Understanding the role of intramolecular ion-pair interactions in conformational stability using an ab initio thermodynamic cycle

Sabyasachi Chakraborty 1, Kalyaneswar Mandal 1, and Raghunathan Ramakrishnan 1

1 Tata Institute of Fundamental Research Hyderabad, Hyderabad 500046, India

(Dated: December 22, 2022)

Intramolecular ion-pair interactions yield shape and functionality to many molecules. With proper orientation, these interactions overcome steric factors and are responsible for the compact structures of several peptides. In this study, we present a thermodynamic cycle based on iso-electronic and alchemical mutation to estimate intramolecular ion-pair interaction energy. We determine these energies for 26 benchmark molecules with common ion-pair combinations and compare them with results obtained using intramolecular symmetry-adapted perturbation theory. For systems with long linkers, the ion-pair energies evaluated using both approaches deviate by less than 2.5% in vacuum phase. The thermodynamic cycle based on density functional theory facilitates calculations of salt-bridge interactions in model tripeptides with continuum/microsolvation modeling, and four large peptides: 1EJG (crambin), 1BDK (bradykinin), 1IL2Y (a mini-protein with a tryptophan cage), and 1SCO (a toxin from the scorpion venom).

I. INTRODUCTION

Ion-pair interactions are ubiquitous and are the fifth most prevalent in biological systems as salt-bridges[1]. There are many definitions of salt-bridge based on the constitution and spatial orientation of the interacting moieties[2–3]. In peptides, such interactions are predominantly found in guanidinium-carboxylate or ammonium-carboxylate ion pairs[7]. Presence of only a few salt-bridges can have a significant impact on both the structure and the function of bio-macromolecules[8–19]. The stability of a salt-bridge depends on its local environment and exposure to polar solvents. Depending on these factors, they can be stabilising[20–25], insignificant[26–29], or even destabilising[30]. Experimentally, the presence of a salt-bridge can be inferred through nuclear magnetic resonance spectroscopy[31] or infrared vibrational spectroscopy[32, 33]. While vibrational spectroscopy also detects different salt-bridging patterns (such as end-on/ side-on or monodentate/polydentate[32]), they cannot provide any insight on their relative thermodynamic stabilities. The absolute value of the salt-bridge interaction energy of a folded protein is not an experimental observable. Between a folded and an unfolded state of a macromolecule, the Gibbs free energy of folding, \( \Delta G_{\text{folding}} \), can be experimentally determined in terms of stability constants. This quantity is essentially the sum of contributions due to structural relaxation or strain, \( \Delta G_{\text{strain}} \), salt-bridge (or ion-pair) interaction, \( \Delta G_{\text{ip}} \), and reorganization due to interactions with solvent molecules. The contribution due to salt-bridge, \( \Delta G_{\text{ip}} \), can compensate for the other two terms resulting in a net negative \( \Delta G_{\text{folding}} \). Experimentally, \( \Delta G_{\text{ip}} \) is determined using denaturation titrations of the wild-type protein and its single and double mutants[31, 34]. These experiments have been extremely useful to shed light on the strength of different salt-bridges across many different proteins. The two salt-bridges (E6/R92 and E62/K46) in thermostable ribosomal protein L30e stabilize the system by 0.5–1.2 kcal/mol[35]. A double mutant cycle analysis revealed the K11/E34 surface salt-bridge in ubiquitin to have a strength of 0.24 kcal/mol[36], while the conserved salt-bridges (R69/D93 and R83/D75) in barnase revealed an interaction energy of 3.0–3.5 kcal/mol[37]. White et al. used molecular dynamics simulations to investigate charged amino-acids in water, when interaction strength (\( \Delta G_{\text{int}} \)) was found to vary from 0.5–2.0 kcal/mol[38]. Typically, the stabilization due to salt-bridges in most proteins are within the 4–5 kcal/mol range[7]. However, studies have shown that this value can be influenced by interactions with osmolytes[39]. In general, engineered surface salt-bridges are destabilizing but conserved salt-bridges present inside the protein are stabilizing[40].

