The interaction between phenylalanine rings in proteins

J. Singh and J.M. Thornton*

Laboratory of Molecular Biology, Crystallography Department, Birkbeck College, Malet Street, London WC1E 7HX, England

Received 16 August 1985

An analysis has been made of the geometry of phenylalanine-phenylalanine interactions in proteins of known structure. 162 Phe-Phe interactions were found with C-C distances less than 4.6 Å. Three angles were used to define the geometry of interaction, P = the angle betwen ring planes, and polar coordinates, $T\theta$, $T\varphi$ to specify the relative spatial disposition of the two rings. The overall distribution of P values is the same as that expected for a random distribution of planes in 3 dimensions; i.e. the majority of interactions have P approaching 90°. However, for high $T\theta$ values (when one Phe lies directly above the ring of the other Phe) the distribution is non-random, and a preference for perpendicular interactions is expressed. This preference is in accord with recent quantum-mechanical calculations.

Aromatic-aromatic interaction *Phenylalanine-phenylalanine interaction* *Protein structure* *Protein stability* *Drug design*

1. INTRODUCTION

In globular proteins close interactions between the aromatic amino acids phenylalanine, tyrosine and tryptophan are common [1,2]. Furthermore these residues often from part of a hydrophobic pocket designed to bind an aromatic substrate [3]. The geometry and energetics of such interactions will be crucial for protein structure, specificity and activity.

The classical herringbone pattern observed for benzene in the crystal structure [4] involves perpendicular, cogwheel and parallel interactions between the benzene rings (see table 1). In contrast in B-DNA the purine and pyrimidine bases are stacked with their planes parallel.

Warme and Morgan [1,2] have shown that of all the aromatic interactions Phe-Phe are the most common and, unlike interactions between the different aromatic amino acids (eg. Phe-Tyr), occur more often than expected by chance. Therefore to begin a study of the preferred geometry of aromatic interactions in a protein environment, we have extracted from the Protein Data Bank [5] coordinates for all the interacting phenylalanine residues and analysed their geometries.

2. METHODS

Two sets of coordinate data were used. The first comprised 29 proteins solved to a resolution of 2 Å or better; the second included the less well refined structures (≧3 Å resolution) giving 62 non-homologous proteins. The results obtained from the two sets were essentially identical.

To define an interaction between 2 phenylalanines, the closest carbon-carbon approach distance was calculated for every pair of phenylalanines in the data set, and the observed distribution of d values is shown in fig.1. From this plot an operational definition of an interacting Phe-Phe pair was chosen as a cut-off of $d \leqq 4.6$ Å. The van der Waals contact distance between 2 aromatic -CH groups is approx. 3.6 Å and allowing 1 Å coordinate error, d = 4.6 Å is the maximal value for contacting Phe rings.

* To whom correspondence should be addressed

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

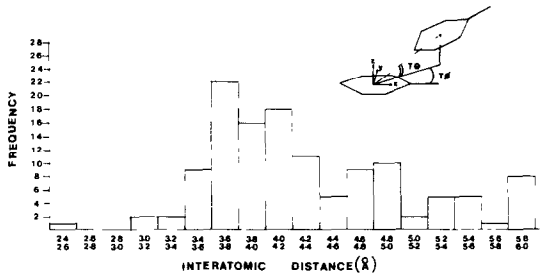


Fig.1. Distribution of closest interatomic (C-C) distances (d) between phenylalanines in 29 proteins. The inset shows the coordinate axes for the reference phenylalanine, and the definition of the angle of elevation $T\theta$, and equatorial angle $T\phi$ (see text).

Given a pair of interacting Phe rings F1 and F2, their geometry was calculated as follows:

(1) Calculate 2 parameters, which are independent of the coordinate system:
D = distance between centres of rings (Å)
P = angle between the ring planes

(2) Define a 'reference' phenylalanine (PHE), placed with the centre of the ring at the origin 0, with the x-axis along 0-Cγ; the z-axis perpendicular to the ring-plane and y in the plane of the ring, orthogonal to x and z (see fig.1).

