HHS Public Access

Author manuscript

Biochemistry. Author manuscript; available in PMC 2019 March 27.

Published in final edited form as:

Biochemistry. 2018 March 27; 57(12): 1852–1867. doi:10.1021/acs.biochem.7b01006.

A Comprehensive Analysis of Anion–Quadrupole Interactions in Protein Structures

Suvobrata Chakravarty[*,†,§], **Adron R. Ung**[#†], **Brian Moore**[#©], **Jay Shore**[#†], and **Mona Alshamrani**[†]

[†]Chemistry & Biochemistry, South Dakota State University, Brookings, SD, USA, 57007.

[§]BioSNTR, Brookings, SD, USA, 57007.

[©]University Networking and Research Computing, South Dakota State University, Brookings, SD, USA, 57007.

[#] These authors contributed equally to this work.

Abstract

The edgewise interactions of anions with phenylalanine (Phe) aromatic rings in proteins, known as anion–quadrupole interactions, have been well studied. However, the anion–quadrupole interactions of the tyrosine (Tyr) and tryptophan (Trp) rings have been less well studied, probably because these have been considered weaker interactions than anions hydrogen bonded to Trp/Tyr sidechains. Distinguishing from such hydrogen bonding interactions, we comprehensively surveyed the edgewise interactions of certain anions (aspartate, glutamate, and phosphate) with Trp, Tyr, and Phe rings in high-resolution, non-redundant protein single chains and interfaces (protein-protein, DNA/RNA–protein, and membrane–protein). Trp/Tyr anion–quadrupole interactions are common, with Trp showing the highest propensity and average interaction energy for this type of interaction. The energy of an anion–quadrupole interaction (−15.0 −0.0 kcal/mol, based on quantum mechanical calculations) depends not only on the interaction geometry but also on the ring atom. The phosphate anions at DNA/RNA–protein interfaces interact with aromatic residues with energies comparable to that of aspartate/glutamate anion–quadrupole interactions. At DNA–protein interfaces, the frequency of aromatic ring participation in anion–quadrupole interactions is comparable to that of positive charge participation in salt bridges, suggesting an underappreciated role for anion–quadrupole interactions at DNA–protein (or membrane–protein) interfaces. Although less frequent than salt bridges in singlechain proteins, we observed highly conserved anion–quadrupole interactions in the structures of remote homologs, and evolutionary covariance-based residue contact score predictions suggest that conserved anion–quadrupole interacting pairs, like salt bridges, contribute to polypeptide folding, stability, and recognition.

[*]**Corresponding Author**: suvobrata.chakravarty@sdstate.edu, Tel: 1-605-688-5694; Fax: +1-605-688-6364, Box-2202 SAV367, Department of Chemistry & Biochemistry, South Dakota State University, Brookings, SD, USA, 57007.

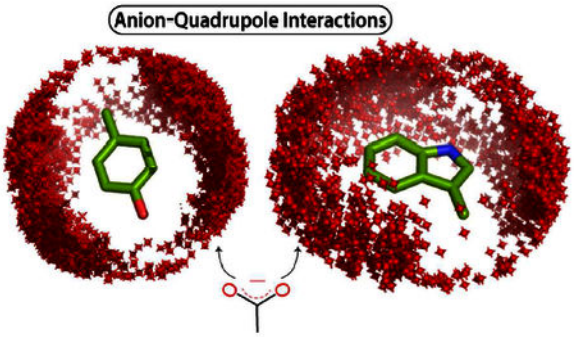
The authors declare no competing financial interest.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of authors and do not necessarily reflect the views of the National Science Foundation and National Institutes of Health.

Supporting Information (SI)

The supporting information (SI) attached with the text contains: (i) Extended experimental section, (ii) Table S1, and (iii) Supplemental Figures S1 – S12.

TOC GRAPHICS



INTRODUCTION

Non-covalent interactions, such as Van der Waals forces, the hydrophobic effect, hydrogen bonding, and electrostatic interactions, may be weak individually, but collectively play an important role in the folding, assembly, and recognition of macromolecules[1-3]. Interactions involving aromatic functional groups (commonly known as π interactions), such as π–π[4–6], CH–π[713], lone pair–π[14–17], cation–π[18–22], and anion–π[22–25] interactions, are important components of weak, non-covalent interactions and have gained considerable interest in chemistry[26], for example, in the design of catalysts. In addition to synthetic chemistry, a number of recent reports suggest that weak interactions are critical in regulating protein function. For example, the CH–π switch in protein-tyrosine phosphatase 1B (PTP1B, also known as PTPN1) is critical in regulating phosphatase function[27]. Similarly, the repair function of the DNA glycosylase AlkD is critically dependent on CH–π interactions for the recognition of N3 methyladenosine in alkylated DNA[28]. Enhanced stability of engineered miniproteins[29] and G-quadruplexes[30] are other notable examples of CH–π interactions, not just in proteins but also in nucleic acids. While CH–π interactions are gaining prominence in the field of biological macromolecules, investigation of another potentially important non-covalent interaction in biological macromolecules, the anion–quadrupole interaction, has been rare, possibly due to the only recent recognition of its presence in proteins in the seminal study of Howell et al.[25]. In general, aromatic ring π-systems have a quadrupole moment as a result of the two opposing dipoles originating from each face[31]. The electron cloud on each face of the aromatic ring is therefore attracted to cations (by a cation–π interaction), and the edges of the ring are attracted to anions (by an anion–quadrupole interaction). In the same study[25], anion–quadrupole interactions in proteins were modeled on the interactions between aspartate (Asp) or glutamate (Glu) sidechain carboxylate ions and the aromatic ring of the phenylalanine (Phe) residue in monomeric proteins. Anions could also interact with tryptophan and tyrosine ring-systems by anion–quadrupole interaction but these ring systems also have dipole moment and they make strong hydrogen bonds with anions that could influence their anion–quadrupole interactions. In general, we wanted to further understand the roles of anion–quadrupole interactions in proteins in the following categories, which were not discussed: (i) anion–quadrupole interactions involving the rings of tyrosine (Tyr) and tryptophan (Trp) residues; (ii) deconvolution of the energetic contributions of anion–quadrupole and anion-dipole interactions (iii) anion–quadrupole

interactions involving the phosphate ion, an important anion in biological systems; (iv) anion–quadrupole interactions at protein–protein interfaces; (v) a comparison of the behavior of anion–quadrupole interactions with that of salt bridges in proteins and at protein–protein interfaces; and (vi) the evolutionary conservation of anion–quadrupole interactions in proteins. As salt bridges in protein structures have been well studied[32–35], the behavior of anions in anion–quadrupole interactions has been compared with their behavior in salt bridges to gain a better understanding of how nature, through evolutionary conservation, utilizes anions in a range of polar interactions. Since the report on anion–quadrupole interactions in proteins referred to above, there have been other reports[36–38]. But these have been typically focused on a small set of a specific group/family of highly homologous proteins (or predicted interfaces) and do not cover all of the abovementioned categories. In addition, unlike the recent reports[36–38], we distinguish an anion's edgewise interactions with Trp/Tyr ring from that of an anion's hydrogen bonded interactions with Trp/Tyr sidechains to capture the influence of different ring atoms on anion–quadrupole interactions, and the orientation preferences of aromatic rings while interacting with anions. Therefore, we systematically performed a comprehensive analysis of anion–quadrupole interactions in proteins that should also be useful in the study of other weak, non-covalent interactions in biological macromolecules. As novel weak interactions (e.g., RNA ribose–nucleobase lone pair–π stacking[39], Lys-NH_3^+ − carbonyl linear arrangement[40]) in biological macromolecules continue to be reported, our analysis will be useful in the study of weak molecular interactions in general.

METHODS

Detection Procedure for Anion–quadrupole Interactions.

The procedure for detecting anion–quadrupole interactions in protein structures was developed using the PSD dataset (*see* Protein Datasets, SI page S3–S4). Residues with missing sidechain atoms were ignored. Sidechain–sidechain hydrogen bonded pairs of Asp/Glu and Trp/Tyr were not counted as *regular* anion–quadrupole interactions. We determined the distance-and-angle criterion for anion–quadrupole interactions in the following steps: (a) defined the vector $\mathbf{n_r}$ as normal to the ring plane and passing through the ring center (Figure S1A) for each of the aromatic residues in a coordinate file; (b) defined the vector $\mathbf{r_c}$ as the line joining the ring center and the center of carboxylate oxygen atoms of Asp/Glu sidechains (Figure S1A) for every anion-aromatic residue pair in the coordinate file; (c) determined $\boldsymbol{\theta_N}$ the angle between $\mathbf{n_r}$ and $\mathbf{r_c}$, and $\boldsymbol{\theta_P}$ (90° - $\boldsymbol{\theta_N}$), the angle the anion makes with the ring plane (Figure S1A); (d) for each anion–aromatic residue pair, determined all interresidue interatomic vectors $\mathbf{d_r}$ (the vector between the carboxylate oxygen and a ring atom); and (e) determined $\mathbf{d_S}$, the smallest value of $\mathbf{d_r}$, that is, the shortest interresidue interatomic distance (Figure S1A). The distribution of $\boldsymbol{\theta_P}$ for $\mathbf{d_S}$ < 5 Å (greater than the sum of van der Waal contact + 1.75 Å) showed that $\boldsymbol{\theta_\rho}$ decreases from 0° to 90°, as expected. Seventy-five percent of the population of $\boldsymbol{\theta_P}$ values for Trp/Tyr/Phe, taken together, was < 35° (Figure S1b). The Howell study[25], using quantum mechanical calculations on a benzene-formate (BF) pair, determined the interaction energies to be negative for $\boldsymbol{\theta_P}$ < 35°. Therefore, for determining a distance criterion, we considered $\boldsymbol{\theta_P} \leq$ 35°, and the distribution of $\mathbf{d_S}$ for $\boldsymbol{\theta_P} \leq$ 35° (Figure S1C, D) showed a bell-shaped

appearance with a peak at 3.75 Å for $\mathbf{d_S} \leq 4.5$ Å. For $\mathbf{d_S} > 4.5$ Å, the non-specific pairs kept increasing (Figure S1C, D). This behavior was similar for Trp, Tyr, and Phe residues (Figure S1C, D). Based on the distribution of $\mathbf{d_S}$ for Trp/Tyr/Phe taken together, we set a 4.5-Å cutoff for $\mathbf{d_S}$ for the detection of anion–quadrupole interactions. Thus, a pair of (Asp/Glu)–(Trp/Tyr/Phe) residues with $\mathbf{d_S} \leq 4.5$ A and $\boldsymbol{\theta_P} \leq 35°$, as well as not having a sidechain–sidechain hydrogen bond between them, was considered a *regular* anion–quadrupole interaction (Figure 1A, *see* main text). We also added two additional criteria: (i) if the $\mathbf{d}_S$ for a carboxylate ion was > 4.5 Å and the $\mathbf{d}_r$ distance from the anion center was ≤ 4.5 Å, there was a ~13% increase in the number of anion–quadrupole pairs; and (ii) if the $\boldsymbol{\theta_P}$ value was > 35° and the angle with either of the oxygen atoms was ≤ 35°, there was a ~7% increase in anion–quadrupole interactions. We ignored the C-terminal carboxyl group (except for the PDZ–peptide interaction, Figure S5D).

Energy Calculations.

The energy of interactions was computed using NWChem[41] 6.6 quantum mechanical calculations. For computing the energy of interaction (e.g., for an anion–quadrupole pair) the coordinates of side chain atoms of the interacting pair were first extracted from a protein *d*ata *b*ank (PDB) file. Hydrogen atoms for these side chains were added using the Open Babel version 2.3.1 chemical toolbox[42]. Each CB atom had three protons and setting the pH to 7.0 (in Open Babel) ensured that Asp/Glu carboxyl groups remained deprotonated. An interacting pair of the type **A–B** (e.g., Asp–Phe) had three input files: (i) **A**, (ii) **B**, and (iii) **A–B**, where the **A** input file has the coordinates of the atoms of **A**. Each of the three input files for a pair was imported to the NWChem 6.6 program for energy calculations. Single-point energies were calculated using DFT with the Grimme DFT-D3 dispersion correction[43] with the AUG-CC-PVDZ basis set and B3LYP functional. The energy (in E_h units) was extracted from the corresponding NWChem log files. The energy of interaction ($\Delta E_{\mathbf{AB}}$) was obtained as, $\Delta E_{AB} = E(\mathbf{A}\text{-}\mathbf{B}) - [E(\mathbf{A}) + E(\mathbf{B})]$, and $\Delta E_{\mathbf{AB}}$ in E_H units was then converted to kcal/mole. There were 13,308 anion–quadrupole and 6,256 HB-anion–quadrupole **A–B** pairs in total for the energy calculations. We also compared the interaction energy of aromatic ring-carboxylate anion–quadrupole to that of other types of ring-edgewise interactions (see Figure 8A). For example, we considered the edgewise interactions of aromatic rings with side chain nonpolar $-CH_2-/-CH_3$ groups (Phe-Val), polar -OH groups (Phe-Ser), polar $-CONH_2$ groups (Phe-Asn) and the polar nucleic acid backbone anion (Phe-$[-O-PO_2-O^-]^-$). There were 4399 Phe–Val, 1079 Phe–Ser, 1037 Phe–Asn, and 176 (Phe-$[-O-PO_2-O^-]^-$) **A–B** pairs in total for the energy calculations. Energies calculated using only the dispersion corrected DFT (computed with MWChem 6.6) method were used in this study (e.g., Figures 1–4 and 7). Single-point energies were also calculated using DFT with the B3LYP functional and the 6–31G* [i.e., 6–31(d)] basis set using Gaussian 09[44], but these energies were not used here. A correlation between the anion–quadrupole energies calculated by either approach is shown in Figure S12.

