FUNDAMENTAL PERIODICITIES IN THE AMINO ACID SEQUENCE OF THE COLLAGEN α1 CHAIN

David J.S. Hulmes[1], Andrew Miller[1], David A.D. Parry[2] and John Woodhead-Galloway[3]

1 European Molecular Biology Laboratory, C.E.N.G., 85X, 38041 Grenoble Cedex, France

2 Dept. Chemistry, Biochemistry + Biophysics, Massey University, Palmerston North, New Zealand

3 Medical Research Council, 20 Park Crescent, London W1N 4AL

Received February 18,1977

Summary

The distributions of certain types of residues along the α1 chain of rat/calf skin collagen are considered in terms of fundamental periodicities of D/5 and D/6, where D is the axial repeat of the collagen fibril. By Fourier analysis, charged residues are D/6 periodic, proline residues are D/5 periodic, whilst both these periodicities are necessary to describe the hydrophobic residue distribution. The relative positions of these periodicities are determined from the phases of the Fourier components.

A. INTRODUCTION

The axial D (~670 Å) periodic structure in collagen fibrils is due to displacements of integral multiples of D between the ~2990 Å long molecules (1,2). For these intermolecular staggers, the α1 amino acid sequence of rat/calf skin collagen (3) is such that possible charged and hydrophobic interactions are maximised (4,5,6). The specificity for axial self assembly is therefore contained in the amino acid sequence.

Both charge and hydrophobic interactions can be regarded as the alignment of complementary side-chains, i.e. residues with opposite charges or both with large apolar groups. If there is an interaction maximum when two molecules are staggered by, for example, 1D, this will lead to a D periodically repeating pattern of complementary residues along the molecule (7,8). Any random

ISSN 0006-291X

distribution of interacting residues in a single D segment will conform to this principle. A regular pattern within D is not necessary; indeed, if there is a pattern which divides D into n complementary segments, a D/n stagger would be specified.

On the above basis, it is curious that there should be such a highly regular distribution of large apolar (hydrophobic) residues in the collagen α1 sequence, such that residue positions conform to the D periodic pattern $|(2D/11)_4, (D/11), (2D/11)|_n$ (4). Evidence has also been presented for an approximate D/6 periodicity in the distribution of charged residues (9). In this paper we will show that these and other regular patterns of residues in the sequence can be thought of in terms of fundamental periodicities of D/5 and D/6 which, in combination, give rise to the D specificity.

B. FOURIER ANALYSIS OF THE SEQUENCE

1. General analysis

Any distribution of residues or residue types can be Fourier analysed as a summation of cosine waves each with a given wavelength, amplitude and phase. If certain periodicities dominate the residue distribution, these will be revealed as waves with large amplitudes whose wavelengths are integral submultiples of the periodicity. This kind of analysis has already clearly demonstrated the presence of both 7 residue and 19.7 residue periodicities in the tropomyosin sequence (10, 11). It is particularly appropriate for fibrous proteins, where the residue translation along the molecular axis is constant.

In the following analysis the positions of certain residues or residue types in the triple-helical region of the α1 sequence were "fast" Fourier transformed (FFT) using a Cooley-Turkey algorithm. An array size of 8192 was used to permit fine sampling of the Fourier transform and hence determine the wavelengths of strong components with some precision. The peaks in the FFT which do and do not correspond to integral submultiples of D (232 to 235 residues)(4,5,6,12,13) are called "D-maxima" and "noise-maxima" respectively. The results are summarised in Table 1. The wavelengths listed are those for which the D-maxima are greater than the noise maxima, D/21 being the shortest wavelength

Table 1. Strong D-maxima in the Fourier Transform of certain residue positions in the collagen α1 sequence

	Large apolar			Charged			Imino		
Wavelength D/n	MET PHE VAL LEU ILE	LEU	PHE	LYS HYL ARG ASP GLU	ASP GLU	LYS	PRO HYP	PRO	HYP
1									
2									
3			X						
4									
5	X	X						X	
6	X	X		X	X	X			
7									
8									
9									
10							X		
11	X								
12								X	
13									
14								X	
15									
16									
17		X							
18									
19									
20		X							
21							X		X

considered. Within the three general categories of large apolar, charged and imino acid side-chains, there are no D-maxima greater than noise maxima for the following residue groupings : MET, VAL, ILE, LYS+ARG, ASP, GLU, ARG. Furthermore, there are no overall D-maxima in the FFT for the distribution of either ALA or hydrophilic (SER, THR, ASN, GLN) residues.

