# Amino-aromatic interactions in proteins

S.K. Burley*+ and G.A. Petsko+

+*Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139* and **Harvard Medical School (Health Sciences and Technology Division), 25 Shattuck Street, Boston, MA 02115, USA*

Received 19 May 1986

Geometric analysis of 33 refined high-resolution protein crystal structures (2 Å or higher) demonstrates that side-chain amino groups interact with aromatic side chains. Positively charged or $\delta(+)$ amino groups of lysine, arginine, asparagine, glutamine and histidine are preferentially located within 6 Å of the ring centroids of phenylalanine, tyrosine and trytophan, where they make van der Waals' contact with the $\delta(-)$ $\pi$-electrons and avoid the $\delta(+)$ ring edge. This geometric pattern is different from the distribution expected due to random close packing of side chains in a protein. It is opposite to oxygen- and sulfur-aromatic interactions, similar to aromatic-aromatic interactions, and almost certainly electrostatic in origin.

*Amino-aromatic interaction* *Protein crystal structure* *Side-chain contact* *Packing* *Electrostatic interaction*

## 1. INTRODUCTION

Ab initio calculations of benzene and ammonium suggest that positively charged side chains make enthalpically favorable interactions with the $\pi$-electron cloud of aromatic side chains [1]. In addition, protein-crystallographic studies of compounds bound by deoxyhemoglobin A suggest that $\delta(+)$ side-chain amino groups make similar interactions with the $\pi$-cloud of an aromatic ring [2]. We have studied the frequency and geometry of interactions between the side chains of the aromatic amino acids phenylalanine, tyrosine and tryptophan and the amino groups of lysine, arginine, asparagine, glutamine and histidine.

## 2. PROCEDURES

### 2.1. *Geometric analysis*

33 high-resolution (2 Å or higher), refined protein crystal structures were examined for the proximity of aromatic side chains of Phe, Tyr and Trp, and the side-chain amino groups of Lys, Arg, Asn, Gln and His (see fig.1 legend). Their packing geometry was analyzed using a standard polar coordinate system (see fig.1A), and the position ($r$, $\theta$, $\phi$) of each amino group was calculated for all amino-aromatic contact distances less than 10 Å in the 33 proteins. These data were compared with the results of an identical calculation for all atoms in the protein < 10 Å from the centroid of an aromatic side chain. Amino-aromatic contacts meeting criteria described below were retained for subsequent geometric and statistical analysis.

## 3. RESULTS

### 3.1. *Amino-aromatic contacts in proteins*

The normalized frequency distribution of 1556 amino-aromatic contact distances, illustrated in fig.1B, reaches a maximum at about 4.75 Å and is more sharply peaked than the corresponding normalized frequency distribution for contacts between all protein atoms and aromatic groups (see fig.1C). Beyond 6 Å the observed frequency of amino-aromatic contacts is nearly constant, which suggests that amino groups prefer van der Waals' contact with aromatic groups. When displayed as a fraction of all-atom-aromatic contacts (see fig.1D) the distribution of amino-aromatic contacts confirms that the amino and aromatic groups in proteins are preferentially separated by between

*Published by Elsevier Science Publishers B.V. (Biomedical Division)*
00145793/86/$3.50 