Hydrogen Bonds and Salt Bridges.

All hydrogen bond calculations were carried out using the HBPLUS[45] program with the default settings. We considered a salt bridge pair to occur when Lys/Arg side chain atoms

were hydrogen bonded to Asp/Glu side chain atoms. That is, electrostatically interacting opposite charges that were not hydrogen bonded were not considered here as salt bridges. Although this criterion underestimated the number of salt bridges, a “strict” criterion was imposed to ensure a fair comparison with anion–quadrupole interactions, which were also selected with a strict criterion. Similarly, phosphoryl group salt bridges were considered to occur when Lys/Arg side chain atoms were hydrogen bonded to a phosphoryl group.

RESULTS AND DISCUSSION

Anion–quadrupole Interactions of Tryptophan and Tyrosine.

Using highresolution (≤ 1.6 Å) crystal structures, a simple procedure was developed for detecting anion-quadrupole interactions in proteins (Figure S1 and *see* Methods). We considered the concept of an anion–quadrupole interacting pair to apply only when the following conditions are satisfied (Figure 1A): (a) one of the Asp/Glu carboxylate oxygen atoms is within 4.5 A of any of the ring atoms, (b) the center of carboxylate oxygen atoms subtends an angle ≤ 35° with the ring plane, and (c) a pair satisfying (a) and (b) should not also be engaged in sidechain-sidechain hydrogen bonding between them. The hydrogen bonded pairs Trp/Tyr (side chain donor) and Asp/Glu (side chain acceptor) were excluded (Figure S1A, C). Condition (c) ensured that the treatment of Trp/Tyr was similar to that of Phe, and the distance and angle distributions of Tyr and Trp were very similar to that of Phe (Figure S1B, D). Therefore, in the same manner as the Phe–Asp/Glu anion–quadrupole pairs were studied in the previously mentioned report[25], Tyr and Trp were examined here for anion–quadrupole interactions. We observed 13,308 anion–quadrupole interactions in 3700 non-redundant single chains of the PISCES[46] dataset. The number of anion–quadrupole interactions with Phe (~2.06 per protein) reported in the earlier study[25] is higher than we observed here (~1.55 per protein), as we implemented a “strict” distance-and-angle detection criterion. Among the anion–quadrupole interactions, 43%, 36%, and 21% were respectively associated with Phe, Tyr, and Trp residues. However, the normalized occurrence frequencies (propensities, Figure S2A) were 1.38 (Trp), 0.98 (Phe), and 0.88 (Tyr), suggesting that a larger surface area and/or the indole ring dipole moment of Trp likely contributes to the interaction energy and therefore a higher propensity.

A typical upper limit for the hydrogen bond donor-acceptor distance is 3.5 A, while the distance range for the anion–quadrupole interactions is 3.25 – 4.5 A. This suggests that carboxylateions making hydrogen bonds with tryptophan (through the NE1 atom) and tyrosine (through the OH atom) will be positioned closer to the ring atoms than do the carboxylate ions engaged only in *regular* anion–quadrupole interactions (Figure S1E). This shorter distance and the high electronegativity of the NE1 and OH atoms would obviously make the interaction energy between an anion and an aromatic ring through hydrogen bonds much greater than that of a *regular* anion–quadrupole. This energy difference is consistent with the following. In the dataset, we observed 6,256 anions hydrogen bonded to tryptophan and tyrosine (i.e., with just two atoms, Trp-NE1 and Tyr-OH). By comparison, in the same dataset, we observed 7,585 anions engaged in anion–quadrupole interactions with tryptophan and tyrosine residues in which 12 ring atoms are capable of participating. If the anions hydrogen bonded to Trp-NE1 and Tyr-OH were included, the distance distribution

(Figure S1E) of anion–quadrupole interactions would also have been different. Therefore, for determining the distance-angle criterion for detecting anion–quadrupole interactions, we opted for a criterion that is uniform for all the aromatic rings (Phe, Tyr and Trp). For a fair comparison of anion–quadrupole interaction energies between phenylalanine, tyrosine and tryptophan, we (unlike recent reports[36–38]) avoided anions hydrogen bonded to Trp-NE1 and Tyr-OH atoms. A number of structural parameters (see Supplemental Information, SI, pp. S4–S6) of the anion–quadrupole interactions of Phe residues were compared with those of Trp and Tyr residues (Figure S3). These parameters were: (a) the specific contacts, (b) the interplanar angle, (c) the presence of hydrogen bonding for anions, (d) the depth and packing, (e) the energy, and (f) the sequence separation (Figure S3A–F). Consistent with our approach of treating tryptophan and tyrosine like phenylalanine, a comparison of structural parameters showed that the behavior of tryptophan or tyrosine anion–quadrupole interactions is similar to that of phenylalanine (Figure S3), and therefore the anion–quadrupole interactions discussed here include those associated with all three aromatic residues.

Energies of Anion–quadrupole Interactions.

Interaction Energy.—We utilized dispersion-corrected density *f*unctional *t*heory (DFT)[43-47] based quantum mechanical calculations (see Methods) for computing the energy of anion–quadrupole interactions. Dispersion-corrected DFT calculations have been shown to be more accurate in capturing the behavior of the noncovalent interactions of aromatic ring systems[47]. The interacting functional groups were represented as a simplified organic scaffold at pH 7.0 (Figure 1A). Energy calculations carried out with the 13,308 anion–quadrupole pairs are shown in Figure 1B. The distributions of the energies suggest that the strength of the interactions, on average, is similar for all of the aromatic residues and that the average energy of Trp interactions is slightly greater (*more negative*) than that of the others. The observed energies are also consistent with those observed earlier with phenylalanine's anion–quadrupole interactions[25]. Based on the calculated energy, we loosely defined weak (≥ −3.0 kcal/mol) and strong (≤ −8.0 kcal/mol) interactions (Figure 1B). The geometrical features of strong and weak interactions are distinct (Figure 1C). Stronger interactions tend to: (a) take place over shorter distances, (b) subtend angles closer to zero, and (c) occur preferentially in a perpendicular interplanar orientation. In addition to steric effects, the stability of conformations (e.g., trans-proline[48], the polyproline II helix[49], and Z-DNA[15]) within macromolecules is also strongly supported by stereoelectronic effects (i.e., conformational preferences that arise due to hyperconjugative delocalization), for example, $n_{CO} \rightarrow \pi^*$ of the succeeding >C=O in trans-proline[48] and cytidine $n_{O4'} \rightarrow \pi^*$ of the guanine base in CpG Z-DNA[15]. For anion–quadrupole interactions with a distance ≥ 3.5 Å (i.e., within orbital mixing distances), the preferences for angle (~0°, Figure 1C, bottom left) and interplanar orientation (~90°, Figure 1C, bottom right) are far stronger than in those interactions with distances > 4.0 Å (Figure 1C). Stereoelectronic effects could also contribute to the preferred geometries of anion–quadrupole interactions.

Deconvolution of Dipole-Quadrupole Energy Contributions.—The benzene ring possesses only a quadrupole moment, while the indole ring possesses a ring dipole moment in addition to a quadrupole moment (Figure 2B). Therefore, anion–quadrupole interactions

of the tryptophan ring are expected to also be influenced by the dipole moment of the indole ring. To quantify the influence of a dipole moment, we analyzed anion–quadrupole interactions energy in the following manner. We expected that the dipole moment would influence the polarity of the ring-CH atoms, and a difference in the resulting ring-CH atom polarity also suggested that the energy of interactions would be dependent on the ring-CH atoms. We, therefore, analyzed the energy of anion–quadrupole interactions based on the ring atom that is closest to the anion. For example, the anion–quadrupole interaction with the Phe-CZ atom (Figure 2A) refers to the interactions, among all the Phe anion–quadrupole interactions, in which the CZ atom is closest to the anion. The distance dependency of the interaction energy (ion–ion, ion–dipole, dipole–dipole, ion–quadrupole) is characteristic of an interaction type, and therefore, we determined the distance dependency of the anion–quadrupole interaction energy for each of the ring atoms within the distance range 3.25 – 4.50 Å (Figure 2). For convenience, we used the distance between the ring carbon atom and the center of the carboxylate oxygen atoms. We found that the interaction energies of anions with the phenylalanine ring atoms were not identical and that the distance dependency (within this distance range) of the interaction energy for each atom type was linear and nearly parallel to that of the other atoms (Figure 2A). For benzene's anion–quadrupole interactions, we expected all atoms to behave identically. For phenylalanine, even with a small ring dipole moment, the behavior of the interaction energies of its atoms differed from that of benzene; that is, the interaction energy of the Phe-CZ atom is different from that of the Phe-CD1/2 atom (Figure 2A). Instead of a single energy value, we considered a range of interaction energies (spanning between the Phe-CZ and Phe-CD1/CD2 parallel lines, Figure 2A) as the baseline range for anion–quadrupole interactions. We refer to this energy range as the "**F**-range" (Figure 2B). For convenience, we consider a deviation from the F-range to be a result of the contributions from other factors such as the ring dipole moment (Figure 2B). Consistent with the dipole moment of the indole ring, we found that the interaction energies for tryptophan ring atoms deviate from the **F**-range (Figure 2B). That is, anions interact with Trp-CD1 and Trp-CZ3 atoms respectively with energies much higher (*more negative*) and lower (*more positive*), respectively, than those observed in the **F**-range (Figure 2B, left). The value of the deviation from the F-range depends on the ring atom and provides a rough estimate of the contribution from the dipole moment for each ring atom. In other words, we propose that, in the absence of a dipole moment, the expected energies would be within the **F**-range, and the presence of a dipole moment would result in deviations of the energy values from the **F**-range.

As mentioned above, we do not consider anions hydrogen bonded to Trp-NE1 as participating in for *regular* anion–quadrupole interactions. However, given the polarity of the Trp-NE1 and Trp-HE1 atoms, anions within the distance range 3.25 – 4.50 Å (i.e., no hydrogen bond with Trp-NE1) can still interact with Trp-NE1 or Trp-HE1 atoms with energies greater than that of Trp-CD1 atom interactions (Figure 2C). We consider this interaction to be a part of the anion–quadrupole interactions. To distinguish this interaction from the hydrogen bonds, the interaction energy of anions hydrogen bonded to Trp-NE1 (distance range, 2.75 – 3.50 Å) is also overlaid in Figure 2C. The value and the distance dependence of the energies of hydrogen bonds and atom-specific anion–quadrupole interactions differ from one another (Figure 2C). That is, the slope and the intercept depend

on the polarity and the interaction distance range. Due to the dipole moment, one end of the tryptophan ring (Trp-CZ2, Trp-CH2, Trp-CZ3) interacts with anions with energies lower than that of the phenylalanine ring, while the other end (Trp-NE1, Trp-CD1) interacts with energies greater than that of the phenylalanine ring. The anion–quadrupole interaction energy averaged over all tryptophan atoms is, however, greater (*more negative*) than that of phenylalanine. This is because of the higher polarity of the protons (Trp-HE1, Trp-HD1), and tryptophan's anion–quadrupole energetics is dominated by Trp-NE1 and Trp-CD1 atoms. If the interaction energies of anions hydrogen bonded to Trp-NE1 were added, then tryptophan's anion–quadrupole interaction energies would be even more different from that of phenylalanine. This logic is also consistent with the reason for excluding anions hydrogen bonded to the ring for a fair comparison with phenylalanine. A comparison of the energy-distance relationships for the symmetry-related positions of tyrosine and phenylalanine (Figure 2D) also shows that the presence of the −OH group in tyrosine shifts the intercept of the energy-distance plot, and the shift is more prominent for positions closer to the −OH group (Figure 2D).