(3) Superpose F1 on this reference PHE, apply the same matrix to F2 and calculate the polar coordinates $T\theta$, Tϕ of the centroid of the second ring 02, where: $T\theta_1$ = angle of elevation of 02 from the plane of F1 (azimuthal angle)
$T\phi_1$ = equatorial angle of 02 in the plane of F1

(4) Repeat step 3, using F2 to define the coordinate system and calculate $T\theta_2$, $T\phi_2$, the polar coordinates of the centroid of F1 relative to F2. Note: these will be different from $T\theta_1$, $T\phi_1$

3. RESULTS

The interatomic distribution shown in fig.1 has a pronounced peak between 3.6 and 4.2 Å, illustrating that phenylalanine rings close pack in the interior of proteins. We found 162 interacting pairs (84 at high resolution) with $d < 4.6$ Å, and the interaction geometries for these pairs are given in figs.2–4. The angle between the planes (fig.2a) shows a striking distribution, with very few planar stacked structures ($P = 0°$). and many pairs in which the planes are approximately perpendicular

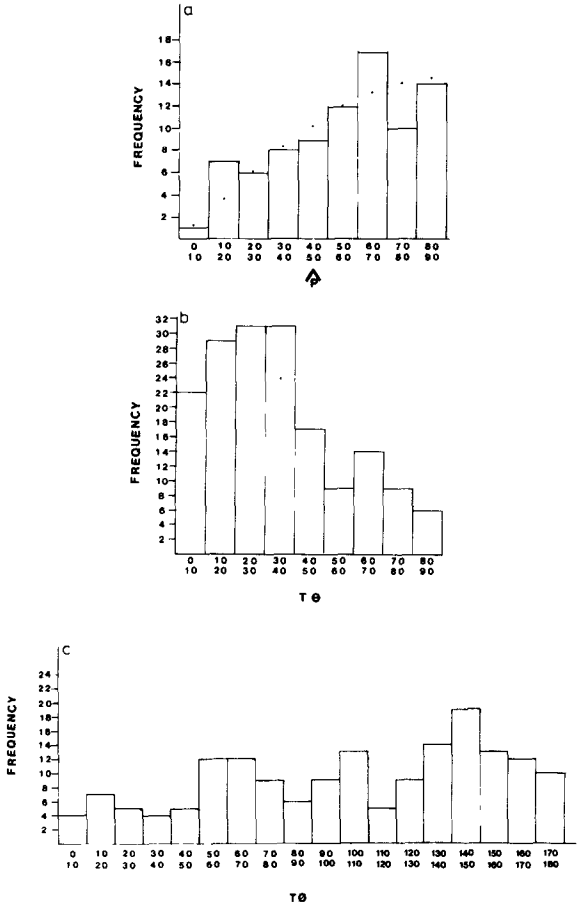


Fig.2. Distributions calculated for 84 interacting phenylalanine pairs ($d \leq 4.6$ Å) from 29 proteins, with resolution ≤ 2 Å. (a) Distribution of the angle between aromatic ring planes. Each plane is defined as the least-squares plane of the 6 ring carbon atoms. The distribution expected for a random orientation of 2 planes [frequency α $\delta(\cos P)$] is shown as dots. (b) $T\theta$ distribution – the angle of elevation of the centre of one ring relative to the plane of the other ring. Each pair contributes 2 points $T\theta_1$, $T\theta_2$. The frequencies expected for a random distribution of $T\theta$, as calculated from volume considerations [frequency α $\delta(\sin \theta)$] are shown as dots. (c) $T\phi$ distribution – the equatorial angle measured by taking the projection of 02 onto the x-y plane of the reference Phe and calculating the angle $02_{x,y}$-01-Cγ_1. Each pair contributes two points $T\phi_1$, $T\phi_2$. $T\phi$ has no normalisation factor.