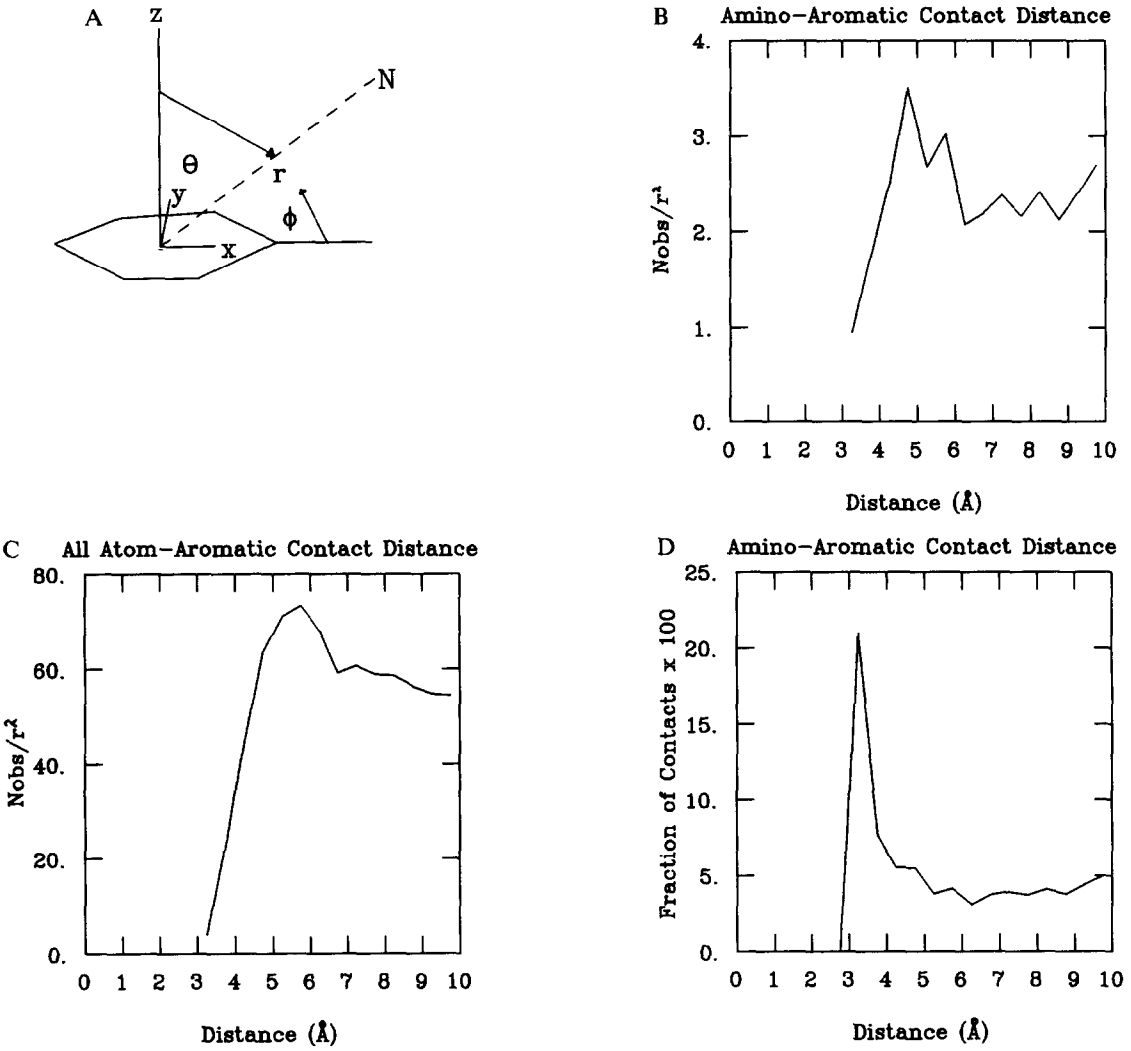



Fig.1. Thirty-three coordinate data sets were used in analysis of proteins: actinidin (P2ACT), avian pancreatic polypeptide (P1PPT), carbonic anhydrase C (P1CAC), carboxypeptidase A (P1CPA), concanavalin A (P2CNA), crambin (P1CRN), cytochrome $b_5$ (P2B5C), cytochrome *c* (P4CYT), cytochrome *c*-551 (P251C), erythrocrourin (P1ECD), immunoglobulin-FAB fragment (P3FAB), ferredoxin (P1FDX, P2FD1), flavodoxin (P4FXN), hemoglobin (P1LHB), hemoglobin $\alpha$-chain (P2MHB), hemoglobin $\beta$-chain (P2MHB), high-potential iron protein (P1HIP), insulin (P1INS), lactate dehydrogenase (P4LDH), leghemoglobin (P1HBL), lysozyme (P2LYZ), myoglobin [9], neurotoxin [10], parvalbumin (P3CPV), phospholipase $A_2$ (P1BP2), plastocyanin (P1PCY), prealbumin (P2PAB), pancreatic trypsin inhibitor (P3PTI), Bence-Jones REI protein (P1REI), ribonuclease A [11], superoxide dismutase (P2SOD) and trypsin (P1PTN). The abbreviations used to identify each protein correspond to Brookhaven protein data bank codes [12]. (A) Coordinate axes for the definition of the polar coordinate system ($r$, $\theta$, $\phi$), which places the center of mass of the 6-membered ring at the origin. Its 6-fold symmetry axis is colinear with the *Z*-axis, and the $C_\beta$–$C_\gamma$ bonds of Phe and Tyr are colinear with the *X*-axis. Tryptophan (Trp) was treated as a single 6-membered ring and the *X*-axis was placed parallel to the vector connecting the atoms $C_{\epsilon 3}$ and $C_{\zeta 2}$. (B) The distance distribution function for amino-aromatic contacts (< 10 Å). Each value of the distribution function was normalized for sample size by dividing by $r^2$. The expected distribution is shown in C. (C) The distance distribution function for all protein atom-aromatic contacts (< 10 Å). Each value of the distribution function was normalized for sample size. (D) The frequency of amino-aromatic contacts (< 10 Å) displayed as a percentage of all protein atom-aromatic contacts (< 10 Å). Like the distribution displayed in B there is a sharp peak due to the preference for amino-aromatic contact distances between 3.4 and 6 Å.