Boltzmann Statistics from Protein Structures.—In evaluating the performances of alternative methods (e.g., DFT, dispersion-corrected DFT, MP2) for computing the energy of noncovalent interactions, typically an agreement between the calculated results, the experimental structure (crystalline state), and the energetic (e.g., sublimation) data is sought[47]. In a similar spirit, we have also attempted to correlate the computed energies of anion–quadrupole interactions to crystalline state properties. For example, we investigated whether the computed energies are consistent with the frequency with which interactions occur in protein crystal structures. In this regard, we assumed that the *p*rotein *d*ata *b*ank (PDB) represents an equilibrium distribution of protein conformations. This assumption is based on the success of the Boltzmann hypothesis in explaining the energetics of protein structures[50]. All of the energy calculations described above were carried out with isolated side chain atoms (see Methods), while these interactions in real proteins actually take place among a myriad of other interactions. For example, an Asp side chain engaged in anion–quadrupole interaction may also be involved in a salt bridge interaction as well as be in contact with surface solvent water molecules by hydrogen bonds (see *below*). Despite these differences in context (i.e., isolated functional groups versus the entire protein), consistency between the calculated energy and observed statistics in protein structures would provide a strong rationale for basing our analyses on high-resolution protein structures. Given that the energy of anion–quadrupole interactions depends on ring atoms (as we noted above), we wanted to determine whether the occurrence frequency of anions in the vicinity of a specific ring atom is consistent with the calculated atom-specific energy. For the convenience of discussion about the occurrence frequency of anions in the vicinity of ring atoms, we utilized a measure termed $\mathbf{F_{ATOM–ds}}$ (*see* SI pp. S6–S7). The $\mathbf{F_{ATOM–ds}}$ value for a ring atom of a given residue type represents the percentage of time it is present at the shortest distance among all anion–quadrupole interactions of that residue type, and the sum of the $\mathbf{F_{ATOM–ds}}$ values of all ring atoms of a given residue type is 100% (Figure 3). It is interesting to note that the $\mathbf{F_{ATOM–ds}}$ values of even phenylalanine atoms (who show small energy differences, Figure 2A) differ among the phenylalanine atoms, and phenylalanine $\mathbf{F_{ATOM–ds}}$ values are qualitatively consistent with the atom-specific energies. For example,

the $\mathbf{F_{ATOM\_ds}}$ for Phe-CZ is 18% while that for Phe-CD1 is 22% (Figure 3A, Figure 2A). Similarly, the $\mathbf{F_{ATOM\_ds}}$ values for tyrosine and tryptophan atoms are also consistent with the computed energy (Figure 3A, Figure 2). Due to the dipole moment of the indole ring, more than 30% of tryptophan residue interactions with anions are with its CD1 atom alone (Figure 3A). In other words, the anion–quadrupole interactions of tryptophan residues have an orientation preference. Consistent with its $\mathbf{F_{ATOM\_ds}}$ value, the magnitude of the interaction energy of the Trp-CD1 (Figure 3B) atom was also higher than that of other Trp ring carbon atoms. In general, the computed interaction energy for ring atoms using isolated functional groups was consistent with their respective $\mathbf{F_{ATOM\_ds}}$ values (Figure 3C). This observation is also useful in the context of molecular mechanics force fields. For example, the partial charges of Trp ring protons in the CHARMM36[51] force field were better correlated with the observed ring atom participation frequency (Figure S6, and SI S6–S7) than those of the CHARMM22[52] force field. Other than the HE1 atom, all of the Trp ring protons in the CHARMM22 force field are assigned a uniform partial charge equal to +0.115 (Figure S6C), while in the CHARMM36 force field, the Trp ring atom partial charges are consistent with the indole dipole moment[51]. As a result, a better correlation was observed with the CHARMM36 force field.

HB-anion–quadrupole Interactions.

Anion–quadrupole Interaction Types and The Energy Gap.—Because of the shorter distance (Figure S1D), anions hydrogen bonded to tryptophan and tyrosine side chains also satisfy the distance-angle criteria (≤ 4.5 Å and ≤ 35.0°; *see above*) for the detection of anion–quadrupole interactions (Figure 4A). Therefore, by default, anions hydrogen bonded to tryptophan and tyrosine side chains are also considered here for anion–quadrupole interactions. We refer to these pairs as HB-anion–quadrupole interacting pairs to distinguish them from the *regular* anion–quadrupole pairs discussed above. There are 6,256 HB-anion–quadrupole pairs in the dataset. As expected, the energies of these interactions are much greater than those of *regular* anion–quadrupole interactions (Figure 4B), i.e., *regular* anion–quadrupole interactions are typical weak interactions. Due to dynamic motion in proteins (e.g., bond rotation), is it possible for a *regular* anion–quadrupole interaction to convert to an HB-anion–quadrupole interaction and *vice versa*? In other words, can a *regular* anion–quadrupole pair (distance ≤ 4.50 Å) be within hydrogen bonding distance (≤ 3.50 Å) and satisfy the hydrogen bond angle criterion during dynamic motion? Because of the significant energy gap between the two types of anion–quadrupole interactions (Figure 4B), we expected a low frequency for such conversions. To test this prediction, we created a dataset of proteins whose structures have been determined by NMR as well as X-ray crystallography and for which the NMR structure ensemble is believed to approximate the dynamics of protein motion (Figure S4). In this dataset, in only ~7% of the time are *regular* anion-quadrupole pairs (tryptophan, tyrosine) in the X-ray structure is observed as an HB-anion–quadrupole pair in the NMR structure of the same protein molecule. The interconversion frequency for specific ring atoms (adjacent to the ring hydrogen bond donor position) is shown in Figure S4. Other than the Trp-NE1 atom, no other ring atom has an interconversion frequency > 10% (Figure S4). The energy of the *regular* anion–quadrupole interaction of only the Trp-NE1 atom comes close to that of the HB-anion–quadrupole interaction (Figures 2C, 4B), and the observed higher interconversion frequency of Trp-NE1

is likely due to the small energy gap between its *regular* anion–quadrupole and HB-anion-quadrupole interactions. The small interconversion frequency for other ring atoms is, in general, likely due to the large energy difference between the *regular* anion–quadrupole and HB-anion–quadrupole interactions, and therefore we have treated these anion–quadrupole interaction types separately.

Orientation Preference of HB-anion–quadrupole Interactions.—The energy difference between the two types of interactions is also relevant for mutational analysis. For example, mutating tyrosine → phenylalanine when the tyrosine residue is engaged in an HB-anion–quadrupole interaction in the wild type protein will likely result in a *regular* anion–quadrupole interaction in the mutant protein. As *regular* anion–quadrupole interactions are weak, there can be a significant difference in stability between the wild type and the mutant in this example. As another example, an aspartate-tyrosine-mediated HB-anion–quadrupole interaction between the CDR loops of the monoclonal antibody MR78 contributes significantly to the binding affinity for its antigen[53] (i.e., the Marburg virus glycoprotein), and substitution of this tyrosine residue with phenylalanine significantly compromises the binding affinity[53], justifying the perception of the weak nature of anion–quadrupole interactions. Even though the energy of the anion–quadrupole component of the HB-anion–quadrupole interaction is low, an anion–quadrupole interaction can still make a significant contribution to the behavior of an HB-anion–quadrupole interaction. Here we show that an anion–quadrupole interaction contributes to the orientation preferences of HB-anion–quadrupole interactions (Figures 5). For example, tyrosine residues have an orientation preference for Asp/Gly carboxylate ions when hydrogen bonded to the Tyr-OH group (Figure 5). The orientation preference can be assessed by a measure referred to here as the pseudo-dihedral angle (Figures 5, S7A). The distribution of pseudo-dihedral angles for Tyr-OH was quite different from that of the Thr/Ser-OH group, suggesting that the carboxylate oxygen atoms prefer proximity to tyrosine ring edges (Figures 5, S7B). This is likely due to anion–quadrupole interactions, as we observed that the distances of anions to the tyrosine ring edges were shorter than the respective carbon-anion distances of Thr/Ser residues (Figure S7B).

Investigation of the above orientation preference was prompted by the following observation. Asp/Glu sidechain carboxylates were the most frequent hydrogen bond acceptors for each type of side chain hydrogen bond donor in proteins (Figure S7C). The Asp/Glu-acceptor frequency for Tyr was much higher than that of the uncharged side chain donors (Figure S7C), that is to say, the frequency was midway between that of the positively charged and uncharged side chain donors. Although the partial charges of the Tyr–OH group (Figure S7C) may contribute to this, the observed orientation preferences (Figure 5) suggest that anion–quadrupole interactions also contribute to the observed higher Asp/Glu-acceptor frequency for the Tyr–OH group. The observed higher frequency of the Tyr hydrogen bonds with Asp/Glu also suggested that the strength of these hydrogen bonds (HB-anion-quadrupole) is likely to be greater than typical hydrogen bonds, and the following examples support this notion: (i) in the anti-Marburg virus glycoprotein monoclonal antibody (MR78)[53], a kink in the HCDR3 loop, critical for antigen recognition, is stabilized by an Asp–Tyr HB-anion–quadrupole interaction in place of a canonical Asp–Arg/Lys salt

bridge[53] (Figure S5F), (ii) the nanobody Nb60, by virtue of an interfacial Glu–Tyr HB-anion–quadrupole interaction, locks the ligand-(carazolol)-bound conformation of the G protein coupled receptor β_2AR[54] (Figure S5G). In general, we expect that the strength of noncovalent interactions is likely to follow the order: salt bridge > HB-anion–quadrupole > typical hydrogen bond > anion–quadrupole interactions.

Hydrogen Bonds of Anion and Interaction Clusters.

Interaction Clusters.—Anions participating in HB-anion–quadrupole interactions are always hydrogen bonded to the side chain atoms of the partner aromatic residue (Tyr–OH and Trp-NE1); i.e., the anions of 100% of HB-anion–quadrupole pairs are hydrogen bonded. By comparison, anions of ~70% and ~50% of *regular* anion–quadrupole and salt bridge pairs, respectively, were also hydrogen bonded to other hydrogen bond donors. In addition to the hydrogen bonds with Tyr–OH and Trp-NE1 atoms, 32% of anions of HB-anion–quadrupole pairs also engage in additional hydrogen bonds. These observations suggest that all of these interactions tend to be part of interaction networks (Figure 6A, left). In these networks, the anion and the aromatic ring of one anion–quadrupole interaction can be part of another anion–quadrupole interaction. For example, ~7% of aromatic residues of anion-quadrupole interactions interact with more than one anion (Figure 6A, right), and ~7% of anions of anion–quadrupole interactions similarly interact with more than one aromatic residue. Even though the energy difference between the *regular* anion–quadrupole and HB-anion–quadrupole interactions is significant, in certain homologous families we find that a *regular* anion–quadrupole in one member of the family can be replaced by an HB-anion–quadrupole in another member (Figure 6B). Given that these interactions are part of a cluster of interactions, such amenability to substitution seems reasonable, because the loss/gain of energy due to replacement of one type of interaction is likely compensated by a corresponding replacement in another interaction within the interaction network.

Buried Anions.—Polar groups, buried within the protein interior, must be hydrogen bonded to satisfy the hydrogen bonds with solvent water molecules that were broken upon folding[55]. As anions of ~30% *regular* anion–quadrupole interactions are not hydrogen bonded, we determined the solvent accessible surface area (ASA) of these anions to look for buried anions without a hydrogen bond. We found an ASA < 2.0 $Å^2$ for each of its side chain carboxyl-oxygen atoms of 158 anions, and we considered these anions to be buried with an unsatisfied hydrogen bond. The remaining anions of the 30% pairs are considered solvent accessible and expected to be hydrogen bonded to solvent water molecules. How does one account for the lost or missing hydrogen bonds for these 158 buried anions? For the convenience of discussion on this we refer to these 158 anions as BAUH_AQ (i.e., *b*uried *a*nions with *u*nsatisfied *h*ydrogen bonds participating in anion–quadrupole interactions). We examined the distance distribution of BAUH_AQ pairs and found that 50% of them are ≤ 3.5 Å away from the ring, and therefore there is a possibility of forming CH—O hydrogen bonds (see Figure 6C, top left). By comparison, the distances of between 36% of the *regular* anion–quadrupole interacting pairs in the dataset are ≤ 3.5 Å. Relaxing the hydrogen bond detection criterion could lower the number of BAUH_AQ pairs. However, the fact that the observed proportion (50%) of distances of BAUH_AQ pairs is ≤ 3.5 Å instead of the expected proportion (36%) suggests that *regular* anion–quadrupole interactions could be a

solution for satisfying the lost hydrogen bonds of buried anions. To see whether this could be true, we found that there are 252 BAUH anions in this dataset and 158 of them are BAUH_AQ (~63%). We also checked whether the anions of BAUH_AQ pairs participate in more than one *regular* anion–quadrupole interaction, such that the sum of two or more weak interactions could make up for the lost hydrogen bond. We found that ~22% of the BAUH_AQ anions are involved in two or more interactions of the anion–quadrupole. This proportion is much higher than the expected ~7% in the dataset (see above). Thus, anion–quadrupole interaction clusters are expected to be more common for BAUH_AQ anions (Figure 6C), and we have observed as many as five aromatic rings surrounding a single BAUH_AQ anion engaged in as many anion–quadrupole interactions (Figure 6C). These observations suggest that such strategies could be employed to stabilize buried anions in the design of "non-ideal" proteins, as discussed recently in the context of designing binding antibodies[56], in which buried charges and unpaired polar group are common[57].

Comparison with Salt Bridge.—Anions predominantly interact with cations, typically forming salt bridges. The number of salt bridges in the dataset was 28,589. By comparison, anions interacting with aromatic residues in *regular* anion–quadrupole and HB-anion–quadrupole interactions in the dataset was 19,561 (13,308 + 6,256). On average, we observed ~5 anion–quadrupole and ~8 salt bridge interacting pairs per protein, and 17% of aromatic residues (Phe, Tyr, Trp) were engaged in *regular* anion–quadrupole interactions, and 14% of Tyr/Trp residues were engaged in HB-anion-quadrupole interactions. The fraction of aromatic residues engaged in anion–quadrupole and HB-anion–quadrupole interactions taken together (17% + 14%) was comparable to the 32% of Lys/Arg residues engaged in salt bridges. Like salt bridges, anion–quadrupole interactions may also play an important role in macromolecular folding, assembly, and recognition, which is supported by the set of examples in Figure S5. These include apoptosis-inducing factor (AIF)[58], redesigned Tiam1 PDZ[59] (Figure S5A), the extracellular receptor of the kinase WalK (*er*Walk) of the *Staphylococcus aureus* two-component signal-transduction system[60], and others. In the NADH-driven allosteric regulation of AIF, the release of the C-loop region from the rest of the polypeptide chain is a critical step[58] (Figure S5B). AIF residues, crucial for the C-loop release[58], interact by the anion–quadrupole interaction (Figure S5B). It is also interesting to see that two residues (Y165, D119) of the three key signal-transducing residues of *er*WalK are engaged in anion–quadrupole interactions (Figure S5C) and that substitutions of these residues impair WalKR function[60].

