Hypothesis

Structural model for interferons

O B Ptitsyn, A.V. Finkelstein and A.G. Murzin

Institute of Protein Research, Academy of Sciences of the USSR, 142292 Pushchino, Moscow Region, USSR

Received 5 May 1985

Secondary structures of leucocyte α_1- and α_2-interferons and of fibroblast β-interferon are calculated using the molecular theory of protein secondary structures The common secondary structure calculated for α- and β-interferons is used to predict the three-dimensional structures of fragments 1–110 and 111–166 of the chains (which are supposed to be quasi-independent domains) The predicted structure of the active domain I (1–110) is an 'up-and-down' tetrahelical complex (in which the second helix is shorter than the others and can be absent in α_1-interferon) similar to the mirror image of myohaemoerythrin The predicted structure of domain II (111–166) is either a three-stranded β-sheet screened from one side by two α-helices or a three-helical complex (similar to that in the N-domain of papain), the first structure being more consistent with the circular dichroism data of α-interferon and its C-end fragment

Interferon *Protein secondary structure prediction* *Protein tertiary structure prediction*

1. INTRODUCTION

Interferons attract considerable attention as potential anti-viral and anti-tumour agents. The primary structures of a number of interferons have been elucidated [1–6], however, their three-dimensional structures remain completely unknown. This has given rise to numerous attempts to predict the three-dimensional structures of interferons from their amino acid sequences [7–9]. The secondary structures of α-, β- and γ-interferons were predicted earlier by the combination of empirical methods [10–12] with the method proposed by Lim [13] and were used as a basis for the prediction of their three-dimensional structures [7–9,14]. In this paper we calculate the secondary structures of α- and β-interferons using our molecular theory of protein secondary structure [15] and predict the three-dimensional structures of their domains using the algorithm developed recently in our laboratory

2. METHODS

2.1. *Secondary structures*

The secondary structures of α- and β-interferons have been calculated by the computer program [16] based on our molecular theory [15]. This theory takes into account both local interactions inside each chain region (which are approximately taken into account in empirical methods) and long-range interactions between different regions (which are qualitatively taken into account in Lim's method). According to the theory, α-helices and β-strands are formed in the chain regions enriched in non-polar residues, the choice between α- and β-structures is determined by the local interaction inside each region, and the lengths of α-helices or β-strands are determined mainly by the lengths of their continuous non-polar surfaces. The theory has been compared with the X-ray data for 62 globular proteins (A.V. Finkelstein, unpublished) and it has been shown that it fits the experimental data remarkably better than both empirical

Published by Elsevier Science Publishers B V (Biomedical Division)
00145793/85/$3 30

methods and Lim's method For example, for α-helical proteins the theory correctly gives the number of α-helices as 4 in uteroglobin [15], 6 in parvalbumin [15], 7 in bacteriorhodopsin [17] and 8 in erythrocruarin [17] The theory has been used to predict unknown secondary structures of some proteins [15,17,18], the X-ray data obtained afterwards generally confirming these predictions The quality of predictions can be improved by their averaging over proteins with homologous primary structures. For example, though some regions have been predicted incorrectly for separate globins, the averaging of results over several globins correctly predicts the approximate localization of all 8 helices of these proteins without predicting any 'excess' secondary structure (A.V Finkelstein and O B. Ptitsyn, unpublished).

2 2. *Three-dimensional structures*

The prediction of three-dimensional structures of helical complexes (see below) has been made using the algorithm of this prediction developed recently (A.G. Murzin, unpublished) This is an algorithm for choosing the most stable folding pattern of α-helices from the limited number of folding patterns obtained earlier [19] For three-helical complexes it is the choice between only 10 folding patterns (fig.1) This choice is based on the following rules:

(1) α-Helices with continuous non-polar surfaces inclined to the left relative to the helix axis (cf. [20]) must form left-handed complexes and α-helices with the right-handed surfaces must form right-handed complexes

(2) Short irregular connections between adjacent α-helices along the chain cannot cross the surface of the complex and therefore must connect the ends of the helices in the shortest possible way