(P = 90°). However, the expected distribution for a random orientation of 2 planes (see fig.2) is very similar (probability > 0.5 from a χ^2 calculation) suggesting that in a protein core there is no inherent preference for any specific P value, but by chance the majority will interact perpendicularly, rather than in a parallel stacked fashion. Similarly, the $T\theta$ distribution (the azimuthal angle) broadly follows the random distribution as calculated from volume considerations, which decreases monotonically from 0 to 90° (fig.2b). The slight dip at $T\theta$ = 0° may be explained by the occluded volume of the hydrogen atoms, which lie in the plane of the ring. The $T\phi$ plot (fig.2c), showing the radial distribution, indicates that few phenylalanines have $T\phi$ < 50°, which would be expected because of the Cβ and backbone atoms. The majority of interactions occur at the distal end of the Phe with a pronounced peak between 140 and 150° where a 2-hydrogen interaction can be made.

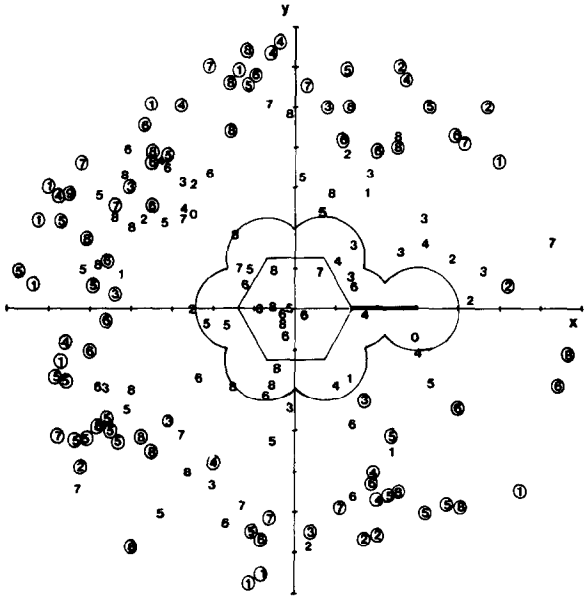


Fig.3. Spatial distribution of P angles for 84 Phe-Phe interactions. The P value is plotted at the x,y coordinate of 02, the centre of the second Phe, after superposition of the first ring on the reference PHE. The single numerals represent the following angular ranges 0 = 0–9°; 1 = 10–19°; 2 = 20–29°, etc. The reference PHE, and its van der Waals surface is shown. Each pair of phenylalanines contribute 2 points derived from the superposition of each Phe on the reference PHE. The encircled numbers represent phenylalanines for which the z coordinate of the centroid 02 ≦ 3 Å.

However, although the P and T distributions individually appear essentially random, fig.3 shows that the angle between the planes is influenced by the relative spatial displacement of the Phe rings. In this figure the angle P between planes is plotted at the x, y coordinate of the centre of the second Phe, once the first Phe has been superposed on the reference PHE (shown in the diagram). In the volume directly 'above' the reference PHE ($|x|$; $|y| \leqq 1$ Å) all the interactions have $P \geqq 50°$. There are no fully overlapped parallel stacked structures, instead all the rings are almost perpendicular. Proceeding radially outwards from the centre, up to 4 Å in the x-y plane, there is an annulus, which is dominated by P values ≦ 60°, with several examples in which the planes of the 2 rings are essentially parallel. These structures correspond to staggered stacked phenylalanines, with partial overlap of the rings. At distances more than 5 Å from the centre (in the x-y plane) the whole range of P values is observed (see table 1). There are perpencidular interactions (P ~ 90°), tilted planes (P ~ 60°) and also several examples of parallel in-plane contacts (P ~ 0°). Fig.4 shows the distribution of P values for different angles of elevation $T\theta$. Most striking are the number of parallel in-plane interactions for low $T\theta$ and the absence of fully stacked structures at $T\theta$ > 67.5°.

The most significant deviation from the 'random' distribution occurs for $T\theta$>67.5°. There are remarkably fewer parallel structures (P <40°) than expected and more perpendicular interactions. The probability of this distribution (about P = 40°) occurring by chance is less than 0.005, as calculated by a χ^2 analysis. A classification of the different types of interactions and their frequency of occurrence are given in table 1. The interactions found in benzene crystals are also described.

