Salt Bridges: Geometrically Specific, Designable Interactions

Jason E. Donald[†], **Daniel W. Kulp**[†,‡], and **William F. DeGrado**
Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, PA

Abstract

Salt bridges occur frequently in proteins, providing conformational specificity and contributing to molecular recognition and catalysis. We present a comprehensive analysis of these interactions in protein structures by surveying a large database of protein structures. Salt bridges between Asp or Glu and His, Arg, or Lys display extremely well-defined geometric preferences. Several previously observed preferences are confirmed and others that were previously unrecognized are discovered. Salt bridges are explored for their preferences for different separations in sequence and in space, geometric preferences within proteins and at protein-protein interfaces, cooperativity in networked salt bridges, inclusion within metal-binding sites, preference for acidic electrons, apparent conformational side chain entropy reduction upon formation, and degree of burial. Salt bridges occur far more frequently between residues at close than distant sequence separations, but at close distances there remain strong preferences for salt bridges at specific separations. Specific types of complex salt bridges, involving three or more members, are also discovered. As we observe a strong relationship between the propensity to form a salt bridge and the placement of salt-bridging residues in protein sequences, we discuss the role that salt bridges might play in kinetically influencing protein folding and thermodynamically stabilizing the native conformation. We also develop a quantitative method to select appropriate crystal structure resolution and B-factor cutoffs. Detailed knowledge of these geometric and sequence dependences should aid de novo design and prediction algorithms.

Keywords

Structural bioinformatics; protein design; protein structure; protein stability; protein folding; arginine; lysine; histidine; side chain rotamers; networked salt bridges

Introduction

Although significant strides have been made[1–5], salt bridges are difficult to accurately predict and model. A salt bridge can be defined as an interaction between two groups of opposite charge in which at least one pair of heavy atoms is within hydrogen bonding distance. Salt bridges can contribute to protein stability[6–8], although the effect depends on the environment[9–11]. Both the high cost of dehydrating a basic residue and a carboxylate to form a salt bridge and the stringent geometric restraints placed by the electrostatic and hydrogen-bonding interactions make predicting salt bridge interactions uniquely challenging.

Contact information for corresponding author, William F. DeGrado: Phone: 215-898-4590, Fax: 215-573-7229, wdegrado@mail.med.upenn.edu.
[†]These authors contributed equally to this work.
[‡]Current address: Biosciences Division, Los Alamos National Laboratory, Los Alamos, NM
Institution at which work was performed: University of Pennsylvania

From a design perspective, salt bridges can contribute to conformational specificity, as well as in positioning critical functional groups. For example, a complex lysine-aspartate-histidine interaction was employed to position a metal ligand in a family of designed metalloproteins[12,13]. In addition, salt bridges can serve as keystone interactions, in much the same way as disulfide bonds[14]. In membrane proteins, one expects salt bridges to be particularly important due to a smaller dehydration penalty (loss of favorable contacts with water) upon salt bridge formation. Charged groups become largely dehydrated when inserted into membranes[15], and therefore experience a smaller change in hydration between non-salt-bridging and salt-bridging states. There should also be a smaller effect due to solvent screening, strengthening salt bridge interactions. The T-cell receptor appears to use these features of membrane salt bridges for proper assembly and function[16].

Previously, several researchers have investigated the geometry and sequence dependence of salt bridges. When involving arginine, there are several potential interactions of the guanidinium group with a carboxylate of aspartate/glutamate. The side on and end on interactions (Figure 1A), are bidentate configurations involving the formation of a ring of six heavy atoms. These interactions have been observed experimentally and predicted to be the lowest energy states based on QM calculations[2,3]. An additional interaction, termed here "backside", is monodentate with respect to the oxygen engaging the $N\eta1$ hydrogens closest to $N\varepsilon$. In a similar manner, histidine can form a salt bridge using either of its two side chain nitrogens, $N\delta$ or $N\varepsilon$ (Figure 1B).

For lysine, quantum mechanical calculations show distinct conformational states with respect to rotation about the $C\varepsilon$-$N\zeta$ bond. The terminal ammonium hydrogens occupy staggered orientations relative to the $C\varepsilon$ substituents[17] (Figure 1C). We therefore expect to observe staggered hydrogen bond acceptors that interact with these hydrogens, although to the best of our knowledge this has not been observed in previous analyses of salt bridges.

The carboxylate of aspartate and glutamate has two non-bonded lone pairs of electrons (Figure 1D). The syn lone pair is more than four orders of magnitude more basic than the peripheral anti lone pairs[18], and only syn lone pairs can be used to make a bidentate hydrogen bonding interaction. One would therefore expect the syn lone pair to be used primarily in hydrogen bonding interactions. Structural studies[19] and quantum mechanical calculations[20], however, suggest that carboxylates in water and protein environments tend to donate hydrogen bonds from both types of lone pairs nearly equally, with only a slightly greater propensity to use syn lone pairs.

In addition to geometric descriptions, one can consider how separation in primary sequence affects salt bridges formation. As with any geometrically constricting interaction one expects a preference for forming the interaction between residues that are close in sequence, thereby minimizing the entropy of loop closure[14]. Strong biases towards the formation of salt bridges between sequentially proximal side chains have been observed[4,21], but details of these preferences remain to be elucidated.

Given recent increases in the size of the protein database, it is now possible to ask much more refined, specific questions including how distance in structure and sequence influences the geometry of salt bridge formation. Here, we expand on previous analyses of salt bridges by including the third most populated configuration for arginine as well as the location of hydrogen bond acceptors around both histidine and lysine. These interactions are contrasted with those of the carboxyamide functional group of aspargine/glutamine around basic residues. The geometry of interaction is found to depend on both the distance between the residue backbones and the separation in primary sequence. We observe strong, specific biases for salt bridges between residues that are close in sequence, which we define as "local

salt bridges." There are strong main chain conformation biases as well as side chain torsional angle biases associated with the formation of local salt bridges. Salt bridges involving three members are often designated as "complex" or "networked" salt bridges. We find distinct sequence preferences for networked salt bridges that bridge secondary structure elements. We anticipate this information will be of utility in a wide range of protein design problems.

Materials and Methods

Salt Bridge Database

To determine an acceptable crystal structure resolution and Debye-Waller B-factor (atomic mean-square displacement in a solved crystal structure) cutoff, a dataset of 8,410 proteins monomers with resolution no less than 3.0 Å, an R-factor of no greater than 0.25, and sequence identity of 30% or less was used. This dataset was selected using the PISCES server[22] on September 16, 2010. In analyzing the crystal structures, the mean angle of carboxylate oxygens out of the guanidinium plane, θ, was used as a metric of salt bridge quality. To remove bias to the calculation due to stacking interactions that should not occur within the guanidinium plane, oxygens that are more than 36° (π/5 radians) outside of the guanidinium plane were excluded.

The B-factors in each structure were normalized by calculating the number of standard deviations an atomic B-factor is from the mean value for all non-hydrogen atoms in the structure[23]. The B-factor is an atomic property, so to define a B-factor for the basic residues, a representative carbon atom from the functional group was selected for each residue (Cζ of Arg, Cε of His, and Cε of Lys). For acidic and polar residues, the B-factor of the atom of interest (e.g. Oδ of Asp) was likewise normalized. Residues or atoms with a normalized B-factor less than zero have a B-factor that is smaller than the all-atom average value for the protein.

Using the selected resolution limit of 1.8 Å (see Results), the primary dataset of protein structures is a set of 3,644 protein monomers. These structures have a resolution of no less than 1.8 Å, an R-factor of no greater than 0.25, and sequence identity of 30% or less. This dataset was selected using the PISCES server on May 2, 2009[22].

Acidic (Asp and Glu) and isosteric polar (Asn and Gln) residues were defined as interacting with a basic residue if any atoms are within 4 Å of the basic residue side chain nitrogens[1]. Barlow and Thornton used a cutoff of 4 Å between N-O atom pairs as a definition of salt bridge formation[1], and later work from that group used a similar distance-based definition using van der Walls radii[2,3]. Kumar and Nussinov have also used a more restrictive cutoff of 4 Å between functional group centroids[4,5]. This allowed them to consider only "good" salt bridge geometries[4] in their analysis. As we are interested in the distribution of all salt bridges within folded proteins, we preferred the definition of Barlow and Thornton. In addition, the centroid distance restriction removes almost all "backside" Arg salt bridges, one of the major types of Arg based salt bridges, from the dataset (see Results: Arginine).

Results of lattice contacts were calculated using the primary database where contacts were defined as basic residues on the protein monomer contacting acidic residues in different asymmetric units. To compare to salt bridges found in protein-protein interfaces, 601 protein dimer structures were selected using the Dockground server[24]. Structures were only included if they met the resolution, R-free, and sequence identity cutoffs of the monomeric dataset. For His, because of the difficulty of differentiating nitrogen and carbon atoms in protein crystal structures, the program reduce[25] was used to systematically select orientations of the imidazole ring of His for analysis.

Environmental descriptors

Secondary structure and solvent accessibility were defined using DSSP[26]. To determine the solvent accessibility that residues are experiencing in the crystal lattice, all asymmetric units near the chain of interest were included in the solvent accessibility calculation. Relative accessible surface area (ASA) calculations[27] for a residue X were calculated by dividing by the ASA observed in a Gly-X-Gly extended conformation. This allows different residues to be compared on the same relative ASA scale.

Acidic residue syn/anti ratio

Each salt-bridging oxygen was classified into one of four categories based on which lone pair it brought closest to a basic residue: syn lone pair(s) only, anti lone pair(s) only, both syn and anti lone pairs, or neither type of lone pair. To qualify as syn or anti, the salt bridging nitrogen was required to be within 3.25 Å. Use of syn and anti lone pairs was determined by the side of the $C\gamma$-$O\delta$ or $C\delta$-$O\varepsilon$ bond that the nitrogen was found on (Figure 1D). Syn lone pairs are on the side closest to the other oxygen of the acidic residue while anti lone pairs are on the opposite side[18] (Figure 1D).