(3) Different folding patterns give rise to different screening of side groups at the ends of α-helices and in irregular connections (cf. [21])

3. RESULTS AND DISCUSSION

3.1. *Secondary structures*

Fig.2 shows the results of our calculations of the secondary structures of α_1-, α_2- and β-interferons. As α_1- and α_2-interferons have ~80% homology between their primary structures [2,3] and even the homology between α- and β-interferons is as large

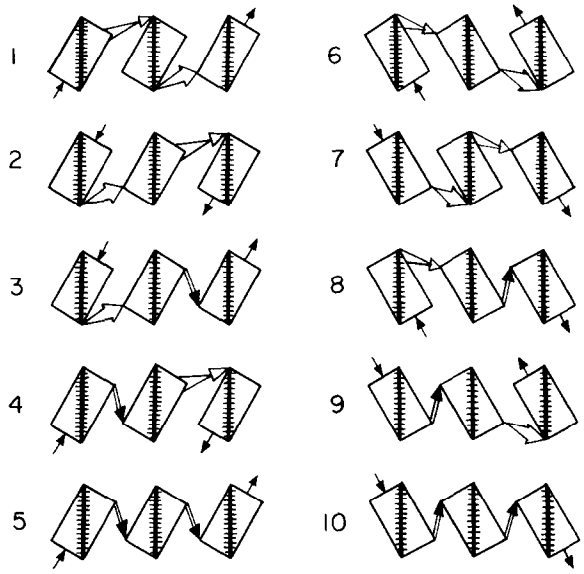

Fig 1 Scheme on plan of all 10 possible structures of a three-helical complex (viewed from the side of the hydrophobic core) Structures 1–5 are left-twisted α-helical bundles, structures 6–10 right-twisted bundles α-Helices are shown by rectangles. Oriented arrows show the connections Bands facing the bundle axis are marked on the helices (in real structures these bands are in contact)

as ~30% [1], one can also expect similarity between the secondary structures of these proteins In fact, fig 2 clearly shows this similarity α-Helices A and C–F (as well as strand β_0) are common for all 3 proteins, and α-helix B is common for α_2- and β-interferons It is also possible that an additional short α-helix exists at the N-end of the chain The situation is less clear for the chain region between α-helices E and F, as the calculations for α- and β-interferons give different results despite high homology between their primary structures (~50%). Therefore, we consider two variants of the predicted secondary structure of this region β-sheet from 3 strands (β_1, β_2 and β_3) and α-helix E'.

Fig 3 shows the common secondary structure of α- and β-interferons which follows from our calculations (including two variants for the region between helices E and F) Secondary structure predictions made by other authors [7,8] are shown for comparison It has been shown recently [22] that limited proteolysis of α_2-interferon gives rise to fragment 1–110 which is resistant to further

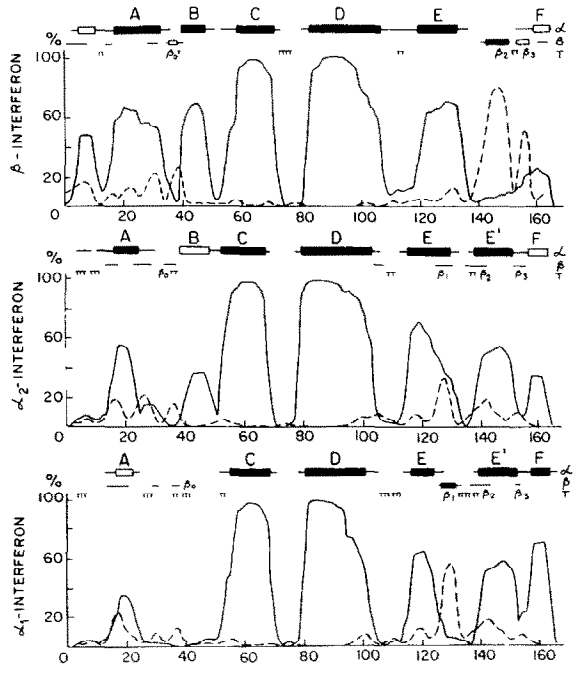

Fig 2. Computer calculations of the secondary structures of α_1-, α_2- and β-interferons Abscissa: residue number in the amino acid sequence, ordinate, probability of α-helical (——) and β-structural (– – –) states Automatic computer predictions of α-helices (α), β-strands (β) and β-turns (T) are shown at the top of each figure. Filled rectangles denote definitely predicted α- ( ) and β- ( ) regions, open rectangles denote predicted α- ( ) and β- ( ) regions, lines indicate possible α- and β-regions (or possible continuations of these regions) TTT denote predicted β-turns.

