

# A Survey of Aspartate–Phenylalanine and Glutamate–Phenylalanine Interactions in the Protein Data Bank: Searching for Anion– $\pi$ Pairs


Vivek Philip,<sup>†</sup> Jason Harris,<sup>†,||</sup> Rachel Adams,<sup>‡</sup> Don Nguyen,<sup>†</sup> Jeremy Spiers,<sup>†,||</sup> Jerome Baudry,<sup>†,||</sup> Elizabeth E. Howell,<sup>†,\*</sup> and Robert J. Hinde<sup>§</sup>

<sup>†</sup>Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, Knoxville, Tennessee 37996-0840, United States

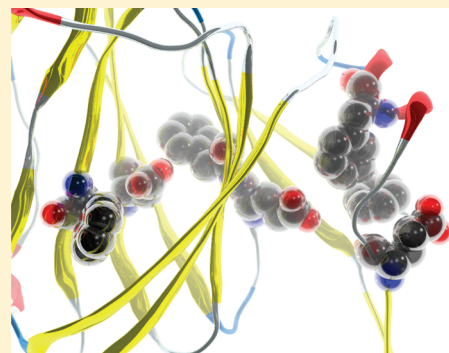
<sup>‡</sup>Genome Science and Technology Program, University of Tennessee-Oak Ridge National Laboratory, 1060 Commerce Park, Oak Ridge, Tennessee 37830-8026, United States

<sup>§</sup>Department of Chemistry, University of Tennessee, Knoxville, Tennessee 37996-1600, United States

<sup>||</sup>Center for Molecular Biophysics, University of Tennessee-Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830-8026, United States

 Supporting Information

**ABSTRACT:** Protein structures are stabilized using noncovalent interactions. In addition to the traditional noncovalent interactions, newer types of interactions are thought to be present in proteins. One such interaction, an anion– $\pi$  pair, in which the positively charged edge of an aromatic ring interacts with an anion, forming a favorable anion–quadrupole interaction, has been previously proposed [Jackson, M. R., et al. (2007) *J. Phys. Chem. B* 111, 8242–8249]. To study the role of anion– $\pi$  interactions in stabilizing protein structure, we analyzed pairwise interactions between phenylalanine (Phe) and the anionic amino acids, aspartate (Asp) and glutamate (Glu). Particular emphasis was focused on identification of Phe–Asp or –Glu pairs separated by less than 7 Å in the high-resolution, nonredundant Protein Data Bank. Simplifying Phe to benzene and Asp or Glu to formate molecules facilitated in silico analysis of the pairs. Kitaura–Morokuma energy calculations were performed on roughly 19000 benzene–formate pairs and the resulting energies analyzed as a function of distance and angle. Edgewise interactions typically produced strongly stabilizing interaction energies (–2 to –7.3 kcal/mol), while interactions involving the ring face resulted in weakly stabilizing to repulsive interaction energies. The strongest, most stabilizing interactions were identified as preferentially occurring in buried residues. Anion– $\pi$  pairs are found throughout protein structures, in helices as well as  $\beta$  strands. Numerous pairs also had nearby cation– $\pi$  interactions as well as potential  $\pi$ – $\pi$  stacking. While more than 1000 structures did not contain an anion– $\pi$  pair, the 3134 remaining structures contained approximately 2.6 anion– $\pi$  pairs per protein, suggesting it is a reasonably common motif that could contribute to the overall structural stability of a protein.



Analysis of protein structure and ligand binding has been traditionally understood based on hydrogen bonds, van der Waals interactions, hydrophobic interactions, and ion pairs. However, other types of nonbonded interactions have been suggested to play a role in the stabilization of protein structures and of protein–ligand interactions. For example, hydrogen bonds involving CH as a donor group and  $\pi$  systems as acceptor groups have been described, including formation of CH– $\pi$  pairs<sup>1,2</sup> and CH–O pairs,<sup>3,4</sup> where the aliphatic hydrogen forms weak bonds with nearby aromatic and carbonyl groups, respectively. Other groups have proposed S– $\pi$  bonds between cysteine and aromatic amino acids.<sup>5</sup>

An additional type of nonbonded interaction includes the cation– $\pi$  pair between aromatic side chains (e.g., phenylalanine, tyrosine, and tryptophan) and positively charged side chains (lysine and arginine), where the cation interacts with the  $\pi$  electron cloud on the face of the aromatic ring. These cation– $\pi$

pairs have been proposed to stabilize protein structures by  $\sim 3 \pm 1.5$  kcal/mol.<sup>6,7</sup> Recently, an n– $\pi$ <sup>8</sup> interaction whereby a delocalized, lone pair of electrons from a backbone carbonyl atom interacts with the antibonding orbital of the next carbonyl oxygen has been proposed to exist. This interaction is predicted to provide up to 0.5–1.3 kcal/mol of stability to helices when it occurs in protein structures. Additional calculations suggest that  $\pi$ – $\pi$  stacking is likely important.<sup>9–11</sup> While all these interactions are weak, numerous such occurrences can produce a substantial stabilizing effect on protein structure.

Lastly, we have proposed that an anion– $\pi$  interaction that can contribute energetically to protein stabilization, ligand binding, or protein–protein association exists.<sup>12</sup> In this work, we focus on

**Received:** January 14, 2011

**Revised:** February 25, 2011

**Published:** March 02, 2011

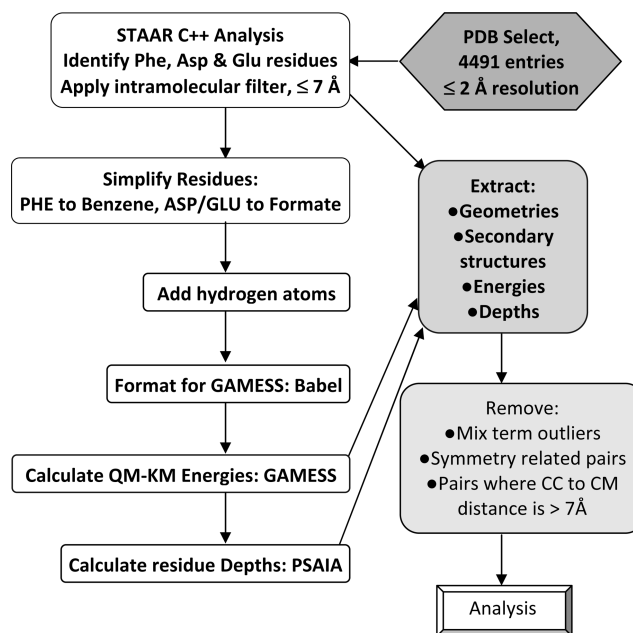
the prevalence of anion- $\pi$  interactions in protein structures where negatively charged amino acids (aspartate or glutamate) can form energetically favorable pairs utilizing the positively charged ring edge of aromatic groups. The positive charge on the aromatic ring edge arises from the quadrupole moment of the side chain, leading to the anion-quadrupole or anion- $\pi$  name. As the description suggests, this interaction is related to the cation- $\pi$  pair; however, anion- $\pi$  pairs facilitate an interaction between an anion and the aromatic ring edge rather than between a cation and the ring face. Other groups have studied anion- $\pi$  interactions in small molecules and have focused on electron-deficient  $\pi$  rings by incorporating strong electron-withdrawing substituents such as fluorobenzene derivatives, fluoro-*s*-triazine, and tetrafluoroethene as well as other aromatic molecules.<sup>13–25</sup> Experimental evidence of the existence of anion- $\pi$  interactions in small molecules includes spectroscopic, NMR, and crystallographic data of anion binding sites in electron deficient aromatics and host-guest molecular complexes, as well as other compounds found by screening the Cambridge Structural Database.<sup>26–30</sup>

Early examples of anion- $\pi$  interactions in biology appear to have been noticed but not identified as such. For example, oxygen atoms and cysteines display a high propensity to occur at the ring edge of aromatic amino acids.<sup>31–34</sup> Additionally, the Atlas of Protein Side chain Interactions indicates a statistical preference for an edgewise interaction between aromatic amino acids and Asp or Glu residues.<sup>35</sup> Studies by Kallenbach et al.<sup>36–38</sup> using short  $\alpha$ -helical peptides with glutamate-phenylalanine pairs positioned at *i* and *i* + 4 spacing indicate this pairwise interaction provides ~0.5 kcal/mol additional stability to the helix. For intermolecular binding interactions, Jouglin et al. found that tryptophan and tyrosine (as well as histidine and arginine) residues show the most enrichment at phospho-residue binding sites.<sup>39</sup> Soga et al. propose aromatic residues are “binding site-philic” as phenylalanine, histidine, and tryptophan are most commonly found in binding of druglike molecules.<sup>40</sup> Most recently, studies of a limited four-amino acid code employed at antibody binding sites find tyrosine as the dominant residue that provides tight binding to a host of antigens.<sup>41–43</sup> For the latter two examples, aromatic groups can play many roles during binding as they are amphipathic and can provide cation- $\pi$  and anion- $\pi$  interactions and the tyrosine and tryptophan side chains can provide hydrogen bonds. The relevance of the anion- $\pi$  interaction is further recognized as being important in more recent experimental as well as theoretical studies.<sup>11,44–49</sup>

Our statistical analysis expands our previous theoretical study by searching for anion- $\pi$  pairs in a nonredundant, high-resolution subset of the Protein Data Bank (PDB). We focus on contributions involving phenylalanine as the aromatic partner and either aspartate or glutamate as the anion. First, angles and distances are calculated between these partners identified from the PDB screening. Then, using simplified models for each chemical group, interaction energies are obtained. We find a substantial number of such anion- $\pi$  pairs to be present in the PDB with stabilizing energetics (more negative than -2 kcal/mol).