4. DISCUSSION

These results show that in a protein environment the geometry of Phe-Phe interaction is very varied. The relative spatial displacement of the 2 rings is critical in determining the preferred angle between the planes. Only when one Phe lies above the ring of the other Phe, is the preference for perpendicular interaction as found from quantum-

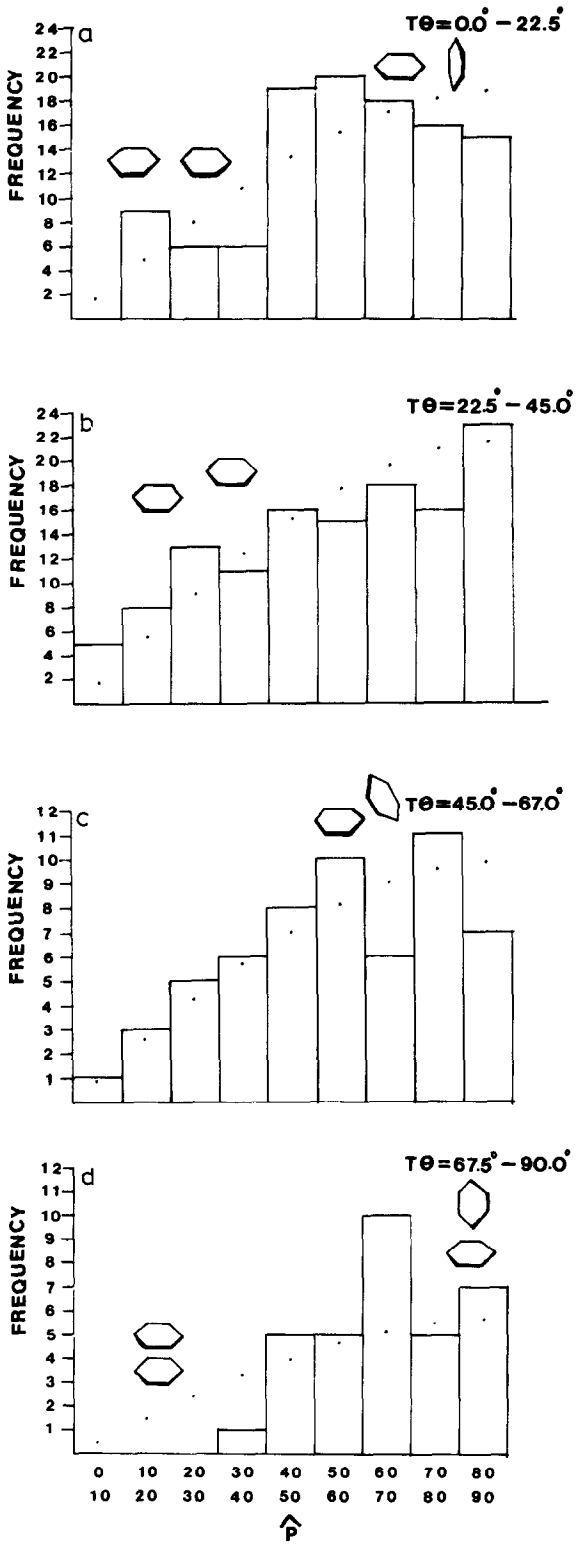


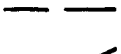
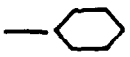
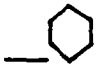
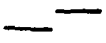
Fig.4. Observed distribution of the angle between ring planes, P, for different spatial positions of the 2 rings, as measured by $T\theta$ ranges. Data for 162 Phe-Phe pairs. The expected values for a random distribution of 2 planes are shown by dots. (a) $T\theta$ = 0–22.5°. Note the peak for parallel in-plane interactions (P = 10–20°) and the peak for tilted planes (P = 50–60°). (b) $T\theta$ = 22.5–45°. (c) $T\theta$ = 45–67.5°. (d) $T\theta$ = 67.5–90°. There are no stacked structures with full overlap of the rings. In this range all the interactions have $P > 30°$, and the majority approach perpendicular. Tilted planes ($P \sim 60°$) are unusually common.

mechanical calculations [7,8] expressed. Elsewhere the P distribution is essentially a random distribution of planes in 3-dimensional space, which will be dominated by perpendicular interactions.