proteolysis and has an antiviral activity. On the other hand, the synthetic fragment 111–166 of α_1-interferon has a pronounced secondary structure which can be reversibly destroyed in 8 M urea [29]. This suggests that α- (and probably also β-) interferons consist of two quasi-independent domains (1–110 and 111–166). Table 1 compares different predictions of the secondary structures of the whole chain and its C-end fragment (domain II) with the available circular dichroism data for the α_2-interferon [24] and the C-fragment of α_1-interferon [23]. Table 1 shows that the first variant of our prediction fits the experimental data much better than the second. Nevertheless, it would be dangerous at this stage to neglect the second variant, as circular dichroism [25] and Raman spectroscopy [26] data obtained previously had given a larger α-helical content and a smaller β-structure content for α_2-interferon. The predictions made in [7,8] fit the experimental data worse than our predictions.

3.2. *Three-dimensional structures*

The structural frame of domain I most likely consists of 3 α-helices (A, C and D) which are common for all 3 proteins. In this case the application of the prediction algorithm for three-helical complexes gives the following results:

(1) Helix A in α-interferons has a *left-handed* continuous non-polar surface and helices C and D (as well as helix A in β-interferon) have both left- and right-handed surfaces. Therefore the *common* three-dimensional structure of domain I of α- and

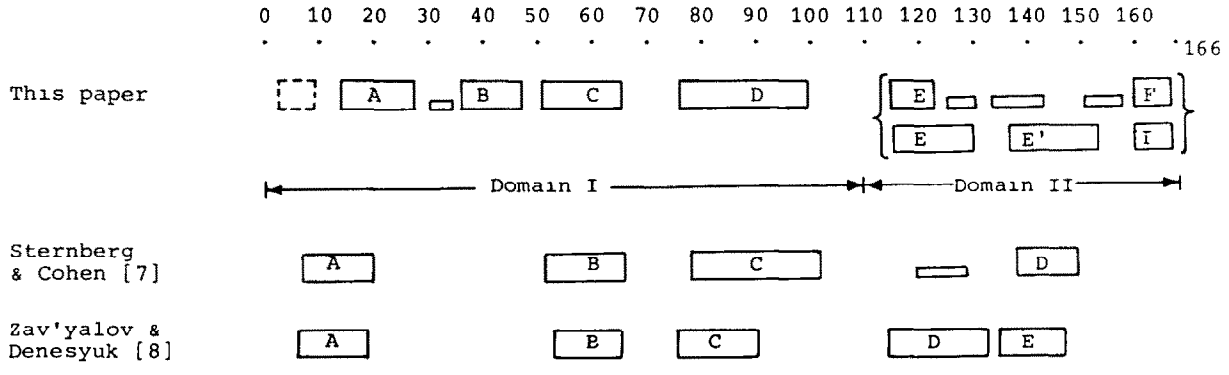

Fig 3 Common secondary structure predicted for α- and β-interferons ( ) α-Helices; ( ) β-strands; ( ) α-helices which can be absent in some interferons The predictions made in [7,8] are shown for comparison

Table 1

Total number of residues in α- and β-structure

	Domain I (1–110)		Domain II (111–166)		In total	
	α	β	α	β	α	β
Circular dichroism [23,24]	–	–	13	20	97	25
This paper variant 1	72	3	16	17	88	20
variant 2	72	3	35	0	107	3
Sternberg and Cohen [7]	54	5	12	9	66	14
Zav'yalov and Denesyuk [8]	43	0	32	0	75	0

α- and β-contents for α-interferon [24] and its C-terminal fragment [23] are recalculated to the total numbers of residues included in α- and β-regions. Predicted numbers of residues are averaged for α_1-, α_2- and β-interferons

β-interferons must be *left-handed* (see structures 1–5 in fig.1).