## EXPERIMENTAL PROCEDURES

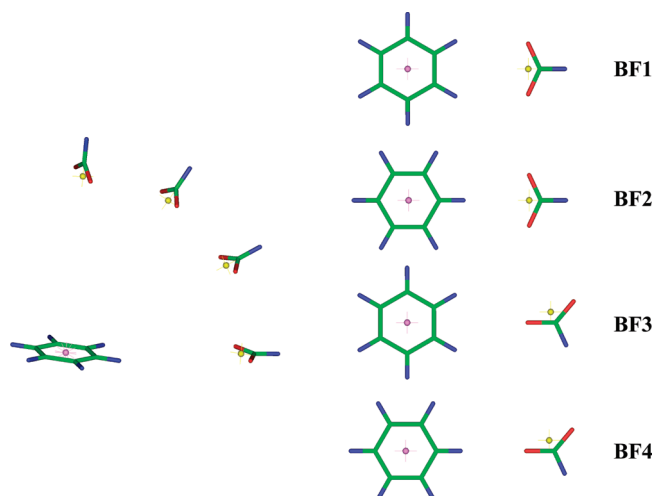
This section describes the steps taken to calculate energies associated with potential anion- $\pi$  interactions in proteins. Only interactions between phenylalanine, simplified to a benzene ring,



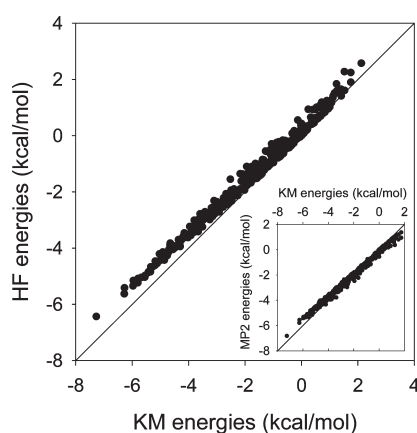
**Figure 1.** Flowchart of the calculations. CC denotes the center of charge of the formate and CM the center of mass of the benzene molecule.

and aspartic acid or glutamic acid, simplified to formate, were studied. While tyrosine and tryptophan are aromatic amino acids that can be simplified to phenol and indole, respectively, both molecules contain dipole moments and can participate in the formation of hydrogen bonds. Therefore, it is difficult to parse out the role of the anion- $\pi$  interaction in their energetics, and they were not analyzed. Figure 1 shows a general flowchart of the steps in our analysis of biological anion- $\pi$  interactions. The first step used a C++ program named STAAR (STatistical Analysis of Aromatic Rings)<sup>12</sup> to analyze the Hobohm and Sander subset<sup>50,51</sup> of the PDB, which includes nonredundant, high-resolution structures. In our approach, only crystal structures with a resolution of  $\leq 2$  Å were analyzed; this corresponded to 4491 entries (March 2006 release). STAAR locates phenylalanine rings and determines their centers of mass (CM). For each aromatic ring, STAAR then calculates the distance *r* between the ring's center of mass and the nearest oxygen atom in a Glu or Asp carboxylate group, as well as the angle  $\theta$  between the plane of the ring and the vector connecting the ring center of mass with this oxygen atom. We adopted the cutoff criteria of Gallivan and Dougherty;<sup>7</sup> i.e., those pairs possessing a distance *r* of  $\leq 7$  Å were chosen for analysis to eliminate cases in which a water molecule could fit between the two residues and diminish the interaction energy. While STAAR can identify intermolecular pairs, in this study, we focus only on intramolecular pairs.

The next steps in our analysis simplified the Phe and Asp or Glu pairs to benzene-formate (BF) pairs. This was followed by addition of hydrogens to the BF pairs using ProDrg2 at <http://davapc1.bioch.dundee.ac.uk/prodrg/>.<sup>52</sup> Because the *pK<sub>a</sub>* values for Asp and Glu are low (i.e., 3.5–4.5<sup>53</sup>), we assume Asp and Glu are always ionized. The resulting file was converted from PDB coordinates to an xyz format using BABELwin.<sup>54</sup> PC GAMESS (June 1999 version<sup>55</sup>) was used to calculate the pairwise energies. Perl scripts and Excel spreadsheets were used to sort through the data. Later calculations used a Perl script to calculate the center of



**Figure 2.** Benzene–formate pairs. On the right, four different coplanar dimers of benzene and formate were constructed and labeled BF1–BF4.<sup>12</sup> On the left, the dihedral angle  $\theta$  (between the planes containing the benzene and formate monomers) of the BF1 pair was varied between  $0^\circ$  and  $90^\circ$  in  $30^\circ$  increments. The center of mass for benzene is colored purple, while the center of charge for formate is colored yellow. The atoms are colored as follows: green for carbon, red for oxygen, and blue for hydrogen.



**Figure 3.** Correlation between Hartree–Fock (HF) and Kitaura–Morokuma energies. From more than 4000 PDB files, more than 19000 phenylalanine–Asp or –Glu pairs were identified and simplified to benzene–formate pairs. PC GAMESS was used to calculate interaction energies using a Kitaura–Morokuma (KM) energy decomposition approach. The pairs were sorted according to their interaction energies, and approximately every 50th pair was selected for HF and MP2 energy calculations using a BSSE correction error.<sup>63</sup> The HF and KM energies from these 322 pairs are shown.

charge (CC) for formate. This allowed us to redefine  $r$  and  $\theta$  as the distance and angle between the formate center of charge and the benzene center of mass, respectively.

An important step in our study of anion– $\pi$  interactions was to identify a tractable but energetically accurate calculation. Using Gaussian 03,<sup>56</sup> we previously performed quantum chemical calculations, corrected for basis set superposition error (BSSE), for optimized benzene–formate pairs BF1–BF4 (see Figure 2) at both the Hartree–Fock (HF) and second-order Møller–Plesset (MP2) levels of theory.<sup>12</sup> To screen a large number of

pairs, we hoped to identify an approach that would be more rapid than these first-principles quantum chemical calculations. We therefore looked for a linear correlation between the HF energies of the BF1–BF4 pairs with energies calculated using several semiempirical approaches. Although Gallivan and Dougherty found the OPLS (optimized potentials for liquid simulations) force field worked well for identifying cation– $\pi$  interactions,<sup>7</sup> we did not observe a linear correlation between OPLS and HF energies. While reasonable correlations were observed using AM1 and PM3 force fields, we ultimately used PC GAMESS running a Kitaura–Morokuma (KM) energy decomposition analysis.<sup>57–60</sup> Using the optimized BF1–BF4 pairs, a reasonably linear correlation (slope = 0.92, and  $\chi^2 = 0.988$ ) was observed when the HF and KM energies were compared. To compare these two treatments for benzene–formate pairs derived from actual Phe with Asp or Glu residues that occur in the PDB, we calculated both KM and HF energies for 322 different pairs, as shown in Figure 3. The aug-cc-pVTZ atom-centered basis set was employed in all calculations;<sup>61,62</sup> spherical Gaussians were used except where otherwise indicated. Interaction energy calculations were performed with the benzene center of mass held at the origin of the spherical coordinate system and the benzene molecule stationed in the  $(x, y)$  plane, with C–H bonds oriented along the positive and negative  $y$  axes. The counterpoise method<sup>63</sup> was used to correct for basis set superposition error in all HF benzene–formate calculations.

**CHARMM22 Calculations.** One hundred forty pairs from the PDB that were found to exhibit negative (i.e., attractive) anion– $\pi$  ab initio interaction energies were analyzed by the CHARMM22 force field<sup>64</sup> implemented in MOE version 2009.10 (Chemical Computing Group, Ltd., Montreal, QC). Calculations were performed in the gas phase with no cutoff value for nonbonded interactions. For each of the 140 pairs, the residues were isolated from their protein context and backbone atoms were deleted with the exception of the  $\alpha$ -carbon; i.e., only side chain and  $\alpha$ -carbon atoms were considered. Hydrogen atoms were added, and an energy minimization was performed on the hydrogen atoms only (heavy atoms were held fixed). Interaction energies were calculated by subtracting the sum of the CHARMM22 potential energies of the two individual amino acids from the CHARMM22 potential energy of the amino acid pair. The van der Waals and electrostatic contributions to the nonbonded interaction energies were also recorded. In another set of calculations, side chain atoms were deleted from the models until only the functional groups remained. The potential interaction energy between these functional groups was calculated as described for the side chains for all the 140 anion– $\pi$  pairs. Correlation coefficients were calculated between the energies derived from CHARMM22 and those interaction energies obtained from HF, MP2, and KM calculations for both side chain and functional groups.

## RESULTS

**Screening the PDB.** A C++ program, STAAR,<sup>12</sup> was used to identify Phe, Asp, and Glu residues and to calculate the distances and angles between the aromatic ring and the carboxylate moiety. This was the first step in investigating the energetics associated with formation of the aromatic–anionic pairs. Our approach follows the general procedure of Gallivan and Dougherty, who examined cation– $\pi$  interactions in the PDB.<sup>7</sup> Their rationale for including an energy calculation step was based on both the



complex electrostatic potential surfaces of aromatic residues and the inability of geometric criteria to differentiate between attractive and repulsive interactions.