Anion–quadrupole Interactions of Histidine Residues.—For the aromaticity of the imidazole ring, histidine residues are also expected to participate in anion–quadrupole interactions (like the cation–π interactions of histidine residues). However, the protonation status of the imidazole ring can significantly influence the energy of cation-π interactions of a histidine residue[61]. Similarly, we expect that protonation status of a histidine ring will likely influence the energy its anion–quadrupole interactions due to electrostatics. Therefore, here we had focused on analyzing the anion–quadrupole interactions of histidine residues that were made only with the ring carbon atoms (Figure 7A). That is, we ignored the interactions of anions with the histidine-ring nitrogen atoms for: (a) the uncertainty of their protonation status, (b) hydrogen bonding as discussed above. Using the same distance-angle

detection criterion, we observed 3445 histidine-mediated (i.e., with CG, CD2 and CE1 ring-atoms) anion–quadrupole interactions in the same dataset. The distance and angle distributions of the histidine-anion–quadrupole interactions were respectively very similar to those of phenylalanine, tyrosine and tryptophan suggesting that anions interact with histidine ring edges in a manner very similar to those of the three aromatic rings. In this dataset, the frequency of histidine, phenylalanine and tryptophan residues respectively were 2.4%, 4.1% and 1.5%. Normalizing the occurrence frequency of anion–quadrupole interactions of the ring carbon atoms with the respective amino acid frequency (Figure 7B) suggested that histidine-ring carbon (CD2 and CE1) atoms participate in anion–quadrupole interactions more frequently than the others. Thus, we expect the energies of histidine-anion–quadrupole interactions to be higher (*more negative*) than those of the other aromatic amino acids. This is also consistent with the shorter distances of histidine-anion–quadrupole interactions in comparison with others (Figure 7C). For the following sections, however, we discuss anion-quadrupole interactions of only the three aromatic amino acids (phenylalanine, tyrosine and tryptophan).

Other Functional Groups and Interfaces.

Other Functional Groups.—Like carboxylate ions, the phosphoryl group, –(PO_4^-)–, may also participate in anion–quadrupole interactions. We also estimated the energy of the phosphoryl group interacting with the Phe ring edges (Figure 8A). For comparison, the energy of interaction of other amino acid sidechain functional groups (–OH, –$CONH_2$, and –CH_2–/CH_3–) with Phe edges was also calculated (Figure 8A). Like the carboxylate anion–quadrupole interaction, we imposed the same distance-and-angle geometric criteria when selecting these functional groups from protein structures. The phosphoryl diester, CH_3–O–(PO_2−)–O–sCH_3, functional groups included here for the calculations are those present at DNA/RNA–protein interfaces, and they were selected by imposing the geometric criterion on the OP1 and OP2 atoms (*see* Methods). The average energy of interaction of the phosphoryl group was comparable with that of the carboxylate ion (Figure 8A). As expected, the energy of interaction for other functional groups was much smaller (Figure 8A), justifying the conclusion that, irrespective of the nature of the anion, the energetic preference of ring edges is to interact with anions. Therefore, using the same detection criterion, we checked for the presence of anion–quadrupole and HB-anion–quadrupole interactions of nucleic acid backbone phosphate ions with protein aromatic residues (Phe, Tyr, and Trp) (Figure 8B) in a high-resolution (≤ 2.20 Å), non-redundant dataset of 252 DNA–protein and 126 RNA–protein interfaces. For the DNA-protein dataset, there were 347 anion–quadrupole and 179 HB-anion–quadrupole pairs present at the interface, and for comparison there were 904 pairs of salt bridges with the backbone phosphate (Table S1). Even though the number of anion–quadrupole and HB-anion–quadrupole interactions for the DNA–protein interfaces was fewer than the number of salt bridges, it is interesting to note that the fraction of interfacial aromatic residues engaging in anion–quadrupole interactions (30.0%, Table S1) at DNA–protein interfaces was considerably higher than that seen in single-chain proteins (17.0%, *see* above). The same was true for HB-anion–quadrupole interactions (Table S1). The fraction of interfacial Arg/Lys residues at DNA-protein interfaces engaging in salt bridges (32.6%, Table S1) was comparable to that seen in single-chain proteins (31.9%). Even though the percentage of Arg/Lys residues more than doubled

(10.6% → 25.2%) going from the single-chain set to the DNA–protein interfaces, the fraction of Arg/Lys residues in physical contact with negative changes did not seem to alter. The percentage of aromatic residues (9.2%) in the single-chain dataset was comparable to that in the DNA–protein interfaces (10.5%), and yet the fraction of aromatic residues in physical proximity to negative charges (phosphoryl groups) was nearly double (16.9% → 30.0%), suggesting a role for anion–quadrupole interactions in DNA recognition. The numbers for RNA–protein interfaces, 130 (anion–quadrupole) and 19 (salt bridge) interactions (Table S1), and, similarly, the enhanced fraction of aromatic residues engaged in anion–quadrupole interactions both support this conclusion (Table S1).

The role of aromatic residues in membrane protein structure at the lipid-water interface is well known[62]. The abovementioned interactions of phosphate ions with aromatic residues may therefore also be important for the structure of membrane proteins at the lipid–water interface. Four recent 3.2-Å structures of SERCA1 (a Ca^{2+}-ATPase pump), which were elucidated using X-ray solvent contrast modulation, captured the structure of the lipid-bilayer-like membrane around a protein for the first time[63] (Figure 9). In these four structures, representing the four conformational states of SERCA1, we observed 26 anion–quadrupole and 6 HB-anion–quadrupole interactions of the SERCA1 aromatic residues with the O1P and O2P atoms of the phospholipid head groups (Figure 9). This study highlighted the roles of the Arg/Lys residues when forming salt bridges with the phospholipid head groups. By comparison, in these four structures we observed 17 salt bridges between SERCA1 and the O1P/O2P atoms of the phospholipids. Thus, our observation of anion–quadrupole interactions at the bilayer–protein interface may also suggest an underappreciated role for anion–quadrupole interactions in membrane proteins. With the recent report on the functional contact between the broadly neutralizing anti-HIV-1 antibody (4E10) and the membrane-proximal external region (MPER) of the HIV gp41 protein, which extends into the membrane[64], it is tempting to speculate that anion-quadrupole interactions between the Trp residues of the “long” antibody CDRH3 loop and the membrane phospholipid head groups contribute to this recognition (Figure S8A). Like other membrane anion transporters (e.g., the fluoride channel Phe box[65]), we also observed anion–quadrupole interactions between the multidrug resistance protein MRP1[66] and its physiological substrate, leukotriene C4 (a negatively charged small molecule, Figure S8B). Similarly, other organic small molecules may also utilize the anion–quadrupole mechanism to interact with proteins, for example, the cereblon:GSPTI assembly[67] and kinase suppressor of RAS (KSR2)[68] in complex with their respective small-molecule modulators (Figure S8C).

Interfaces.—We also checked a set of non-redundant, high-resolution protein interfaces (Table S1) for the presence of anion–quadrupole interactions. In all of the different interfaces, anion–quadrupole interactions were invariably present (Table S1, Figure 10), even though the fraction of interfacial aromatic residues engaging in interchain anion–quadrupole and HB-anion–quadrupole interactions was lower than that of intrachain interactions. The examples in Figure 10, however, suggest that key interfacial contacts in many complexes do involve anion–quadrupole interactions. Some of these examples are: (i) the *E. coli* periplasmic molecular chaperone SurA PPIase P1 domain, which binds the consensus Ar-x-Ar-x motif[69] (Ar, aromatic residue) using anion–quadrupole interactions

(Figure 10A); (ii) bapineuzumab[70], a humanized antibody (3D6 Fab) that targets amyloid plaques and binds amyloid peptide engaged in anion–quadrupole interactions (Figure 10B); and (iii) the interface between the S and L chain of *Desulfovibrio desulfuricans* iron hydrogenase, which exhibits an unusually high number of anion–quadrupole interactions (Figure 10C). Redox-active proteins are known to utilize aromatic residues to transport oxidizing equivalents (holes) to the protein surface[71] to mitigate protein interior damage, and the *D. desulfuricans* hydrogenase likely utilizes the interfacial aromatic residues, oriented by anion–quadrupole interactions, to protect from such damage. It is also interesting to observe that at the interfaces of antibody-protein complexes (Table S1), the number of anion–quadrupole pairs was comparable to the number of salt bridges. This observation on anion–quadrupole interactions at the antibody interfaces is also consistent with the recent antibody-antigen interface amino acid compositional analysis[72] (e.g., enriched Trp, Asp, Arg, etc.). In general, other than peptide-binding interfaces, the number of interchain salt bridges was ~2-fold higher than the number of anion–quadrupole and HB-anion–quadrupole interactions combined. However, like interfacial aromatic residues, the fraction of interfacial Arg/Lys residues participating in interchain salt bridges was also lower than that of intrachain salt bridges (Table S1), suggesting that interactions of aromatic and Arg/Lys residues with Asp/Glu residues are likely to be more prominent in the context of polypeptide chain folding, and therefore we considered probing the behavior of their conservation in the context of intramolecular interactions (*see* below).

Interaction Conservation.

Conservation of Weak Interactions.—How conserved is a weak interaction? How does the conservation of a weak interaction compare with that of a salt bridge, an interaction considered be to be energetically stronger than an anion–quadrupole interaction. A non-redundant set of 4835 high-resolution protein structures belonging to 776 pfam domains was chosen to address this in our conservation analysis (see SI pp. S8–S10). A position in a structure in this set was first mapped to a column of the corresponding pfam domain alignment (Figure S9). For an anion–quadrupole interaction observed in a structure, the occurrence frequency of the residue types (e.g., Trp/Tyr/Phe) in the corresponding alignment columns was determined. We represented the occurrence frequency as a *probability* ($0.0 \leq p \leq 1.0$), and the distributions of the occurrence frequencies of Asp/Glu in anion–quadrupole, HB-anion–quadrupole, and salt bridge interactions were not only similar but also tended to show low levels of conservation ($p \leq 0.3$) in general (Figure 11A). The trend was similar for the aromatic residues of anion–quadrupole and HB-anion–quadrupole interactions and the positively charged residues of salt bridges (Figure S9C), suggesting that an interaction observed in one structure can be replaced by a different type of interaction or be absent in another structure. Even though less conserved in general, we wanted to probe the correlation between the interacting positions. We utilized a simple measure termed the *conditional probability (detailed* in SI pp. S9–S10 and Figure S10) for probing the correlation between interacting positions. We have also used *conditional probability* in describing the coevolving nature of the binding-site residues of the PHD-finger module[73]. The probability of aromatic residues (W75 in the structure, Figure 11B) in column number 73 of the pfam CMD domain alignment increases ($0.24 \rightarrow 0.67$) when the alignment was perturbed by column number 91 (D93 in the structure), in which the W75 and D93 residues interacted by the HB-anion-

quadrupole interaction (Figure 11B). In the pfam CMD example, the *conditional probability* ($p = 0.67$) was greater than the *probability* ($p = 0.24$), and we considered an enhancement of the occurrence frequency as an indication of positional correlation; that is, if a residue is retained, the interacting partner is likely to be retained as well. For anion–quadrupole, HB-anion–quadrupole, and salt bridge residue pairs, the *conditional probability* on average was always greater than the probability (Figure 11C), suggesting that there is a preference for retention of the interacting partner. The average *conditional probability* generally followed the order salt bridge > HB-anion–quadrupole > anion-quadrupole interaction (Figure 11C), and the observed energetics (Figure 4B) likely contributed to the observed order, in which energetically stronger interactions were more likely to retain the interacting partner during evolution.

Conservation Among Remote Homologs.—From the set of 4835 protein structures, we created a subset of structures, termed pfam_nr40, for each of the pfam domains (see SI pp. S3). The pfam_nr40 subset for each domain contained at least two structures, and the sequences of all the structures within pfam_nr40 shared less than 40% sequence identity, and we considered pfam_nr40 to consist of structures of remote homologs. We observed many instances in which an interacting pair was retained at the respective positions of the pfam_nr40 structures, for example, anion–quadrupole interactions of the pfam Peripla_BP_4 (Figure 12A) and MR_MLE_C (Figure S11C) families. Despite structural deviations among the pfam_nr40 structures, conserved interacting residues ($p \leq 0.5$) participated in anion–quadrupole interactions in each of the structures of Peripla_BP_4 (Figure 12A). In the pfam_nr40 analysis, anion–quadrupole interactions were first detected in each structure and then mapped to the pfam alignment columns (see *above*, and Figure S9A) to check whether an alignment column was common among the pfam_nr40 structures. This means that the distance-and-angle detection criteria first need to be satisfied in the corresponding positions of the pfam_nr40 structures for a common column to be observed for a pfam domain family (Figure S9). This strict criterion restricted accounting for positions, not merely for the amino acid's presence but for its presence with a well-defined geometry, even among remote homologs. For the common alignment columns of pfam_nr40 structures (Figure S9), we also obtained the occurrence frequency (p). A distribution of these frequency values for the interacting residues (anion–quadrupole, HB-anion–quadrupole, and salt bridge interactions combined) in this case (i.e., conserved among pfam_nr40 structures) showed a much higher level of conservation, $p \leq 0.35$ (Figure 12B).

