PREDICTION OF THE SECONDARY STRUCTURE OF THE L7, L12 PROTEINS OF THE *E. COLI* RIBOSOME

O.B. PTITSYN, A.I. DENESYUK, A.V. FINKELSTEIN and V.I. LIM
Institute of Protein Research, Academy of Sciences of the USSR, Poustchino, Moscow Region, USSR

Received 15 May 1973

1. Introduction

The primary structure of the first ribosomal proteins – the L7, L12 ("A-protein") of the *E. coli* ribosome 50 S subparticle was recently published [1]. The primary structure of these two proteins is almost identical, the only difference being that the N-terminal serine in the L7 protein is acetylated. The present paper is devoted to a prediction of the secondary structure of the L7, L12 protein by the theory of unfolded protein chain helical structure [2] and the theory of helical and β-structures of compact protein globules [3] developed by the authors earlier. The theories predict in the unfolded chain the presence of four fluctuating helical regions with the most likely end locations 7–42, 44–73, 80–88 and 91–119 and in the compact globule five fixed helical regions (4–30, 32–40, 65–73, 82–88, and 104–112). According to a theoretical evaluation the degree of helicity of the L7, L12 protein in the unfolded state is 42%, while in the compact globular state it is 53%; the β-structure is absent. Theoretical evaluations of the degree of helicity are in satisfactory agreement with circular dichroism studies of the L7, L12 protein in an aqueous solution [4–6].

2. Helical structure of the unfolded protein chain

The partition function of the unfolded protein chain, of which each of the n residues can have either a fixed helical or an unfixed (non-helical) conformation is expressed as [2]

$$Z = (1.0) \prod_{i=1}^{n-1} \left(\tilde{\sigma}_i^{\frac{1}{2}} \tilde{s}_i \quad \frac{1}{\tilde{s}_i} \right) \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad (1)$$

here $\tilde{s}_i = s_i \exp\left[-(\Delta F_N^i - \Delta F_N^{i+1})/kT\right]$ and $\tilde{\sigma}_i = \sigma' \exp\left[-(\Delta F_N^{i-1} + \Delta F_C^i)/kT\right]$ (s_i is the helix-coil equilibrium constant for the i^{th} residue, σ' is the initiation constant of the helical region and ΔF_N^i and ΔF_C^i are supplementary free energies correspondingly of the N-end of the helical region beginning with the i^{th} residue and the C-end of the helical region terminating with the i^{th} residue). Eq. (1) does not take into account interactions between side groups of amino acid residues (in particular, electrostatic interactions between charged side groups).

From eq (1) by usual methods (see [2, 7, 8]) it is possible to obtain for each i^{th} residue of the given amino acid sequence the probability w_H^i that it is in the helical state as well as probabilities w_N^i and w_C^i that it is correspondingly at the N- or C-end of the helical region.

On the basis of experimental data on helix-coil transitions in water-soluble synthetic polypeptides we ascribed the following $\bar{s}_i$ constant values to amino acid residues: "helix-forming" residues ($\bar{s} > 1$): Glu ($\bar{s} = 1.3$); Ile, Leu, Lys, Met, Phe, Val ($\bar{s} = 1.2$); Ala, Arg, His ($\bar{s} = 1.1$), "helix-breaking" residues ($\bar{s} < 1$): Asn, Asp, Cys, Ser, Thr, Tyr ($\bar{s} = 0.75$); Gly ($\bar{s} = 0.60$). The values of $\bar{s} = 1.0$ were ascribed to Gln and Trp for which experimental data are absent. The values of $\bar{s}$ and the contributions to ΔF_N have been ascribed to Pro ensuring the advantages of its building-in into the first position from the N-end of the helix (or the second and third position after Gly) as well as ensuring the impossibility of its building-in into all the other positions of the helix [9]. The contributions to ΔF_N and ΔF_C of all the other residues were considered to be equal to zero. The σ' value was assumed to be 5×10^{-4}. The basis of this choice of thermodynamic parameters is given in paper [2].

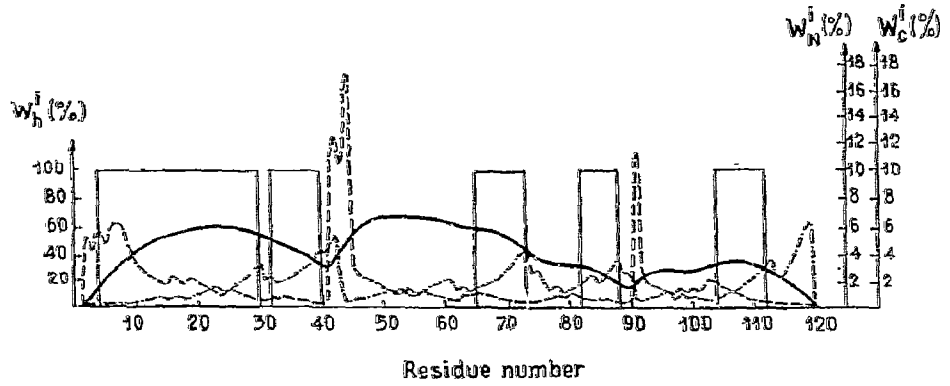

Fig. 1. The helical structure of the L7, L12 protein. The residue number (from the N-end of the chain) is on the abscissa, the probability of the helical state of the residue w_h^i (to the left) and probabilities of the residue being on the N- and C-ends of the helical region w_N^i, w_C^i (to the right) are on the ordinate. (——) w_h^i in the unfolded chain [2]; (——) w_h^i in the compact globule [3]; (– – –) w_N^i in the unfolded chain [2], (–·–·–) w_C^i in the unfolded chain [2].

