RESEARCH ARTICLE

Aromatic-Aromatic Interaction: A Mechanism of Protein Structure Stabilization

S. K. Burley and G. A. Petsko

Abstract. *Analysis of neighboring aromatic groups in four biphenyl peptides or peptide analogs and 34 proteins reveals a specific aromatic-aromatic interaction. Aromatic pairs (<7 Å between phenyl ring centroids) were analyzed for the frequency of pair type, their interaction geometry (separation and dihedral angle), their nonbonded interaction energy, the secondary structural locations of interacting residues, their environment, and their conservation in related molecules. The results indicate that on average about 60 percent of aromatic side chains in proteins are involved in aromatic pairs, 80 percent of which form networks of three or more interacting aromatic side chains. Phenyl ring centroids are separated by a preferential distance of between 4.5 and 7 Å, and dihedral angles approaching 90° are most common. Nonbonded potential energy calculations indicate that a typical aromatic-aromatic interaction has an energy of between −1 and −2 kilocalories per mole. The free energy contribution of the interaction depends on the environment of the aromatic pair. Buried or partially buried pairs constitute 80 percent of the surveyed sample and contribute a free energy of between −0.6 and −1.3 kilocalories per mole to the stability of the protein's structure at physiologic temperature. Of the proteins surveyed, 80 percent of these energetically favorable interactions stabilize tertiary structure, and 20 percent stabilize quaternary structure. Conservation of the interaction in related molecules is particularly striking.*

Globular proteins often contain phenylalanine, tyrosine, and tryptophan, all of which have aromatic side chains. As a general rule, these residues are buried in the interior of proteins near other nonpolar side chains, whereas hydrophilic groups tend to be more exposed to solvent. In some proteins, such as parvalbumin (*1*), aromatic side chains occupy a substantial volume of the hydrophobic core and appear to interact to mediate helix-helix contacts (*2*). Although various workers have reported statistical studies of side-chain interactions in proteins (*3–5*), little is known about interactions between aromatic side chains. X-ray studies of four biphenyl compounds that inhibit sickle cell hemoglobin gelation or homozygous erythrocyte sickling (*6*) suggest that aromatic groups in small peptides can engage in specific, energetically favorable interactions. Such interactions can maintain the peptides in characteristic compact conformations, even in aqueous solution, thereby suggesting that they could participate in stabilizing protein structure.

We have examined interactions between aromatic side chains in some well-refined, high-resolution protein structures and compared these interactions to those in the four biphenyl peptide or peptide analog structures. (i) We established criteria by which interacting pairs of aromatic amino acids could be selected. (ii) A detailed characterization of selected pairs was made to define the interaction more precisely. (iii) The contribution of the interaction to protein structure stability was assessed by studies of the locations of interacting pairs and by nonbonded potential energy calculations. Further corroboration of the significance of the aromatic-aromatic interaction was sought by examining its conservation in structurally and functionally related macromolecules. (iv) The importance of aromatic-aromatic interactions in protein stability and function has been considered with reference to some biological examples.

S. K. Burley is a postdoctoral fellow and G. A. Petsko is an associate professor in the Department of Chemistry, Massachusetts Institute of Technology, Cambridge 02139. S. K. Burley is also a student in the Health Sciences and Technology Division of Harvard Medical School, Boston, Massachusetts 02115.

Aromatic-aromatic interaction definition. In order to define an interacting aromatic pair we studied the proximity of aromatic groups in proteins. A preferred distance of aromatic side-chain separation was identified by examining 34 protein structures (Fig. 1) at 2-Å resolution (or higher). For each protein, the centroids of the phenyl ring portion of each aromatic side chain were determined, and all separations less than 10 Å were calculated. The results on each macromolecule were pooled to give a distance distribution function that was normalized to correct for sample size.

The plot of the normalized distance distribution function derived from 580 pairs of nearby aromatic side chains (Fig. 1A) shows a single prominent peak centered at about 5.5 Å, suggesting that pairs of aromatic side chains have a preferred separation distance of between 4.5 and 7.0 Å. Beyond 7 Å, the distribution function shows a low, background incidence of aromatic pairs that has also been observed for pairs of charged groups in proteins (*7*), and reflects the spatial distribution of all amino acids in globular proteins. At separation distances below about 4.5 Å, aromatic pairs are rarely observed, a result of obvious physical constraints (that is, van der Waals contacts). We defined an interacting aromatic pair as one for which the distance between phenyl ring centroids is less than 7 Å.

Geometric analysis and selection of small-molecule peptide structures. Unlike most proteins, small-molecule peptide structures can be solved to very high accuracy by x-ray crystallography, and, therefore, serve as useful models of protein structure. We selected biphenyl peptides and peptide analogs, from structures having R factors of 6 percent or less and containing aromatic pairs for comparison with the previously described proteins. Four such structures are described in the legend to Fig. 1. Further examples of the interaction were sought by surveying a peptide subset of the Cambridge Crystallographic Data Base according to the above criteria. Although some small-molecule structures were identified, none were retained because they did not have bond and torsion angles typical of amino acids in proteins.