Protein Folding.—Based on the observation of a highly conserved pair of interacting residues (anion–quadrupole, HB-anion–quadrupole, and salt bridge) observed among diverse sequences that retain the same fold (Figures 12A, S11C), it is tempting to believe that interacting residues or residue positions conserved among remote homologs contribute towards folding and fold determination. Probing this question would, however, require detailed experimentation and has already been the subject of folding studies in the past decade[74-76]. The availability of *in silico* folding algorithms[77]–[78] based on evolutionary covariance information encouraged us to check the predicted residue–residue contact scores for the interacting residues, particularly for interactions that were conserved in two structures of the remote homologs (Figure 13C). We utilized the GREMLIN[79, 80]

interresidue contact scores (see SI pp. S10-S11) available for the pfam domain sequences to map the scores onto the structures of pfam_nr40 (Figure S11A). The distribution of the interresidue contact scores for interacting residues (anion–quadrupole, HB-anion–quadrupole, and salt bridge interactions) that were conserved among remote homologs was shifted towards higher scores when compared with the background distribution of the contact scores (Figure 13C). This suggests that, at least in the in silico folding of a protein sequence, the distance restraints obtained from the conserved anion–quadrupole, HB-anion–quadrupole, and salt-bridge pairs can make predominant contributions, even though these interactions are on the surface. These observations are physically meaningful in the context of the pfam_nr40 structures (Figure 13B). For example, in the pfam_nr40 structures of the bromodomain and MR_MLE_N domains, the anion–quadrupole and HB-anion–quadrupole interacting pairs with large interresidue contact scores were in the regions having the least structural deviation among pfam_nr40 structures (Figure 13B). In general, anion–quadrupole, HB-anion–quadrupole, and salt bridge pairs that were conserved among remote homologs appeared within regions showing small structural deviations (Figures 13B, S11B). This observation was quantified by the distribution of the root-mean-square deviation (RMSD) of Cα atoms of conserved anion–quadrupole, HB-anion–quadrupole, and salt bridge pairs, which was shifted to lower RMSD values when compared with that of equivalent Cα atoms of remote homologs (Figure 13A). These observations suggest that anion–quadrupole and HB-anion–quadrupole interactions likely play important roles in folding and stability.

Concluding Remarks.

Like the Phe-Asp/Glu pairs, anion–quadrupole interactions with His, Tyr and Trp residues are also very common, and such interactions often occur within networks of other interactions. The energies of anion–quadrupole interactions depend on the geometry between the functional groups and the ring atom type (e.g., on the presence or absence of a dipole moment). Highly conserved anion–quadrupole interactions can be observed, even among remote homologs, suggesting that such interactions could contribute to folding and stability of proteins. The recent success in designing miniature monomeric proteins based on conserved and stabilizing structural motifs (e.g., cation–π in TrpPlexus[81], CH–π in PPα-Tyr[29]) suggests that the conserved anion–quadrupole structural motifs observed here could also be exploited for similar design studies. In general, standard protein design algorithms typically lack energy functions for interactions such as cation–π or water-mediated stabilization. To function as well as natural proteins, in many cases designed proteins require multiple rounds of laboratory evolution in which such interactions can be introduced (e.g., in a recently redesigned ubiquitin variant[82]). Several examples of anion–quadrupole interactions in natural proteins cited in this study encourage the incorporation of energy functions for this kind of interaction in design algorithms. As anion–quadrupole interactions are a weak in nature, they likely contribute cooperatively with other interaction. For example, the highly conserved anion–quadrupole pairs (among remote homologs) appear in a network of other interactions (Figure 13B) and parsing the energy contributions of an anion–quadrupole interaction in a network of other interactions may not be straightforward.

However, the consistency between computed energy (calculated using isolated functional groups) and ring atom participation frequency and geometry (both obtained from folded proteins) suggests that statistics on the structures of weak interactions reasonably approximate interaction energies obtained from quantum mechanical calculations. The statistics on the orientation preferences of Tyr/Trp residues with anions observed here could be useful for describing such interaction potentials[83-86] in protein design, refinement, and structure prediction applications. Since aromatic scaffolds are common in small organic molecules such as the Trp residue, substituted/unsubstituted organic aromatic rings with a net dipole moment can also exploit edgewise interactions with anions on the surface of a protein molecule. It is also encouraging to observe the correlation between molecular mechanics force field parameters and the structural statistics of weak interactions (e.g., Trp partial charges in the CHARMM36 force field[51]). Our observation thus provides a rationale for probing the performance of CHARMM36[87] (for proteins) compared with other force fields in the study of anion–quadrupole-like interactions by molecular dynamics simulations. In general, reports about a novel structure or mode of interaction (e.g., Fcab[88]) often involve analysis of residue interaction networks (e.g., using the RING 2.0 server[89]). Such analysis does not typically include anion–quadrupole interactions. However, since we observed that anion–quadrupole interactions can be important components of interaction networks, inclusion of such interactions will enrich the analysis of residue interaction networks. For example, buried anions lacking hydrogen bonds were observed to be frequently involved in anion–quadrupole interaction networks, and one or more edgewise interactions with aromatic residues could stabilize buried anions lacking a hydrogen-bonded partner. Finally, anion–quadrupole interactions are observed not just in proteins but also in DNA/RNA–protein and lipid–water–membrane-protein interfaces. Although less frequent than salt bridges, our conservation analysis, especially among remote homologs, suggests that in many instances anion–quadrupole-interacting residues remain highly conserved and likely contribute to the folding of the polypeptide chain. Many critical contacts at interfaces mediated by anion–quadrupole interactions have also been observed here, and, in general, the numerous examples we came across provide a good source of information with which one can experimentally further probe the role of anion–quadrupole interactions in molecular recognition, folding, and assembly.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This study has received funding support from the National Science Foundation/EPSCoR Award No. AII-1355423 and the state of South Dakota, through BioSNTR (Biochemical Spatiotemporal NeTwork Resource). Additional support was provided by the National Institutes of Health (NIH) grant 1R15GM116040–01A1 to SC and the Department of Chemistry & Biochemistry startup funds to SC. Authors thank the SDSU academic computing facility for the access to the Linux cluster for all the energy calculations. Authors also thank the anonymous referees for their valuable suggestions.

REFERENCES

(1). Kim KS, Tarakeshwar P, Lee JY. Molecular Clusters of pi-Systems: Theoretical Studies of Structures, Spectra, and Origin of Interaction Energies. Chem. Rev. 2000; 100:4145–4186. [PubMed: 11749343]

(2). Jain A, Ramanathan V, Sankararamakrishnan R. Lone pair … pi interactions between water oxygens and aromatic residues: quantum chemical studies based on high-resolution protein structures and model compounds. Protein Sci. 2009; 18:595–605. [PubMed: 19241386]

(3). Dougherty DA. Cation-pi interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. Science. 1996; 271:163–168. [PubMed: 8539615]

(4). Sinnokrot MO, Sherrill CD. High-accuracy quantum mechanical studies of pi-pi interactions in benzene dimers. J. Phys. Chem. A. 2006; 110:10656–10668. [PubMed: 16970354]

(5). McGaughey GB, Gagne M, Rappe AK. pi-Stacking interactions. Alive and well in proteins. J. Biol. Chem. 1998; 273:15458–15463. [PubMed: 9624131]

(6). Martinez CR, Iverson BL. Rethinking the term “pi-stacking”. Chemical Sci. 2012; 3:2191–2201.

(7). Bartlett GJ, Choudhary A, Raines RT, Woolfson DN. n→ π * interactions in proteins. Nat. Chem. Biol. 2010; 6:615–620. [PubMed: 20622857]

(8). Nishio M, Umezawa Y, Fantini J, Weiss MS, Chakrabarti P. CH-pi hydrogen bonds in biological macromolecules. Phys. Chem. Chem. Phys. 2014; 16:12648–12683. [PubMed: 24836323]

(9). Ozawa T, Okazaki K. CH/pi hydrogen bonds determine the selectivity of the Src homology 2 domain to tyrosine phosphotyrosyl peptides: an ab initio fragment molecular orbital study. J. Comp. Chem. 2008; 29:2656–2666. [PubMed: 18484636]

(10). Tsuzuki S, Honda K, Uchimaru T, Mikami M, Tanabe K. The magnitude of the CH/π interaction between benzene and some model hydrocarbons. J. Am. Chem. Soc. 2000; 122:3746–3753.

(11). Steiner T, Koellner G. Hydrogen bonds with pi-acceptors in proteins: frequencies and role in stabilizing local 3D structures. J. Mol. Biol. 2001; 305:535–557. [PubMed: 11152611]

(12). Brandl M, Weiss MS, Jabs A, Suhnel J, Hilgenfeld R. C-H...pi-interactions in proteins. J. Mol. Biol. 2001; 307:357–377. [PubMed: 11243825]

(13). Mitchell JB, Nandi CL, McDonald IK, Thornton JM, Price SL. Amino/aromatic interactions in proteins: is the evidence stacked against hydrogen bonding? J. Mol. Biol. 1994; 239:315–331. [PubMed: 8196060]

(14). Singh SK, Das A. The n → pi* interaction: a rapidly emerging non-covalent *interaction*. Phys. Chem. Chem. Phys. 2015; 17:9596–9612. [PubMed: 25776003]

(15). Egli M, Gessner RV. Stereoelectronic effects of deoxyribose 04’ on DNA *conformation*. Proc. Natl. Acad. Sci. U. S. A. 1995; 92:180–184. [PubMed: 7816812]

(16). Egli M, Sarkhel S. Lone pair-aromatic interactions: To stabilize or not to stabilize. Acc. Chem. Res. 2007; 40:197–205. [PubMed: 17370991]

(17). Mooibroek TJ, Gamez P, Reedijk J. Lone pair-π interactions: a new supramolecular bond? CrystEngComm. 2008; 10:1501–1515.

(18). Gallivan JP, Dougherty DA. Cation-pi interactions in structural biology. Proc. Natil. Acad. Sci. U. S. A. 1999; 96:9459–9464.

(19). Crowley PB, Golovin A. Cation-pi interactions in protein-protein interfaces. Proteins. 2005; 59:231–239. [PubMed: 15726638]

(20). Wintjens R, Lievin J, Rooman M, Buisine E. Contribution of cation-pi interactions to the stability of protein-DNA complexes. J. Mol. Biol. 2000; 302:395–410. [PubMed: 10970741]

(21). Zhong W, Gallivan JP, Zhang Y, Li L, Lester HA, Dougherty DA. From ab initio quantum mechanics to molecular neurobiology: a cation-pi binding site in the nicotinic receptor. Proc. Natl. Acad. Sci. U. S. A. 1998; 95:12088–12093. [PubMed: 9770444]

(22). Xiu X, Puskar NL, Shanata JA, Lester HA, Dougherty DA. Nicotine binding to brain receptors requires a strong cation-pi interaction. Nature. 2009; 458:534–537. [PubMed: 19252481]

(23). Estarellas C, Frontera A, Quinonero D, Deya PM. Relevant anion-pi interactions in biological systems: the case of urate oxidase. Angew. Chem. Int. Ed. Engl. 2011; 50:415–418. [PubMed: 21132687]

(24). Zhao Y, Domoto Y, Orentas E, Beuchat C, Emery D, Mareda J, Sakai N, Matile S. Catalysis with anion-π interactions. Angewandte Chemie. 2013; 125:10124–10127.

(25). Philip V, Harris J, Adams R, Nguyen D, Spiers J, Baudry J, Howell EE, Hinde RJ. A survey of aspartate-phenylalanine and glutamate-phenylalanine interactions in the protein data bank: searching for anion-pi pairs. Biochemistry. 2011; 50:2939–2950. [PubMed: 21366334]

(26). Neel AJ, Hilton MJ, Sigman MS, Toste FD. Exploiting non-covalent pi interactions for catalyst design. Nature. 2017; 543:637–646. [PubMed: 28358089]

(27). Choy MS, Li Y, Machado LE, Kunze MB, Connors CR, Wei X, Lindorff-Larsen K, Page R, Peti W. Conformational Rigidity and Protein Dynamics at Distinct Timescales Regulate PTP1B Activity and Allostery. Mol. cell. 2017; 65:644–658 e645. [PubMed: 28212750]

(28). Mullins EA, Shi R, Parsons ZD, Yuen PK, David SS, Igarashi Y, Eichman BF. The DNA glycosylase AlkD uses a non-base-flipping mechanism to excise bulky lesions. Nature. 2015; 527:254–258. [PubMed: 26524531]

(29). Baker EG, Williams C, Hudson KL, Bartlett GJ, Heal JW, Porter Goff KL, Sessions RB, Crump MP, Woolfson DN. Engineering protein stability with atomic precision in a monomeric miniprotein. Nat. Chem. Biol. 2017; 13:764–770. [PubMed: 28530710]

(30). Tateishi-Karimata H, Ohyama T, Muraoka T, Podbevsek P, Wawro AM, Tanaka S, Nakano SI, Kinbara K, Plavec J, Sugimoto N. Newly characterized interaction stabilizes DNA structure: oligoethylene glycols stabilize G-quadruplexes CH-pi interactions. Nucleic Acids Res. 2017; 45:7021–7030. [PubMed: 28453855]

(31). Gallivan JP, Dougherty DA. A Computational Study of Cation-π Interactions vs Salt Bridges in Aqueous Media: Implications for Protein Engineering. J. Am. Chem. Soc. 2000; 122:870–874.

(32). Waldburger CD, Schildbach JF, Sauer RT. Are buried salt bridges important for protein stability and conformational specificity? Nat. Struct. Mol. Biol. 1995; 2:122–128.