Local sequence separation

The expected number of local salt bridges was calculated by assuming that the probability of forming a local salt bridge between two residues is constant, then multiplying by an expected number of potential salt bridge partners given the propensity of a residue to be in a particular secondary structure and sequence separation:

$$\text{Exp}_{a,b,ssa,ssb,sep} = \left(\frac{SB_{a,b}}{Pairs_{a,b}} \right) Frac_{a,ssa} Frac_{b,ssb} Spacing_{ssa,ssb,sep}$$

$\text{Exp}_{a,b,ssa,ssb,sep}$ is the expected number of salt bridges given the acidic residue (a), basic residue (b), acidic residue secondary structure (ssa), basic residue secondary structure (ssb), and sequence separation (sep). $SB_{a,b}$ is the number of a, b local salt bridges (sequence separation less than 5). $Pairs_{a,b}$ is the number of a, b pairs that are local. The ratio of these terms gives the expected frequency of salt bridges given that residues a and b are local. $Frac_{a,ssa}$ is the fraction of residues in secondary structure ssa that are the acidic residue, a. $Spacing_{ssa,ssb,sep}$ is the number of all 20 amino acids found at a given sequence separation with the specified secondary structures. The last three terms define an expected number of times that residues a and b would be local given their identity, secondary structure, and sequence separation.

Enrichment in Sequence

In correlating the percentage salt bridge formation to enrichment in local sequence, we extend the work of Meier and Burkhard on coiled coils[28]. In addition to a correction for chain length, we also correct for the frequency of amino acids within secondary structures. Only residues with a substantial number of side chain-side chain contacts were included: i−4, i−3, i+1, i+3, and i+4 for helices and i−3, i−2, i+2, and i+3 for sheets.

To calculate the enrichment in local sequence for a salt bridge pair (e.g. $Arg_i Asp_{i+4}$ in helices), we first calculate the observed ratio of the number of the residue pairs at that sequence separation to the number of the residue pairs at non-local (greater than four) sequence separations in the given secondary structure. Using this ratio controls for the different amino acid frequencies within secondary structures.

$$Obs(a,b,sep,ss) = \frac{NumberLocal(a,b,sep,ss)}{NumberNonlocal(a,b,ss)}$$

where a is the acidic residue, b is the basic residue, sep is the sequence separation, ss is the secondary structure, NumberLocal(a,b,sep,ss) is number of a, b pairs at the sequence separation sep and secondary structure ss, and NumberNonlocal(a,b) is the number of a, b pairs with secondary structure ss that are greater than four residues apart in the database. This value is then compared to an expected value that controls for the size of the chains:

$$Exp(a,b,sep,ss) = \frac{1}{\sum\limits_{P=1}^{M} N(a,b)} \sum_{P=1}^{M} \sum_{Q=1}^{N(a,b)} \frac{NumberLocal_p(sep,ss)}{NumberNonlocal_p(ss)}$$

where M is the number of proteins in the database, N is the total number of a, b pairs in protein P, and the counts are for the protein P. The enrichment in local sequence is then taken as the fraction change of the observed versus the expected:

$$Enrichment = \frac{Obs(a,b,sep,ss) - Exp(a,b,sep,ss)}{Exp(a,b,sep,ss)}$$

Rotamer pseudo-entropy

A pseudo-entropy term estimating part of the side-chain conformational entropy (SCCE) loss was calculated for basic residue side chains in different secondary structures and sequence separations based upon Gibb's entropy equation:

$$S = -R \sum p_i \log p_i$$

where p_i is the probability of being in state i, in this case, to have rotamer i[29,30] and R is the gas constant. This term estimates a component of the SCCE loss for the longer basic residue that is due to the formation of the salt bridge. The basic residue side chain dihedral angles were calculated and each residue classified by its rotamer[31−33]. Each dihedral angle was classified as gauche minus, gauche plus, and trans, and the rotamer identified by the combination of these dihedral angles. The change in pseudo-entropy was calculated by subtracting the pseudo-entropy value for all Arg, His, or Lys with the same secondary structure (helix, sheet, or coil).

Cooperativity

The expected number of networked salt bridges and His-Asp/Glu-metal interactions is given by the following formula:

$$Exp_{ij} = N f_i f_j$$
$$f_i = \frac{Obs_i}{N}$$

where Exp_{ij} is the number expected, f_i is the probability of observing an interaction of type i, N is the total number of basic residues, and Obs_i is the number observed of type i. For this

calculation, only the basic residue was constrained to have a B-factor below 2. Otherwise, an interacting acidic side chain or metal that has a higher B-factor would not be counted.

Results

Selecting an appropriate dataset

Some atoms are positioned with higher confidence in x-ray crystallographic structures than others. A detailed analysis of salt bridge geometry will depend both on the quality of the dataset and the quantity of data points. In statistical analyses, a cutoff in resolution and/or Debye-Waller B-factor is generally selected, but the appropriate cutoffs can be challenging to select and often are chosen without justification. To be more quantitative, we selected a variable deriving from salt bridges themselves that can act as a guide of the quality of the structure.

To help choose an appropriate cutoff of resolution for inclusion in our database, we examined a number of parameters, and found that the angle of interaction between Arg and Asp/Glu was particularly sensitive to resolution. Arg-based salt bridges prefer carboxylate oxygens to be in the same plane as the Arg guanidinium[3]. In preparing the dataset, we noticed that low resolution salt bridge data tend to place oxygen atoms further out the Arg guanidinium plane than high resolution data. Such a trend could aid in determining an appropriate cutoff in resolution and B-factor. The mean angle of carboxylate oxygens out of the guanidinium plane, θ (Figure 1A legend), was calculated for structures in 0.3 Å resolution bins. As resolution decreases there is a slight, linear shift in the mean deviation of the angle up to approximately 2.1 Å. The change is, however, small, increasing from 7.5 degrees to 9.9 degrees over this increment. After 2.1 Å there is both an increase in the deviation to 11.0 degrees and an increase in the slope (change in deviation with respect to resolution). As an extra safeguard, only structures with resolution of 1.8 Å or better were included in the dataset.

We next considered a local measure of the quality of the resolved coordinates, the Debye-Waller B-factor (atomic mean-square displacement in a solved crystal structure). A larger B-factor reflects a greater variance in the positions of the atom within the protein crystal, such as a flexible solvent-exposed position with weak electron density[34]. To compare B-factor in different structures, B-factors were normalized by measuring the distance in standard deviations from the mean value for all non-hydrogen atoms in the structure (Methods). Analogous to the case for crystal structure resolution, as the normalized B-factor values increase, the value of $|\theta|$ increases (Figure 2B). A comparable trend is observed for raw B-factors (Supporting Figure S1). Using a similar cutoff to that used for structure resolution, salt bridges with B-factors with less than 2 standard deviations above the all-atom mean were retained in the dataset.

At this point, it is important to note that the number of salt bridges observed in the dataset, especially on the surface, may be biased by crystallization conditions such as cryo-freezing[35,36]. Likewise, as a static structure, it cannot provide dynamic information about the duration of a salt bridge in solution. However, as shown below, buried and surface salt bridges follow the same geometric constraints and thus inclusion of surface salt bridges should not bias the analysis of salt bridge geometry.

Arginine

Following Thornton and co-workers, we used spherical coordinates to describe the interaction of arginine (Arg) with aspartate/glutamate (Asp/Glu, Figure 1A)[3]. The interaction geometry was defined by a vector from the central gauanidino carbon (Cε) to a carboxylate oxygen atom. An angle, $\boldsymbol{\psi}$, describes the rotation of the vector projected onto

the plane of the guanidinium group. Another angle, **θ**, describes the degree to which the oxygen atom is out of the guanidinium plane. The two symmetrical carboxylate oxygens of Asp and Glu were treated identically unless otherwise stated.

As expected from previous work, we observed a strong preference for the end on and side on orientations of carboxylates with Arg. A plot of ψ and θ (Figure 3A) showed three strong peaks: a single peak at ψ = 90°, and two clusters centered at ψ = 180° and ψ = 300°, each consisting of a doublet with a small central feature. Each cluster peaks at the in-plane value, θ = 0°. The lack of data points in other regions of the ψ/θ plot speaks to the geometric specificity of this interaction.

Although the ψ/θ plot shows only the position of individual oxygen atoms and could simply reflect two separate monodentate interactions at each doublet (ψ = 180° and = 300°), other information suggested that the doublets are largely bidentate end on and side on interactions. In a bidentate interaction, the guanidinium and carboxylate planes should be close to coplanar to allow the hydrogens of the guanidinium to hydrogen bond directly to the syn lone pairs of the carboxylate. After correcting for the random expectation, we saw a strong bias for co-planar orientations, as described previously for Arg[3]. The strong geometric bias was observed only for the side on and end on interactions, not the backside interaction which cannot accommodate a bidentate interaction (Supporting Figure S2). Likewise, if the side on and end on interactions are largely bidentate than there should be a strong preference to use the syn lone pairs on the acidic residues (Figure 1D). While early analyses suggested a large preference for the more basic syn lone pairs[18], later work has found a more even distribution of syn and anti salt bridges in proteins[19]. As shown in Supporting Figure S3, the side on and end on interactions show a large preference for syn lone pairs, but the backside interaction shows almost no preference for syn lone pairs. Consistent with these observations, side on and end on interactions have a large number of bidentate interactions, comparable to the number of monodentate interactions, while backside interactions are almost exclusively monodentate (Supporting Figure S4).

Figure 3B shows the variation in ψ with respect to the distance from the central Cζ of the guanidinium (ρ). Here a previously unidentified feature is clearly observed between the doublets, which appears as a small, closer central peak. These central peaks between the doublets reflect single oxygen atoms that form hydrogen bonds with the hydrogens of two nitrogens in the guanidinium, resulting in a particularly close approach of the oxygen to the central Cζ (Figure 3B). This feature has not been recognized previously in parameterizations of Arg-based salt bridge interactions. The differential distance dependence likely should be represented in hydrogen bond parameterizations of Arg and negatively charged side chains.