Aromatic pairs in selected small-molecule and protein structures were surveyed for dihedral angle, phenyl ring center-to-center separation, and detailed interaction geometry (all values are

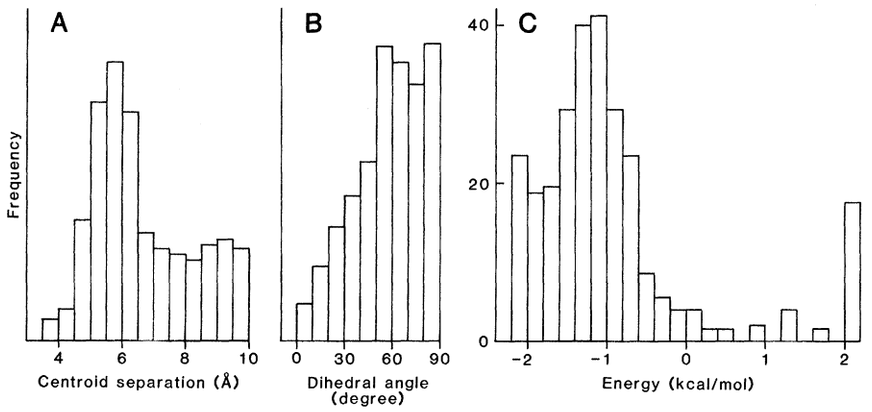

Fig. 1. Thirty-four coordinate data sets were used in analysis of proteins. These were actinidin (P2ACT), avian pancreatic polypeptide (P1PPT), carbonic anhydrase C (P1CAC), carboxypeptidase A (P1CPA), concanavalin A (P2CNA), crambin (P1CRN), cytochrome b_5 (P2B5C), cytochrome c (P4CYT), cytochrome c_{551} (P251C), erythrocruorin (P1ECD). Immunoglobulin-Fabl fragment (P3FAB), ferredoxin (P1FDX, P2FD1), flavodoxin (P4FXN), hemoglobin (P1LHB), hemoglobin α chain (P2MHB), hemoglobin β chain (P2MHB), high potential iron protein (P1HIP), insulin (P1INS), lactate dehydrogenase (P4LDH), leghemoglobin (P1HBL), lysozyme (P2LYZ), myoglobin (*33*), neurotoxin (*34*), parvalbumin (P3CPV), phospholipase A_2 (P1BP2), plastocyanin (P1PCY), prealbumin (P2PAB), pancreatic trypsin inhibitor (P3PTI), Bence-Jones REI protein (P1REI), ribonuclease A (*35*), rubredoxin (P3RXN), superoxide dismutase (P2SOD), and trypsin (P1PTN). The abbreviations used to identify each protein correspond to Brookhaven protein data bank codes (*36*). Four coordinate data sets were used for analysis of small-molecule structures as follows: benzoyl ester of L-Phe (*37*); L-Lys-L-Phe-L-Phe (*38*), L-Phe-Gly-Gly-D-Phe (*39*), *N*-phenylacetyl-L-Phe (*40*). These are small-molecule structure data sets provided by S.K.B. (A) The distance distribution function for phenyl ring centroid separation (<10 Å). Each value of the distribution function was normalized for sample size by dividing by $4\pi r^2$. (B) The dihedral angle distribution function for all aromatic pairs separated by less than 7 Å. (C) A histogram of the calculated potential energy of interaction for all aromatic pairs separated by less than 7 Å. Small-molecule analysis results are given with the corresponding data for crystalline benzene for comparison: Mean phenyl ring centroid separation, 5.05 Å (range between 4.86 and 5.18 Å), and the benzene value was 5.02 Å. The mean dihedral angle, 77° (range between 70° and 82°), benzene value, 87°. Mean potential energy of interaction, 1.28 kcal/mol (range between −0.92 and −1.70 kcal/mol), benzene value, −1.14 kcal/mol.

Table 1. The parameter set used for the nonbonded potential energy calculations.

Element	Coefficients as in Eq. 1		
	a	*b*	*c*
H	11.66	108.06	1.87
C	49.13	606.01	1.80
N	37.12	504.51	1.89
O	33.52	479.65	1.98
	Effective X . . . H length (Å)		
C	1.03		
N	0.92		
O	0.90		

Phe	δq*	Tyr	δq*	Trp	δq*
CG	0.0	CG	0.0	CG	0.0
CD1	-0.15_3	CD1	-0.15_3	CD1	-0.01_1
HCD1	0.15_3	HCD1	0.15_3	HCD1	0.15_3
CD2	-0.15_3	CD2	-0.15_3	CD2	-0.05_5
HCD2	0.15_3	HCD2	0.15_3		
CE1	-0.15_3	CE1	-0.15_3	NE1	-0.54_0
HCE1	0.15_3	HCE1	0.15_3	HNE1	0.44_0
CE2	-0.15_3	CE2	-0.15_3	CE2	0.14_2
HCE2	0.15_3	HCE2	0.15_3	CE3	-0.17_4
CZ	-0.15_3	CZ	0.21_8	HCE3	0.15_3
HCZ	0.15_3	O	-0.65_5	CZ2	-0.16_4
		HO	0.43_7	HCZ2	0.15_3
				CZ3	-0.16_4
				HCZ3	0.15_3
				CH2	-0.19_7
				HCH2	0.15_3

*As in Eq. 1.

quoted as ± 1 standard deviation where meaningful). The irregular phenyl rings often found in refined protein structures were regularized in order to preserve ring centroid and least-squares plane. Where coordinates for hydrogen atoms were not available, they were generated from ideal geometry with effective X–H bond lengths given in Table 1.