(33). Sindelar CV, Hendsch ZS, Tidor B. Effects of salt bridges on protein structure and design. Protein Sci. 1998; 7:1898–1914. [PubMed: 9761471]

(34). Kumar S, Nussinov R. Salt bridge stability in monomeric proteins. J. Mol. Biol. 1999; 293:1241–1255. [PubMed: 10547298]

(35). Tomlinson JH, Ullah S, Hansen PE, Williamson MP. Characterization of salt bridges to lysines in the protein G B1 domain. J. Am. Chem. Soc. 2009; 131:4674–4684. [PubMed: 19281232]

(36). Estarellas C, Frontera A, Quiñonero D, Deyà PM. Anion? π Interactions in *Flavoproteins*. Chemistry-An Asian Journal. 2011; 6:2316–2318.

(37). Breberina LM, Milčić MK, Nikolić MR, Stojanović SĐ. Contribution of anion-π interactions to the stability of Sm/LSm proteins. JBIC J. Biol. Inorg. Chem. 2015; 20:475–485. [PubMed: 25502146]

(38). Zlatović MV, Borozan SZ, Nikolić MR, Stojanović SĐ. Anion-π interactions in protein-porphyrin complexes. RSC Advances. 2015; 5:38361–38372.

(39). Chawla M, Chermak E, Zhang Q, Bujnicki JM, Oliva R, Cavallo L. Occurrence and stability of lone pair-pi stacking interactions between ribose and nucleobases in functional RNAs. Nucleic Acids Res. 2017; 45:11019–11032. [PubMed: 28977572]

(40). Rogacheva ON, Izmailov SA, Slipchenko LV, Skrynnikov NR. A new structural arrangement in proteins involving lysine NH3(+) group and carbonyl. Scientific Rep. 2017; 7:16402.

(41). Valiev M, Bylaska EJ, Govind N, Kowalski K, Straatsma TP, Van Dam HJ, Wang D, Nieplocha J, Apra E, Windus TL. NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations. Comp. Phys. Comm. 2010; 181:1477–1489.

(42). O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. J Cheminform. 2011; 3:33. [PubMed: 21982300]

(43). Grimme S, Antony J, Ehrlich S, Krieg H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J. Chem. Phys. 2010; 132:154104. [PubMed: 20423165]

(44). Frisch M, Trucks G, Schlegel H, Scuseria G, Robb M, Cheeseman J, Scalmani G, Barone V, Mennucci B, Petersson G. Gaussian 09, Revision C. 01, Gaussian, Inc., Wallingford CT, 2009 Search PubMed;(b) Y. Zhao and DG Truhlar. Theor. Chem. Acc. 2008; 120:215.

(45). McDonald IK, Thornton JM. Satisfying hydrogen bonding potential in proteins. J. Mol. Biol. 1994; 238:777–793. [PubMed: 8182748]

(46). Wang G, Dunbrack RL Jr.. PISCES: a protein sequence culling server. Bioinformatics. 2003; 19:1589–1591. [PubMed: 12912846]

(47). Ehrlich S, Moellmann J, Grimme S. Dispersion-corrected density functional theory for aromatic interactions in complex systems. Acc. Chem. Res. 2013; 46:916–926. [PubMed: 22702344]

(48). Hinderaker MP, Raines RT. An electronic effect on protein structure. Protein Sci. 2003; 12:1188–1194. [PubMed: 12761389]

(49). Horng JC, Raines RT. Stereoelectronic effects on polyproline conformation. Protein Sci. 2006; 15:74–83. [PubMed: 16373476]

(50). Shortle D. Propensities, probabilities, and the Boltzmann hypothesis. Protein Sci. 2003; 12:1298–1302. [PubMed: 12761401]

(51). Macias AT, Mackerell AD Jr.. CH/pi interactions involving aromatic amino acids: refinement of the CHARMM tryptophan force field. J. Comp. Chem. 2005; 26:1452–1463. [PubMed: 16088926]

(52). MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FT, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B. 1998; 102:3586–3616. [PubMed: 24889800]

(53). Sangha AK, Dong J, Williamson L, Hashiguchi T, Saphire EO, Crowe JE Jr., Meiler J. Role of Non-local Interactions between CDR Loops in Binding Affinity of MR78 Antibody to Marburg Virus Glycoprotein. Structure. 2017; 25:1820–1828 e1822. [PubMed: 29153506]

(54). Staus DP, Strachan RT, Manglik A, Pani B, Kahsai AW, Kim TH, Wingler LM, Ahn S, Chatterjee A, Masoudi A, Kruse AC, Pardon E, Steyaert J, Weis WI, Prosser RS, Kobilka BK, Costa T, Lefkowitz RJ. Allosteric nanobodies reveal the dynamic range and diverse mechanisms of G-protein-coupled receptor activation. Nature. 2016; 535:448–452. [PubMed: 27409812]

(55). Panasik N Jr., Fleming PJ, Rose GD. Hydrogen-bonded turns in proteins: the case for a recount. Protein Sci. 2005; 14:2910–2914. [PubMed: 16251367]

(56). Baran D, Pszolla MG, Lapidoth GD, Norn C, Dym O, Unger T, Albeck S, Tyka MD, Fleishman SJ. Principles for computational design of binding antibodies. Proc. Natl. Acad. Sci. U. S. A. 2017; 114:10900–10905. [PubMed: 28973872]

(57). Ferreiro DU, Komives EA, Wolynes PG. Frustration in biomolecules. Q. Rev. Biophys. 2014; 47:285–363. [PubMed: 25225856]

(58). Brosey CA, Ho C, Long WZ, Singh S, Burnett K, Hura GL, Nix JC, Bowman GR, Ellenberger T, Tainer JA. Defining NADH-Driven Allostery Regulating Apoptosis-Inducing Factor. Structure. 2016; 24:2067–2079. [PubMed: 27818101]

(59). Liu X, Speckhard DC, Shepherd TR, Sun YJ, Hengel SR, Yu L, Fowler CA, Gakhar L, Fuentes EJ. Distinct Roles for Conformational Dynamics in Protein-Ligand Interactions. Structure. 2016; 24:2053–2066. [PubMed: 27998539]

(60). Ji Q, Chen PJ, Qin G, Deng X, Hao Z, Wawrzak Z, Yeo WS, Quang JW, Cho H, Luo GZ, Weng X, You Q, Luan CH, Yang X, Bae T, Yu K, Jiang H, He C. Structure and mechanism of the essential two-component signal-transduction system WalKR in Staphylococcus aureus. Nature Commun. 2016; 7:11000. [PubMed: 26987594]

(61). Tencer AH, Gatchalian J, Klein BJ, Khan A, Zhang Y, Strahl BD, van Wely KHM, Kutateladze TG. A Unique pH-Dependent Recognition of Methylated Histone H3K4 by PPS and DIDO. Structure. 2017; 25:1530–1539 e1533. [PubMed: 28919441]

(62). Killian JA, von Heijne G. How proteins adapt to a membrane-water interface. Trends Biochem. Sci. 2000; 25:429–434. [PubMed: 10973056]

(63). Norimatsu Y, Hasegawa K, Shimizu N, Toyoshima C. Protein-phospholipid interplay revealed with crystals of a calcium pump. Nature. 2017; 545:193–198. [PubMed: 28467821]

(64). Rujas E, Insausti S, Garcia-Porras M, Sanchez-Eugenia R, Tsumoto K, Nieva JL, Caaveiro JM. Functional Contacts between MPER and the Anti-HIV-1 Broadly Neutralizing Antibody 4E10 Extend into the Core of the Membrane. J. Mol. Biol. 2017; 429:1213–1226. [PubMed: 28300601]

(65). Stockbridge RB, Kolmakova-Partensky L, Shane T, Koide A, Koide S, Miller C, Newstead S. Crystal structures of a double-barrelled fluoride ion channel. Nature. 2015; 525:548–551. [PubMed: 26344196]

(66). Johnson ZL, Chen J. Structural Basis of Substrate Recognition by the Multidrug Resistance Protein MRP1. Cell. 2017; 168:1075–1085 e1079. [PubMed: 28238471]

(67). Matyskiela ME, Lu G, Ito T, Pagarigan B, Lu CC, Miller K, Fang W, Wang NY, Nguyen D, Houston J, Carmel G, Tran T, Riley M, Nosaka L, Lander GC, Gaidarova S, Xu S, Ruchelman AL, Handa H, Carmichael J, Daniel TO, Cathers BE, Lopez-Girona A, Chamberlain PP. A novel cereblon modulator recruits GSPT1 to the CRL4(CRBN) ubiquitin ligase. Nature. 2016; 535:252–257. [PubMed: 27338790]

(68). Dhawan NS, Scopton AP, Dar AC. Small molecule stabilization of the KSR inactive state antagonizes oncogenic Ras signalling. Nature. 2016; 537:112–116. [PubMed: 27556948]

(69). Xu X, Wang S, Hu YX, McKay DB. The periplasmic bacterial molecular chaperone SurA adapts its structure to bind peptides in different conformations to assert a sequence preference for aromatic residues. J. Mol. Biol. 2007; 373:367–381. [PubMed: 17825319]

(70). Miles LA, Crespi GA, Doughty L, Parker MW. Bapineuzumab captures the N-terminus of the Alzheimer's disease amyloid-beta peptide in a helical conformation. Scientific Rep. 2013; 3:1302.

(71). Gray HB, Winkler JR. Hole hopping through tyrosine/tryptophan chains protects proteins from oxidative damage. Proc. Natl. Acad. Sci. U. S. A. 2015; 112:10920–10925. [PubMed: 26195784]

(72). Nguyen MN, Pradhan MR, Verma C, Zhong P. The interfacial character of antibody paratopes: analysis of antibody-antigen structures. Bioinformatics. 2017; 33:2971–2976. [PubMed: 28633399]

(73). Boamah D, Lin T, Poppinga FA, Basu S, Rahman S, Essel F, Chakravarty S. Characteristics of a PHD Finger Subtype. Biochemistry. 2018; 57:525–539. [PubMed: 29253329]

(74). Zarrine-Afsar A, Larson SM, Davidson AR. The family feud: do proteins with similar structures fold via the same pathway? Curr. Opin. Struct. Biol. 2005; 15:42–49. [PubMed: 15718132]

(75). Plaxco KW, Larson S, Ruczinski I, Riddle DS, Thayer EC, Buchwitz B, Davidson AR, Baker D. Evolutionary conservation in protein folding kinetics. J. Mol. Biol. 2000; 298:303–312. [PubMed: 10764599]

(76). Gunasekaran K, Eyles SJ, Hagler AT, Gierasch LM. Keeping it in the family: folding studies of related proteins. Curr. Opin. Struct. Biol. 2001; 11:83–93. [PubMed: 11179896]

(77). Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, Zecchina R, Sander C. Protein 3D structure computed from evolutionary sequence variation. PloS One. 2011; 6:e28766. [PubMed: 22163331]

(78). Anishchenko I, Ovchinnikov S, Kamisetty H, Baker D. Origins of coevolution between residues distant in protein 3D structures. Proc. Natl. Acad. Sci. U. S. A. 2017; 114:9122–9127. [PubMed: 28784799]

(79). Kamisetty H, Ovchinnikov S, Baker D. Assessing the utility of coevolutionbased residue-residue contact predictions in a sequence- and structure-rich era. Proc. Natil. Acad. Sci. U. S. A. 2013; 110:15674–15679.