The fully overlapped parallel stacked structure has not been found once in 162 Phe-Phe interactions, although stacking involving partial overlap of the 2 rings is found (12 out of 162). Such partial stacking occurs frequently in aromatic crystal structures [6] and indeed between the bases in B-DNA. Quantum-mechanical calculations show that the fully stacked parallel planes structure is less stable than the perpendicular interaction by 1–2 kcal/mol [8,9]. The in-plane parallel structure (P = 0°, $T\theta$ = 0°) is found to be even less stable, but this packing is found for phenylalanines in proteins (12 out of 162) where interactions with other residues can occur. Both 'edge-to-ring face' and cogwheel type of perpendicular interactions occur in proteins (17/162 and 36/162, respectively) as are found in the benzene crystal structure (see table 1). However interactions where the ring planes are tilted at ~ 60° are most common (74 out of 162), but the energies for this type of structure have not been evaluated. The major contribution to the energy differentials is apparently electrostatic involving interactions between the positively charged hydrogens and the cloud of π-electrons [7,8].

Burley and Petsko [9] have recently suggested that aromatic-aromatic interactions make a special contribution to protein structure stabilisation. Whilst these interactions will undoubtedly contribute to stability, as do the interactions between almost all the atoms in the protein's hydrophobic core, our results do not support the idea of a separate class of interaction. Except for a small

Table 1
Classification of the Phe-Phe interactions and observed frequencies in 62 proteins (162 Phe-Phe pairs)

Type	Structure[c]	Angle between planes P (°)	Angle of elevation(°)		No. observed[b]
			$T\theta_1$	$T\theta_2$	
Fully stacked		<30	>60		0
Staggered stacking		<30	30–60		12
Parallel in-plane		<30	<30		12
Tilted		30–70			74
Edge-ring-face		>70	>60	>30	17
Cogwheel	a  b 	>70	0–60		37
Benzene[a]					
I–IV		83	3	77	
I–II		89	14	37	
I–III		25	12	35	

[a] 3 different ring-ring interactions occur in the benzene crystal specified as in [4]
[b] 10 of the 162 Phe-Phe interactions involve the Cβ atom of one Phe
[c] Lines represent ring-planes edge-on

sub-group, with a special disposition of the phenylalanines, the interaction geometry is essentially random. In proteins the preferred 'edge-to-face' aromatic-aromatic interaction [7,8] is found, but so are many other less favourable orientations. Interactions with other side chains can interfere with and obviously overcome the preference for a perpendicular interaction between the aromatic rings.

However, the majority of interacting phenylalanines in proteins do lie with their planes perpendicular (for reasons of 3-D space, rather than energy considerations) and the parallel stacked structure (with full overlap rings) is not observed and must be energetically unfavourable. These observations are important for drug design and for the prediction of possible binding conformations for an aromatic ligand. We are currently developing a data base of such side chain interactions, to be used in drug design.

ACKNOWLEDGEMENTS

J.S. is supported by a Case studentship, in collaboration with Dr Simon Campbell at Pfizer, Sandwich, England. We thank Professor T.L. Blundell and G. Kaur for their interest and encouragement, D. Barlow, A. Mackay and Willie Taylor for useful discussion and help with computing, and Freddie Elliot for excellent photography.

REFERENCES

[1] Warme, P.K. and Morgan, R.S. (1978) J. Mol. Biol. 118, 273–287.
[2] Warme, P.K. and Morgan, R.S. (1978) J. Mol. Biol. 118, 289–304.
[3] Hangauer, D.G., Monzingo, A.F. and Matthews, B.W. (1984) Biochemistry 23, 5730–5741.
[4] Cox, E.G., Cruickshank, D.W.J. and Smith, J.A. (1958) Proc. Roy. Soc. 247, 1–21.
[5] Bernstein et al. (1978) J. Mol. Biol. 122, 535–542.
[6] Thornton, J.M. (1973) PhD Thesis, University of London.
[7] Pawliszyn, J., Szezesniak, M.M. and Scheiner, S. (1984) J. Phys. Chem. 88, 1726–1730.
[8] Karlstrom, G., Linse, P., Wallquist, A. and Jönsson, B. (1983) J. Am. Chem. Soc. 105, 3777–3782.
[9] Burley, S.K. and Petsko, G.A. (1985) Science 229, 23–28.