Further analysis of each aromatic pair was made by calculating a nonbonded interaction energy by a method used in studies of crystal packing interactions. The pairwise nonbonded potential energy model of Kitaigorodsky (*8*)

$$E = \sum_{jk} b_j b_k \exp[-(c_j + c_k) r_{jk}] - a_j a_k r_{jk}^{-6} + q_j q_k r_{jk}^{-1} \quad (1)$$

was chosen for this calculation on the strength of its careful development and successful application to organic aromatic compounds (*9–13*). The model was constructed from quantum mechanical calculations based on known crystal structures of relatively simple organic compounds, which provided a transferrable potential parameter set. Williams and co-workers further showed that the calculated parameter sets could be transferred to studies of other crystal structures, and especially that of benzene (*14, 15*). Each interacting atom pair i,k is given parameters a_j, a_k, b_j, b_k, c_j, and c_k, depending on atom type, partial charges q_j and q_k, and separation r_{jk} (Table 1). The physical significance of the terms is (i) a short-range repulsive energy due to overlapping electron clouds; (ii) a medium-range attractive dispersion energy; and (iii) the long-range coulombic energy. In practice, the calculation compares the difference between the energy of interaction of the observed distribution of aromatic side chains and that of an identically positioned pair of alanine residues.

Unlike many of the energy calculations used to analyze protein structures, the model given above does not explicitly contain terms describing dielectric and polarizability. We believe, however, that such data were implicitly included when the transferrable potential parameter sets were estimated from known small-molecule crystal structures. The hydrophobic interior of a protein should not differ significantly from a tightly packed crystal of an organic compound.

Small-molecule peptide structures. Four aromatic pairs were detected in the surveyed peptide structures (Fig. 2A), and a summary of the analysis is given in the legend to Fig. 1. The aromatic pair shown in Fig. 2A and those described quantitatively in Fig. 1 are very similar

to those identified in crystalline benzene (provided for comparison). Unlike the structure of each aromatic molecule, which is determined by strong interatomic forces, the packing arrangement is determined entirely by weak nonbonded interactions. Because H atoms are positively charged and C atoms are negatively charged with respect to one another in aromatic molecules, the coulombic term in the potential energy function favors close C · · · H approaches. This effect gives rise to edge-to-face interactions, where H atoms at the edge of one molecule point toward negatively charged C atoms on the faces of the adjacent molecule (Fig. 2). Hence, the dihedral angles between phenyl-ring planes are often close to the perpendicular, and when repeated in a crystal give rise to a characteristic "herringbone" packing of aromatic molecules such as that seen with benzene (*16, 17*).

Further insight into the nature of the aromatic-aromatic interaction is provided by the energy data given in the legend to Fig. 1. Calculated nonbonded energies for the four biphenyl compounds vary between −0.92 and −1.70 kcal/mol. These data indicate that the aromatic pairs form in the four small-molecule structures because they are energetically favorable. This structural feature is present in all four of the biphenyl compounds despite both the rather sizable variation in the number of backbone atoms connecting the two phenyl groups, and the differing types of intervening amino acid residues.

Proteins. A total of 272 examples of aromatic pairs were identified in the 34 high-resolution protein structures surveyed. If such pairs occurred at random over the entire range of separation distances, only 147 pairs would be anticipated to have separation distances of between 4.5 and 7.0 Å. The dihedral angle distribution for interacting aromatic pairs is shown in Fig. 1B. A majority (67 percent) of the dihedral angles lie between 50° and 90° (mean = 57°, median = 61°). These data suggest that, in addition to a preferred separation distance, there is a preferred dihedral angle between interacting aromatic side chains that is not far from perpendicular. Hence, interacting aromatic pairs in proteins do not resemble DNA with its parallel, 3.4-Å base-stacking, but, instead, resemble aromatic pairs found in benzene and many small-molecule crystal structures (as above and in Fig. 2). The analogy between aromatic pairs in proteins and those in benzene and other small-molecule structures is particularly apropos when they occur in networks in hydrophobic regions of the macromolecule. Networks are defined as groups of three or more residues, each of which make aromatic pairs with one or more members of the group [Fig. 2B shows the stereotypic aromatic pair network in carp parvalbumin (P3CPV)]. The significance of the aromatic-aromatic interaction for protein structure stabilization is reflected in the strikingly high incidence of such groups: 80 percent of the aromatic pairs identified in proteins are involved in networks.