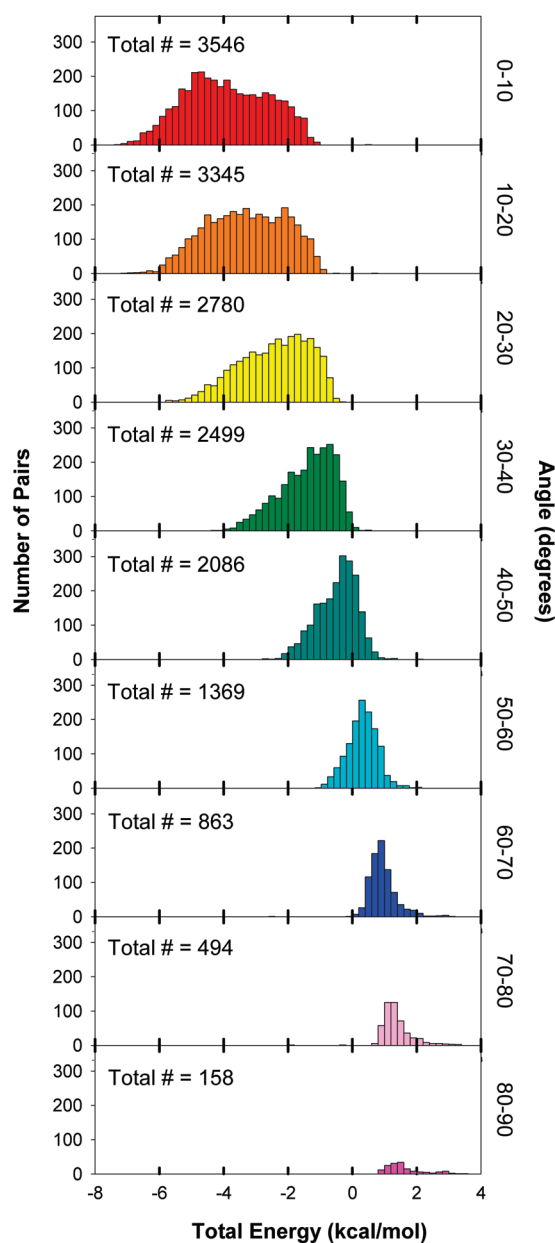
The high-resolution, nonredundant Hobohm and Sander subset of the PDB (March 2006 release) was screened using STAAR.<sup>50,51</sup> A distance filter was employed to identify those aromatic–carboxylate pairs that were separated by  $\leq 7$  Å. An initial search found approximately 40000 pairs that satisfied this distance criterion; however, because STAAR recognizes identical pairs from different but symmetrical subunits, subsets of the identified pairs were redundant. For those PDB files that contained information regarding monomer identity (A, B, C, etc.), a Perl script identified and removed the duplicates, retaining the pairs from only nonredundant monomers. After this step, approximately 22000 pairs between Phe and either Asp or Glu in 3995 PDB files remained.

Next, the amino acid side chains were simplified to benzene and formate; hydrogens were added using ProDrg2.<sup>52</sup> As PC GAMESS requires an xyz file format, the PDB coordinates were reformatted using BABELwin.<sup>65</sup> PC GAMESS<sup>55</sup> was then used to calculate the energies associated with pair formation using the Kitaura–Morokuma energy decomposition analysis.<sup>57–60</sup> We previously found this treatment provided good estimates of the interaction energy at low angles between the plane of the ring and the carboxylate but did not accurately reflect the energetic contributions at larger angles, particularly at 90°.

The resulting pairs were analyzed by various criteria. The KM calculation deconvolutes the total energy into electrostatic (ES), polarization (PL), charge transfer (CT), exchange repulsion (EX), and mix terms. The “mix” term balances any differences that arise between the sum of the ES, PL, EX, and CT terms with the total interaction energy. Figure S1 of the Supporting Information plots the total energy versus the “mix” value. As the bulk of mix values occur between  $-0.25$  and  $0.25$  kcal/mol, approximately 500 outliers with mix values of  $>|0.25|$  kcal/mol were removed from further analysis. This step was performed because when the mix term is large in magnitude, it is difficult to interpret the physical significance of the other energy terms.

The distance  $r$  calculated using STAAR is the distance from the center of mass of benzene to the closest oxygen in the carboxylate group. Using a Perl script, the position for the formate CC was calculated, allowing recalculation of the  $r$  value as the distance between the formate CC and the benzene center of mass. After this step, another distance filter step allowed removal of approximately 5000 pairs with  $r$  values of  $>7$  Å. A final Perl script counted the number of BF pairs per PDB file. Three files (1A9X, 1E6Y, and 1K8K) exhibited  $>50$  occurrences of anion– $\pi$  pairs. Upon inspection of the PDB file, symmetry-related monomers were found to be present in 1A9X and 1E6Y, although not specified by a chain ID. These were manually removed, leaving 17042 pairs for analysis.

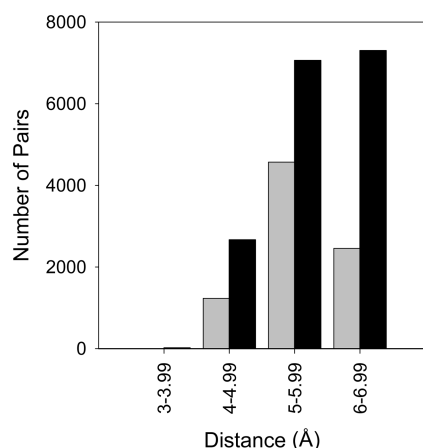
**Angle and Distance Analysis.** From our previous QM calculations of optimized BF pairs,<sup>12</sup> we found that the strongest interaction energies were associated with edgewise interactions. This pattern arises from the positive electrostatic potential at the ring edge compared to a negative electrostatic potential at the ring face associated with the  $\pi$  electron clouds. The coplanar or edgewise interaction trend continues to be observed in crystal structures from the PDB, as shown in Figure 4 where a stacked bar plot displays the number of intramolecular BF pairs compared to the total interaction energy for increasing values of  $\theta$  in



**Figure 4.** Histogram analysis of the interaction energies as a function of angle. Interaction energies for the BF pairs derived from the PDB were binned according to their dihedral angle  $\theta$  and plotted. As the angle increases, the number of pairs decreases, as does the energy.

10° increments. The total number of pairs is noted for each bin of  $\theta$  values. Figure S2 of the Supporting Information compares the fraction of pairs residing in a particular 10° bin with the fraction that would be observed for a random, uniform distribution of BF separation vectors. We see that the number of pairs increases as  $\theta$  decreases, and that more pairs are observed at small  $\theta$  values (and fewer pairs at large  $\theta$  values) than would be expected from a uniform distribution. Similar trends were observed in our earlier study<sup>12</sup> of a smaller ensemble of proteins from the PDB.

To focus on those pairs with stabilizing interaction energies from  $-8$  to  $-2$  kcal/mol, we yet again filtered the data, leaving 8260 pairs in 3134 PDB files. Only six pairs were identified with energies more negative than  $-7$  kcal/mol. The strongest



**Figure 5.** Distance relationships of the BF pairs. The gray bars describe the number of pairs possessing interaction energies from  $-8$  to  $-2$  kcal/mol as a function of distance. The black bars provide a count of the number of total pairs for these distances for all the BF pairs in the  $-8$  to  $2$  kcal/mol energy range.

attractive interaction ( $-7.27$  kcal/mol) arises for the Phe97–Asp23 pair in the lignin peroxidase structure (1LLP). Most pairs with energies more negative than  $-7$  kcal/mol resemble the BF1 configuration, with slight variations in the distance and angle  $\theta$ . The energies calculated for many of the anion– $\pi$  interactions are substantially stabilizing, with roughly 39% of this 8260-pair subset having predicted interaction energies of less than or equal to  $-4.0$  kcal/mol.

As the radius of a sphere increases, so does the volume. Assuming a random distribution, one would expect increasing numbers of pairs to be found as the radius between the benzene center of mass and the carboxylate center of charge increases.<sup>7,35</sup> This is seen in Figure 5 (black bars) for the 17042 BF pairs encompassing the  $-8$  to  $5.6$  kcal/mol energy range. However, when the 8260 pairs with energies between  $-8$  and  $-2$  kcal/mol are analyzed, there are fewer pairs at longer distances (gray bars in Figure 5), consistent with closer contact correlating with more negative interaction energies.

The strongest attractive interactions for the BF1 and BF2 pairs occurred with an approximately  $4$  Å separation distance (see Figure 2 of ref 12), while the optimal energies for the BF3 and BF4 pairs occurred at roughly  $5.5$  Å. When the separation distances between the Phe and Asp or Glu pairs identified in this study are considered (Figure 5), the largest number of pairs occurs between  $5$  and  $6$  Å. This suggests that either the pairs found in nature mimic the BF3 and BF4 pairs or the protein context plays a role in residue positioning, allowing other configurations of benzene and formate. To determine if a large population of BF3 and BF4 pairs are present in the data set, we noted that the distance between oxygen atom 1 (oxy1) of formate and the benzene CM should be similar to the distance between oxygen atom 2 (oxy2) of formate and the benzene CM for the BF1 and BF2 pairs. In contrast, the difference in these distances for the BF3 and BF4 pairs should be larger. Therefore, we calculated the absolute value of the difference of these distances and plotted these values as a function of energy. We did not observe a bimodal distribution as might be expected if the pairs segregated into BF1- and BF2-like and BF3- and BF4-like groups; rather, the pairs sample a large range of sequence space (see Figure S3 of the Supporting Information). This analysis

indicates proteins show more structural variation than the optimized BF pairs initially considered.