We also observed a sharp distribution in ψ at 90° representing the backside interaction. These salt bridges are removed from the dataset if the centroid definition of Kumar and Nussinov[4,5] is used (Supporting Figure S5). As these are salt bridge interactions, the original, non-centroid definition was retained. Notably, the ψ/θ plot for the oxygen of the side chain carboxamide of asparagine and glutamine (Asn/Gln) shows that the backside interaction, which is least favored for Asp/Glu becomes the most favored (Figure 3C, D), along with the central peaks between the side on and end on doublets. This effect is likely due to the single oxygen in the carboxyamide which cannot form a bidentate end on or side on interaction.

Finally, we examined how solvent accessibility affected Arg based salt bridges. First we looked at the fraction of Arg residues forming a salt bridge at each level of burial as measured by the accessible surface area (ASA) distribution (Methods, Supporting Figure S6). As expected, a significant fraction of buried Arg residues form salt bridges. Likewise, a

larger fraction of Asp and Glu residues form salt bridges with Arg when buried than when exposed. This does not, however, imply that the preferred interaction geometry is necessarily different in buried and exposed salt bridges. In fact, the geometry of the interaction and the relative populations of side on, end on, and backside, as reflected in ψ/θ plots, remains largely invariant for different solvent accessibilities (Supporting Figure S7). While Arg forms salt bridges with carboxylate more frequently in the interior, the fundamental interaction geometry does not change.

Histidine

Unlike Arg, the geometric preferences of salt bridges of histidine (His) and lysine (Lys) have not been studied in detail. The specific geometry of His-carboxylate interactions can also be of great importance when rationally engineering metal active site enzymes. His has two nitrogen atoms that, when protonated, are expected to form monodentate salt bridges with Asp/Glu. Like the Arg guanidinium group, the His imidazole is planar. For His, the midpoint of the two nitrogen atoms was taken as the origin of the spherical coordinate frame (Figure 1B).

Because salt bridges with His are monodentate, we expected that His salt bridges would be more similar to backside than side on or end on Arg salt bridges. As shown in Figure 4A, there are two clear peaks in the ψ/θ distribution, termed delta (for Nδ) and epsilon (for Nϵ). The distribution is more varied for the Asn/Gln carboxyamide (Figure 4B) perhaps reflecting a weaker interaction. For comparison, Supporting Figure S8A,B shows the relationship between distance and ψ, which for His shows only a single peak each for Nδ and Nϵ. A lower fraction of His residues (compared to Arg) form salt bridges when buried, but, like Arg, the more buried the His the higher the fraction of salt bridges formed (Supporting Figure S7).

The angle between the imidazole and carboxylate planes shows a small preference for being nearly co-planar despite the salt bridges being monodentate (Supporting Figure S2). Although carboxylate oxygens can approach the Cϵ atom to form a pseudo-bidentate interaction[37,38], this is not common, as there are very few oxygens observed between the Nδ and Nϵ atoms in the ψ/θ distribution (Figure 4A). The preference for co-planarity does not appear to be due primarily to the delta and epsilon interactions (Supporting Figure S2). Instead, this preference arises from interactions where Asp or Glu stack above or below the imidazole plane (Supporting Figure S9). For these interactions, no hydrogen bond involving the delta or epsilon nitrogens is formed. In regards to interactions with acidic syn and anti lone pairs, the delta and epsilon interactions are most similar to the Arg backside interaction, with a slight preference for syn (Supporting Figure S3).

Because His often interacts with metals, we next considered the geometry of these interactions. His interacts with metals with the same spherical angles as Asp/Glu oxygens (Supporting Figure S10A), but the interaction is closer, reflecting ligation as opposed to hydrogen bonding (Supporting Figure S10B). The majority of metal binding interactions involve the epsilon nitrogen (Supporting Figure S10A). Metal-mediated interactions often involve multiple residues, therefore we find that there is significant cooperativity between metal binding and salt bridge formation. If a His engages in an interaction with a metal ion it is approximately two-fold more likely to form a salt bridge between its other nitrogen and an Asp or Glu (Supporting Figure S10C, Methods).

Lysine

Unlike Arg and His, the basic group of Lys is nonplanar; therefore we instead described the tetrahedral geometry (Figure 1C) by a torsional angle (defined by Cδ-Cϵ-Nζ-O(carboxylate))

and an angle (defined by Cε-Nζ-O(carboxylate)). As expected, the torsion angle has a strong preference for the three staggered configurations of the carboxylate oxygen, termed gauche plus (g+, 60 degrees), trans (t, 180 degrees), gauche minus (g−, 300 degrees), at least partially due to the preference for the ammonium hydrogen atoms to be staggered[17] (Figure 5A, Supporting Figure S8C). Steric interactions between the carboxylates and the Cε of lysine might also influence the distribution. The lack of data points in other regions of the plot speaks to the geometric specificity of Lys salt bridges. Surprisingly, the trans conformation is disfavored relative to the gauche plus and gauche minus conformation. By contrast, within the Lys side chain, the torsion angles between carbon atoms prefer the trans conformation[31] because this is the most favorable interaction sterically. Interactions of Lys and the oxygen of carboxamide of Asn and Gln show the same three configurations are observed as for Asp and Glu (Figure 5B, Supporting Figure S8D), including the preference for gauche plus and gauche minus.

We next examined the effect of solvent accessibility and syn-anti preference on Lys-Asp/Glu interaction geometries. Lys salt bridges are intermediate between Arg and His in terms of fraction forming salt bridges when buried (Supporting Figure S6). A similar fraction of acidic residues form salt bridges with His and Lys when buried. The monodentate salt bridges of lysine, like the His salt bridges, show a preference for syn electon lone pairs that is weaker than the bidentate Arg interactions. This preference is larger than that of the His and Arg monodentate interactions, possibly due to higher charge density or the more flexible side chain of Lys. As with Arg and His, the geometry of the interaction and relative populations of the different conformations remains largely invariant for different solvent accessibilities (Supporting Figure S7).

Salt Bridge Geometry Dependence on Backbone Distance

We next sought to discover how the structure of the protein backbone affects the interaction geometry of salt bridges. In protein design, the backbone conformation of the protein is often fixed and therefore the distance between residue backbones is known beforehand. If the salt bridge interaction geometries depend on the separation in space between the residues, this information could be used to design interactions that are more similar to those observed in folded proteins.

The ratios of the different salt bridge geometries show a striking exponential dependence over the range of backbone Cα-Cα distances (Figure 6, left side, note log scale). All salt bridge combinations show a strong bias at short distances that exponentially decays to the opposite bias at large distances. For each interaction, the preference for one geometry versus another changes several orders of magnitude over the range of Cα-Cα distances.

Another way of quantifying these differences is to consider the mean Cα-Cα separations of the different interaction types (Table I). Each salt bridge has interactions that favor close and distant Cα-Cα separations. When engineering salt bridges at closer distances, one should introduce fewer end on (Arg), epsilon (His), or trans (Lys) interactions. In terms solely of Cα-Cα distance, a backside Arg salt bridge may be most suitably replaced by a His delta interaction, and an end on Arg salt bridge by a Lys trans interaction.

Salt Bridge Geometry Dependence on Sequence Distance

One of the most interesting early observations in Arg-based salt bridges[2] was that side on interactions were most common within a protein fold, but end on interactions were most common in protein-protein interactions. The primary dataset used for this study considers interactions within protein folds (Methods), and we confirm the earlier finding that side on interactions are most common within these folds (Figure 3). We also considered two very

different sets of protein-protein interfaces. One is a curated, biologically-relevant protein-protein dataset[24] (Methods) and another is the crystal lattice contacts of our original dataset. These datasets shows a slight bias for end on interactions (solid and dashed-straight lines, respectively, in Figure 6, right side), confirming the observations almost twenty years ago. Likely due to the much larger database now available, a much smaller bias in protein-protein interactions is observed[2]. Although the difference is not statistically significant according to Fisher's exact test, it is interesting to note that the non-biological lattice contacts make a larger number of the more distant interaction for each of three basic residues (dashed-straight lines in Figure 6).

For residues sufficiently separated in sequence, we expected that the interaction preference would be similar to that of protein-protein interactions. The large database in this study allows us to address this question. Arg-based salt bridges show a large bias for side on interactions when the groups are close in sequence (Figure 6, right side). A seven fold preference is observed when residues are separated by fewer than five amino acids, but the bias gradually declines to the protein-protein value (solid, straight line) when residues are separated by 150 residues or more.

Within a protein chain, Lys prefers the gauche plus and gauche minus salt bridge conformations over trans (Figure 5). To measure this preference, the ratio of the number of guache conformations over the number of trans conformations was plotted for different sequence separations, protein-protein contacts, and lattice contacts. In protein-protein and lattice contacts, all three interaction types are found in nearly equal amounts (Figure 6, where the ratio of the two gauche conformations to trans is approximately two). The gauche conformations are much more common than trans at sequence separations of 5 or less. To our knowledge, this change in geometric interaction preferences for Lys has not been previously noted and should be considered when designing Lys salt bridges within and between proteins.

His, like the other basic residues, shows a preference for an interaction type that is strongest at short sequence separations and approaches the protein-protein value at large sequence separations (Figure 6B). For His, salt bridges with Nδ are favored at sequence separations of 5 or less, but at other separations Nϵ is favored. Of note, the pattern is more complex than the simple decay observed for Arg and Lys, with a second peak in the ratio at a separation of 30–40 residues and then again at the largest sequence separations. Also, moderate sequence separations (50–200) are very similar to the ratio observed for lattice contacts, not protein-protein contacts. The pattern may be due to the manner in which histidine salt bridges are used within protein folds.