(2) The irregular connections between helices C and D are short in all 3 proteins. Therefore, they must connect the ends of these helices in the shortest possible way (structures 1, 2 and 4 in fig.1). Only in structure 4 does the connection between helices A and C cross the surface of the complex. Thus helix B (which has been predicted in α_2- and β-interferons) can be incorporated only in this structure without a drastic change of the position of helix A. Therefore only structure 4 can be the common structure for all 3 interferons.

(3) The terminal effects of α-helices also favour structure 4.

The predicted structure of domain I is shown in fig.4. It is similar to the mirror image of the structure of myohaemoerythrin [27]; the main difference is that in myohaemoerythrin helix B is not shorter than the other helices. The disulphide bond between Cys 1 and Cys 99 (in enumeration of α_2) in α-interferons [28] connects in this model the N-end of the chain with the C-end of the last helix D, which are neighbouring in space, and therefore supports the structure of the domain.

The application of the same algorithm to the second variant of the predicted structure of domain II (α-helices E, E′ and F) gives the following results:

(1) Helix F in α-interferons has a *right-handed* continuous non-polar surface and helices E and E′ (as well as helix F in β-interferon) have both left- and right-handed surfaces. Therefore the *common* three-dimensional structure of domain II must be *right-handed* (see structure 6–10 in fig.1).

(2) The irregular connections between helices E,

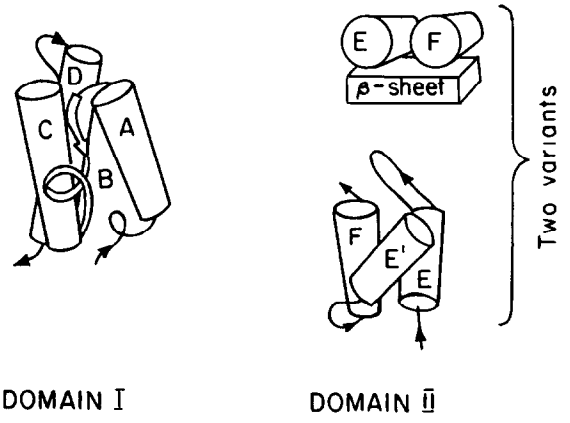

Fig 4 Predicted three-dimensional structures for domains I and II of α- and β-interferons

E′ and F are short in all 3 proteins. Therefore, they must connect the ends of these helices in the shortest possible way (see structures 6 and 7 in fig 1).

(3) The non-polar surface of α-helix F starts at residue 153 (in the enumeration of α_2-interferon). The connection between helices E and F enters this helix from the side of the hydrophobic core in structure 6, while in structure 7 it enters from the side of the surface (see fig.1). Therefore the conservatively non-polar residue 151 can be screened in structure 6 but remains unscreened in structure 7. This makes structure 6 more favourable.

The second variant of the predicted structure for domain II is shown in fig.4 It is similar to the structure of the three-helical complex in the N-domain of papain [29].

As to the first (more probable) variant of domain II (helix E, 3 β-strands and helix F), all the predicted β-strands have only one non-polar surface and therefore can be screened from water only by one side This gives rise to a β-sheet screened from one side by two α-helices (see fig.4), i.e. to a bilayer structure which is typical for small proteins or domains [30,31]. We have not tried to predict the topology of this β-sheet as the strict algorithm for prediction of these topologies is not yet available.

It would be dangerous to speculate about the relative positions of two domains in the overall interferon structure. The disulphide bond Cys 29–Cys 139 (in enumeration of α_2) common for α- and β-interferons [27] connects, in our model, the region between helix A and strand β_0 with the region between strands β_1 and β_2 (or between α-helices E and E′). In the first variant of our model for domain II this bond makes strand β_0 close in space to the β-sheet of domain II suggesting that it enters this β-sheet.