Further analysis. On average, 61 percent of the Phe residues, 54 percent of the Tyr residues and 59 percent of the Trp residues found in a protein are involved in aromatic-aromatic interac-

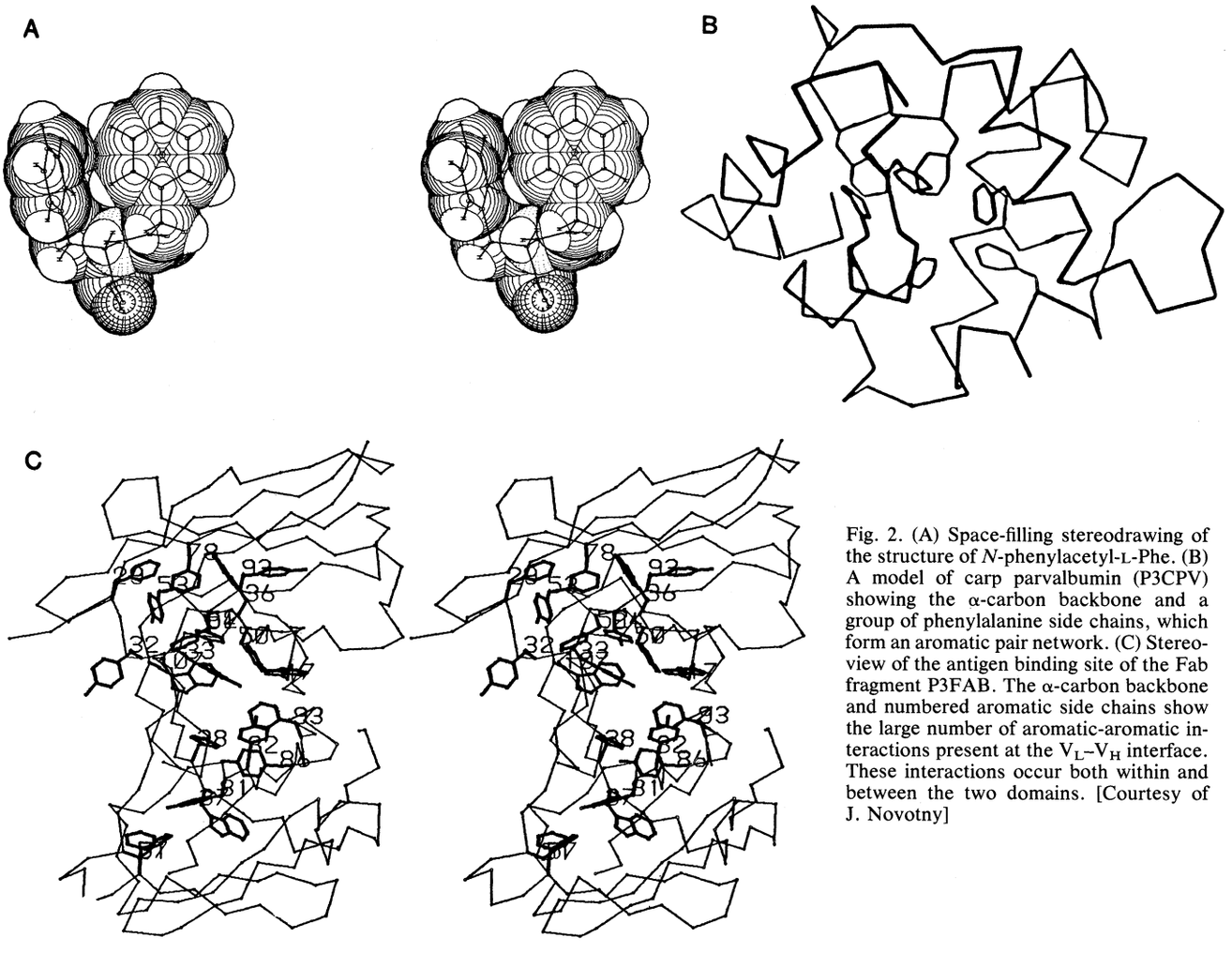

Fig. 2. (A) Space-filling stereodrawing of the structure of *N*-phenylacetyl-L-Phe. (B) A model of carp parvalbumin (P3CPV) showing the α-carbon backbone and a group of phenylalanine side chains, which form an aromatic pair network. (C) Stereoview of the antigen binding site of the Fab fragment P3FAB. The α-carbon backbone and numbered aromatic side chains show the large number of aromatic-aromatic interactions present at the V_L–V_H interface. These interactions occur both within and between the two domains. [Courtesy of J. Novotny]

Table 2. Summary of the different aromatic pair types identified in proteins. The number of observed pairs and (in parentheses) the predicted number of pairs are shown. In addition, the mean phenyl-centroid separation, dihedral angles and fraction of side chains acting as a σ acceptor are included.

	Phe	Tyr	Trp
Phe	96 (51) d = 5.58; $\measuredangle$ = 56°	72 (91) d = 5.74; $\measuredangle$ = 57° f_a = 49%	39 (39) d = 5.93; $\measuredangle$ = 59° f_a = 55%
Tyr		28 (40) d = 5.56; $\measuredangle$ = 57°	31 (35) d = 5.54; $\measuredangle$ = 57° f_a = 69%
Trp			4 (8) d = 5.42; $\measuredangle$ = 77°

tions. The number of interacting aromatic pairs in a protein depends on the number of aromatic residues in the protein (linear correlation coefficient for 32 degrees of freedom is 0.81).

The frequency distribution of the number of residues separating interacting aromatic side chains was calculated. Of the interacting aromatic groups 78 percent are separated by seven or more amino acids in the polypeptide chain. The mean and standard deviation of this distribution are 38.6 and 41.9, respectively.

We found that of the 272 aromatic-aromatic interactions identified, approximately 80 percent of them link distinct secondary structural elements. In addition, the seven two-domain proteins surveyed contain 86 instances of the aromatic-aromatic interaction, of which only 20 percent connect domains. We conclude, therefore, that aromatic-aromatic interactions contribute largely to stabilizing a protein's tertiary and not secondary or quaternary structures (Fig. 2, B and C).