**Structural Analysis.** Figure 6A indicates most anion– $\pi$  pairs are widely separated in the primary sequence of their respective proteins. However, 8% of the total pairs occur next to each other with an  $i$  and  $i + 1$  spacing. Additionally, 11% show an  $i$  and  $i + 2$  spacing, consistent with  $\beta$  sheet interactions. Finally, there is a reasonable percent (14%) showing an  $i$  and  $i + 3$  spacing or an  $i$  and  $i + 4$  spacing, consistent with residues interacting in an  $\alpha$  helix.

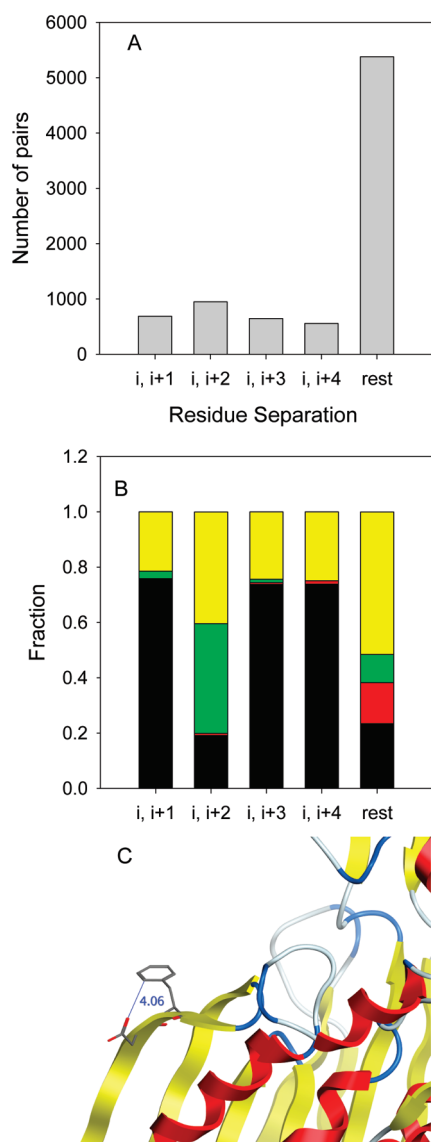
The secondary structure information of the proteins exhibiting anion– $\pi$  pairs was extracted for a subset of 6934 pairs using a Perl script. Figure 6B shows interactions involving an undesignated structure are most common, followed by helix–helix interactions. Strand–strand and helix–strand interactions also occur. As Phe, Glu, and Asp have strong propensities to be in helices,<sup>66,67</sup> these results are not surprising. Additionally, we note Phe has a strong propensity to be found in a  $\beta$  sheet; however, Asp and Glu do not.<sup>68–70</sup>

Because the periodicity of a  $\beta$  strand is 2, it was surprising to observe  $\beta$  strands in the  $i$  and  $i + 1$  as well as  $i$  and  $i + 3$  categories. For the  $i$  and  $i + 1$  pairs, the residues often occur on the surface of the protein and near the end of a strand. The side chain of the anionic residue folds back on top of the Phe residue. Figure 6C shows an example of one such pair in 1VRM. This could potentially be important in “edge protection strategies” preventing amyloid formation.<sup>71</sup>

PSAIA was used to address the question of whether the BF pairs are buried or occur on the surface of the protein. This program calculates the average residue depth, defined as the distance in angstroms from the closest solvent accessible atom.<sup>72</sup> A value of 0 describes a fully accessible atom, while values greater than 0 describe buried atoms, with larger values describing more deeply buried atoms. Use of PSAIA allows automation of these calculations.<sup>73</sup> The average residue depths for each member of the pair were added and plotted against the total KM interaction energy, as shown in Figure S4 of the Supporting Information. The points are randomly scattered and do not show a trend between the depth of the roughly 17000 BF pairs and their interaction energy. However, as shown in Figure 7, a plot of the average residue depths for those 100 BF pairs possessing the most negative energies does show a preference for partial burial of the residues. As an electrostatic component of the anion– $\pi$  interaction exists, burial of the pair would minimize potential disruption of the anion– $\pi$  geometry by water.<sup>74</sup> Protection from solvent appears to be important as a recent computational analysis of cation– $\pi$  interactions indicates weaker energies in the presence of water.<sup>75</sup> In addition, some site-directed mutagenesis studies of surface cation– $\pi$  interactions in four different proteins indicate that cation– $\pi$  interactions are “at best weakly stabilizing and in some cases are clearly destabilizing” at room temperature.<sup>76</sup>

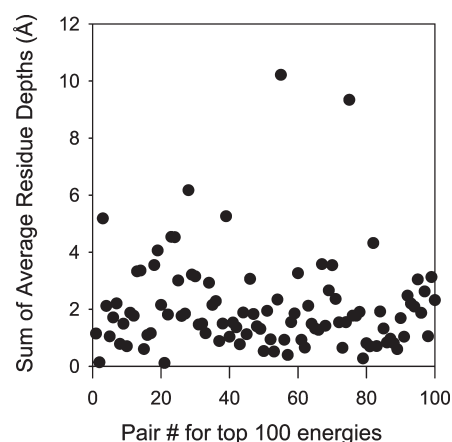
No apparent preference for Glu or Asp exists in their interaction with phenylalanine. The ratio of Asp to Glu involved in anion– $\pi$  pairs is very close to the ratio of Asp to Glu residues observed in our entire protein structure sample.

How often do BF pairs occur? We did not identify energetically significant pairs ( $-8$  to  $-2$  kcal/mol range) in 1357 PDB files. For the 3134 remaining PDB files, 8260 BF pairs were found, corresponding to approximately 2.6 pairs per structure. For comparison, Gallivan and Dougherty found an average of 1 energetically significant cation– $\pi$  interaction per 77 residues in a protein.<sup>7</sup>



**Figure 6.** Structural analysis of 8260 BF pairs possessing energies from  $-8$  to  $-2$  kcal/mol. In panel A, the primary sequence separation of Phe and Asp or Glu residues in the BF pair is shown. A separation of more than four amino acids is preferred. Separations of  $i$  and  $i + 2$  may indicate placement of the BF pair in a  $\beta$  strand, while separations of  $i$  and  $i + 3$  or  $i + 4$  may indicate the BF pair occurs in an  $\alpha$  helix. We were able to extract the secondary structure designation from a subset of 6934 pairs. Fractional preferences for these BF pairs are presented in panel B. Black denotes  $\alpha$  helix- $\alpha$  helix interactions. Green depicts  $\beta$  strand- $\beta$  strand interactions. Red depicts  $\alpha$  helix- $\beta$  strand interactions. Yellow denotes one or both of the structural elements is not designated. Panel C shows the structure of one  $\beta$  strand- $\beta$  strand  $i$  and  $i + 1$  interaction occurring in 1VRM between D50 and F51. The distance between the CD and OD1 atoms is 4.06 Å. Strands are colored yellow, helices red, and turns blue, and undesignated structure is colored white.

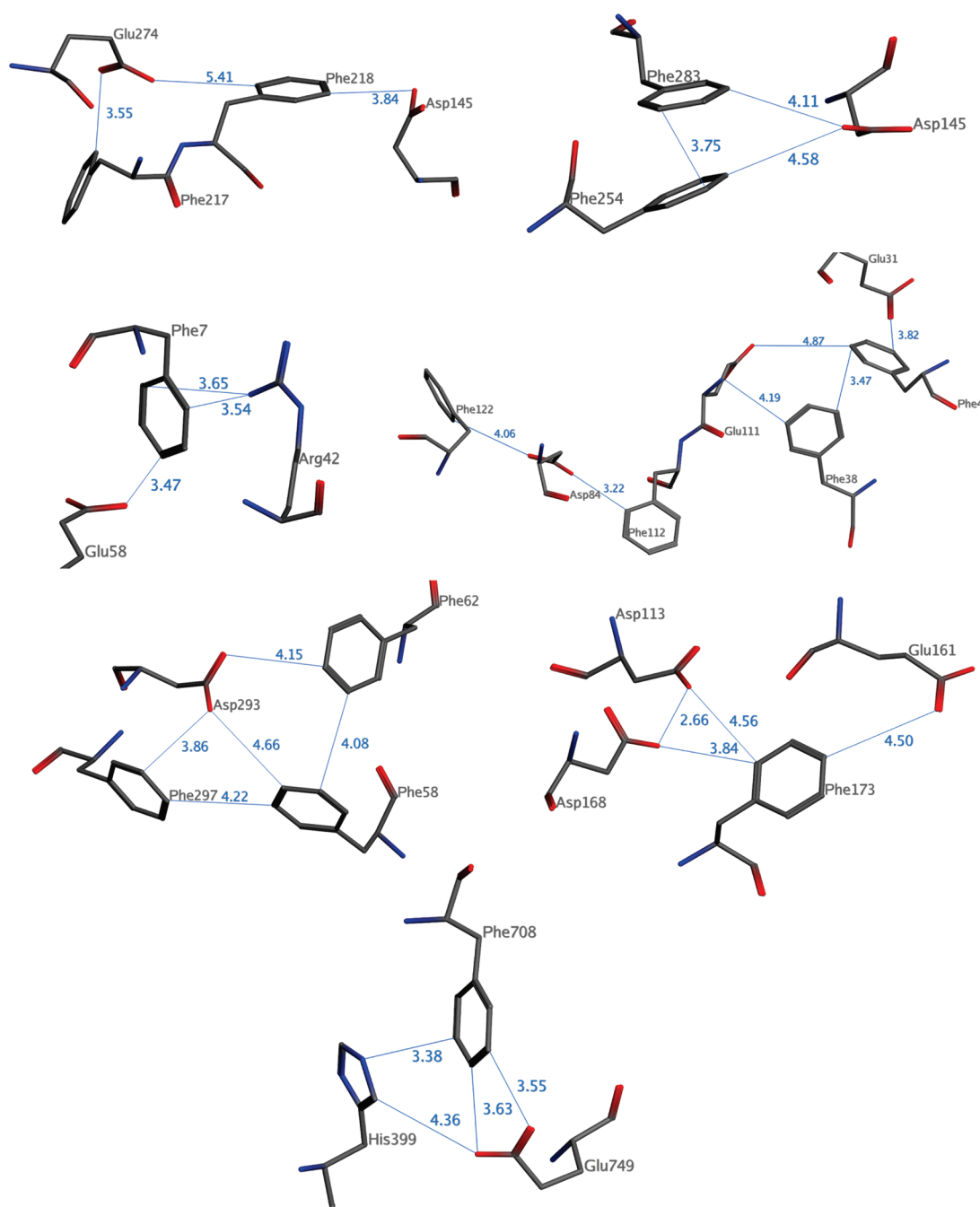
**Interesting Clusters.** Clusters of nonbonded interactions (involving more than one anion- $\pi$  pair) were found in the protein structures. Figure 8 shows several potentially interesting arrangements. For example, panels A and B show the presence of several aromatic groups surrounding one anion. This configuration was reasonably common as 365 clusters were identified with