We note that in denaturation experiments the interacting moieties in ion-pairs are in two very different paradigms: In the folded state, the ion-pair is not exposed to any solvent molecules, while in the unfolded state, it is exposed to both solvent molecules and ions. Furthermore, when ion-pairs are exposed to the solvent media, the interaction strength is significantly quenched due to the dielectric constant of the solvent. Perhaps, the ideal scenario of understanding unfolding must be one where contributions due to solvent can be completely ignored, but at this stage it is experimentally unfeasible. Again, in double mutant experiments, the single mutants involve replacing one of the charged residues with a neutral one. The assumption is that the effects due to mutation will exactly cancel out. However, the wild-type and mutated peptides may exhibit different conformational preferences resulting in non-cancellation of energy terms due to structural relaxation. Another computational protocol would be to access a conformer of the biomolecule where the interacting groups are well-separated[41]. Not only does this exercise become increasingly challenging when it comes to estimating salt-bridges in large peptides, but more fundamentally, the strength of the salt-
bridge is now dependent on the type of elongated conformer accessed.

Focusing only on the enthalpic contributions, a theoretical estimation of intramolecular ion-pair interaction energy, $E_{\text{ip}}^{\text{intra}}$, is more challenging than that of the intermolecular counterpart, $E_{\text{ip}}^{\text{inter}}$. This is because the latter is a well-defined quantity related to the dissociation energy, while such analogy is not suitable for intramolecular bonding. $E_{\text{ip}}^{\text{inter}}$ can be estimated using the supermolecule approach as the difference between the total energies of the complex and its isolated monomers. $E_{\text{ip}}^{\text{inter}}$ determined for model salt-bridges provides insights into $E_{\text{ip}}^{\text{intra}}$ of proteins.

Besides the supermolecule approach, intramolecular interaction energies can also be estimated using energy decomposition analysis (EDA) or symmetry-adapted perturbation theory (SAPT). Both EDA and SAPT provide a breakdown of interaction energies quantifying various physical effects. These methods have been benchmarked using results from the supermolecule approach based on highly accurate wavefunction methods as references. The accuracy of SAPT is limited by the reference wavefunction used to model the interacting fragments and the order of perturbation corrections. In the intramolecular analog of SAPT (I-SAPT), Hartree–Fock (HF) molecular orbitals (MOs) are localized on the interacting fragments and the linker. Collectively, this procedure is denoted as F/ISAPT0, where 0 indicates the use of HF states for the fragments prior to the application of perturbation theory. The initial applications of F/ISAPT0 to organic molecules such as pentanediol have yielded physically meaningful partitioning of the interaction energy, of $E_{\text{ip}}^{\text{intra}}$. In general, the SAPT formalism delivers excellent predictions for weak non-covalent interactions in charge-neutral systems, while they underperform for ionic systems. Hence, for zwitterions stabilized by intramolecular ion-pair interactions, it is desirable to compare F/ISAPT0’s predictions to results from other approaches.

In this study, we present a thermodynamic cycle to estimate $E_{\text{ip}}^{\text{intra}}$ by introducing isoelectronic atomic mutations. For model dimers with intermolecular ionic bonding, we benchmark SAPT0 with two basis sets against CCSD(T) references. We use the best performing basis set in combination with F/ISAPT0 method, and determine intramolecular ion-pair interactions for a set of 26 molecules. These results are compared with $E_{\text{ip}}^{\text{intra}}$ predictions using the thermodynamic cycle based on total energies from the domain-based local pair natural orbital coupled cluster with singles doubles and perturbative triples method, DLPNO-CCSD(T) and various density functional theory (DFT) approximations. We probe the convergence of $E_{\text{ip}}^{\text{intra}}$ with different fragmentation procedures in vacuum and implicit solvation regimes. We explore the effect of microsolvation on structural preferences, salt-bridge interaction energies, and conformer stabilities of four model tripeptides. Finally, we estimate salt-bridge interaction energies in four large peptides and rationalize their experimentally noted geometries.

II. THEORY

Here, we focus on the ion-pair interaction energy, $E_{\text{ip}}$, and its contribution to the stability of conformers. An idealized potential energy profile of an ion-pair system connecting a compact conformer and a longer one is presented in FIG.1. The net ion-pair bonding in the compact conformer, $\Delta E_{\text{ip}}$, overcomes the steric strain associated with folding.