Local salt bridges interactions

The large preference for certain salt bridge geometries at small sequence separations needs to be understood in more detail. We define the name "local salt bridges" to denote interactions separated by fewer than five residues. If all contacts were equally likely, only 2.7% of all salt bridges in the dataset would be expected to local. In contrast, 34% of Arg, 26% of His, and 31% of Lys salt bridges in the database are local salt bridges. The overall greater percentage of local salt bridges is likely due to the entropy of loop closure[14]. There is a greater entropic cost to connect more distance regions of a protein.

The short sequence separations suggest that they might occur within elements of secondary structure. We therefore determined the likelihood of forming a given salt bridge at a specific sequence separation (e.g. DxxxR) within secondary structure elements (helix, sheet, or coil, see Methods). We describe sequence separations relative to the basic group. For example DxxxR is labelled $Asp_{i-4}Arg_i$ while RxxxD is labelled Arg_iAsp_{i+4}.

The ratio of the observed and expected observations (termed here propensity) highlights sequence separations with more frequent placement of salt bridging amino acids and/or increased probability of forming a salt bridge. The expected number of observations was calculated by combining the overall frequency of forming local salt bridges using the specified residues along with the number of times the residues would be expected to be at that sequence separation given their mole fractions (Methods).

To focus on the strongest biases, the salt bridge motifs with at least 25 examples in the database and a propensity of at least 1.8 were considered significant (Supporting Table I). 32 of the 432 possible local salt bridges qualify by these criteria. As we define residues to be helix, sheet, or coil, these motifs can be categorized into four general categories: helix-helix, sheet-sheet, coil-coil, and between secondary structure elements.

Interactions in Helices and Sheets

Nearly half (15 of 32) of the local salt bridge motifs are found within helices. The interacting groups are frequently separated by one turn of the helix (e.g. $Glu_{i-4}Arg_i$). Experimental studies vary[39–41] on the details of the most preferred sequence separations. In one study, the Lys_iAsp_{i+4} was found to be the most favorable Lys based salt bridge[39], while in another study the most favorable interaction was found to be $Asp_{i-4}Lys_i$, $Glu_{i-4}Lys_i$[40], $Asp_{i-4}Arg_i$, and $Glu_{i-4}Arg_i$[41]. On the other hand, a helix-coil transition statistical mechanics analysis[42] showed preferences for $Glu_{i-4}Arg_i$ and $Glu_{i-4}Lys_i$. A statistical survey of salt bridge motifs comparing coiled-coils and non-coiled-coil helices[28] found the most common salt bridge sequence separations were the Arg_iGlu_{i+4}, $Glu_{i-3}Arg_i$, and $Glu_{i-4}Arg_i$ motifs.

Studying the large salt bridge dataset compiled here allows us to compare these results to the propensities observed in crystal structures of folded proteins. The local Arg salt bridge motifs with the highest propensity are $Glu_{i-3}Arg_i$, $Asp_{i-3}Arg_i$, and $Glu_{i-4}Arg_i$ (Figure 7, Supporting Table I). Interestingly the propensities for Lys are different; the salt bridges with the highest propensity are Lys_iAsp_{i+4}, Lys_iGlu_{i+4}, $Asp_{i-4}Lys_i$, and $Glu_{i-3}Lys_i$ (Figure 7, Supporting Table I). Helical, local His salt bridges show smaller propensities; the two significant motifs are His_iGlu_{i+4} and His_iAsp_{i+1}. The Lys_iAsp_{i+4} interaction has the largest propensity (14.1), pointing to a particularly strong tendency for this motif to form a salt bridge. Overall, it appears that Arg based salt bridges favor the acidic residue to be N-terminal while His and Lys based salt bridges have a similar propensity for N- and C-termini. In addition to contacts with residues on the same face of the helix, Arg and His make salt bridges with Asp at i+1 (Arg_iAsp_{i+1} and His_iAsp_{i+1}).

In beta sheets, the primary interactions are expected to lie on the same face of the beta strand (i−2 and i+2). Arg and His salt bridges have higher propensities than Lys salt bridges on sheets (Figure 7, Supporting Table I), and all significant sequence separations are i−2 or i+2. One particular interaction, $Glu_{i-2}His_i$, shows a particularly high propensity (9.2). This salt bridge almost exclusively involves a Glu anti lone pair and the $N\epsilon$ atom of histidine.

To further explore these local salt bridge motifs, the two components of the propensity (frequency of placement of salt bridging amino acids and frequency of forming a salt bridge) were separated and two terms correlated. To calculate the first component (frequency of placement), the ratio of the local sequence separations to non-local sequence separations was calculated (Methods). This was compared to an expected frequency that controls for the length of the protein sequences in the structural database.

If salt bridge formation is important for local protein structure, a salt bridge that has an increased probability of forming would be expected to occur more frequently at that sequence separation[28]. This is precisely what is observed for helical and sheet Arg, sheet

His, and helical Lys (Supporting Figure S11). Two positions that appear to be outliers on these plots are important to note: Lys_iAsp_{i+4} on helices and $Asp_{i-2}Arg_i$ on sheets. Lys_iAsp_{i+4} forms salt bridges at a higher frequency than any other salt bridge and more than twice as often as any other Lys position on a helix. It is enriched, but not commensurately with its high frequency of salt bridge formation, perhaps reflecting a non-linearity in the relationship between the variables. $Asp_{i-2}Arg_i$ is observed less frequently in sheets for a different reason: this amino acid pair can form three favored interactions, type I beta turns (Figure 8A), helix caps (Figure 8B), and C-termini of sheets (Figure 8F) all of which would disrupt the sheet secondary structure.

Interactions Outside of Helices and Sheets

There are several other local salt bridge motifs that are not found wholly within helices and sheets. The majority of these motifs are found at the termini of secondary structures, such as at the N-cap of helices[43,44]. At the N-cap, for example, an Asp can interact at a sequence separation of two with an Arg ($Asp_{i-2}Arg_i$) or a sequence separation of three with a His ($Asp_{i-3}His_i$) (Figure 8B, C, Supporting Table I) while also hydrogen bonding to the amide nitrogen of the N-terminal residue of the helix. We find that it is common for proteins to reverse the Arg- Asp interaction at the N-cap, as well (Arg_iAsp_{i+2}, Figure 8D). At the C-terminus of helices, there is a high propensity of $Glu_{i-3}Lys_i$ interactions, but not $Glu_{i-4}Lys_i$ (Figure 8E). Terminals of beta strands also utilize salt bridges. For instance, at the C-terminus of beta strands, there is a significant propensity to form an $Asp_{i-2}Arg_i$ salt bridge despite the backbone of Arg no longer having dihedral angles that match that of canonical beta sheets (Figure 8F). Instead, the geometry often resembles that of a type I beta turn (14/64, 22%, see below for the definition of type I turns).

In coil regions, away from helix and sheet secondary structure elements, the interaction with the highest propensity is $Asp_{i-2}Arg_i$ (Figure 8A). Many of these salt bridges are part of a type I beta turn[45] (136/284, 48%). Type I beta turns contain four residues, where the first and fourth residues have Cα atoms within 7 Å and the second and third residues have a specific range of non-helical backbone dihedral angles. For the significant motifs, the basic and acidic residues are found at the first and third positions. In these turns, most Asp residues make a side on salt bridge where one of the Asp oxygens forms an additional hydrogen bond with the Arg amide nitrogen (inset of Figure 8A). This hydrogen bonding pattern is identical to the $Asp_{i-2}Arg_i$ interaction at helix N-caps and at the C-terminus of beta strands (insets of Figure 8B and 8F). The salt bridge interaction likely contributes to the stability of the dramatic direction change observed in these loops.

Salt bridges affect side-chain dihedrals

For a particular salt bridge, it is important to know which rotamers are most frequently used as these may be more likely to succeed in a design. It is also helpful to know the entropic consequences of forming a salt bridge interaction as these are expected to differ between salt bridge interactions and may be important to consider in protein design[29]. For a particular secondary structure and sequence separation, only a subset of side chain conformations will allow the acidic and basic residues to form a salt bridge. The loss of available side-chain conformations is known as the side-chain conformational entropy (SCCE) loss and can be estimated by considering the change in distribution of rotamers in x-ray crystal structures[30,46,47]. Entropy estimates using Gibb's entropy equation (Methods), despite using static x-ray crystal structures, have been shown to correlate surprisingly well with experimental measures of SCCE in protein folding[46], protein-peptide[48], and protein-protein interactions[49].

We extend these calculations to estimate the component of the SCCE loss due to salt bridge formation for the basic residue. The basic residue has a larger number of rotatable bonds and thus should make a larger contribution to the SCCE[30]. To estimate the SCCE loss, the rotamer distribution for basic residues in a particular salt bridge type is compared to the overall rotamer distribution for the basic residue given the secondary structure. We term this value as a pseudo-entropy to emphasize its approximate nature. The pseudo-entropy approximates the component of the SCCE that is due solely to salt bridge formation. This is done by comparing the rotameric distribution of a free basic residue to one that is in the particular salt bridge type. However, it is important to note that for a particular salt bridge, interactions with the rest of the protein will also constrain the side-chain rotamers, meaning that value calculated here approximates only one component of the total side-chain conformational entropy loss.

Arginine and lysine have four side chain torsion angles[32] while histidine has two, meaning that the pseudo-entropy loss will tend to be smaller for histidine because there are fewer angles to constrain. Example calculations of the difference in rotamer population are shown in Figure 9. For the local salt bridges motifs in Supporting Table I within helices and sheets, the calculated decreases in pseudo-entropy range from 1.7–4.7 cal/mol K for Arg, 1.0–2.8 cal/mol K for His, and 2.1–3.2 cal/mol K for Lys. For comparison, the gas constant R is 2.0 cal/mol K. Within helices, the Arg and His salt bridge motifs that are most constrained are those with an Asp at i+1 (Arg_iAsp_{i+1} and His_iAsp_{i+1}, 4.7 and 2.8 cal/mol K, respectively).