3.4 and 6 Å. Values below 3.4 Å separation were rarely observed since the amino and aromatic groups would make unfavorable van der Waals' contacts. Hence, we define an amino-aromatic interaction as $3.4 < r < 6$ Å.

The distribution of values of the polar coordinate angle $\phi$ was nearly uniform over its entire range ($0 < \phi < 2\pi$) with the following notable exceptions: The presence of the $C_\beta$–$C_\gamma$ bond in Tyr and Phe makes values of $\phi = 0°$ unlikely, and $\phi = 180°$ in Tyr is partially blocked by the hydroxyl group. The 5-membered ring of Trp partially blocks values of $\phi$ between 180 and 360°.

Table 1 documents the frequency of amino-aromatic interaction by residue. Approx. 50% of each of the aromatic side chains make close contacts with amino groups. About 25% (45/170) of the Lys amino groups make 49 interactions with aromatic groups, and about 50% (44/94) of the Arg residues, which bear 2 side-chain amino groups, make 80 amino-aromatic interactions. The values for Asn and Gln are 31% (43/137) and 40% (35/88) making 49 and 44 close amino-aromatic contacts, respectively. Finally, 40% (44/111) of the His amino groups make 111 amino-aromatic interactions.

### 3.2. *Lysine and arginine amino-aromatic interactions*

Fig.2A illustrates the distribution of the polar coordinate angle $\theta$ for Lys and Arg amino-aromatic contacts <6 Å. Unlike the distribution expected from randomly occurring packing interactions between these two chemical groups, given by $\sin\theta$ [3] and also shown in fig.2A, the distribution is sharply peaked at $\theta \simeq 45°$.

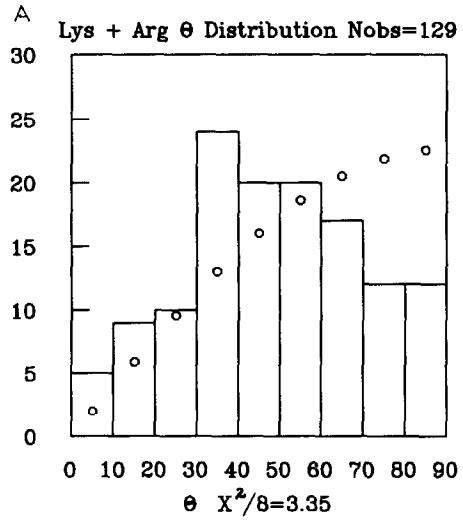
### 3.3. *Asparagine and glutamine amino-aromatic interactions*