(80). Ovchinnikov S, Kamisetty H, Baker D. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. eLife. 2014; 3:e02030. [PubMed: 24842992]

(81). Craven TW, Cho MK, Traaseth NJ, Bonneau R, Kirshenbaum K. A Miniature Protein Stabilized by a Cation-pi Interaction Network. J. Am. Chem. Soc. 2016; 138:1543–1550. [PubMed: 26812069]

(82). Biel JT, Thompson MC, Cunningham CN, Corn JE, Fraser JS. Flexibility and Design: Conformational Heterogeneity along the Evolutionary Trajectory of a Redesigned Ubiquitin. Structure. 2017; 25:739–749 e733. [PubMed: 28416112]

(83). Kortemme T, Morozov AV, Baker D. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. J. Mol. Biol. 2003; 326:1239–1259. [PubMed: 12589766]

(84). Misura KM, Morozov AV, Baker D. Analysis of anisotropic side-chain packing in proteins and application to high-resolution structure prediction. J. Mol. Biol. 2004; 342:651–664. [PubMed: 15327962]

(85). Zhang J, Zhang Y. A novel side-chain orientation dependent potential derived from random-walk reference state for protein fold selection and structure prediction. PloS One. 2010; 5:e15386. [PubMed: 21060880]

(86). Zhou H, Skolnick J. GOAP: a generalized orientation-dependent, all-atom statistical potential for protein structure prediction. Biophys. J. 2011; 101:2043–2052. [PubMed: 22004759]

(87). Huang J, MacKerell AD Jr.. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. J. Comp. Chem. 2013; 34:2135–2145. [PubMed: 23832629]

(88). Lobner E, Humm AS, Goritzer K, Mlynek G, Puchinger MG, Hasenhindl C, Ruker F, Traxlmayr MW, Djinovic-Carugo K, Obinger C. Fcab-HER2 Interaction: a Menage a Trois. Lessons from X-Ray and Solution Studies. Structure. 2017; 25:878–889 e875. [PubMed: 28528777]

(89). Piovesan D, Minervini G, Tosatto SC. The RING 2.0 web server for high quality residue interaction networks. Nucleic Acids Res. 2016; 44:W367–374. [PubMed: 27198219]

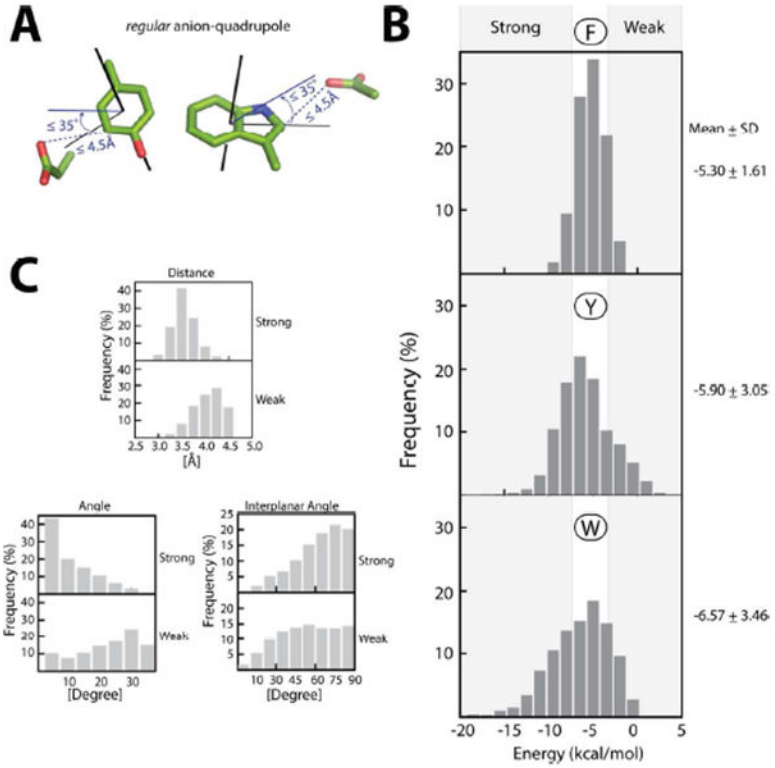


Figure 1. *Regular* anion–quadrupole interactions and their energies:

(A) The distance (≤ 4.5 Å) and angle (≤ 35°) detection criteria. **(B)** The distribution of the interaction energies for Phe (top), Tyr (middle), and Trp (bottom). Energy ranges for the strong and weak anion–quadrupole interactions are indicated by the gray background. **(C)** A comparison of the distribution of properties between strong and weak interactions: distances (top), angles (left, bottom), and interplanar angles (right, bottom).

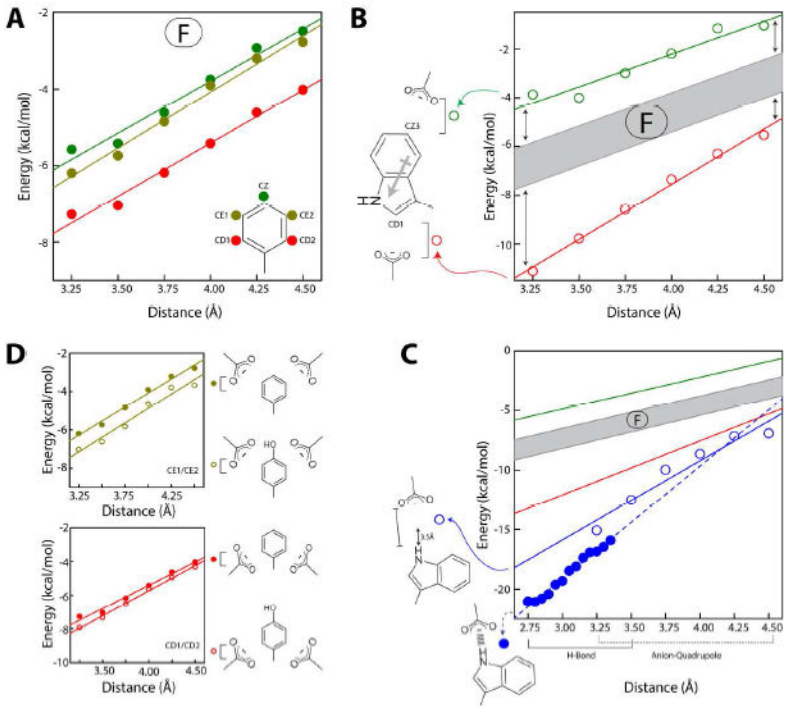


Figure 2. Distance dependence of the anion–quadrupole interaction energy:

The dependencies for Phe, Trp and Tyr are shown in **(A)**, **(B) - (C)**, and **(D)**, respectively, in which Phe ring atoms are represented by filled circles (green, pear and red) and Trp and Tyr are represented by empty circles. The area between the green and red lines in **(A)** is defined as the **F**-range which is shaded in gray in **(B)** and **(C)**. In **(C)**, the empty blue circles represent anion–quadrupole interactions of the Trp-NE1 atom while filled blue circles represent the energies of hydrogen bonds between the Trp-NE1 atom and anions. The green and red lines in **(C)** are the same as in **(B)**. The distance range for hydrogen bonds and anion–quadrupole interactions are also indicated in **(C)**. A comparison between the symmetry-related positions of Tyr and Phe are shown in **(D)**.

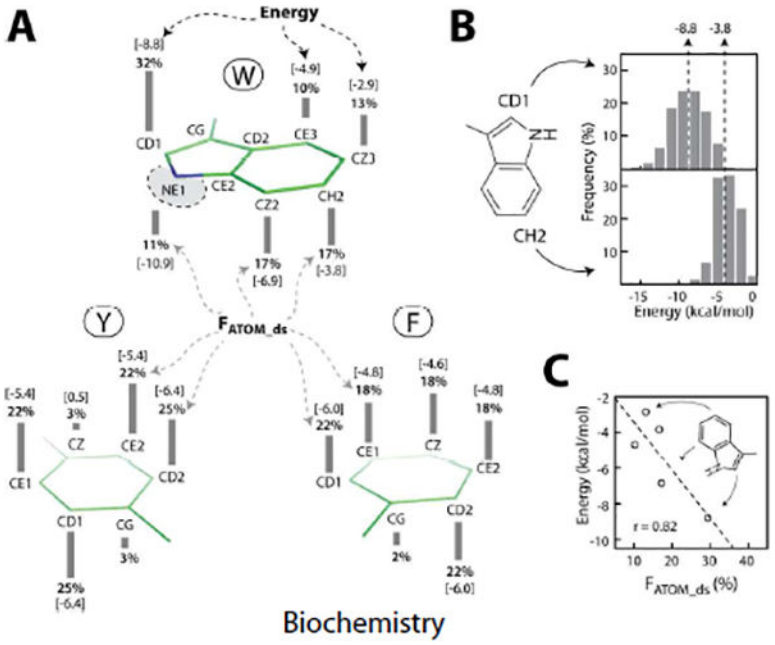


Figure 3. Ring atom participation frequency and interaction energy:

(A) The height of the bar ($F_{ATOM–ds}$) adjacent to the name of each atom represents the frequency (%) of its participation in anion–quadrupole interactions for that residue type, Trp (top), Tyr (bottom left), and Phe (bottom right). The number within the bracket above or below each bar represents the interaction energy (kcal/mol) when the indicated atom is closest to the interacting anion. **(B)** The distribution represents the energy of anion–quadrupole interactions when Trp-CD1 (upper) and Trp-CH2 (lower) atoms are closest to the interacting anion. The dotted lines with an arrow represent the mean interaction energy in kcal/mol. **(C)** The linear correlation (r = 0.82) between the energy and $F_{ATOM–ds}$ for tryptophan ring carbon atoms. For a qualitative agreement between the energy and the frequency, we use F_{ATOM_ds} in place of $\log[F_{ATOM\_ds}]$.

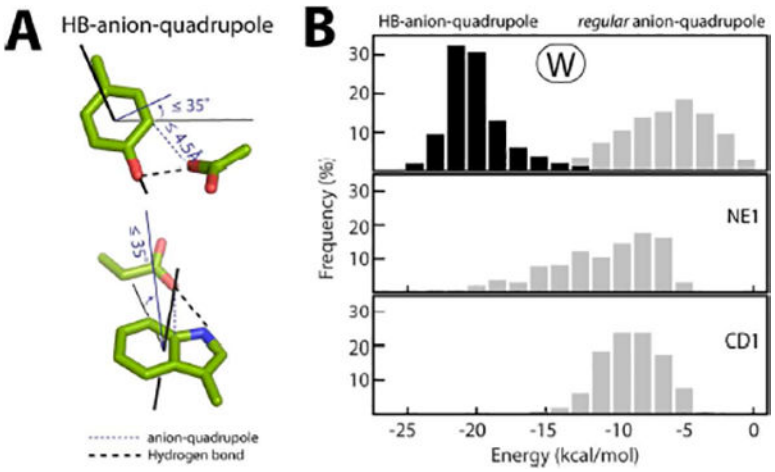


Figure 4. HB-anion–quadrupole interactions:

(A) HB-anion–quadrupole interactions satisfy the distance-angle detection criterion used here for anion–quadrupole interactions. **(B)** Comparison between the distribution of the interaction energy of anion–quadrupole and HB-anion–quadrupole interactions for Trp residues (top). The distribution of anion–quadrupole interaction energies for the Trp-NE1 (middle) and Trp-CD1 (bottom) atoms are shown below for comparison.

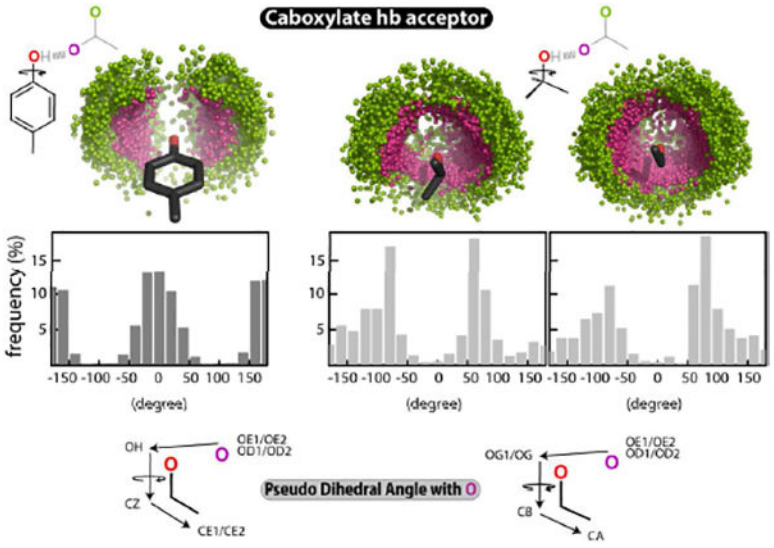


Figure 5. Tyr HB-anion–quadrupole orientation:

The spatial distribution of the carboxylate oxygen atoms (magenta and green, top) when hydrogen bonded with the donor -OH of Tyr (left), Thr (middle), and Ser (right). The acceptor carboxylate O-atom is in magenta, while the other carboxylate O-atom is in green. The distribution of the *pseudo-dihedral* angle **(θ_D)** between the 4 atoms **O-O-CZ-CE1/CE2** (Tyr, left) and **O-O-CB-CA** (Thr/Ser, center right) is shown in the middle. A drawing representing the θ_D is shown at the bottom, with Tyr (left) and Thr/Ser (right).

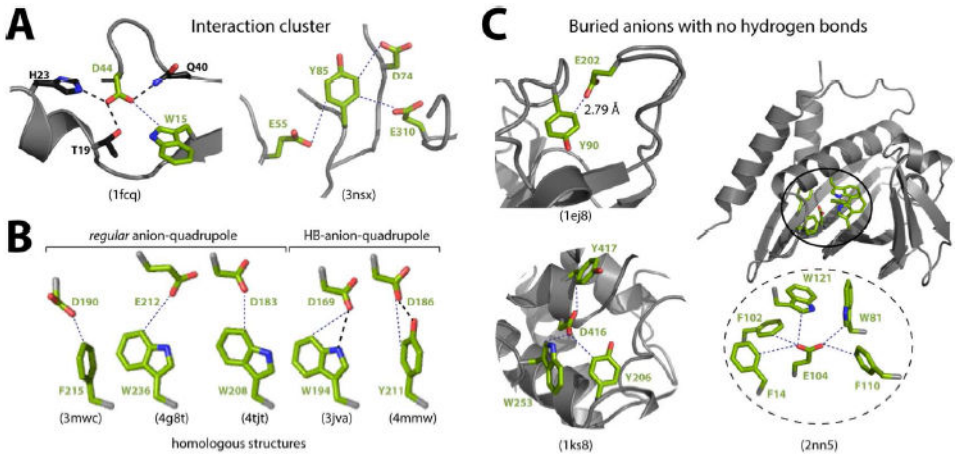


Figure 6. Interaction network and hydrogen bonding of anion–quadrupole interactions:
(A) The anion of an anion–quadrupole pair involved in three hydrogen bonds (left) and the aromatic ring of an anion–quadrupole pair interacting with three anions (right). **(B)** Regular anion–quadrupole and HB-anion–quadrupole interactions observed between conserved positions in homologous protein structures of Pfam MR_MLE_C domain. **(C)** Buried anions having no hydrogen bonds are surrounded by aromatic rings engaged in anion–quadrupole interactions. The pdb IDs are in braces.