The nonbonded potential energy of interaction was calculated for every nearby aromatic pair, and a histogram of the results is shown in Fig. 1C. Most of these interactions (85 percent) have attractive potential energies between 0 and −2 kcal/mol, and 54 percent fall between −1 and −2 kcal/mol. Of the remaining 15 percent, 4 percent have absurdly high attractive calculated potential energies (less than −2 kcal/mol) and 11 percent have repulsive potential energies, most of which are unreasonable (less than 2 kcal/mol). About one-third of these energetically suspect aromatic pairs have unusually short phenyl ring centroid separations, which probably result from incomplete refinement of that portion of the protein's structure. These data indicate that most aromatic-aromatic interactions in proteins are energetically favorable, as in the model compounds described earlier. More important, we can estimate from these data the energetic contribution to a protein's tertiary structure stability made by such aromatic pairs.

Aromatic pair type. The frequencies of the six possible aromatic pairs (Phe-Phe, Phe-Tyr, Phe-Trp, Tyr-Tyr, Tyr-Trp, Trp-Trp) were calculated from the pooled interaction data and compared with their expected values (Table 2). The expected frequency for a given aromatic pair is given by:

$$\frac{N_a \cdot N_b}{N_{all} \cdot N_{all}} \cdot K_{ab} \times \text{the total number of aromatic pairs}$$

where N_a, N_b, and N_{all} = the number of residues of type a, b, and all, respectively. If a and b are the same residue K_{ab} = 1; otherwise, K_{ab} = 2.

Significant differences between observed and expected values were found for Phe-Phe, Phe-Tyr, and Tyr-Tyr interactions. The observed values for both Phe-Tyr and Tyr-Tyr are lower than expected, while that for Phe-Phe is correspondingly greater than expected. These discrepancies presumably reflect the hydrogen bonding properties of Tyr; the differing partial charge distributions of the two side chains; and the degree to which the two residues are likely to be found buried within the protein's interior. Table 2 also provides some indication of the role usually played by each side-chain type in the interaction. All three aromatic residues show no significant preference for acting either as the σ acceptor or as the π donor.

Aromatic side chains that do act as σ-acceptor molecules were also examined for closest H-atom preferences. For example, Phe was most likely to have its ortho-hydrogen atom nearest the π-electron cloud (o = 47 percent, m = 38 percent, and p = 15 percent), as was Tyr (o = 46 percent, m = 34 percent, and OH = 20 percent). Trp was almost equally likely to have either HCH2 or HCZ3 closest to the π-electron cloud (HCE3 = 14 percent, HCZ2 = 21 percent, HCH2 = 34 percent, and HCZ3 = 31 percent). Three of the biphenyl small-molecule structures have their carboxyl-most ortho-hydrogen atom near the amino-most phenyl-ring π-electron cloud. In L-Phe-Gly-Gly-D-Phe, it is the carboxyl-most para-hydrogen atom that approaches the amino-most phenyl ring π-electron cloud.

Aromatic pair environment. Although no quantitative estimates of solvent accessibility were attempted because of the difficulty of comparing interacting pairs of side chains, a clear picture of the environment of aromatic pairs was obtained by examining surveyed protein structures with a molecular graphics system. Most, but not all, of the interactions occur between either buried or partially buried aromatic residues. We suggest that this distribution reflects the low probability of finding two nearby aromatic side chains on the surface of a protein. There is no reason to believe that completely exposed aromatic residues cannot form interacting pairs, especially since three of the four small-molecule structures described earlier were crystallized by slow evaporation of aqueous solution.

Ravishanker *et al.* (*18*) have reported a Monte Carlo study of aqueous hydration of benzene, which indicates that the first-shell of hydration of the π cloud consists solely of two water molecules per benzene with their H atoms pointing at opposite faces of the aromatic ring, giving a mean pair energy of −1.64 kcal/mol. Therefore, solvent-accessible aromatic side chains can only pair by displacing at least one of the water molecules of the first shell of hydration, and, thereby, breaking the hydrogen bond between the displaced water molecule and its nearest neighbor (*19*) (ΔE = 2.5 kcal/mol). This requirement suggests that the potential energy change on pair formation in the presence of solvent is about $\Delta E \doteq 1.3 + 0.8 + 2.5$ kcal/mol $\doteq$ 2.0 kcal/mol. However, the entropic change resulting from water molecule displacement is positive (*19*) ($T\Delta S$ at 37°C = −2.64 kcal/mol), and favors aromatic pair formation. Estimates of the free energy contribution of an exposed aromatic pair are necessarily uncertain; but, a reasonable value derived from the estimates given above would seem to be ΔG° = 2.0 − 2.6 = −0.6 kcal/mol at 37°C.

Conservation of aromatic pairs. If aromatic pairs play an important role in determining the structure and function of proteins, one might reasonably expect to find such interactions conserved in relat-

ed molecules. Members of the globin, calcium-binding, and immunoglobulin families were examined for conservation of aromatic-aromatic interactions.

