with other macromolecular components is indispensable for biological efficiency of enzymes, while incompatibility is necessary for formation of phase-separated vesicles for collective transportation of protein molecules, for digestion, immunity, and for storage of proteins in the cell, and for natural selection of macromolecules.

Formation and dense packaging of helical structures could also diminish the accessibility of peptide bonds to various reagents. This could create preferences in molecular evolution by natural selection since the direction of hydrolysis is opposite to polymerisation. Perhaps for an individual molecule the formation of globular structures is the best way to improve chemical stability. The formation of a three-dimensional solid globule with the interior isolated from water requires a suffi-

ciently long polypeptide chain. However, the increase in polymer chain length exponentially increases the number of randomly different macromolecules. The increased length of polypeptides implies an increase in the length of the genes and possibly corresponds to the origin of introns and shuffling of exons[27,28].

In conclusion, a compact globular structure is possibly preferential from the viewpoint of protein interactions with both low- and high-molecular-weight components of biological systems.

Acknowledgements: The author is indebted to Dr Elizabeth Prior and Mr Stephen Collyer for their help in preparing the manuscript.

References

[1] Doi, N. and Yanagawa, H. (1998) FEBS Lett. 430, 150–153.
[2] Tolstoguzov, V.B. (1986) in: Functional Properties of Food Macromolecules (Mitchell, J.R. and Ledward, D.A., Eds.), pp. 385–415, Elsevier Applied Science, London.
[3] Tolstoguzov, V. (1997) Food Hydrocolloids 11, 181–193.
[4] Tolstoguzov, V.B. (1997) in: Food Proteins and Their Applications in Foods (Damodaran, S. and Paraf, A., Eds.), pp. 171–198, Marcel Dekker, New York.
[5] Tolstoguzov, V.B. (1998) Nahrung 42, 205–209.
[6] Beijerinck, M.W. (1896) Zent.bl. Bakteriol. Parasitenkd. Infekt.-krankh. 2, 697–699.
[7] Grinberg, V.Ya. and Tolstoguzov, V.B. (1997) Food Hydrocolloids 11, 145–158.
[8] Polyakov, V.I., Grinberg, V.Ya. and Tolstoguzov, V.B. (1997) Food Hydrocolloids 11, 171–180.
[9] Tolstoguzov, V. (1988) Food Hydrocolloids 2, 339–370.
[10] Tolstoguzov, V. (1991) Food Hydrocolloids 4, 429–468.
[11] Semenova, M.G., Pavlovskaya, G.E. and Tolstoguzov, V.B. (1991) Food Hydrocolloids 4, 469–479.
[12] Semenova, M.G., Bolotina, V.S., Grinberg, V.Ya. and Tolstoguzov, V.B. (1990) Food Hydrocolloids 3, 447–456.
[13] Tanford, Ch. (1961) Physical Chemistry of Macromolecules, John Wiley and Sons, New York.
[14] Krause, S. (1978) in: Polymer Blends (Paul, D.R. and Newman, S., Eds.), Vol. 1, pp. 16–113, Academic Press, San Diego, CA.
[15] Kwei, T.K. and Wang, T.T. (1978) in: Polymer Blends (Paul, D.R. and Newman, S., Eds.), Vol. 1, pp. 141–184, Academic Press, San Diego, CA.
[16] Osborne, T.B. (1924) Vegetable Proteins, Longmans, Green, New York.
[17] Danilenko, A.N., Vetrov, V.Yu., Dmitrochenko, A.P., Leontev, A.L., Braudo, E.E. and Tolstoguzov, V.B. (1992) Nahrung 36, 105–111.
[18] Bychkova, V.E., Pain, R.H. and Ptitsyn, O.B. (1988) FEBS Lett. 238, 231–234.
[19] Kuwajima, K., Semisotnov, G.V., Finkelstein, A.V., Sugai, S. and Ptitsyn, O.B. (1993) FEBS Lett. 334, 265–268.
[20] Matveev, Y.I., Grinberg, V.Ya., Sochava, I.V. and Tolstoguzov, V.B. (1997) Food Hydrocolloids 11, 125–133.
[21] Tolstoguzov, V. (1998) in: Proceedings ISOPOW 7 Symposiun on Water Management in the Design and Distribution of Quality Foods, Technomic, Lancaster, CA.
[22] Tolstoguzov, V. (1993) in: Food Colloids and Polymers: Structure and Dynamics (Walstra, P. and Dickinson, E., Eds.), pp. 94–102, Royal Society of Chemistry, Cambrige.
[23] Tolstoguzov, V. (1993) in: Food Proteins. Structure and Functionality (Schwenke, H.D. and Mothes, R., Eds.), pp. 203–209, VCH, Weinheim.
[24] Walter, H. and Brooks, D. (1995) FEBS Lett. 361, 135–139.
[25] Rothman, J.E. (1994) Nature 372, 55–63.
[26] Ellis, R.J. (1997) Biochem. Biophys. Res. Commun. 238, 687–692.
[27] Gilbert, W. (1978) Nature 271, 501.
[28] Bork, P. and Doolittle, R.F. (1992) Proc. Natl. Acad. Sci. USA 89, 8990–8994.