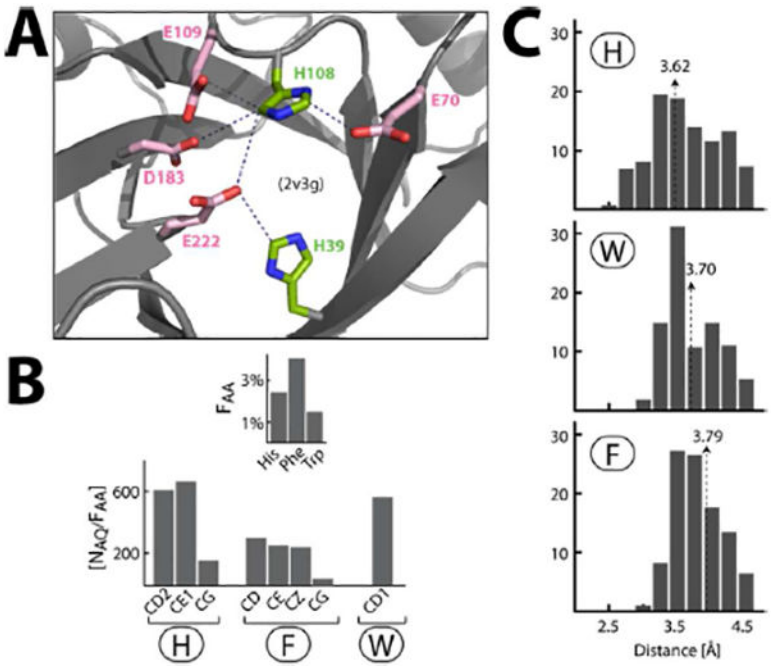


Figure 7. Anion–quadrupole interactions of histidine residues:

(A) Examples of histidine ring carbon atoms engaged in anion–quadrupole interactions. **(B)** Number of *a*nion–*q*uadrupole [N_{AQ}] interaction for each ring atom is normalized by the amino acid frequency [N_{AQ}/F_{AA}] in the dataset. The value of [N_{AQ}/F_{AA}] of ring atoms is compared between histidine, phenylalanine and tryptophan. **(C)** The distribution of the anion–quadrupole interaction distances for histidine ring carbon atoms (top), tryptophan-CD1 atom (middle) and phenylalanine ring-carbon atoms (bottom). The arrows indicate the mean of the respective distributions. The tryptophan-CD1 serves as a reference here as the energy of anion–quadrupole interaction of tryptophan-CD1 atom was observed to be the largest among all ring carbon atoms of phenylalanine, tyrosine and tryptophan.

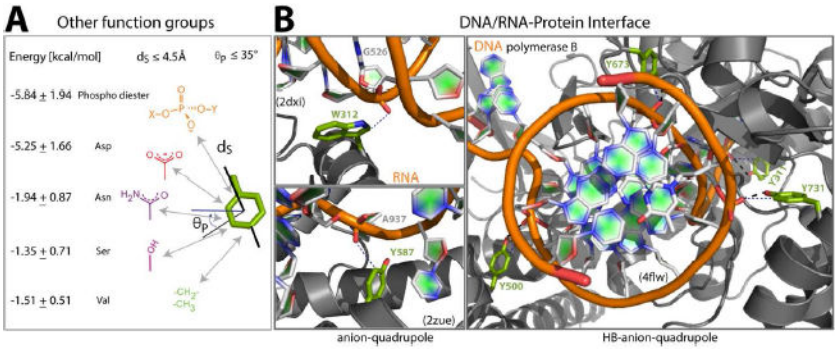


Figure 8. Other functional groups:

(A) The average energy of interaction (left) of other functional groups (right) with the Phe residue ring atoms. **(B)** Anion–quadrupole (left) and HB-anion–quadrupole (right) interactions with phosphoryl groups at DNA/RNA–protein interfaces.

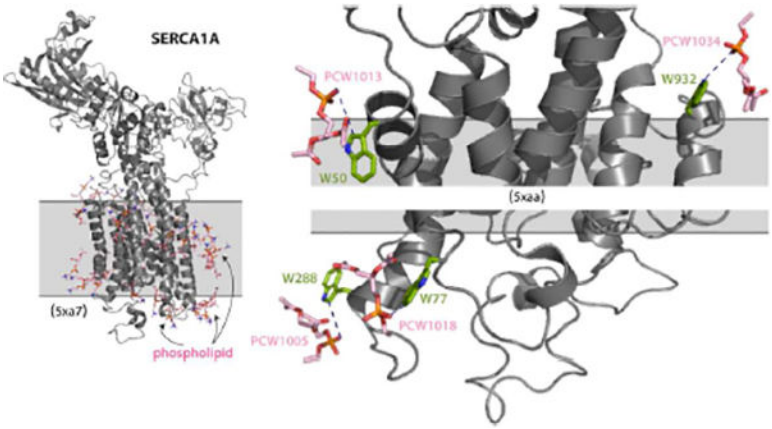


Figure 9. Anion–quadrupole interactions at the lipid-water interface of membrane proteins: Anion–quadrupole interactions between the phospholipid phosphoryl anion and the aromatic residues (green) at the lipid-water interface of SERC1A. Only one of the four structures of SERCIA is shown. The membrane bilayer is represented as a shaded region.

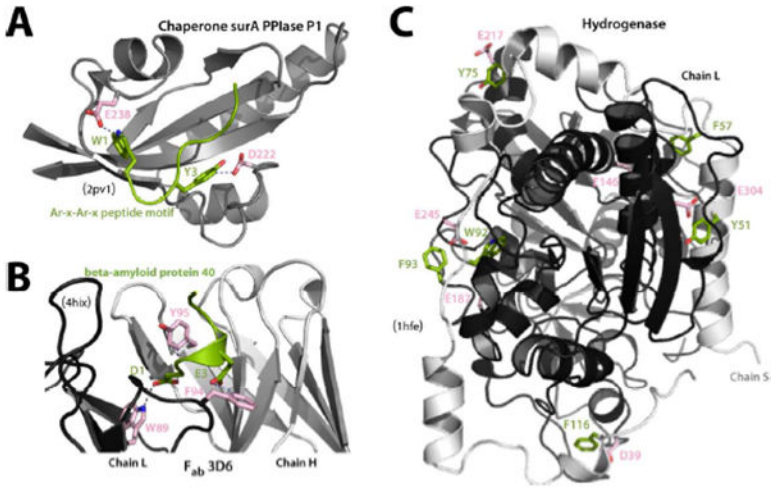


Figure 10. Anion–quadrupole interactions at protein-protein interfaces:

(A) Peptide recognition by the SurA PPIase domain and **(B)** antibody 3D6 involving anion–quadrupole interactions. **(C)** The hydrogenase small chain (chain S, light gray) interacts with the long chain (chain L, dark gray) by a number of anion–quadrupole interactions.

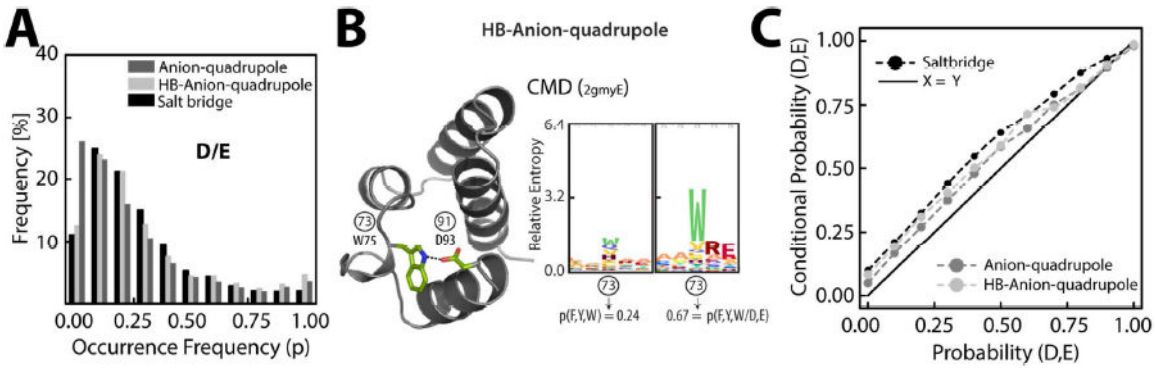


Figure 11. Conservation of anion–quadrupole, HB-anion–quadrupole, and salt bridge interactions:

(A) The distribution (bars) of the occurrence frequency ($0.0 \leq p \leq 1.0$) of residues (D/E) engaged in anion-quadrupole (dark gray), HB-anion–quadrupole (light gray), and salt bridge (black) interactions. **(B)** Sequence logos (right) representing the correlation between the positions of the CMD pfam domain (left) that are engaged in HB-anion–quadrupole interactions. The right panel of the sequence logos represents the increase in the occurrence frequency of aromatic residues [p(F,Y,W)] upon sequence perturbation (see Figure S9). The left panel of the logos represents unperturbed alignments. Circled numbers (73 and 91) represent alignment positions. **(C)** The correlation between the probability and the conditional probability of anion–quadrupole (dark gray), HB-anion–quadrupole (light gray) and salt bridge (black) interactions.

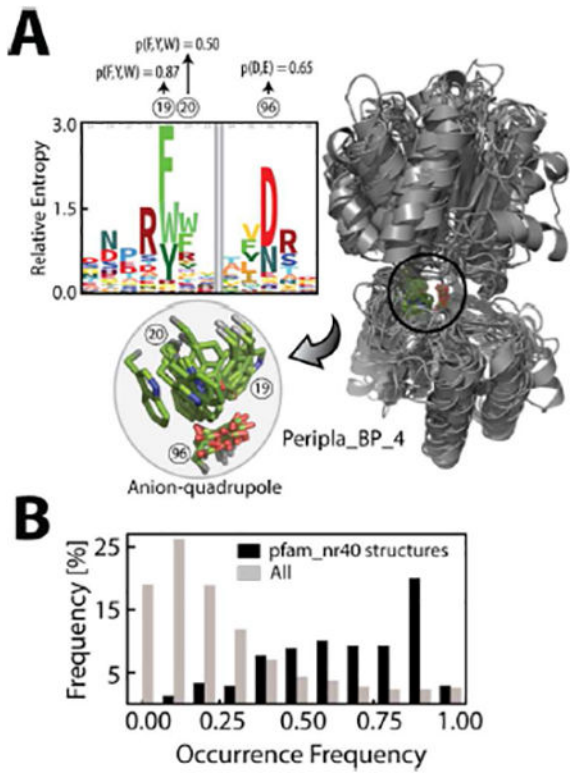


Figure 12. Conservation among remote homologs:

(A) Conserved anion–quadrupole-interacting residues (inset) in the pfam_nr40 structures of the Peripla_BP_4 domain. The occurrence frequency of these residues is displayed above the sequence logo. **(B)** The combined distribution (black bar) of the occurrence frequency of anion–quadrupole, HB-anion–quadrupole (D/E and F/Y/W), and salt bridge (D/E and R/K) residue positions when they are observed in the corresponding positions in at least two of the pfam_nr40 structures. The gray bars represent the combined background distributions of Figure 11A.

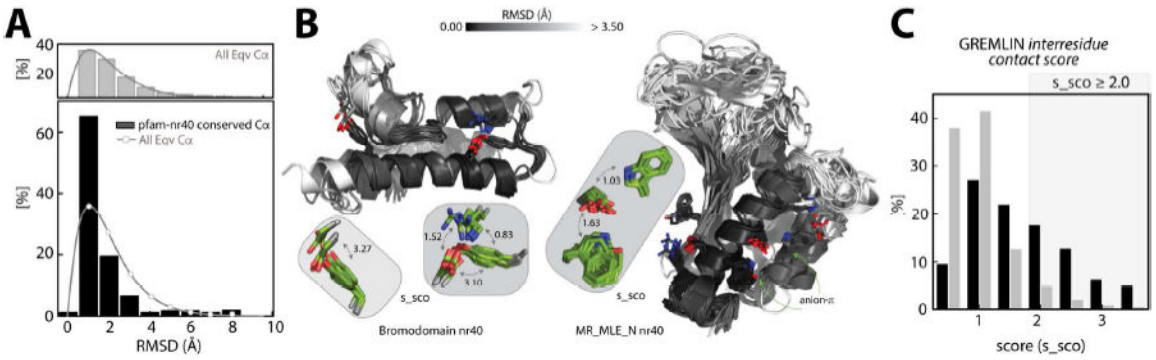


Figure 13. Structural deviations and residue-residue contact score:

(A) Root-mean-square deviation (RMSD) distribution (top, gray bars) of all equivalent Cα atoms in the structural alignment of pfam_nr40 structures. The RMSD distribution of Cα atoms (bottom, black bars) of conserved interacting positions (anion-quadrupole, HB-anion–quadrupole, and salt bridge) in the pfam_nr40 structures. The background distribution (gray line) of all equivalent Cα atoms is overlaid. **(B)** Structural deviation by RMSD (white, ≥ 3.5 Å and black, 0.0 A) in the pfam_nr40 structures of the bromodomain (left) and MR_MLE_N (right) domain. Anion–quadrupole-interacting positions in these structures are also shown in the inset in green with their GREMLIN s_sco scores indicated with curved, double-headed arrows. **(C)** The distribution of the s_sco scores (black bars) for positions conserved in at least two of the pfam_nr40 structures. Gray bars represent the background distribution of the GREMLIN s_sco scores in the dataset.