Significant salt bridge motifs outside of helices and sheets have similar pseudo-entropy losses: 2.3–5.4 cal/mol K for Arg, 3.7 cal/mol K for Lys, and 1.9–2.2 cal/mol K for His. $Asp_{i-2}Arg_i$ interactions in type I turns, at the N-terminus of helices, and the C-terminus of strands show several of the largest pseudo-entropy losses observed (4.5, 5.4, and 5.4 cal/mol K, respectively). The constraint of these basic residue rotamers is likely due to the hydrogen bond that the Asp makes with amide nitrogen (insets of Figure 8A, 8B, and 8F). It appears that this hydrogen bond is favorable enough enthalpically to compensate for the true SCCE loss. Preferred rotamers for all other significant salt bridge motifs can be found in Supporting Figure S12.

Networked Salt Bridges

Design of higher order interactions is an important focus in rational protein design. Such interactions bring together secondary structure elements that are distant in sequence, giving geometric specificity to the native state. Salt bridges can provide one avenue for achieving this goal in designed proteins. Interactions with one basic residue and multiple acidic residues are commonly called "complex" or "networked" salt bridges. Networked salt bridges have been investigated for their role in the stability of protein structure[50], their energetic (anti-) cooperativity[51,52], and their geometric distributions[53].

Using a small set of 94 proteins[53], the same interactions were found in both simple and networked salt bridges. We also observe the same interaction patterns in the current study of a greatly expanded database (3,644 proteins, see Methods). We do not, however, observe marked overall preferences for forming networks of salt bridges. As shown in Supporting Figure S13, there are approximately the number of networked Arg, His, and Lys salt bridges one would expect given the frequency of simple salt bridges (Methods). The different hydrogen bonding geometries are distinct in space and appear to be nearly independent of each other.

We next analyzed local networked salt bridges to see if there is a preference for particular sequence separations. Two-tailed p-values were calculated using Fisher's exact test to determine if the two salt bridges occur more or less frequently than one would expect given

the overall frequency of these salt bridges. Three significant pairs occur: $Arg_iAsp_{i+2}Glu_{i+3}$, found in type I turns ($p = 2 \times 10^{-7}$), $Arg_iGlu_{i+1}Glu_{i+4}$, found in helices ($p = 6 \times 10^{-4}$), and $Glu_{i-3}Arg_iGlu_{i+4}$, found in helices ($p = 4 \times 10^{-2}$) (Figure 10A-C).

The $Arg_iAsp_{i+2}Glu_{i+3}$ interaction occurs exclusively in type I turns ($n = 7$). The i+2 Asp typically stacks on the Arg residue while the i+3 Glu forms a side interaction. The i+2 Asp can also hydrogen bond with the amide backbone of the turn. Because all examples of this salt bridge network form a type I turn, this motif may preferentially form these turns.

The $Glu_{i-3}Arg_iGlu_{i+4}$ interaction is particularly interesting as the most frequent of these networks ($n = 45$). The i−3 Glu forms a side-on interaction with Arg, while the i+4 Glu interacts in an end-on or backside manner. Arg is found primarily in the t,t,g+,g+ rotamer, which is the preferred rotamer for both simple i−3 side on and simple i+4 end on or backside interactions. Among the possible combinations of simple local salt bridges, the i−3 and i+4 interactions are the only combination that shares the same preferred rotamer. It will be interesting to determine if the $Glu_{i-3}Arg_iGlu_{i+4}$ networked salt bridge motif provides additional stability to helices.

An alternative type of networked salt bridge occurs when a locally stabilized salt bridge interacts with a residue distant in sequence (termed here bridging networked salt bridges). These interactions may be used to stabilize tertiary interactions within proteins. P-values were calculated by comparing to the frequency of observing both local and non-local salt bridges individually. Six different types of networked salt bridges were significant ($p < 0.05$). Of these, four often, but not exclusively, involve metal binding sites (Supporting Figure S14), including $Glu_{i-2}Arg_i$ diiron binding sites[54].

Two motifs involve only amino acid side chains (Figure 10D, E). The first non-metal binding motif, $Asp_{i-2}His_i$ found in sheets, is networked to a Glu that is distant in sequence ($p = 6 \times 10^{-3}$). Within helices, the only significant non-metal binding bridging networked salt bridge is Arg_iGlu_{i+3} networked to a Glu ($p = 3 \times 10^{-3}$). The local salt bridge typically involves a backside interaction with Arg, while a more distant Glu makes a side on interaction (Figure 10D). The local interaction, i+3, is the least favored of the four positions on the same face of the helix as Arg (Figure 7). It is striking that different sequence separations are preferred by Arg to form local networked salt bridges on helices (i−3/i+4) and bridging networked salt bridges (i+3).

Discussion

Implications for Folding and Stability

When salt bridge interactions are considered in the context of their secondary structure, we observed many simple (Figure 7 and 7) and networked (Figure 10) salt bridges that are over-represented in the structural database. The reason that these salt bridges are over-represented in non-homologous protein structures is likely that they contribute to the fitness of the protein by stabilizing the fold or contributing to folding kinetics[55]. For Arg and Lys in helices and His in sheets, we also observe a strong correlation between the frequency that salt bridges are formed at a sequence separation and the frequency that these residues are placed at that sequence separation (Supporting Figure S11). This suggests that salt bridge formation is in general beneficial to proteins, likely because it can stabilize the local structure or contribute to folding kinetics.

Bridging networked interactions bring regions that are distant in sequence into close spatial proximity. We found six over-represented motifs, four of which bind metal ligands. While the metal binding sites may be over-represented because of their role as active sites [54], the

two networked (non-metal binding) significant networked interactions, Arg_iGlu_{i+3} in helices and $Asp_{i-2}His_i$ in sheets, may be over-represented because of their ability to form the tertiary contacts necessary for folding and stability. These motifs are promising targets to study their effect of networked interactions on protein folding and stability.

Implications for Design

The geometric and sequence biases discovered in this paper should have an immediate and clear impact for rational design of proteins. Recently, we designed a new enzyme based on the *due ferri* family of designed di-iron proteins[56]. Only after iterations of traditional design procedures did we succeed in incorporating a three-histidine, two-glutamate active site into the four-helix bundle scaffold with proper salt bridge geometry. Along the way, many designs failed where interactions were improperly placed in the second shell around histidine. Information on proper formation of salt bridges would have clearly expedited and focused our design process.

We anticipate at least three ways in which detailed geometric information about salt bridge formation could aid protein design. First, with a detailed description of the geometric orientation of salt bridging residues relative to one another, it may be possible to improve empirical energy functions. Second, over-represented salt bridge motifs could be used as starting points in design, such as a fixed keystone interaction or an initial conformation for a rotamer search. Third, salt bridge geometry could be used to filter designs for reasonable salt bridge geometry, much like good interior packing has been used as a filter[57,58].

Non-bonded energy terms lacking an orientational component, such as those found in many molecular dynamics and knowledge-based potentials, are unlikely to fully recapitulate the preferences observed here. Likewise, distance is an important consideration, so orientational energy terms that lack a distance component are also unlikely to fully describe the observed preferences. While these energy terms are very useful in modeling and design, when the detailed interaction of a salt bridge is being modeled it is important to consider the geometric preferences observed in folded proteins. Orientation-dependent hydrogen bond potentials, for example, are an important step in this direction and have been shown to improve the predictions needed for protein structure prediction and design[59]. The information provided in this work could be used to further update and improve empirical energy terms used in protein design. Regardless, it should provide a point of reference for salt bridges in designed proteins.

There are also several ways that geometric and motif information could be used within the design process. The division of both Arg, His, and Lys salt bridges into distinct conformational types could aid in selecting the appropriate salt bridge type. Designed salt bridges not found in preferred conformations could be discarded because they will be unlikely to form in the folded state. The most commonly observed salt bridges for a given sequence and backbone separation could be used as starting points in design. For example, if an Arg residue is distant in sequence or backbone distance from the target acidic residue, an end on interaction would be most likely to be formed and could be tried first in a design.

Local and networked salt bridge motifs can also be incorporated into the design process. Salt bridges within and between secondary structures could consider the propensities that different residues have for particular sequence separations. The most favorable rotamers could serve as initial models of a newly designed salt bridge. Favorable networked interactions, because they involve three residues interacting, are substantially more challenging to correctly design, but they offer the potential to make critical keystone interactions[12,13]. The salt bridge information provided here should prove useful both as a benchmark for predicted salt bridges and as a guide in the design process.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Barlow DJ, Thornton JM. Ion-pairs in proteins. J Mol Biol. 1983; 168(4):867–885. [PubMed: 6887253]

2. Mitchell JB, Thornton JM, Singh J, Price SL. Towards an understanding of the arginine-aspartate interaction. J Mol Biol. 1992; 226(1):251–262. [PubMed: 1619654]

3. Singh J, Thornton JM, Snarey M, Campbell SF. The geometries of interacting arginine-carboxyls in proteins. FEBS Lett. 1987; 224(1):161–171. [PubMed: 3678490]

4. Kumar S, Nussinov R. Salt bridge stability in monomeric proteins. J Mol Biol. 1999; 293(5):1241–1255. [PubMed: 10547298]

5. Kumar S, Nussinov R. Relationship between ion pair geometries and electrostatic strengths in proteins. Biophys J. 2002; 83(3):1595–1612. [PubMed: 12202384]

6. Hendsch ZS, Tidor B. Do salt bridges stabilize proteins? A continuum electrostatic analysis. Protein Sci. 1994; 3(2):211–226. [PubMed: 8003958]

7. Kumar S, Nussinov R. Close-range electrostatic interactions in proteins. Chembiochem. 2002; 3(7): 604–617. [PubMed: 12324994]

8. Sindelar CV, Hendsch ZS, Tidor B. Effects of salt bridges on protein structure and design. Protein Sci. 1998; 7(9):1898–1914. [PubMed: 9761471]

9. Dominy BN, Minoux H, Brooks CL 3rd. An electrostatic basis for the stability of thermophilic proteins. Proteins. 2004; 57(1):128–141. [PubMed: 15326599]