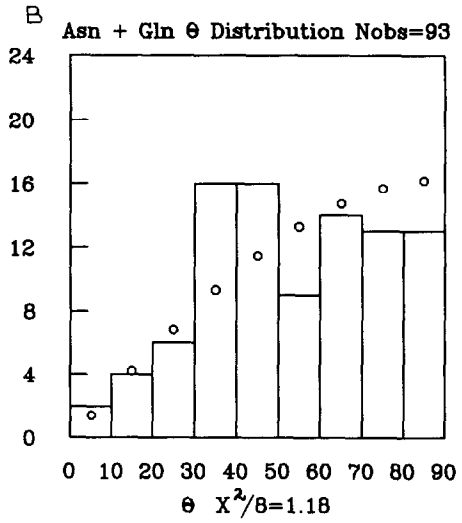
Fig.2B displays the distribution of the angle $\theta$ for these 93 contacts with aromatic side chains. Again, the distribution deviates substantially from the predicted distribution. For $\theta$ between 30 and 50° the observed distribution exceeds the predicted values by almost 50%, and for $\theta$ approaching 90° the converse is true.

Table 1

Amino-aromatic contacts <6 Å

| | Phe | Tyr | Trp | Lys | Arg | Asn | Gln | His |
|---|---|---|---|---|---|---|---|---|
| Total number of residues | 171 | 156 | 74 | 170 | 94 | 137 | 88 | 111 |
| Number and fraction of residues interacting | 84(0.49) | 86(0.55) | 35(0.47) | 45(0.26) | 44(0.47) | 43(0.31) | 35(0.40) | 44(0.40) |
| Number of amino-aromatic interactions | | | | 49 | 80 | 49 | 44 | 111 |
| Number of amino-aromatic interactions per residue | | | | 1.09 | 1.82 | 1.14 | 1.26 | 2.52 |
| Number of amino-aromatic interactions per amino group | | | | 1.09 | 0.91 | 1.14 | 1.26 | 1.26 |



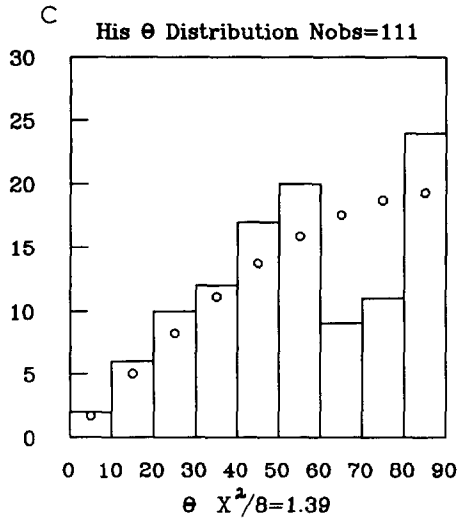




Fig.2. The distribution of values of the polar coordinate angle $\theta$ for amino-aromatic contacts in the protein crystal structures. (○) Predicted distribution of angles given by $\sin\theta$. The value of $\chi^2/8$ between the observed and expected distribution is also given. (A) The positively charged side chains of Lys and Arg. (B) The $\delta(+)$ positive side chains of Asn and Gln. (C) The side chain of His.

### 3.4. *Histidine amino-aromatic interactions*

Fig.2C shows the distribution of $\theta$ for His amino-aromatic contacts <6 Å. Again the observed distribution deviates from the expected distribution, however, there are two angular ranges for which the observed distribution exceeds the predicted values ($40 < \theta < 60°$ and $80 < \theta < 90°$).

### 3.5. *Summary*

These data indicate that the amino groups of Lys, Arg, Asn and Gln show a marked preference for amino-aromatic van der Waals' contacts with the side chains of Phe, Tyr, and Trp. On average 1 amino group participates in only 1 amino-aromatic contact. The two groups are preferentially separated by distances of between 3.4 and about 6 Å, and the preferred interaction geometry places the amino group adjacent to the face of the aromatic ring. Moreover, the amino groups avoid the edge of the aromatic ring. The observed frequency distributions of both contact distance and the polar coordinate angle $\theta$ differ significantly from the distributions expected if random orientations alone determined interaction geometry. The case of histidine is very similar to the other 4 amino-bearing side chains except that a preference for van der Waals' contact with the edges of aromatic rings is also observed.