Six members of the globin family were surveyed for conserved aromatic pairs (Table 3). All six molecules show some degree of aromatic pair conservation. More important, only the aromatic-aromatic interactions in the vicinity of helices B and C and the loop-connecting helices C and D, or C and E, are seen in every structure. This region of the globin molecule makes tight contacts with other helices and the heme, and represents the most highly conserved portion of the protein (*20*). However, only one of the conserved aromatic residues, Phe^{54}, interacts directly with the heme. In all but one of the proteins, lamprey hemoglobin (P1LHB), an incompletely refined structure, the networks created by these conserved pairs were found to have attractive nonbonded interaction energies, and it seems reasonable that the conserved interactions are important in maintaining the structure of this functionally significant region of the globin molecule.

The calcium-binding protein family demonstrates conservation of the dense aromatic pair network found in the hydrophobic core of parvalbumin (*1, 2*). Recently, the crystal structures of calmodulin and two different troponins C were reported (*21–23*). Contrary to numerous predictions made on the basis of considerable sequence homology, these structures consist of two discrete domains separated by a single α helix, and are quite unlike parvalbumin. Despite this significant structural difference, each of the globular domains of the three new calcium-binding proteins has a hydrophobic core rich in neighboring aromatic residues, which is very similar to parvalbumin.

Human immunoglobulin Fab fragments NEW and KOL and mouse immunoglobulin Fab fragment MCPC 603 have been solved to comparable resolution by x-ray crystallography and were extensively compared (*24*). The most striking feature of the antigen binding site is the preponderance of Phe, Tyr, and Trp residues interface of the light and heavy chains (V_L–V_H) (Fig. 2C). Analysis of the Fab fragment NEW P3FAB (legend to Fig. 1) revealed 13 aromatic pairs in the V_L–V_H pseudodimer, 5 of which occur at the bottom of the antigen-binding site where they contribute an attractive nonbonded interaction energy. These five interactions are absolutely conserved in the Fab fragments NEW, KOL, and MCPC 603, and appear to play a role in antigen binding (*25*).

Thermal stabilization of a protein. Temperature-sensitive mutants of enzymes and other proteins have been divided into two classes: TSS (temperature-sensitive for synthesis), and TL (thermolabil) (*26*). The TL mutants often serve as model in studies of the factors determining a protein's thermal stability. But almost all thermolabile mutants are thermally less stable than the wild type and therefore their study provides insight into the problem of thermal instability and not necessarily thermal stabil-

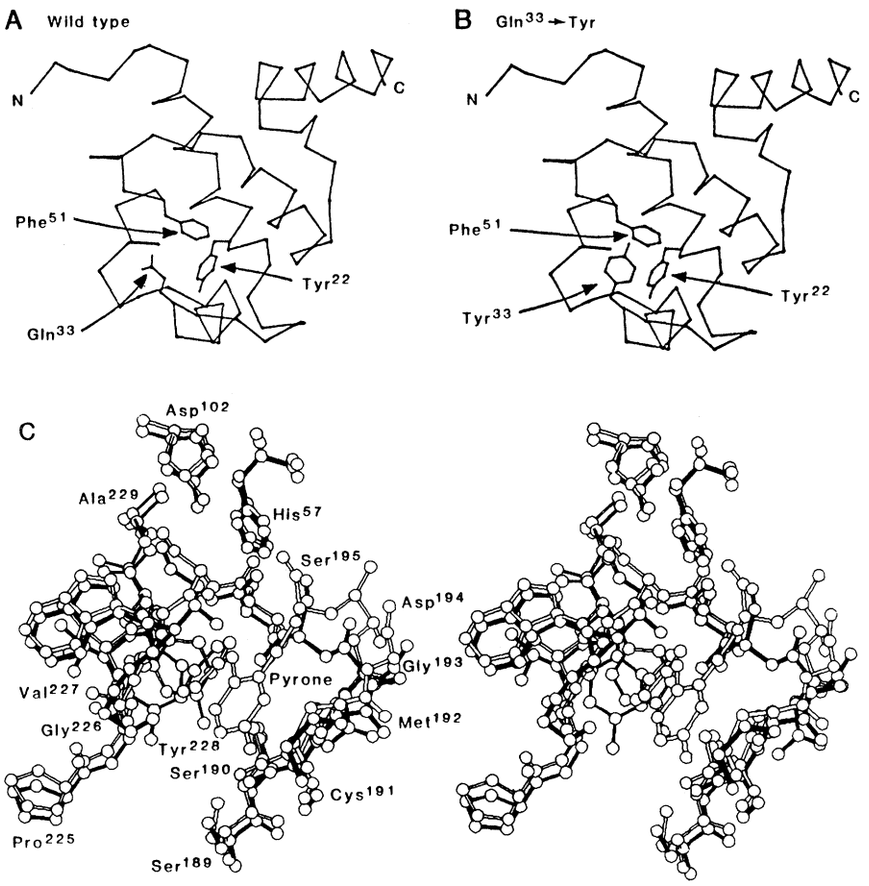

Fig. 3. Computer-generated views of the α-carbon backbone and side chains 22, 33, and 51 for the amino-terminal domain of λ repressor protein. (A) Wild type (*41*). (B) $Gln^{33} \rightarrow$ Tyr mutant. The structure of the $Gln^{33} \rightarrow$ Tyr mutant was generated by model building. [Figures courtesy of M. Hecht] (C) Stereoview of 6-benzyl-3-chloro-2-pyrone bound to the active site of chymotrypsin. The structures of the native enzyme (solid lines) and of the enzyme-inhibitor complex (clear lines) are overlayed. [Courtesy of D. Ringe and *Biochemistry* (*42*)]

Table 3. Comparison of members of the globin family. The number of aromatic pairs detected in each molecule is given in parentheses (ratio of number of pairs to the number of pairs with attractive nonbonded energy). The array tabulates the number of conserved aromatic-aromatic interactions in the smaller of the two proteins. The percentages shown in the square brackets indicate the degree of sequence homology between the two globin molecules. ECD, erythrocruorin; MHB, hemoglobin α and β chains; HBL, leghemoglobin; LHB, lamprey hemoglobin; and MBN, myoglobin. Sequence alignments are from Lesk and Chothia (*20*).