**Figure 7.** Depth of the top 100 BF protein pairs calculated by PSAIAA.<sup>72,73</sup> The BF pairs were sorted according to their energies, and the top 100 are presented, with pair number 1 having the most negative energy ( $-7.27$  kcal/mol). The calculated energy for the 100th pair is  $-6.27$  kcal/mol. The average residue depths for the Phe and Asp or Glu residues were summed, and the sums are plotted. An additional plot for the sum of the average residue depth for Phe and Asp or Glu for the entire data set of roughly 17000 pairs was constructed and is shown in Figure S4 of the Supporting Information. No pattern was observed in the larger data set.

two phenylalanines interacting with one Glu or Asp. Panel B gives an example in which  $\pi$ - $\pi$  stacking may occur between two phenylalanines interacting with one Asp. Panel C shows an example of a cation- $\pi$  interaction occurring concurrently with an anion- $\pi$  interaction, and panel D shows a potential network of five BF pairs in 1EP0. Panel F shows that often several anions may cluster around an aromatic group. Because anion- $\pi$  pairs utilize polarization as an important component of their interaction energy, the stabilizing mechanism involved in this type of cluster is not clear; however, it may be that nearby carboxylates have altered  $pK_a$  values such that some of these residues are protonated. We identified 620 BF pairs that contain one phenylalanine and two Glu or Asp residues. Panel G shows a histidine near a BF pair. While histidine is aromatic and could engage in stacking interactions, it also could participate in a cation- $\pi$  interaction when it is protonated. The number of possible interactions expands if nearby Lys, Arg, and His residues are considered. As support for the context of the anion- $\pi$  interaction in the protein structure affecting the energetics of the system, we note recent quantum mechanical and crystallography studies of small molecules find synergistic effects between anion- $\pi$  interactions and hydrogen bonding networks and lone pair- $\pi$  CH- $\pi$ , cation- $\pi$ , and  $\pi$ - $\pi$  stacking interactions.<sup>77-80</sup> The interplay between these various elements appears to alter the overall energetics, often in a synergistic manner.

**CHARMM22 Calculations.** Figure 9 plots the CHARMM22 interaction energies for the BF functional groups versus the energies for the MP2 calculations. The correlations between CHARMM22 and ab initio functional group interaction energies range from 0.73 to 0.82, depending on the level of theory considered. The correlation coefficients between empirical and ab initio interaction energies are listed in Table 1. The highest correlation (0.82) is obtained when comparing force field calculations with the MP2 level of theory, suggesting that the attractive van der Waals interactions (implicitly included in force field parametrization, but not well-represented in HF and MK

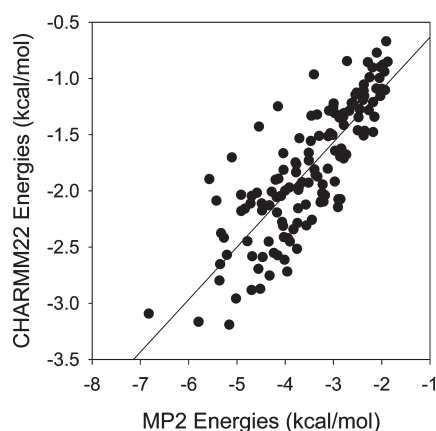


**Figure 8.** Interesting arrangements of anion– $\pi$  pairs. Panel A shows a series of interactions among D145, F217, F218, and E274 residues in 1AYX. In particular, E274 interacts with both phenylalanines. Panel B shows interactions among D145, F254, and F283 in 1XSZ. Here, ring stacking between the phenylalanines appears to be occurring. Panel C shows the juxtapositioning of R42 near the F7 and E58 anion– $\pi$  pair in 2FGQ. The R42 residue engages in a cation– $\pi$  interaction with F7, as per the web-based Capture program (<http://capture.caltech.edu/>), which predicts potential cation– $\pi$  interactions.<sup>7</sup> Polarization is predicted to contribute strongly to both anion– $\pi$  and cation– $\pi$  interactions, suggesting each type of interaction could enhance the other.<sup>7,12</sup> Panel D shows a long series of anion– $\pi$  interactions in 1EP0, involving F4, E31, F38, D84, E111, F112, and F122. Panel E depicts a cluster of three phenylalanines (F58, F62, and F297) interacting with D293 in 1VFL. Panel F illustrates the interactions among F173, D168, D113, and E161 in 1CPO. The rationale for the interaction of several carboxylates with one Phe is not as clear. It may be that a subset of the Asp or Glu residues are protonated. The atoms are colored as follows: gray for carbon, red for oxygen, and blue for nitrogen. Panel G shows the juxtapositioning of H399 near F708 and E749 in 1T3T. Depending on its ionization state, H399 could potentially either engage in ring stacking with F708 or form a cation– $\pi$  interaction with F708.

calculations) significantly contribute to the correlation between classical and quantum chemical models. As found previously,<sup>12</sup> CHARMM22 interactions were undervalued when compared

to MP2 interaction energies. No correlation was found between the angle and distance geometrical parameters between the interacting residues and the difference between MP2 and





**Figure 9.** MP2 vs CHARMM22 interaction energies, for functional groups. A subset of data from Figure 2 in the  $-8$  to  $-2$  kcal/mol energy range (138 BF pairs) was analyzed in CHARMM. A linear fit with a correlation coefficient of 0.82 was observed. The slope of the line is 0.47, indicating CHARMM can capture  $\sim 50\%$  of the MP2 interaction energy.

**Table 1.** Correlation Coefficients between Calculated Interaction Energies with Various Quantum Mechanical Treatments and CHARMM22

	HF	MP2	KM
functional groups	0.73	0.82	0.75
side chains <sup>a</sup>	0.61	0.68	0.65
HF		0.97	0.99
MP2			0.99

<sup>a</sup>After deletion of two pairs of residues of the initial 140 pairs, as described in Results.

CHARMM22 interaction energies. This suggests that the difference between CHARMM22 and MP2 energies is systematic and independent of the geometry of the anion- $\pi$  pairs.

Correlation coefficients between ab initio and CHARMM interaction energies calculated for all the side chains and  $\alpha$  carbons are significantly lower than those for the functional groups only. The side chain interaction energies range from 0.61 to 0.68, the highest correlation being between the empirical force field and MP2 level interaction energies. This was after the exclusion of two of the 140 pairs (the Phe337–Glu295 pair, from 1HLE, and the Phe388–Glu285 pair, from 1TQ4) that exhibited high repulsive interaction energies even after energy minimization of the hydrogen atoms. The highest correlation coefficient of 0.82 found here in the case of functional groups is close to the value of 0.89 obtained<sup>81</sup> when comparing CHARMM22 and MP2 interaction energies for 315 cation- $\pi$  pairs identified from the PDB.<sup>82</sup> It is not known whether the slightly lower correlation coefficient obtained in this study originates from intrinsic parametrization differences that would surface when comparing cation- $\pi$  or anion- $\pi$  interactions by the CHARMM forcefield, or if it is only an effect of the smaller number of pairs considered here. The correlation between CHARMM and ab initio interaction energies obtained from the entire side chain calculation (0.61–0.68), however, is significantly lower than that obtained in a cation- $\pi$  study,<sup>81</sup> which exhibited a correlation coefficient of 0.89 between CHARMM and MP2 interaction energies.

## DISCUSSION

The proposal that anion- $\pi$  interactions exist may seem surprising because at first consideration, this idea predicts an interaction between an electron-donating anion and an aromatic  $\pi$  cloud. However, the quadrupole moment associated with aromatic groups results in points near the ring edge possessing a positive electrostatic potential, while points above and below the ring display a negative electrostatic potential. Anions can thus favorably interact with the ring edge.<sup>12</sup>

**Biological Relevance of the Anion- $\pi$  Interaction.** This study identifies anion- $\pi$  interactions between Phe and Asp or Glu as reasonably common in protein structures. Approximately two energetically favorable Phe–Asp or Phe–Glu pairs, with energies in the  $-8$  to  $-2$  kcal/mol range as calculated by the KM energy decomposition analysis, occur per PDB file. As shown in Figure 4, to achieve these interaction energies, an angle dependence is required, with a strong preference for edgewise interactions. Although favorable stacking interactions are possible when the Asp or Glu residue is positioned above the aromatic ring of the Phe residue,<sup>11</sup> Figure S2 of the Supporting Information shows that in our database of structures, these geometries (with large  $\theta$  values) are less common than would be expected from a uniform distribution.