ACKNOWLEDGEMENT

The authors thank Dr A.I. Denesyuk for useful discussion

REFERENCES

[1] Tanigushi, T , Mantei, N , Schwarzstein, M , Nagata, S., Muramatsu, M. and Weissmann, C. (1980) Nature 285, 547–549

[2] Streuli, M., Nagata, S. and Weissmann, C. (1980) Science 209, 1343–1347.

[3] Goeddel, D V., Leung, D.W , Dull, T J , Gross, M., Lawn, R M , McCandliss, R., Seeburg, P H , Ulrich, A , Gelverton, E and Gray, P W (1981) Nature 290, 20–26.

[4] Lawn, R M., Gross, M , Houck, C.M , Franke, A.E., Gray, P W. and Goeddel, D V (1981) Proc Natl. Acad Sci USA 78, 5435–5439.

[5] Denrynck, F R , Content, J , De Clerq, E., Volckaert, G., Tavernier, J., Devis, R. and Fiers, W (1980) Nature 285, 542–547

[6] Gray, P W , Leung, D W., Pennica, D., Gelverton, E , Narajanan, R., Simonsen, C C., Derynck, R S , Rerwood, P J , Wallace, D M , Berger, S L , Levenson, A D. and Goeddel, D.V (1982) Nature 295, 503–509

[7] Sternberg, M J.E and Cohen, F E (1982) Int J. Biol Macromol 4, 137–144

[8] Zav'yalov, V P and Denesyuk, A I. (1982) Immunol Lett. 4, 7–14.

[9] Denesyuk, A I and Zav'yalov, V P (1982) Immunol Lett 5, 225–226

[10] Chou, P Y and Fasman, G D (1978) Adv Enzymol 47, 45–148

[11] Garnier, J., Osguthrope, D J and Robson, B. (1978) J Mol Biol 120, 97–120

[12] Zav'yalov, V P (1979) Biochim Biophys. Acta 310, 70–75

[13] Lim, V I. (1974) J Mol Biol 88, 873–894

[14] Zav'yalov, V P and Denesyuk, A.I (1984) Dokl Akad Nauk SSSR 275, 242–246

[15] Ptitsyn, O B and Finkelstein, A V (1982) Biopolymers 22, 15–25

[16] Finkelstein, A V (1982) Applied Program Package 'ALB', Institute of Protein Research, Pushchino

[17] Ptitsyn, O B (1985) Proc Int Symp on Biomolecular Structure and Interaction, Bangalore, 1984, in press

[18] Ptitsyn, O B , Finkelstein, A V , Kirpichnikov, M P and Skryabin, K G (1982) FEBS Lett 147, 11–15

[19] Murzin, A G and Finkelstein, A V (1983) Biofizika (USSR) 28, 905–911

[20] Efimov, A V (1982) Mol Biol (USSR) 16, 271–281

[21] Lim, V I , Mazanov, A L and Efimov, A V (1978) Mol Biol (USSR) 12, 206–213

[22] Ackerman, S K., Nedden, B Z , Heintzelman, M., Hunkapiller, M and Zoon, K (1984) Proc Natl Acad Sci USA 81, 1045–1047

[23] Arnheiter, H., Thoman, R M , Leist, T , Fountoulakis, M. and Gutte, B (1981) Nature 294, 278–280

[24] Manavalan, P , Johnson, W.C jr and Johnson, P D (1984) FEBS Lett 175, 227–230
[25] Bewley, T.A , Levine, H.L. and Wetzel, R. (1982) Int J Peptide Protein Res 20, 93–96
[26] Williams, R.W (1983) Biophys J. (Abstr) 41, 269
[27] Hendrickson, W.A. and Ward, K.B (1977) J Biol Chem 252, 3012–3018
[28] Wetzel, R (1980) Nature 289, 606–607
[29] Drenth, J , Jansonius, J N , Koekoek, R and Wolthers, B.C (1971) Adv Protein Chem 25, 79–115
[30] Levitt, M and Chothia, C (1976) Nature 261, 552–558
[31] Richardson, J S (1981) Adv Protein Chem 34, 167–339