	ECD	MHB_α	MHB_β	HBL	LHB	MBN
	(18/18)	(3/3)	(6/5)*	(9/9)	(7/2)	(5/4)†
ECD (18/18)	—	2/2 [16%]	5/4† [19%]	7/7 [15%]	3/1 [14%]	4/4 [18%]
MHB_α (3/3)		—	3/3 [44%]	1/1 [16%]	1/0 [31%]	2/1† [26%]
MHB_β (6/5)*			—	3/3 [19%]	1/1 [23%]	2/0† [24%]
HBL (9/9)				—	7/2 [14%]	3/2† [18%]
LHB (7/2)					—	2/0† [24%]
MBN (5/4)†						—

*One energetically unfavorable interaction is made by Tyr^{35} of the B helix and Trp^{37} of the C helix. The intervening residue Pro^{36} is probably responsible for maintaining this unfavorable configuration. †One energetically unfavorable interaction is made by Phe^{43} of the B helix and Phe^{46} of the C helix. Both these residues interact in a favorable manner with Phe^{33} and are probably maintained in this unfavorable configuration as a result.

ity. A recent study of temperature-sensitive mutants of λ repressor protein has revealed one mutant that is significantly more thermally stable than wild type (*27*). The mutant, which has glutamine changed to tyrosine at position 33, has a melting temperature 6°C higher than wild type and at 51.5°C $\Delta G^\circ = -1.32$ kcal/mol. Examination of the crystal structure of the wild-type protein reveals that this modification positions the Tyr^{33} side chain near those of Tyr^{22} and Phe^{51} (see Fig. 3, A and B). We believe that the mutation creates a three-member aromatic pair network (where before there was a single aromatic pair) and thereby lowers the free energy of the structure by introducing two energetically favorable interactions between aromatic side chains without a substantial compensatory entropic change. The measured change in the protein's free energy is consistent with this speculation. Careful exploitation of this effect offers the possibility of designing thermally stable mutants of a whole variety of globular proteins, which could easily be made by the methods of oligonucleotide-directed site-specific mutagenesis.

Inactivator-enzyme complex. Chymotrypsin is inactivated when bound to various substituted 6-chloro-2-pyrone compounds (*28*). A recent crystallographic study of a complex of 6-benzyl-3-chloro-2-pyrone with the enzyme (*29*) demonstrated that inactivation results from (i) covalent attachment of C-6 to the γ-oxygen of the active site serine; and (ii) conformational changes that fix the 6-benzyl group in the enzyme's hydrophobic specificity pocket. When the inactivator binds to the enzyme, Tyr^{228}, which normally lies at the edge of the pocket, rotates 19° about the C_α–C_β bond (X^1) and 20° about the C_β–C_γ bond ($X^{2.1}$). This structural alteration establishes an aromatic-aromatic interaction between the benzyl group of the inhibitor and the tyrosine side chain in which the meta-H atom of Tyr^{228} acts as the σ acceptor, and the benzyl ring acts as the π donor (Fig. 3C). An analogous aromatic pair is found in a carboxypeptidase A inactivator-enzyme complex (*30*), and may represent a useful model for both drug-protein and enzyme-substrate interactions.

Conclusion. Consideration of the types of interactions possible for amino acids in proteins has identified three broad classes that contribute to the forces stabilizing the native structure. These are hydrogen bonds, electrostatic interactions (of which salt bridges are the best-characterized example), and van der Waals interactions (of which "hydrophobic bonds" may be regarded as a special case).

We believe that aromatic-aromatic interactions form a fourth, and important, class. These energetically favorable nonbonded interactions can be defined formally as those pairs with phenyl ring centroid separations >4.5 Å and <7 Å, dihedral angles within 30° of 90°, and free energies of formation of between −0.6 and −1.3 kcal/mol. Although the total contribution of this interaction to the stability of the protein is small relative to other types of interactions (≐ −10 kcal/mol), it is comparable to the free energy of protein folding (*31*). Where proteins contain a large number of aromatic pairs, such as prealbumin or the calcium-binding proteins, the stabilizing role of the interaction is probably very important.

It is interesting to point out that aromatic residues, and hence aromatic pairs, are absent in regions where the polypeptide chain is disordered. We suggest that aromatic-aromatic interactions may form nucleation sites in the protein folding pathway. This hypothesis could be tested by making carefully selected mutant proteins, again with the use of oligonucleotide-directed site-specific mutagenesis.

