Volume 233, number 1, 105–108 FEB 05910 June 1988

Amino acid sequence template useful for α-helix-turn-α-helix prediction

Boris V. Shestopalov

Institute of Cytology, Academy of Sciences of the USSR, 194064 Leningrad, USSR

Received 7 March 1988

Necessary stereochemical requirements for an amino acid sequence segment to fold into an α-helix-turn-α-helix supersecondary structure are presented in sequence template form. The usefulness of the template is illustrated by α-helix-turn-α-helix predictions consistent with experimental data from the large T antigens of two polyoma viruses, simian virus 40 (segment 143–165) and mouse polyoma virus (segment 297–319), and the yeast transcription activator GCN4 (segment 256–278).

Amino acid sequence; Protein structure prediction; DNA binding protein; (Simian virus 40, Mouse polyoma virus, *Saccharomyces cerevisiae*)

1. INTRODUCTION

The α-helix-turn-α-helix is a protein supersecondary structure [1,2] used by proteins for sequence specific binding to double-stranded regions of DNA [3,4]. The structure is localized experimentally in the Cro protein [1,2] and the repressor [2] of bacteriophage λ, in the catabolite gene activator (CAP) of *Escherichia coli* [1], in the histone like protein of *Bacillus stearothermophilus* [5], in the repressor of bacteriophage 434 [6,7], in the trp repressor [8] and lac repressor [9] of *E. coli*. The structure is predicted, the prediction being supported experimentally, in the Cro protein of bacteriophage 434 [3,10], in the repressor of bacteriophage P22 [4,10], in the gal [10,11] and deo repressors [11,12] of *E. coli*. Approaches used for these predictions as well as the most recent prediction method [13] include sequence comparisons between proteins containing an α-helix-turn-α-helix structure and those in which this structure is searched. Thus results of such predictions are dependent on sequence similarity between compared proteins. To avoid such dependence an approach based on the α-helix-turn-α-helix stereochemistry alone has been developed and is described below, the embryo of the approach being stereochemical requirements incorporated in the prediction scheme reported in [10].

Correspondence address: B.V. Shestopalov, Institute of Cytology, Academy of Sciences of the USSR, 194064 Leningrad, USSR

The preliminary results have been reported at the Ninth All-Union Symposia 'Structure and Functions of Cell Nucleus', Chernogolovka (USSR), May 25–27, 1987 [Shestopalov, B.V. (1987) in: Abstracts of the Ninth All-Union Symposia 'Structure and Functions of Cell Nucleus' (Zbarsky, I.B. and Kuzmina, S.N. eds) pp.136–137, Chernogolovka]

2. MATERIALS AND METHODS

Stereochemical properties of an α-helix-turn-α-helix structure [2,10,14] and amino acid sequences of this structure from the eleven proteins mentioned in section 1 (see the sequences and/or the sequence references in [1,2,5,6,8–10,12]) are confronted using stereochemical properties of amino acid residues [15]. The α-helix-turn-α-helix stereochemical properties and occurrence of amino acid residues in this structure are presented in fig.1, the residues in the column being divided into three groups as in [15]: the most buried in globular proteins, from Cys to Met (no polar groups in side chain); the most exposed, from Asp to Lys (with charge or amide group); the middle group, from Trp to Pro (with polar groups but without charge or amide group and

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

Pro). A length of the structure (23 residues) is derived from [2]. The numbering is proposed herein.

3. RESULTS AND DISCUSSION

A protein conformation amino acid sequence code is unknown. Thus the complete set of (α-helix-turn-α-helix)-determining sequence requirements cannot be indicated. But it is possible to formulate some provisional rules.

It is well known that Pro is unlikely to be found outside the N-terminal positions of α-helix, cannot be fitted into β-structural regions and very seldomly adopts left-handed α-helical conformation (see recent appropriate statistical data in [15]). These facts are fixed in fig.2. The Pro distribution rules are similar but not identical to those from [10]. Other amino acid residue restrictions for the turn positions (-1), $(+1)$ follow from the fact that an α-helix-turn-α-helix structure may be considered as a particular case of an $\alpha\alpha$-corner with short connections (see these restrictions in [14]). All the above formulated rules are consistent with the α-helix-turn-α-helix amino acid residue distribution presented in fig.1. In the earlier α-helix-turn-α-helix stereochemical sequence rules [10] only Gly has been allowed for position (-1). Really, only Gly occurs in this position (fig.1), but it is also known that the λ repressor mutant with Glu instead of the dominant Gly retains the wild-type activity [16]. In the 'buried' positions (-6), (-2), $(+5)$ [2] residues from the most exposed residue group (see section 2) do not occur (fig.1). Also in position $(+5)$ residues from the middle group do not occur. The restrictions for positions (-6), (-2), $(+5)$ fixed in fig.2 are more strict than similar restrictions in [10]. Finally, in [10] the C^{β}-branched residues are not allowed for the crevice position (-5). Here the residues with any C-branched side chains are forbidden which is consistent with the data of fig.1.

The α-helix-turn-α-helix sequence template (fig.2) has been tested both negatively (the fraction of wrong predictions) and positively (accordance

				α-HELIX		●	▽			●	αL ✕	β	β ○				α-HELIX ●						
	-11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11+
Cys												1											
Ile	2	1	1			1			1					1			6			2	2		
Val	1					4				3		6		1	1		4			2	2		
Leu	3	1	2			4			2	3		1							1	2	2		
Gly		1					2				11		2			1		2		1			3
Phe	1																				1		
Ala				2	2	1	8	1		3		1			2	2	1		1	2		1	2
Met	2								1	1		1			1								1
Trp																				1			
Ser		2			1					1			6		2	2		3			1		2
Thr	1	3		2		1		1	1			1	1			5		2					
His		1																				1	
Tyr				1					1				1	1									
Pro																							
Asp					3			1	1							1				1		1	
Asn		1		1				1						1				1				6	3
Gln			6	1				2						4	3			1	2				
Glu	1			1	4			2	1					1	1			1			3		
Arg		1	1	1				1						1				1	5			1	
Lys			1	2	1		1	2	3				1	1	1				2			1	
	-11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11+

Fig.1. Stereochemical properties and occurrence of amino acid residues for an α-helix-turn-α-helix supersecondary structure. Positions are marked as follows: (●) completely buried; (▽) in surface crevice; (×,○) special stereochemical requirements. See text.