Calculated values of probabilities w_h^i, w_N^i and w_C^i for the L7, L12 protein are given in fig. 1. It is seen in the figure that in the unfolded chain of this protein the theory predicts four fluctuating helical regions with the most probable locations of the ends (corresponding to the maxima w_N^i and w_C^i) 7–42, 44–73, 80–88 and 91–119. The predicted by the theory average degree of helicity of the unfolded chain of the L7, L12 protein $\langle w_h \rangle$ is 42%. It should be noted that the theory does not take into account other possible types of secondary structure of the protein chain (the β-structure and bends). The probability of any of these types of secondary structure can happen to be greater for some regions of the chain than the probability of the helical state which can somewhat reduce the dimensions of the helical regions predicted by the theory.

3. Secondary structure of a compact protein globule

The stereochemical theory of helical and β-structural compact protein globules developed by one of the authors [3] is based on recognition of amino acid sequence which can build-in into the compact globule with a tightly packed hydrophobic core and a polar shell being in the α-helical or β-structural state. The theory proceeds from the suggestion that such and only such sequences assume regular secondary structure (α or β) in a compact globule which form hydrophobic regions on the surface of the helix or the β-structure capable of tightly building-in into the hydrophobic core without the simultaneous inclusion into it of hydrophilic side groups. For the α-helix this condition is primarily satisfied by the sequences which have massive hydrophobic side groups (Val, Leu, Ile, Phe, Tyr, Trp, Cys) in positions 1–2, 1–4 and 1–5 (as well as 1–2–5 and 1–4–5) if the stereochemistry of these groups allows their tight building-in into the core. The hydrophilic groups must be located on the given fragment of the chain in such a way that they do not build-in into the core but shield it from contacts with water; especially favourable in this respect is the localization of long hydrophilic side groups Glu, Gln, Arg, Lys and His in positions 1–5 on the opposite to hydrophobic groups side of the α-helix turn. For the β-structure this condition is primarily satisfied by the massive hydrophobic side groups in positions 1–3. Besides this, long and short sequences consisting entirely of massive hydrophobic amino acids, also will have, correspondingly, the α- and β-structure.

Application to the L7, L12 protein of the rules of recognition of helical and β-structural sequences in globular proteins obtained in paper [3] leads to the following five helical regions in the compact globule of this protein: 4–30, 32–40, 65–73, 82–88 and 104–112 (see fig. 1). β-Structural sequences in the L7, L12 protein, in accordance with the mentioned rules, are absent. Thus, according to this theory the L7, L12 protein in the compact globular state must have a 53% helicity and not have the β-structure.

4. Comparison with the experiment and discussion

Secondary structure of the L7, L12 proteins was studied by circular dichroism in papers [4–6]; the following evaluations of the degree of helicity were obtained: 60 ± 5% (from CD at 208 nm) [4], ~50% (from CD at 210 and 222 nm) [5] and 45% (from CD at 222 nm) [6]. The presence of ~20% of the β-structure was reported in paper [4] but, according to paper [5], the β-structure is practically absent in the L7, L12 protein as well as in other ribosomal proteins. These results are in satisfactory agreement with our theoretical evaluations.

It is not known at present whether the L7, L12 protein (as well as other ribosomal proteins) has a globular or unfolded structure in aqueous solution. However, according to our theoretical evaluations, the degree of helicity of the protein and localization of helical regions in both cases do not greatly differ from each other. It should be noted in this connection that the L7, L12 protein differs from the overwhelming majority of other proteins in one respect: the theory predicts for it a high degree of helicity not only in the compact globular state (53%) but in the unfolded state as well (42%). It must be mentioned for comparison that even such a high-helical protein as sperm whale myoglobin has in the unfolded state, according to theoretical evaluation, less than 30% helicity and other globular proteins (including haemoglobins) only about 10% [2]. This circumstance suggests the existence of some peculiarities of self-organization or functioning of the L7, L12 protein.

Acknowledgements

The authors express their gratitude to Prof. A.S. Spirin for interest in the work and reading the manuscript. The authors are grateful to O.A. Rakitina and A.G. Raiher for translating the manuscript into English.

References

[1] Terhorst, C., Möller, W., Laursen, R. and Wittmann-Liebold, B. (1972) FEBS Letters 28, 325.
[2] Denesyuk, A.I., Ptitsyn, O.B. and Finkelstein, A.V. Biofizika (USSR) in press.
[3] Lim, V.I. Biofizika (USSR) in press.
[4] Möller, W., Castleman, H. and Terhorst, C.P. (1970) FEBS Letters 8, 192.
[5] Dzionara, M. (1970) FEBS Letters 8, 197.
[6] Brot, N., Boublik, M., Yamasaki, E. and Weissbach, H. (1972) Proc. Natl. Acad. Sci. U.S. 69, 2120.
[7] Lewis, P.N., Gō, N., Gō, M., Kotelchuk, D. and Scheraga, H.A. (1970) Proc. Natl. Acad. Sci. U.S. 65, 810.
[8] Lewis, P.N. and Scheraga, H.A. (1971) Arch. Biochem. Biophys. 144, 576, 584.
[9] Schimmel, P.R. and Flory, P.J. (1968) J. Mol. Biol. 34, 105.