It is clear from Table 1 that the two major Fourier

Table 2. Phases for the principal Fourier components. Phase origin at the N-terminus of the triplet region.

	Prolines	Large Apolar (PHE+MET+VAL+LEU+ILE)
D/5	92.2	-85.2

	Large Apolar (PHE+MET+VAL+LEU+ILE)	Charged (ASP+GLU+ ARG+LYS+HYL)	Acidic (ASP+GLU)	Basic (ARG+LYS+ HYL)
D/6	20.9	-146.4	-146.1	-146.7

components correspond to wavelengths of D/5 and D/6. The large apolar (hydrophobic) residues have both strong D/5 and D/6 components whilst the charged residues have only a strong D/6 component and the prolines have only a strong D/5 component. It is interesting to compare the phases of these waves, as these give the relative positions of the periodicities along the sequence, see Table 2. The two D/5 waves, hydrophobic residues and prolines, are approximately 180° out of phase and hence mutually displaced by D/10. Similarly the two D/6 waves, hydrophobic residues and charged residues, are also approximately 180° out of phase and hence mutually displaced by D/12.

2. Hydrophobic Residues

As noted above, the principal Fourier components of the hydrophobic residue distribution correspond to wavelengths of D/5 and D/6. Thus if two cosine waves with these wavelengths are added with the appropriate phase (Table 2), the main features of the known residue distribution should be apparent. Fig. 1 shows this simple Fourier synthesis; clearly the positions of the peaks correspond to the $|(2D/11)_4,(D/11),(2D/11)|_n$ pattern. The component waves in Fig. 1 may be regarded as representing the probability of finding hydrophobic residues along the sequence, whence the total probability is given by the summation. The regular D periodic distribution of hydrophobic residues can therefore be thought of as the sum of D/5 + D/6 "probability waves"; there is no need to invoke the more complex

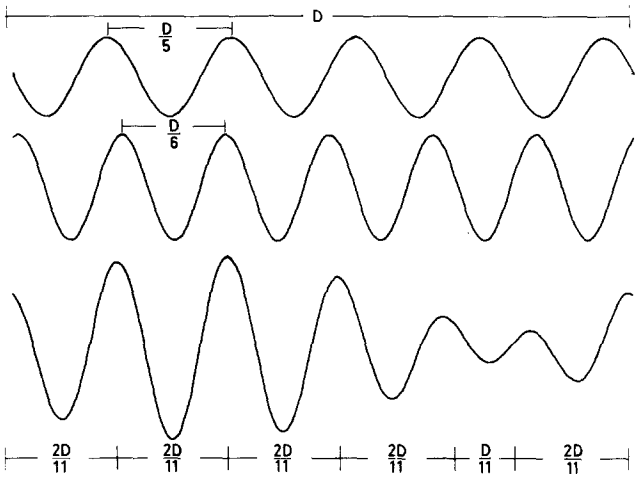

Fig. 1 Synthesis of the distribution of hydrophobic residues from D/5 + D/6 Fourier components.

$|(2D/11)_4,(D/11),(2D/11)|_n$ formula. Though neither a D/5 nor a D/6 periodicity can separately specify favourable intermolecular interactions at multiples of D, the D specificity is a property of their combination.

3. Charged residues

The strong D/6 periodicity in the distribution of charged residue can be separated into contributions from the acidic (ASP + GLU) and basic groups (ARG + LYS + HYL). Table 2 shows that their phases are almost identical; the D/6 charged "probability waves" are therefore in register. This is consistent with the known tendency of oppositely charged residues to occur close together in the sequence (6,14).

The phase of this charged residue D/6 wave is such that the peaks correspond to the positions of the six main positively stained bands 'a','b_1','b_2','c_2','d' and 'e' observed in electron micrographs of collagen fibrils (9).

C. DISCUSSION

McLachlan (15) has recently shown the presence of both D/5 and D/6 periodicities in the autocorrelation function of

similar residue types occupying the X positions in the α1 (GLY-X-Y) triplet region. These periodicities have been also noticed in a previous Fourier transform analysis of the sequence (8), where the segregation into X and Y positions was not considered. This paper has extended the latter observations to demonstrate from the calculated phases, the relative positions of the D/5 hydrophobic and proline periodicities and of the D/6 hydrophobic and charged periodicities. Furthermore the analysis has shown that the $|(2D/11)_4,(D/11),(2D/11)|_n$ pattern in the distribution of hydrophobic residues in the α1 sequence (4) can be regarded simply as the combined effect of the hydrophobic D/5 and D/6 periodicities.