## 4. DISCUSSION

Unlike the aromatic amino acids, which are usually found in the hydrophobic core of a protein, the positively charged amino acids and Asn and Gln have polar side chains and would not normally be thought to be located in van der Waals' contact with aromatic groups in proteins. However, the present results establish that this is indeed the case in globular proteins. The geometry of these amino-aromatic interactions demonstrates

that the positively charged or $\delta(+)$ amino groups are preferentially located near the $\delta(-)$ $\pi$-electron cloud. His reproduces this interaction geometry and also displays a statistical preference for the edge of aromatic ring. We believe that this finding, and the differences for $\theta$ approaching 90° between the Asn + Gln and Lys + Arg distributions, might be explained by misorientation of the quasi-symmetric branched or imidazole side chains in areas of poorly resolved electron density. The side-chain oxygen atoms in Asn and Gln are known to exhibit a preference for $\theta = 90°$ [4]. The carbon atoms of the His side chain exhibit a $\theta$ distribution similar to $\sin\theta$ (not shown). Examination of avian pancreatic polypeptide, trypsin and ribonuclease A, all solved to 1.5 Å resolution or higher and exhaustively refined, revealed only 1 of 14 amino-aromatic interactions involving His with $\theta > 75°$.

Close approach of the amino groups to the $\pi$-electron cloud is opposite to the way in which sulfur and oxygen interact with aromatic side chains in proteins. These atoms have a $\delta(-)$ charge and are preferentially found in van der Waals' contact with the $\delta(+)$ edges of aromatic rings [3,4]. Quantum-mechanical calculations by Thomas et al. [4] are consistent with this electrostatic interpretation, and a similar hypothesis has been advanced to explain the sulfur-aromatic interaction [3]. Moreover, the amino-aromatic interaction is analogous to the enthalpically favorable interaction between aromatic side chains observed in protein and peptide crystal structures [5–8]. In this case the two rings make edge-to-face interactions, which bring a $\delta(+)$ hydrogen atom of one ring into van der Waals' contact with the $\delta(-)$ $\pi$-electron cloud.

We suggest that packing of both polar and non-polar atoms with aromatic side chains in the hydrophobic core of a protein is determined by at least 2 requirements: (i) the need to exclude water molecules, and (ii) the formation of a large number of ethalpically favorable, weakly polar interactions that are almost certainly electrostatic in origin. Although each interaction is only capable of making a small enthalpic contribution to protein structure stability, there are on average many such interactions in a protein and their total enthalpic contribution is not insubstantial.

## ACKNOWLEDGEMENTS

We would like to thank Dr M.F. Perutz who suggested this analysis and Dr R.F. Tilton, jr for his useful comments. In addition, we thank Drs J. Singh and J. Thornton for early transmission of their work on Phe-Phe interactions in proteins. S.K.B. was supported as a Natural Sciences and Engineering Research Council of Canada Postdoctoral Fellow.

## REFERENCES

[1] Deakyne, C.A. and Meot-Ner(Mautner), M. (1985) J. Am. Chem. Soc. 107, 474–479.
[2] Perutz, M.F., Fermi, G., Abraham, D.J., Poyart, C. and Bursaux, E. (1986) J. Am. Chem. Soc. 108, 1064–1078.
[3] Reid, K.S.C., Lindley, P.F. and Thornton, J.M. (1985) FEBS Lett. 190, 209–213.
[4] Thomas, K.A., Smith, G.M., Thomas, T.M. and Feldman, R.J. (1982) Proc. Natl. Acad. Sci. USA 79, 4843–4847.
[5] Gould, R.O., Gray, A.M., Taylor, P. and Walkinshaw, M.D. (1985) J. Am. Chem. Soc. 107, 5921–5929.
[6] Burley, S.K. and Petsko, G.A. (1985) Science 229, 23–29.
[7] Singh, J. and Thornton, J.M. (1985) FEBS Lett. 191, 1–6.
[8] Burley, S.K. and Petsko, G.A. (1986) J. Am. Chem. Soc., submitted.
[9] Frauenfelder, H., Hartmann, H., Karplus, M., Kuntz, I.D., jr, Kuriyan, J., Parak, F., Petsko, G.A., Ringe, D., Tilton, R.F.T., jr, Connolly, M. and Max, N. (1986) Biochemistry, submitted.
[10] Tsernoglou, D. and Petsko, G.A. (1977) Proc. Natl. Acad. Sci. USA 74, 971–974.
[11] Gilbert, W.A. and Petsko, G.A. (1986) Biochemistry, in preparation.
[12] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., jr, Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–558.