10. Robinson-Rechavi M, Alibes A, Godzik A. Contribution of electrostatic interactions, compactness and quaternary structure to protein thermostability: lessons from structural genomics of Thermotoga maritima. J Mol Biol. 2006; 356(2):547–557. [PubMed: 16375925]

11. Waldburger CD, Schildbach JF, Sauer RT. Are buried salt bridges important for protein stability and conformational specificity? Nat Struct Biol. 1995; 2(2):122–128. [PubMed: 7749916]

12. Lombardi A, Summa CM, Geremia S, Randaccio L, Pavone V, DeGrado WF. Inaugural article: retrostructural analysis of metalloproteins: application to the design of a minimal model for diiron proteins. Proc Natl Acad Sci U S A. 2000; 97(12):6298–6305. [PubMed: 10841536]

13. Calhoun JR, Nastri F, Maglio O, Pavone V, Lombardi A, DeGrado WF. Artificial diiron proteins: from structure to function. Biopolymers. 2005; 80(2–3):264–278. [PubMed: 15700297]

14. Bastolla U, Demetrius L. Stability constraints and protein evolution: the role of chain length, composition and disulfide bonds. Protein Eng Des Sel. 2005; 18(9):405–415. [PubMed: 16085657]

15. Johansson AC, Lindahl E. Protein contents in biological membranes can explain abnormal solvation of charged and polar residues. Proc Natl Acad Sci U S A. 2009; 106(37):15684–15689. [PubMed: 19805218]

16. Call ME, Pyrdol J, Wiedmann M, Wucherpfennig KW. The organizing principle in the formation of the T cell receptor-CD3 complex. Cell. 2002; 111(7):967–979. [PubMed: 12507424]

17. Zeroka D, Jensen JO, Samuels AC. Rotation/Inversion study of the amino group of ethylamine. J Phys Chem A. 1998; 102:6571–6579.

18. Gandour RD. On the importance of orientation in general base catalysis by carboxylate. Bioorg Chem. 1981; 10:169–176.

19. Ippolito JA, Alexander RS, Christianson DW. Hydrogen bond stereochemistry in protein structure and function. J Mol Biol. 1990; 215(3):457–471. [PubMed: 2231715]

20. Li Y, Houk KN. Theoretical assessments of the basicity and nucleophilicity of carboxylate syn and anti lone pairs. J Am Chem Soc. 1989; 111:4505–4507.

21. Sarakatsannis JN, Duan Y. Statistical characterization of salt bridges in proteins. Proteins. 2005; 60(4):732–739. [PubMed: 16021620]

22. Wang G, Dunbrack RL Jr. PISCES: a protein sequence culling server. Bioinformatics. 2003; 19(12):1589–1591. [PubMed: 12912846]

23. Gromiha MM, Pujadas G, Magyar C, Selvaraj S, Simon I. Locating the stabilizing residues in (alpha/beta)8 barrel proteins based on hydrophobicity, long-range interactions, and sequence conservation. Proteins. 2004; 55(2):316–329. [PubMed: 15048825]

24. Liu S, Gao Y, Vakser IA. DOCKGROUND protein-protein docking decoy set. Bioinformatics. 2008; 24(22):2634–2635. [PubMed: 18812365]

25. Word JM, Lovell SC, Richardson JS, Richardson DC. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. J Mol Biol. 1999; 285(4):1735–1747. [PubMed: 9917408]

26. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983; 22(12):2577–2637. [PubMed: 6667333]

27. Samanta U, Bahadur RP, Chakrabarti P. Quantifying the accessible surface area of protein residues in their local environment. Protein Eng. 2002; 15(8):659–667. [PubMed: 12364580]

28. Meier M, Burkhard P. Statistical analysis of intrahelical ionic interactions in alpha-helices and coiled coils. J Struct Biol. 2006; 155(2):116–129. [PubMed: 16876432]

29. Hu X, Kuhlman B. Protein design simulations suggest that side-chain conformational entropy is not a strong determinant of amino acid environmental preferences. Proteins. 2006; 62(3):739–748. [PubMed: 16317667]

30. Doig AJ, Sternberg MJ. Side-chain conformational entropy in protein folding. Protein Sci. 1995; 4(11):2247–2251. [PubMed: 8563620]

31. Lovell SC, Word JM, Richardson JS, Richardson DC. The penultimate rotamer library. Proteins. 2000; 40(3):389–408. [PubMed: 10861930]

32. Ponder JW, Richards FM. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. J Mol Biol. 1987; 193(4):775–791. [PubMed: 2441069]

33. Dunbrack RL Jr, Cohen FE. Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Sci. 1997; 6(8):1661–1681. [PubMed: 9260279]

34. Shapovalov MV, Dunbrack RL Jr. Statistical and conformational analysis of the electron density of protein side chains. Proteins. 2007; 66(2):279–303. [PubMed: 17080462]

35. Juers DH, Matthews BW. Reversible lattice repacking illustrates the temperature dependence of macromolecular interactions. J Mol Biol. 2001; 311(4):851–862. [PubMed: 11518535]

36. Halle B. Biomolecular cryocrystallography: structural changes during flash-cooling. Proc Natl Acad Sci U S A. 2004; 101(14):4793–4798. [PubMed: 15051877]

37. Nanda V, Schmiedekamp A. Are aromatic carbon donor hydrogen bonds linear in proteins? Proteins. 2008; 70(2):489–497. [PubMed: 17705268]

38. Schmiedekamp A, Nanda V. Metal-activated histidine carbon donor hydrogen bonds contribute to metalloprotein folding and function. J Inorg Biochem. 2009; 103(7):1054–1060. [PubMed: 19501913]

39. Smith JS, Scholtz JM. Energetics of polar side-chain interactions in helical peptides: salt effects on ion pairs and hydrogen bonds. Biochemistry. 1998; 37(1):33–40. [PubMed: 9425023]

40. Marqusee S, Baldwin RL. Helix stabilization by Glu-..Lys+ salt bridges in short peptides of de novo design. Proc Natl Acad Sci U S A. 1987; 84(24):8898–8902. [PubMed: 3122208]

41. Huyghues-Despointes BM, Scholtz JM, Baldwin RL. Helical peptides with three pairs of Asp-Arg and Glu-Arg residues in different orientations and spacings. Protein Sci. 1993; 2(1):80–85. [PubMed: 8443591]

42. Munoz V, Serrano L. Elucidating the folding problem of helical peptides using empirical parameters. Nat Struct Biol. 1994; 1(6):399–409. [PubMed: 7664054]

43. Aurora R, Rose GD. Helix capping. Protein Sci. 1998; 7(1):21–38. [PubMed: 9514257]

44. Doig AJ, MacArthur MW, Stapley BJ, Thornton JM. Structures of N-termini of helices in proteins. Protein Sci. 1997; 6(1):147–155. [PubMed: 9007987]

45. Wilmot CM, Thornton JM. Analysis and prediction of the different types of beta-turn in proteins. J Mol Biol. 1988; 203(1):221–232. [PubMed: 3184187]

46. Lee KH, Xie D, Freire E, Amzel LM. Estimation of changes in side chain configurational entropy in binding and folding: general methods and application to helix formation. Proteins. 1994; 20(1): 68–84. [PubMed: 7824524]

47. Pickett SD, Sternberg MJ. Empirical scale of side-chain conformational entropy in protein folding. J Mol Biol. 1993; 231(3):825–839. [PubMed: 8515453]

48. Wang J, Szewczuk Z, Yue SY, Tsuda Y, Konishi Y, Purisima EO. Calculation of relative binding free energies and configurational entropies: a structural and thermodynamic analysis of the nature of non-polar binding of thrombin inhibitors based on hirudin55–65. J Mol Biol. 1995; 253(3):473–492. [PubMed: 7473728]

49. Weng Z, Delisi C, Vajda S. Empirical free energy calculation: comparison to calorimetric data. Protein Sci. 1997; 6(9):1976–1984. [PubMed: 9300497]

50. Kumar S, Ma B, Tsai CJ, Nussinov R. Electrostatic strengths of salt bridges in thermophilic and mesophilic glutamate dehydrogenase monomers. Proteins. 2000; 38(4):368–383. [PubMed: 10707024]

51. Gvritishvili AG, Gribenko AV, Makhatadze GI. Cooperativity of complex salt bridges. Protein Sci. 2008; 17(7):1285–1290. [PubMed: 18469176]

52. Li X, Liang J. Geometric cooperativity and anticooperativity of three-body interactions in native proteins. Proteins. 2005; 60(1):46–65. [PubMed: 15849756]

53. Musafia B, Buchner V, Arad D. Complex salt bridges in proteins: statistical analysis of structure and function. J Mol Biol. 1995; 254(4):761–770. [PubMed: 7500348]

54. Summa CM, Lombardi A, Lewis M, DeGrado WF. Tertiary templates for the design of diiron proteins. Curr Opin Struct Biol. 1999; 9(4):500–508. [PubMed: 10449377]

55. Berezovsky IN, Zeldovich KB, Shakhnovich EI. Positive and negative design in stability and thermal adaptation of natural proteins. PLoS Comput Biol. 2007; 3(3):e52. [PubMed: 17381236]

56. Reig AJ, Pires MM, Calhoun JR, Jo H, Kulp DW, Snyder RA, Solomon EI, DeGrado WF. Altering the O2-Dependent Reactivity of de novo Due Ferri Proteins. 2010 in preparation.