The predicted energies describe gas-phase calculations. In vivo, water and other protein atoms will be present, which might be expected to screen the overall interaction energy. However, a significant number of Phe–Asp or Phe–Glu pairs appear to be buried, as calculated by PSAIA (Figure 7 and Figure S4 of the Supporting Information), which would minimize the screening of anion- $\pi$  interactions for direct disruption of the anion- $\pi$  pairs by water molecules.

The overall interaction energies may potentially be modulated by the environment of the Phe–Asp or Phe–Glu pair in the protein. For example, while we focus on Phe–Asp and Phe–Glu pairs, nearby tyrosines, tryptophans, and neutral histidines could provide additional polarization, stacking effects, and/or hydrogen bonds. Nearby arginines, lysines, and protonated histidines could provide cation- $\pi$  interactions that could enhance the anion- $\pi$  interaction. Finally, other nearby anions and/or anion- $\pi$  pairs could either perturb  $pK_a$  values or form a network of interactions (see “interesting clusters” in Figure 8). While the anion- $\pi$  interaction may be weak, it is still significantly above  $k_B T$  (i.e.,  $\sim 0.6$  kcal/mol at 300 K) and can occur frequently, rendering its overall effect on protein structure significant. Also, when a large network of interactions is considered, cooperativity could result, either enhancing or diminishing the overall effect.

How do the energetics of the anion- $\pi$  and cation- $\pi$  interactions compare? We might predict that the edgewise nature of the anion- $\pi$  interaction, and the anisotropy of the polarizability tensor of aromatic rings, would favor the polarization term compared to that in the cation- $\pi$  interaction. In addition, both the electrostatic and polarization effects in the anion- $\pi$  interaction are most favorable for the edgewise approach of the ion. In contrast, for cation- $\pi$  interactions, the electrostatic component of the interaction is most favorable for perpendicular geometries, while the polarization component of the interaction is most favorable for edgewise approaches of the ion. These two elements suggest the energetics of anion- $\pi$  interactions should be at least similar in scope to those of cation- $\pi$  pairs. Using an OPLS force field, Gallivan and Dougherty found the average strength of a cation- $\pi$  interaction involving Lys was  $-3.3 \pm 1.5$



kcal/mol and involving Arg was  $-2.9 \pm 1.4$  kcal/mol.<sup>7</sup> Additionally, in studying the cation- $\pi$  interaction with a simplified benzene-ammonium pair, Aschi et al. predicted an energy of  $-4.4$  kcal/mol using a KM energy decomposition analysis. These values are clearly within the predicted range of the anion- $\pi$  interaction shown in Figure 4.

We can expand our thinking about anion- $\pi$  interactions to include protein-ligand binding. For example, numerous biologically relevant anions exist, such as DNA, RNA, phosphorylated proteins, ATP, NADPH, membrane bilayers, etc. If binding involves ion pair formation, binding specificity becomes an issue as other ions can compete. In contrast, neutral residues can provide contacts, steric constraints, and potentially greater ion selectivity by use of polarization during binding.<sup>83</sup> Both cation- $\pi$  and anion- $\pi$  interactions could participate in this fashion to facilitate both affinity and selectivity in binding of ions. Also of note, the desolvation penalty for formation of anion- $\pi$  and cation- $\pi$  pairs should not be as large as for ion pairs.<sup>84,85</sup> It will be interesting to extend this study to protein-ligand complexes and identify potential anion- $\pi$  interactions between protein receptors and their ligands.

**Computational Prediction Status.** The correlation between CHARMM22 and ab initio interaction energies (using the entire side chain) was 0.61–0.68, as shown in Figure 9. This result suggests that ab initio anion- $\pi$  interaction energies are reproduced correctly by the CHARMM empirical force field for basic functional groups, albeit with a magnitude that underestimates the ab initio results. Because ab initio or semiempirical calculations have not been performed for all the side chains for the anion- $\pi$  pairs studied here, it is not possible to definitively quantify how well force field parametrization reproduces anion-quadrupole interactions for molecules larger than functional groups. However, these preliminary results suggest that the force field calculations are able to correctly assign a global ranking of the relative interaction energies between different pairs, but they underestimate the ab initio anion- $\pi$  interaction energies.

While cation- $\pi$ , anion- $\pi$ , and CH- $\pi$  interactions as well as other interactions appear to be important to the formation of protein structure, stability, and dynamics, the impact of incorporating these results into protein structure prediction remains unclear, in particular with respect to their potential influence on the backbone structure and folding of proteins. Our previous theoretical studies of simplified phenylalanine-Glu or -Asp pairs found the charge-quadrupole term contributes between 30 and 45% of the total MP2 energy, and the rest of the interaction energy arises mostly from polarization contributions.<sup>12</sup> However, most force fields do not include explicit, “on-the-fly” polarization and multipole terms in the calculation of potential energies. While the global effect of polarization on the structure is included in the force field parametrization process, recent studies have either proposed or added polarization terms to force fields.<sup>86,87</sup> These new force fields are expected to improve the agreement between empirical force field and ab initio results for anion- $\pi$  interactions. It will be particularly interesting to see if an improved force field treatment of the anion- $\pi$  interactions will amplify the improving trend of protein structure predictions observed in the Critical Assessment of protein Structure Prediction (CASP) experiments.<sup>88</sup>

## ■ ASSOCIATED CONTENT

**S Supporting Information.** Plot of the KM mix term versus total energy (Figure S1), plot of the observed fractional

occurrence of benzene-formate pairs in the PDB as a function of the angle  $\theta$  compared to the expected occurrence calculated from volume (Figure S2), plot of the distance differences versus total interaction energy where the distance describes the difference between oxygen 1 (oxy1) and the benzene center of mass (CM) subtracted from the distance between oxygen 2 (oxy2) and the benzene CM and plotted as an absolute value (Figure S3), and plot of the the average residue depth of the BF pairs in the protein structures calculated using PSAIAA versus the interaction energy (Figure S4). This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, Knoxville, TN 37996-0840. Phone: (865) 974-4507. Fax: (865) 974-6306. E-mail: [lh@utk.edu](mailto:lh@utk.edu).

## ■ ACKNOWLEDGMENT

We thank Jordan Grubbs for construction of the images in Figure 8. J. H. was supported by an NSF IGERT grant (no. 0801540) entitled “Scalable Computing and Leading Edge Innovative Technologies (SCALE-IT) for Biology”.

## ■ ABBREVIATIONS

HF, Hartree-Fock; MP2, second-order Møller-Plesset energy calculations; BF, benzene-formate; STAAAR, C++ computer program that provides STatistical Analysis of Aromatic Rings; PDB, Protein Data Bank; OPLS, optimized potentials for liquid simulations; KM, Kitaura-Morokuma energy decomposition analysis; ES, electrostatic term; PL, polarization term; CT, charge transfer term; EX, exchange repulsion term; QM, quantum mechanical; MM, molecular mechanics treatment; CC, center of charge; CM, center of mass; BSSE, basis set superposition error.

## ■ REFERENCES

- (1) Brandl, M.; Weiss, M. S.; Jabs, A.; Suhnel, J.; and Hilgenfeld, R. (2001) CH- $\pi$  interactions in proteins. *J. Mol. Biol.* 307, 357–377.
- (2) Weiss, M. S.; Brandl, M.; Suhnel, J.; Pal, D.; and Hilgenfeld, R. (2001) More hydrogen bonds for the (structural) biologist. *Trends Biochem. Sci.* 26, 521–523.
- (3) Jiang, L., and Lai, L. (2002) CH...O hydrogen bonds at protein-protein interfaces. *J. Biol. Chem.* 277, 37732–37740.
- (4) Guo, H.; Beahm, R. F.; and Guo, H. (2004) Stabilization and destabilization of the C<sup>δ</sup>-H...O=C hydrogen bonds involving proline residues in helices. *J. Phys. Chem. B* 108, 18065–18072.
- (5) Ringer, A. L.; Senenko, A.; and Sherrill, C. D. (2007) Models of S/ $\pi$  interactions in protein structures: Comparison of the H<sub>2</sub>S-benzene complex with PDB data. *Protein Sci.* 16, 2216–2223.
- (6) Dougherty, D. A. (1996) Cation- $\pi$  interactions in chemistry and biology: A new view of benzene, Phe, Tyr, and Trp. *Science* 271, 163–168.
- (7) Gallivan, J. P., and Dougherty, D. A. (1999) Cation- $\pi$  interactions in structural biology. *Proc. Natl. Acad. Sci. U.S.A.* 96, 9459–9464.
- (8) Bartlett, G. J.; Choudhary, A.; Raines, R. T.; and Woolfson, D. N. (2010) n $\rightarrow\pi^*$  interactions in proteins. *Nat. Chem. Biol.* 6, 615–620.
- (9) Sinnokrot, M. O., and Sherrill, C. D. (2006) High-accuracy quantum mechanical studies of  $\pi$ - $\pi$  interactions in benzene dimers. *J. Phys. Chem. A* 110, 10656–10668.