Finally, there is some evidence that amino acids with aromatic side chains can participate in interactions with carbonyl oxygen atoms in proteins. Thomas *et al.* (*32*) described a favorable interaction in which the O atom is preferentially in the plane of the benzyl ring of the phenylalanine residue. As with aromatic pairs, this interaction arises because of the characteristic distribution of partial charges in the aromatic ring. The findings of this investigation and those reported by Thomas *et al.* suggest that proteins can derive a substantial amount of structural stabilization and ligand binding capacity from inclusion of one or more aromatic residues.

References and Notes

1. C. E. Nockolds, R. H. Kretsinger, C. J. Coffee, R. A. Bradshaw, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 581 (1972).
2. P. C. Moews and R. H. Kretsinger, *J. Mol. Biol.* **91**, 201 (1975).
3. P. K. Warme and R. S. Morgan, *ibid.* **118**, 273 (1978); *ibid.*, p. 289.
4. ———, *ibid.*, p. 118.
5. G. M. Crippen and I. D. Kunz, *Int. J. Peptide Protein Res.* **12**, 47 (1978).
6. S. K. Burley, A. H.-J. Wang, J. R. Votano, A. Rich, in preparation.
7. D. J. Barlow and J. M. Thornton, *J. Mol. Biol.* **168**, 867 (1983).
8. A. Kitaigorodsky, *Molecular Crystals and Molecules* (Academic Press, New York, 1973).
9. S. R. Cox, L.-Y. Hsu, D. E. Williams, *Acta Cryst.* **A37**, 293 (1981).
10. D. E. Williams, *ibid.* **A28**, 629 (1972).
11. ———, *ibid.* **A30**, 71 (1974).
12. ——— and S. R. Cox, *ibid.*, in press.
13. D. E. Williams and T. L. Starr, *Computers Chem.* **1**, 173 (1977).
14. L.-Y. Hsu and D. E. Williams, *Acta Cryst.* **A36**, 277 (1980).
15. D. E. Williams, *ibid.*, p. 715 (1980).
16. E. G. Cox, D. W. J. Cruickshank, J. A. S. Smith, *Proc. Soc. London, Ser. A* **247**, 1 (1958).
17. R. W. G. Wyckoff, *Crystal Structures, The Structure of Benzene Derivatives* (Interscience, New York, 1969), vol. 6.
18. G. Ravishanker, P. K. Mehrotra, M. Mezei, D. L. Beveridge, *J. Am. Chem. Soc.* **106**, 4102 (1984).
19. S. N. Vinogradov and R. H. Linnell, *Hydrogen-Bonding* (Van Nostrand Reinhold, New York, 1971).
20. A. N. Lesk and C. Chothia, *J. Mol. Biol.* **136**, 225 (1980).
21. Y. S. Babu, J. S. Sack, T. J. Greenough, C. E. Bugg, A. R. Means, W. J. Cook, *Nature (London)* **315**, 37 (1985).
22. M. Sundaralingam *et al.*, *Science* **227**, 945 (1985).
23. O. Herzber and M. N. G. James, *Nature (London)* **313**, 653 (1985).
24. J. Novotny and E. Haber, *Proc. Natl. Acad. Sci. U.S.A*, in press.
25. S. K. Burley *et al.*, in preparation.
26. J. R. Sadler and A. Novick, *J. Mol. Biol.* **12**, 305 (1965).
27. M. H. Hecht, J. M. Strurtevant, R. T. Sauer, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5685 (1984).
28. R. B. Westkaemper and R. H. Abeles, *Biochemistry* **22**, 3256 (1983).
29. D. B. Ringe, D. B. Seaton, M. Gelb, R. H. Abeles, *ibid.* **24**, 64 (1984).
30. W. N. Lipscomb, *Proc. R. A. Welch Found. Conf. Chem. Res.* **15**, 140 (1971).
31. J. L. Finney, B. J. Gellatly, I. C. Golton, J. Goodfellow, *Biophys. J.* **32**, 17 (1980).
32. K. A. Thomas, G. M. Smith, T. B. Thomas, R. J. Feldman, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4843 (1982).
33. H. Frauenfelder *et al.*, in preparation.
34. D. Tsernoglou and G. A. Petsko, *Proc. Natl. Acad. Sci. U.S.A.* **7**, 971 (1971).
35. W. Gilbert and G. A. Petsko, in preparation.
36. F. C. Bernstein *et al.*, *J. Mol. Biol.* **112**, 535 (1977).
37. A. H.-J. Wang and S. K. Burley, in preparation (structure of benzoyl ester of L-Phe).
38. ———, in preparation (structure of L-Lys-L-Phe-L-Phe).
39. S. Fujii, S. K. Burley, A. H.-J. Wang, in preparation (structure of L-Phe-Gly-Gly-D-Phe).
40. S. K. Burley and A. H.-J. Wang, in preparation (structure of *N*-phenylacetyl-L-Phe).
41. C. O. Pabo and M. Lewis, *Nature (London)* **298**, 443 (1982).

42. We thank Drs. R. E. Bruccoleri, M. H. Hecht, D. S. Kemp, J. Novotny, D. Ringe, R. T. Sauer, B. Seaton, R. F. Tilton, Jr., and D. E. Williams for their useful discussion. S.K.B. was supported as a Natural Sciences and Engineering Research Council of Canada postdoctoral fellow.

11 December 1984; accepted 5 June 1985