57. Sood VD, Baker D. Recapitulation and design of protein binding peptide structures and sequences. J Mol Biol. 2006; 357(3):917–927. [PubMed: 16473368]

58. Hu X, Wang H, Ke H, Kuhlman B. High-resolution design of a protein loop. Proc Natl Acad Sci U S A. 2007; 104(45):17668–17673. [PubMed: 17971437]

59. 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(4):1239–1259. [PubMed: 12589766]
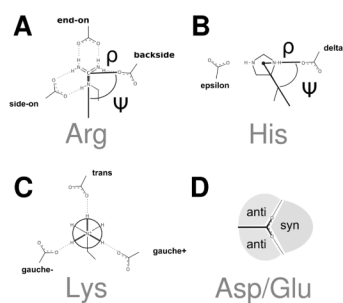
Figure 1. Parameterization of residues
For arginine (A) and histidine (B), spherical coordinates for an interacting oxygen are defined by three geometric parameters (ρ, ψ, θ). ψ measures the angle around the guanidinium (A) or imidazole (B) plane, while θ measures the degree out of that plane (with 0 degrees being in plane). ρ is the distance from the central point, Cϵ for Arg and the midpoint between Nδ and Nϵ for His, to the oxygens. For lysine (C), the relevant parameters are (1) the angle of the oxygen with the ζ nitrogen and ϵ carbon and (2) the dihedral angle of the oxygen relative to the final three heavy atoms of lysine. For the acidic residues asparatate and glutamate (D), the lone pair used to form salt bridges is categorized as syn or anti.
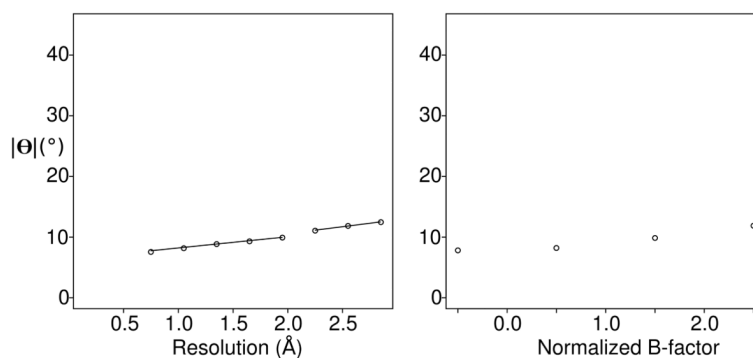
Figure 2. Mean angle, θ, of Asp and Glu carboxylate oxygens out of the guanidinium plane varies with crystal structure resolution and b-factor

(A) Variation in $|\theta|$ with crystal structure resolution. Structures were grouped into 0.3 Å bins shown at the midpoint of the distribution (e.g. 0.9 Å < resolution ≤ 1.2 Å is shown at 1.15 Å). The sharp increase between the 1.8 Å < resolution ≤ 2.1 Å and 2.1 Å < resolution ≤ 2.4 Å bins suggests a cutoff at 2.1 Å. As an extra safeguard to the reliability of the dataset, only structures with resolution of 1.8 Å or better were included in the dataset. Lines are shown to guide the eye. (B) Variation in $|\theta|$ with normalized B-factor. For structures with resolutions of 1.8 Å or better, normalized atomic B-factors were grouped into 1 standard deviation unit bins, again labeled by the midpoint. For a given salt bridge interaction, the larger of the normalized B-factors of the two salt bridging residues was used as the B-factor of the salt bridge. Using a similar cutoff to that in (A), only salt bridges with normalized B-factors below 2.0 were retained in the salt bridge dataset.
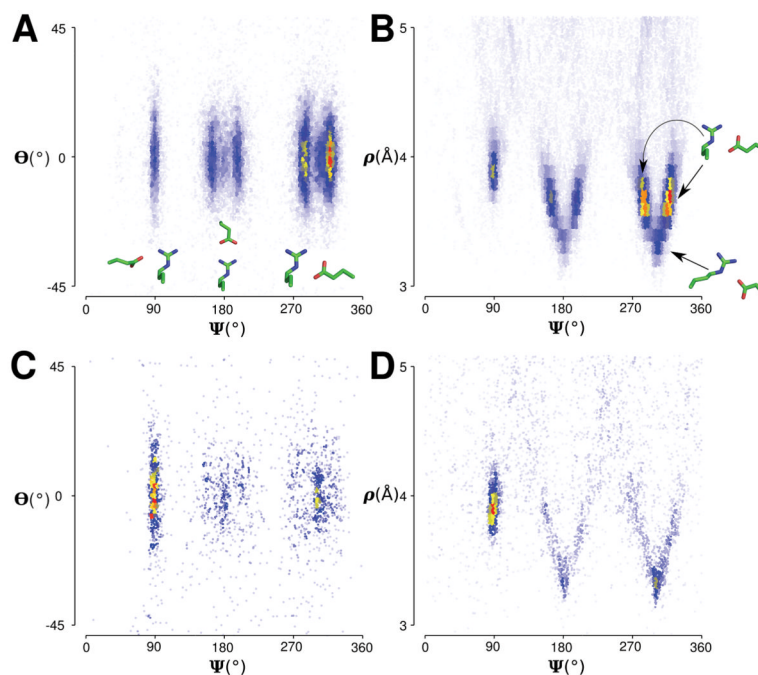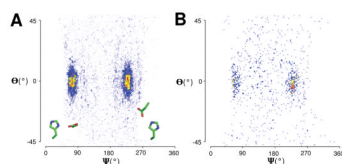
Figure 3. Spherical coordinates of arginine with (A,B) aspartate and glutamate and (C,D) asparagine and glutamine

(A,C) Two spherical coordinate parameters (ψ and θ) are plotted for oxygens interacting with arginine, showing three distinct interaction geometries. Examples of the interactions are shown in (A): backside (left), end on (middle), and side on (right). (B,D) The spherical coordinate parameters (ψ and ρ) show the relationship of distance to in-plane angle. Data points are colored by density, with red being the most dense.

Figure 4. Spherical coordinates of histidine with (A) aspartate and glutamate and (B) asparagine and glutamine
Two spherical coordinate parameters (ψ and θ) are plotted for oxygens interacting with histidine, showing two distinct interaction geometries. Examples of the interactions are shown in (A): delta (left) and epsilon (right).
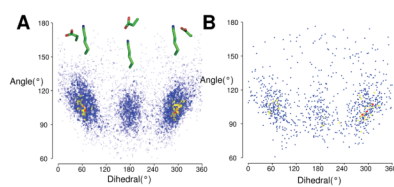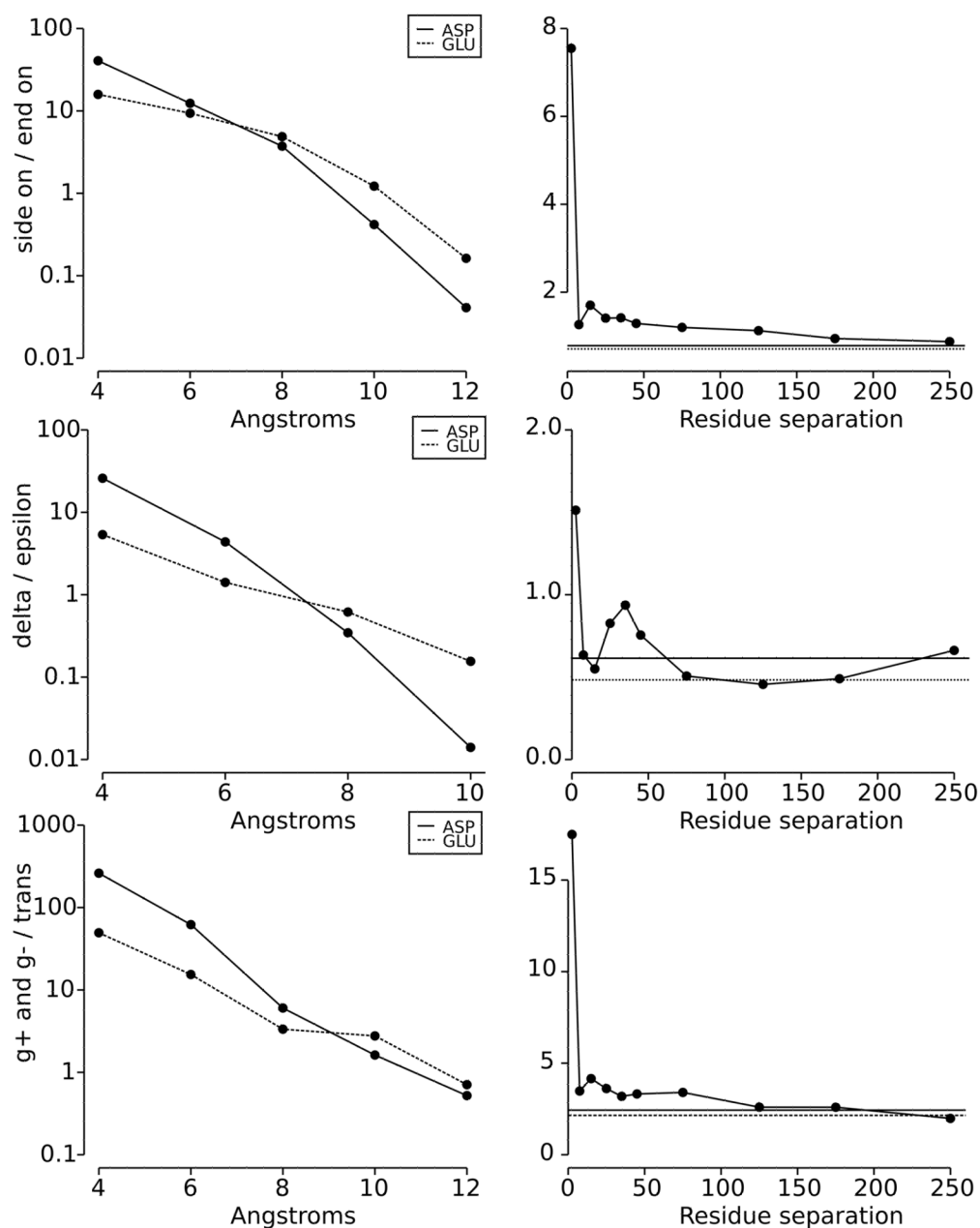
Figure 5. Angles of interaction between lysine and (A) aspartate and glutamate and (B) asparagine and glutamine

The angle and dihedral angle between the oxygen and the terminus of lysine show three distinct interaction geometries. Examples of the interactions are shown in (A): gauche plus (left), trans (middle), and gauche minus (right).