- (10) Jurecka, P., Sponer, J., Cerny, J., and Hobza, P. (2006) Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* 8, 1985–1993.
- (11) Marsili, S., Chelli, R., Schettino, V., and Procacci, P. (2008) Thermodynamics of stacking interactions in proteins. *Phys. Chem. Chem. Phys.* 10, 2673–2685.
- (12) Jackson, M. R., Beahm, R., Duvvuru, S., Narasimhan, C., Wu, J., Wang, H. N., Philip, V. M., Hinde, R. J., and Howell, E. E. (2007) A preference for edgewise interactions between aromatic rings and carboxylate anions: The biological relevance of anion-quadrupole interactions. *J. Phys. Chem. B* 111, 8242–8249.
- (13) Quinonero, D., Garau, C., Rotger, C., Frontera, A., Ballester, P., Costa, A., and Deya, P. M. (2002) Anion- $\pi$  interactions: Do they exist?. *Angew. Chem., Int. Ed.* 41, 3389–3392.
- (14) Garau, C., Quinonero, D., Frontera, A., Ballester, P., Costa, A., and Deya, P. M. (2003) Dual binding mode of s-triazine to anions and cations. *Org. Lett.* 5, 2227–2229.
- (15) Garau, C., Frontera, A., Quinonero, D., Ballester, P., Costa, A., and Deya, P. M. (2003) A topological analysis of the electron density in anion- $\pi$  interactions. *ChemPhysChem* 4, 1344–1348.
- (16) Gallivan, J. P., and Dougherty, D. A. (1999) Can lone pairs bind to a  $\pi$  system? The water  $\cdots$  hexafluorobenzene interaction. *Org. Lett.* 1, 103–105.
- (17) Alkorta, I., Rozas, I., and Elguero, J. (1997) An Attractive Interaction Between the  $\pi$ -Cloud of C<sub>6</sub>F<sub>6</sub> and Electron-Donor Atoms. *J. Org. Chem.* 62, 4687–4691.
- (18) Mascal, M., Armstrong, A., and Bartberger, M. D. (2002) Anion-aromatic bonding: A case for anion recognition by  $\pi$ -acidic rings. *J. Am. Chem. Soc.* 124, 6274–6276.
- (19) Kim, D., Tarakeshwar, P., and Kim, K. (2004) Theoretical Investigations of Anion- $\pi$  Interactions: The Role of Anions and the Nature of  $\pi$  Systems. *J. Phys. Chem. A* 108, 1250–1258.
- (20) Danten, Y., Tassaing, T., and Besnard, M. (1999) On the Nature of the Water-Hexafluorobenzene Interaction. *J. Phys. Chem. A* 103, 3530–3534.
- (21) Garau, C., Frontera, A., Quinonero, D., Ballester, P., Costa, A., and Deya, P. (2004) Cation- $\pi$  versus anion- $\pi$  interactions: Energetic, charge transfer and aromatic aspects. *J. Phys. Chem. A* 108, 9423–9427.
- (22) Quinonero, D., Garau, C., Frontera, A., Ballester, P., Costa, A., and Deya, P. M. (2005) Structure and binding energy of anion- $\pi$  and cation- $\pi$  complexes; A comparison of MP2, RI-MP2, DFT and DF-DFT methods. *J. Phys. Chem. A* 109, 4632–4637.
- (23) Garau, C., Frontera, A., Quinonero, D., Ballester, P., Costa, A., and Deya, P. (2004) Cation- $\pi$  vs anion- $\pi$  interactions: A complete  $\pi$ -orbital analysis. *Chem. Phys. Lett.* 399, 220–225.
- (24) Garau, C., Frontera, A., Quinonero, D., Ballester, P., Costa, A., and Deya, P. M. (2004) Cation- $\pi$  versus anion- $\pi$  interactions: A comparative ab initio study based on energetic, electron charge density and aromatic features. *Chem. Phys. Lett.* 392, 85–89.
- (25) Garau, C., Frontera, A., Quinonero, D., Ballester, P., Costa, A., and Deya, P. M. (2004) Anion- $\pi$  interactions. *Recent Res. Dev. Chem. Phys.* 5, 227–255.
- (26) Rosokha, Y. S., Lindeman, S. V., Rosokha, S. V., and Kochi, J. K. (2004) Halide recognition through diagnostic “anion- $\pi$ ” interactions: molecular complexes of Cl<sup>−</sup>, Br<sup>−</sup>, and I<sup>−</sup> with olefinic and aromatic  $\pi$  receptors. *Angew. Chem., Int. Ed.* 43, 4650–4652.
- (27) de Hoog, P., Gamez, P., Mutikainen, I., Turpeinen, U., and Reedijk, J. (2004) An aromatic anion receptor: Anion- $\pi$  interactions do exist. *Angew. Chem., Int. Ed.* 43, 5815–5817.
- (28) Quinonero, D., Garau, C., Frontera, A., Ballester, P., Costa, A., and Deya, P. M. (2002) Counterintuitive interactions of anions with benzene derivatives. *Chem. Phys. Lett.* 359, 486–492.
- (29) Frontera, A., Saczewski, F., Gdaniec, M., Dziemidowicz-Borys, E., Kurland, A., Deya, P. M., Quinonero, D., and Garau, C. (2005) Anion- $\pi$  interactions in cyanuric acids: A combined crystallographic and computational study. *Chemistry* 11, 6560–6567.
- (30) Berryman, O. B., Hof, F., Hynes, M. J., and Johnson, D. W. (2006) Anion- $\pi$  interaction augments halide binding in solution. *Chem. Commun.*, 506–508.
- (31) Thomas, K. A., Smith, G. M., Thomas, T. B., and Feldmann, R. J. (1982) Electronic distributions within protein phenylalanine aromatic rings are reflected by the three-dimensional oxygen atom environments. *Proc. Natl. Acad. Sci. U.S.A.* 79, 4843–4847.
- (32) Burley, S. K., and Petsko, G. A. (1988) Weakly polar interactions in proteins. *Adv. Protein Chem.* 39, 125–189.
- (33) Duan, G., Smith, V. D., Jr., and Weaver, D. F. (2001) Characterization of aromatic-thiol  $\pi$ -type hydrogen bonding and phenylalanine-cysteine side chain interactions through *ab initio* calculations and protein database analyses. *Mol. Phys.* 99, 1689–1699.
- (34) Meyer, E. A., Castellano, R. K., and Diederich, F. (2003) Interactions with aromatic rings in chemical and biological recognition. *Angew. Chem., Int. Ed.* 42, 1210–1250.
- (35) Singh, J., and Thornton, J. (1992) *Atlas of Protein Side-Chain Interactions*, Vols. 1 and 2, IRL Press at Oxford University Press, Oxford, U.K.
- (36) Shi, Z., Olson, C. A., and Kallenbach, N. R. (2002) Cation- $\pi$  interaction in model  $\alpha$ -helical peptides. *J. Am. Chem. Soc.* 124, 3284–3291.
- (37) Olson, C. A., Shi, Z., and Kallenbach, N. R. (2001) Polar interactions with aromatic side chains in  $\alpha$ -helical peptides: CH $\cdots$ O H-bonding and cation- $\pi$  interactions. *J. Am. Chem. Soc.* 123, 6451–6452.
- (38) Shi, Z., Olson, C. A., Bell, A. J., Jr., and Kallenbach, N. R. (2002) Non-classical helix-stabilizing interactions: C-H $\cdots$ O H-bonding between Phe and Glu side chains in  $\alpha$ -helical peptides. *Biophys. Chem.* 101–102, 267–279.
- (39) Joughin, B. A., Green, D. F., and Tidor, B. (2005) Action-at-a-distance interactions enhance protein binding affinity. *Protein Sci.* 14, 1363–1369.
- (40) Soga, S., Shirai, H., Kobori, M., and Hirayama, N. (2007) Use of amino acid composition to predict ligand binding sites. *J. Chem. Inf. Model.* 47, 400–406.
- (41) Koide, S., and Sidhu, S. S. (2009) The importance of being tyrosine: Lessons in molecular recognition from minimalist synthetic binding proteins. *Chem. Biol.* 4, 325–334.
- (42) Fellouse, F. A., Barthelemy, P. A., Kelley, R. F., and Sidhu, S. S. (2006) Tyrosine plays a dominant functional role in the paratope of a synthetic antibody derived from a four amino acid code. *J. Mol. Biol.* 357, 100–114.
- (43) Fellouse, F. A., Esake, K., Birtalan, S., Raptis, D., Cancasci, V. J., Koide, A., Jhurani, P., Vasser, M., Koide, S., and Sidhu, S. S. (2007) High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries. *J. Mol. Biol.* 373, 924–940.
- (44) Gorteau, V., Bollot, G., Mareda, J., Perez-Velasco, A., and Matile, S. (2006) Rigid Oligonaphthalenediimide Rods as Transmembrane Anion- $\pi$  Slides. *J. Am. Chem. Soc.* 128, 14788–14789.
- (45) Ioanoviciu, A., Mehareenna, Y. T., Poulos, T. L., and Ortiz de Montellano, P. R. (2009) DevS Oxy Complex Stability Identifies This Heme Protein as a Gas Sensor in *Mycobacterium tuberculosis* Dormancy. *Biochemistry* 48, 5839–5848.
- (46) Juszczak, L. J., and Desamero, R. Z. B. (2009) Extension of the Tryptophan  $\chi_{2,1}$  Dihedral Angle-W3 Band Frequency Relationship to a Full Rotation: Correlations and Caveats. *Biochemistry* 48, 2777–2787.
- (47) Dharmarajan, L., Case, C. L., Dunten, P., and Mukhopadhyay, B. (2008) Tyr235 of human cytosolic phosphoenolpyruvate carboxykinase influences catalysis through an anion-quadrupole interaction with phosphoenolpyruvate carboxylate. *FEBS J.* 275, 5810–5819.
- (48) Sartorius, J., and Schneider, H. J. (1995) NMR titrations with complexes between ds-DNA and indole derivatives including tryptophane containing peptides. *FEBS Lett.* 374, 387–392.
- (49) Hobza, P. (2008) Stacking interactions. *Phys. Chem. Chem. Phys.* 10, 2581–2583.
- (50) Hobohm, U., and Sander, C. (1994) Enlarged representative set of protein structures. *Protein Sci.* 3, 522–524.