-11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11+
				Pro	Pro	Pro	Pro	Pro	Pro	Pro	Pro	Pro				Pro	Pro	Pro	Pro	Pro	Pro	Pro
					Asp	Ile			Asp	Ile		Ile				Trp						
					Asn	Val			Asn	Val		Val				Ser						
					Gln	Leu			Gln	Leu		Leu				Thr						
					Glu	Phe			Glu	Phe		Phe				His						
					Arg	Trp			Arg	Trp		Trp				Tyr						
					Lys	Thr			Lys	Thr						Asp						
						His				Tyr						Asn						
						Tyr										Gln						
						Asp										Glu						
						Asn										Arg						
						Gln										Lys						
						Glu																

Fig.2. Amino acid sequence template for an α-helix-turn-α-helix supersecondary structure (distribution of forbidden residues).

of predicted localizations with experimental data for the simian virus 40 and mouse polyoma virus large T antigens and the yeast *Saccharomyces cerevisiae* transcription activator GCN4, all these proteins being selected because other approaches had failed, for example [10,13]). Only 2% of all the non-(α-helix-turn-α-helix) structure 23 residue long segments of the eleven proteins used for the sequence template elaboration satisfies the template and only 0.2% of such segments satisfies the modified template with only Gly allowed in position (−1) as in [10]. In the mouse polyoma virus large T antigen the region 290–310 is necessary for DNA-binding activity [17] and the superimposed segment 297–319 satisfies the template. The corresponding segment 143–165 of the simian virus 40 large T antigen, both T antigens being similar in amino acid sequence [18], also satisfies the template and at the same time is included in the DNA-binding domain 83–214 [19]. Furthermore, the mutant of the simian virus 40 T antigen with Thr-153 instead of Asn in the wild-type protein has no DNA-binding activity [20] which is consistent with the template prohibition of Thr in this position for the α-helix-turn-α-helix localization in the 143–165 segment (Thr-153 is in position (−1) of the template). The DNA-binding activity of GCN4 is localized in the C-terminal domain 222–281 [21], this activity not being connected with a finger-like DNA-binding structure and an attempt to localize an α-helix-turn-α-helix structure having failed because of the absence of any sequence similarity between this region and known α-helix-turn-α-helix structures. Having used the sequence template it has appeared that the segment 256–278 may have folded into an α-helix-turn-α-helix structure.

REFERENCES

[1] Steitz, T.A., Ohlendorf, D.H., McKay, D.B., Anderson, W.F. and Matthews, B.W. (1982) Proc. Natl. Acad. Sci. USA 79, 3097–3100.

[2] Ohlendorf, D.H., Anderson, W.F., Lewis, M., Pabo, C.O. and Matthews, B.W. (1983) J. Mol. Biol. 169, 757–769.

[3] Wharton, R.P., Brown, E.L. and Ptashne, M. (1984) Cell 38, 361–369.

[4] Wharton, R.P. and Ptashne, M. (1985) Nature 316, 601–605.

[5] Tanaka, I., Appelt, K., Dijk, J., White, S.W. and Wilson, K.S. (1984) Nature 310, 376–381.

[6] Anderson, J.E., Ptashne, M. and Harrison, S.C. (1985) Nature 316, 596–601.

[7] Anderson, J.E., Ptashne, M. and Harrison, S.C. (1987) Nature 326, 846–852.

[8] Schewitz, R.W., Otwinowski, Z., Joachimiak, A., Lawson, C.L. and Sigler, P.B. (1985) Nature 317, 782–786.

[9] Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R. and Van Gunsteren, W.F. (1985) J. Mol. Biol. 182, 179–182.

[10] Ohlendorf, D.H., Anderson, W.F. and Matthews, B.W. (1983) J. Mol. Evol. 19, 109–114.

[11] Lehming, N., Sartorius, J., Niemöller, M., Genenger, G., Von Wilcken-Bergmann, B. and Müller-Hill, B. (1987) EMBO J. 6, 3145–3154.

[12] Valentin-Hansen, P., Højrup, P. and Short, S. (1985) Nucleic Acids Res. 13, 5927–5938.

[13] Dodd, I.B. and Egan, J.B. (1987) J. Mol. Biol. 194, 557–564.
[14] Efimov, A.V. (1984) FEBS Lett. 166, 33–38.
[15] Shestopalov, B.V. (1988) Mol. Biol. (USSR), in press.
[16] Hochschild, A., Irwin, N. and Ptashne, M. (1983) Cell 32, 319–325.
[17] Cowie, A., De Villiers, J. and Kamen, R. (1986) Mol. Cell. Biol. 6, 4344–4352.
[18] Griffin, B.E., Soeda, E., Barrell, B.G. and Staden, R. (1981) in: DNA Tumor Viruses (Tooze, J. ed.) pp.869–873, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[19] Campbell, J.L. (1986) Annu. Rev. Biochem. 55, 733–771.
[20] Gluzman, Y. and Ahrens, B. (1982) Virology 123, 78–92.
[21] Hope, I.A. and Struhl, K. (1986) Cell 46, 885–894.