Figure 6. Ratios of distinct interactions versus Cα distance and sequence separation
(Left) The ratio of distinct salt bridge interactions is shown for arginine side on over end on, histidine delta over epsilon, and lysine gauche minus plus gauche plus over trans with a 2.0 Å binning. Asp (solid) and Glu (dashed) are plotted separately. Note the log scale. (Right) The same ratios for the combination of Asp and Glu were calculated at the following sequence separations: 1–10, 11–20, 21–30, 31–40, 41–50, 51–100, 101–150, 151–200, and greater than 200. The solid straight line corresponds to the value for biological protein-protein interactions, and the dashed straight line corresponds to the value for lattice contacts.
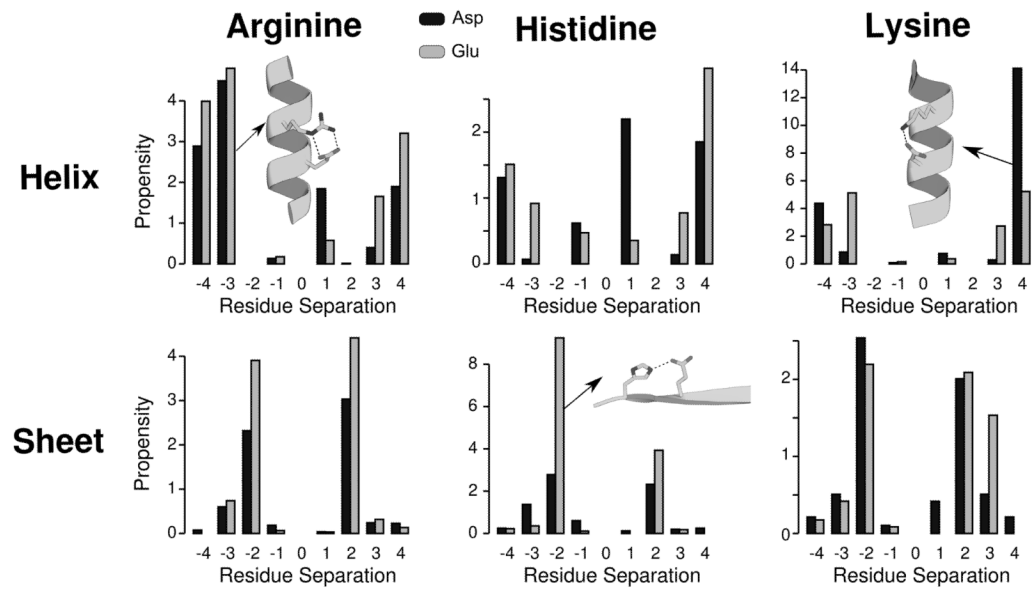
Figure 7. Helical and sheet local salt bridge motifs
For helices and sheets, the ratio of observed to expected interactions (propensity) of arginine, histidine, and lysine with acidic residues is shown. Examples of three prominent motifs are presented in the insets.
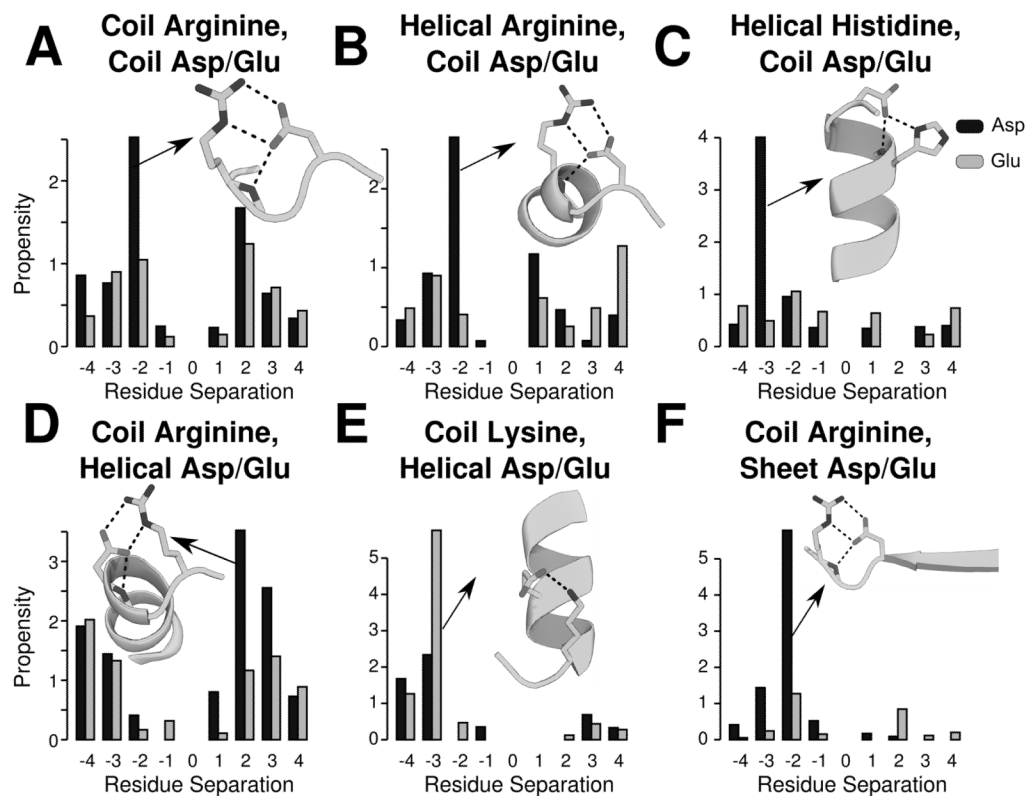
Figure 8. Other local salt bridge motifs
Salt bridges between secondary structure elements and coil regions as well as within coiled regions are shown. The motif with the highest propensity for each combination is shown in the insets.
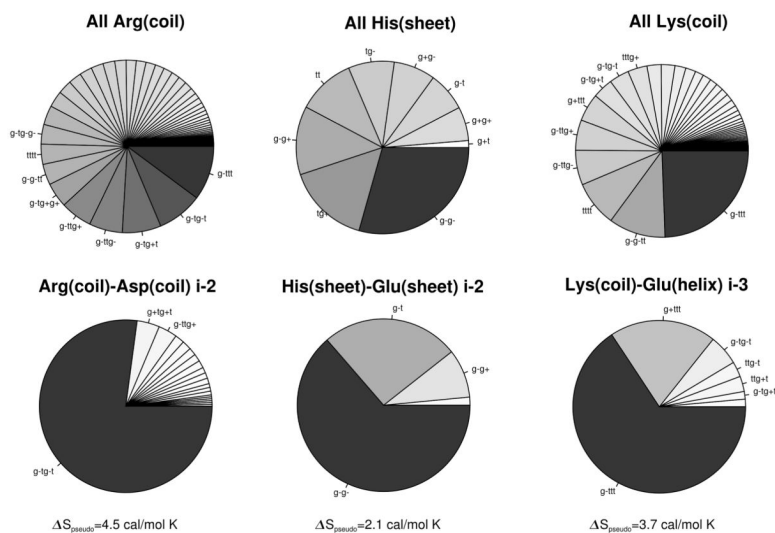
Figure 9. Side chain dihedral angle distribution for basic residue side chains
Pie charts show pictorially the calculation of pseudo-entropy values for three prominent examples. For each example, the overall distribution of the basic residue with the given secondary structure is shown above and the distribution when forming the specific salt bridge is shown on bottom. The pseudo-entropy term involves comparing the variability of these two distributions. The dihedral angles are abbreviated by t (trans), g+ (gauche plus), and g− (gauche minus), and listed in order moving away from the main chain.
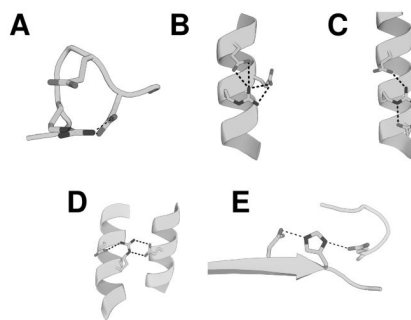
Figure 10. Significant networked salt bridges

The following networks of salt bridges occur more often than would be expected given probability of the two individual salt bridges formed ($p < 0.05$). Local networked salt bridges: (A) $Arg_i Asp_{i+2} Glu_{i+3}$, found in type I turns ($p = 2 \times 10^{-7}$), (B) $Arg_i Glu_{i+1} Glu_{i+4}$, found in helices ($p = 6 \times 10^{-4}$), and (C) $Glu_{i-3} Arg_i Glu_{i+4}$, found in helices ($p = 4 \times 10^{-2}$). Networks that contain local and non-local salt bridges: (D) $Arg_i Glu_{i+3}$, found in helices, networked to a Glu that is distant in sequence ($p = 3 \times 10^{-3}$) and (E) $Asp_{i-2} His_i$, found in sheets, networked to a Glu that is distant in sequence ($p = 6 \times 10^{-3}$).

Table I

Mean Cα-Cα distance of salt bridges

Mean Cα-Cα distances for the different types of salt bridges to Asp and Glu. Distances are given in Ångstroms. Standard deviations are given in the parentheses

Salt Bridge Interaction	Mean Cα-Cα distance to Asp	Mean Cα-Cα distance to Glu
Arg side on	6.7 (1.7)	7.4 (2.1)
Arg end on	10.2 (1.6)	10.9 (2.0)
Arg backside	6.6 (1.6)	6.6 (1.7)
His delta	6.0 (1.3)	6.9 (1.5)
His epsilon	8.6 (1.3)	8.8 (1.7)
Lys gauche plus	7.2 (1.8)	7.7 (2.1)
Lys gauche minus	7.2 (1.8)	7.5 (2.1)
Lys trans	9.5 (1.2)	9.6 (2.0)