As well as specifying the axial D stagger between collagen molecules, it may be that the D/5 and D/6 periodicities are important in the lateral interactions. If, for example, the pitch of the molecular triple helix, P, is such that D/5=3P/2 and D/6=5P/4 (i.e. D/P=7.5) then two and four fold rotational symmetries, respectively, would arise in the azimuthal edge distribution. (Since D is 232 to 235 residues and P is 27 to 40 residues (16) the experimental limits of D/P are 5.8 and 8.7, thus many other integral relationships and consequent edge symmetries are possible). If P deviates slightly from such ideal relationships with D the edges would, of course, be helical rather than parallel to the molecular axis. Hence, any molecular packing scheme which requires rotational symmetries in the azimuthal edge distribution, coupled with the necessary requirement for axial D specificity, could provide an evolutionarily selective pressure for the presence of D/5 and D/6 periodicities in the sequence.

Fourier analysis of the first 393 residues in the α2 chain of calf skin collagen (17) does not show strong D/5 or D/6 periodicities in the distributions of either hydrophobic or charged residues. Further studies are clearly necessary to investigate the role of the α2 chain in intermolecular interactions.

Acknowledgements

We are grateful to Dr. A.D. McLachlan for sending a preprint describing his own analysis of the α1 sequence. We thank Mr. S.W. White for discussion. D.J.S.H. is an EMBL postdoctoral fellow. Some of this work was carried out in the Laboratory of Molecular Biophysics, Oxford.

References

1. Hodge, A.J. & Petruska, J.A. (1963). In 'Aspects of Protein Structure' (ed. G.N. Ramachandran) pp. 289-300, Academic Press, N.Y.
2. Doyle, B.B., Hulmes, D.J.S., Miller, A., Parry, D.A.D., Piez, K.A. & Woodhead-Galloway, J. (1974). Proc. Roy. Soc. B186, 67-74.
3. Fietzek, P.P. & Kuhn, K. (1975). Mol. Cell. Biochem. 8, 141-157.
4. Hulmes, D.J.S., Miller, A., Parry, D.A.D., Piez, K.A. & Woodhead-Galloway, J. (1973). J. Mol. Biol. 79, 137-148.
5. Walton, A.G. (1973). Proc. 5th Int. Symp. Biomaterials, Clemson University.
6. Doyle, B.B., Hukins, D.W.L., Hulmes, D.J.S., Miller, A., Rattew, C.J. & Woodhead-Galloway, J. (1974). Biochem. Biophys. Res. Comm. 60, 858-864.
7. Segrest, J.P. & Cunningham, L.W. (1973). Biopolymer 12, 825-834.
8. Hulmes, D.J.S. (1975). D.Phil.Thesis, Univ. of Oxford
9. Doyle, B.B., Hulmes, D.J.S., Miller, A., Parry, D.A.D., Piez, K.A. & Woodhead-Galloway, J. (1974). Proc. Roy. Soc. B187, 37-46.
10. Stewart, M. & McLachlan, A.D. (1975). Nature 257, 331-333.
11. Parry, D.A.D. (1975). J. Mol. Biol. 98, 519-535.
12. Chapman, J.A. & Hardcastle, R.A. (1974). Conn.Tiss.Res. 2, 151-159.
13. Hulmes, D.J.S., Miller, A., White, S.W. & Brodsky Doyle, B. (1977) J. Mol. Biol., 110, 643-666
14. Doyle, B.B., Hukins, D.W.L., Hulmes, D.J.S., Miller, A. & Woodhead-Galloway, J. (1975). J. Mol. Biol. 91, 79-99.
15. McLachlan, A.D. (1976). J. Mol. Biol. 107, 159-174.
16. Miller, A. (1976) Biochemistry of Collagen (ed. G.N. Ramachandran & A.H. Reddi) pp. 85-136, Plenum, New York.
17. Fietzek, P.P. & Rexrodt, F.W. (1975) Eur. J. Biochem. 59, 113-118.