- (51) Hobohm, U., Scharf, M., Schneider, R., and Sander, C. (1992) Selection of representative protein data sets. *Protein Sci.* 1, 409–417.
- (52) Schuettelkopf, A. W., and van Aalten, D. M. F. (2004) PRODRG: A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D* 60, 1355–1363.
- (53) Pace, C. N., Grimsley, G. R., and Scholtz, J. M. (2009) Protein ionizable groups: pK values and their contribution to protein stability and solubility. *J. Biol. Chem.* 284, 13285–13289.
- (54) Shah, A. V., Walters, W. P., Shah, R., and Dolata, D. P. (1994) BABEL: A tool for converting between molecular structural data formats. In *Computerized Chemical Data Standards: Databases, Data Interchange, and Information Systems*, pp 45–53, ASTM, Philadelphia.
- (55) Schmidt, M., Baldridge, K., Boatz, J., Elbert, S., Gordon, M., Jensen, J., Koseki, S., Matsunaga, N., and Nguyen, K. (1993) et al. General atomic and molecular electronic structure system. *J. Comput. Chem.* 14, 1347–1363.
- (56) Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Zakrzewski, V. G., Montgomery, J. A., Jr., Stratmann, R. E., Burant, J. C., Dapprich, S., Millam, J. M., Daniels, A. D., Kudin, K. N., Strain, M. C., Farkas, O., Tomasi, J., Barone, V., Cossi, M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Petersson, G. A., Ayala, P. Y., Cui, Q., Morokuma, K., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Cioslowski, J., Ortiz, J. V., Baboul, A. G., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Gomperts, R., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Gonzalez, C., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Andres, J. L., Gonzalez, C., Head-Gordon, M., Replogle, E. S., and Pople, J. A. (1998) *Gaussian 03*, Gaussian, Inc., Pittsburgh, PA.
- (57) Kitauro, K., and Morokuma, K. (1976) New energy decomposition scheme for molecular interactions within Hartree-Fock approximation. *Int. J. Quantum Chem.* 10, 325–340.
- (58) Ishida, K., Morokuma, K., and Komornicki, A. (1977) The intrinsic reaction coordinate. An ab initio calculation for  $\text{HNC} \rightarrow \text{HCN}$  and  $\text{H}^- + \text{CH}_4 \rightarrow \text{CH}_3 + \text{H}^-$ . *J. Chem. Phys.* 66, 2153–2156.
- (59) Aschi, M., Mazza, F., and Di Nola, A. (2002) Cation- $\pi$  interactions between ammonium ion and aromatic rings: An energy decomposition study. *J. Mol. Struct.* 587, 177–188.
- (60) Umeyama, H., and Morokuma, K. (1977) The origin of hydrogen bonding. An energy decomposition study. *J. Am. Chem. Soc.* 99, 1316–1332.
- (61) Dunning, T. H., Jr. (1989) Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* 90, 1007–1023.
- (62) Kendall, R. A., Dunning, T. H., Jr., and Harrison, R. J. (1992) Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* 96, 6796–6806.
- (63) Boys, S., and Bernardi, F. (1970) The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* 19, 553–566.
- (64) MacKerell, A. D., Jr., Bashford, D., Bellott, M., Dunbrack, R. L., Jr., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kucera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., III, Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102, 3586–3616.
- (65) Walters, P., and Stahl, M. (1994) *BABELwin*, Department of Chemistry, University of Arizona, Tucson, AZ (modified by J. Gosper at Brunel University, Uxbridge, West London, U.K.) (<ftp://joplin.biosci.arizona.edu/pub/Babel/>).
- (66) Chou, P. Y., and Fasman, G. D. (1974) Conformational parameters for amino acids in helical,  $\beta$ -sheet, and random coil regions calculated from proteins. *Biochemistry* 13, 211–222.
- (67) Chou, P. Y., and Fasman, G. D. (1974) Prediction of protein conformation. *Biochemistry* 13, 222–245.
- (68) Pokkuluri, P. R., Gu, M., Cai, X., Raffin, R., Stevens, F. J., and Schiffer, M. (2002) Factors contributing to decreased protein stability when aspartic acid residues are in  $\beta$ -sheet regions. *Protein Sci.* 11, 1687–1694.
- (69) Niwa, T. S., and Ogino, A. (1997) Multiple regression analysis of the  $\beta$ -sheet propensity of amino acids. *THEOCHEM* 419, 155–160.
- (70) Smith, C. K., and Regan, L. (1997) Construction and design of  $\beta$ -sheets. *Acc. Chem. Res.* 30, 153–161.
- (71) Richardson, J. S., and Richardson, D. C. (2002) Natural  $\beta$ -sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl. Acad. Sci. U.S.A.* 99, 2754–2759.
- (72) Pinter, A., Carugo, O., and Pongor, S. (2002) CX, an algorithm that identifies protruding atoms in proteins. *Bioinformatics* 18, 980–984.
- (73) Mihel, J., Sikic, M., Tomic, S., Jeren, B., and Vlahovicek, K. (2008) PSAIA: Protein structure and interaction analyzer. *BMC Struct. Biol.* 8, 21.
- (74) Fletterick, R. J., Schroer, T., and Matela, R. J. (1985) *Molecular Structure, Macromolecules in 3D*, Blackwell Scientific Publications, Oxford, U.K.
- (75) Xu, Y., Shen, J., Zhu, W., Luo, X., Chen, K., and Jiang, H. (2005) Influence of the Water Molecule on Cation- $\pi$  Interaction: Ab Initio Second Order Møller-Plesset Perturbation Theory (MP2) Calculations. *J. Phys. Chem. B* 109, 5945–5949.
- (76) Prajapati, R. S., Sirajuddin, M., Durani, V., Sreeramulu, S., and Varadarajan, R. (2006) Contribution of Cation- $\pi$  Interactions to Protein Stability. *Biochemistry* 45, 15000–15010.
- (77) Quinonero, D., Dey, P. M., Carranza, M. P., Rodriquez, A. M., Jalon, F. A., and Manzano, B. R. (2010) Experimental and computational study of the interplay between C-H/ $\pi$  and anion- $\pi$  interactions. *Dalton Trans.* 39, 794–806.
- (78) Lucas, X., Estarellas, C., Escudero, D., Frontera, A., Quinonero, D., and Dey, P. M. (2009) Very long-range effects: Cooperativity between anion- $\pi$  and hydrogen-bonding interactions. *ChemPhysChem* 10, 2256–2264.
- (79) Quinonero, D., Frontera, A., Garau, C., Ballester, P., Costa, A., and Dey, P. M. (2006) Interplay between cation- $\pi$ , anion- $\pi$  and  $\pi$ - $\pi$  interactions. *ChemPhysChem* 7, 2487–2491.
- (80) Das, A., Choudhury, S. R., Dey, B., Yalamanchili, S. K., Helliwell, M., Gamez, P., Mukhopadhyay, S., Estarellas, C., and Frontera, A. (2010) Supramolecular assembly of Mg(II) complexes directed by associative lone pair- $\pi$ / $\pi$ - $\pi$ / $\pi$ -anion- $\pi$ / $\pi$ -lone pair interactions. *J. Phys. Chem. B* 114, 4998–5009.
- (81) Gilis, D., Biot, C., Buisine, E., Dehouck, Y., and Rooman, M. (2006) Development of novel statistical potentials describing cation- $\pi$  interactions in proteins and comparison with semiempirical and quantum chemistry approaches. *J. Chem. Inf. Model.* 46, 884–893.
- (82) Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J. D., and Zardecki, C. (2002) The protein data bank. *Acta Crystallogr. D* 58, 899–907.
- (83) Atwood, J., and Steed, J. (1997) Structural and topological aspects of anion coordination. In *Supramolecular Chemistry of Anions* (Bianchi, A., Bowman-James, K., and Garcia-Espana, E., Eds.) pp 147–215, Wiley-VCH, New York.
- (84) Levy, Y., and Onuchic, J. N. (2006) Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.* 35, 389–415.
- (85) Chong, L. T., Dempster, S. E., Hendsch, Z. S., Lee, L. P., and Tidor, B. (1998) Computation of electrostatic complements to proteins: A case of charge stabilized binding. *Protein Sci.* 7, 206–210.
- (86) Ponder, J. W., Wu, C., Ren, P., Pande, V. S., Chodera, J. D., Schnieders, M. J., Haque, I., Mobley, D. L., Lambrecht, D. S., DiStasio, R. A., Jr., Head-Gordon, M., Clark, G. N. I., Johnson, M. E., and Head-Gordon, T. (2010) Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* 114, 2549–2564.
- (87) Lopes, P. E. M., Roux, B., and MacKerell, A. D., Jr. (2009) Molecular modeling and dynamics studies with explicit inclusion of

electronic polarizability. Theory and applications. *Theor. Chem. Acc.* 124, 11–28.

(88) Moult, J., Fidelis, K., Kryshchuk, A., Rost, B., and Tramontano, A. (2009) Critical assessment of methods of protein structure prediction—Round VIII. *Proteins* 77 (Suppl. 9), 1–4.