Hence, on the potential energy surface (PES) of the non-polar mutant, the longer conformer is energetically favored. In principle, it is possible to apply the schematics shown in FIG.1 to estimate the difference in $E_{\text{ip}}$ between a compact and a long conformer ($\Delta E_{\text{ip}}$). In practice, the interest lies in estimating $E_{\text{ip}}$ for an experimentally observed folded geometry such as those collected in the protein data bank (PDB), without depending on a reference unfolded geometry. While determining an unfolded conformer is possible by exploring the torsional PES (as routinely explored in molecular dynamics studies), our thermodynamic scheme discussed below, avoids this step.

In FIG.1, we present a scheme to estimate $E_{\text{ip}}^{\text{intra}}$ of an ion-pair system using a thermodynamic cycle. The specific case of ammonium-carboxylate ion-pair is used as an example. Here, $wt$ denotes the system with an intramolecular ion-pair bonding, and $mwt$ is obtained by combinatorially mutating the polar groups with isoelectronic substitutions. Thus, N atoms in the positively charged polar moiety in $wt$ are substituted with C atoms while in the negatively charged moiety, O atoms are replaced by F. The difference between the total energies of $wt$ and $mwt$ quantifies the ion-pair interaction that is present in $wt$ while absent in $mwt$, and contributions due to the net change in the external potential and nuclear repulsion going from $wt$ to $mwt$:

$$E_{\text{wt}} - E_{\text{mwt},i} = E_{\text{ip},i} + E_{\text{al}}. \quad (1)$$

The alchemical contribution, $E_{\text{al}}$, can be estimated separately by using the total energies of the fragmented monomers of $wt$ and $mwt$, where the Cartesian coordinates of atoms of the polar groups are kept frozen. Hence,

$$E_{\text{fr}} - E_{\text{mfr},i} = E_{\text{al}}. \quad (2)$$

Fragmenting $wt$ to infinitely separated units comprising the charged moieties and a limited degree of their local chemical environment yields two fragmented systems, denoted as $fr_{\text{pos}}$ and $fr_{\text{neg}}$ (commonly denoted as $fr$). These polar moieties are saturated with terminal groups of varying length whose performances are later investigated. Further, for all possible mutation paths in the wild type system, $wt \rightarrow mwt_i$, there are corresponding mutations in the fragments, $fr_{\text{pos/neg}} \rightarrow mfr_{\text{pos/neg},i}$.

Our scheme can be applied to determine $E_{\text{ip}}$ with microsolvation to capture solvent effects. In this regime,
Figure 1. Development of the thermodynamic scheme. 

a) Schematic illustration of typical potential energy profiles of a molecule in the zwitterionic form (bottom) and in its alchemically mutated non-polar form (top). The horizontal axis is denoted by $\theta$, a general reaction coordinate. For simplicity, effect of hydration is not considered. Net stabilization of the compact structure of an ion-pair can be estimated by assuming a cancellation of alchemical changes due to introduction of charge neutral atoms or groups, $E_{\text{alch.}} \approx \tilde{E}_{\text{alch.}}$. 

b) Thermodynamic cycle for estimating intramolecular ion-pair interaction energy, $E_{\text{ip}}$. As an example, we consider the salt-bridging interaction between the guanidinium and carboxylate moieties (highlighted with a red ellipse). Here, $\text{wt}$ denotes the wild-type variant of the molecule. Atomistic mutations to remove charges on moieties leads to mutated wild-type variants denoted by $m\text{wt}_i$, where $i$ runs from 1 to 6, spanning all possible mutation paths. We now fragment $\text{wt}$ to two separate units, one containing the positively charged moiety and the other containing the negatively charged moiety, denoted as $f\text{r}_{\text{pos}}$ and $f\text{r}_{\text{neg}}$ respectively. Dangling bonds in both moieties are saturated. Thus, mutating $f\text{r}_{\text{pos}}$ and $f\text{r}_{\text{neg}}$ leads to $m\text{fr}_{\text{pos},j}$ and $m\text{fr}_{\text{neg},k}$ where $j$ covers two mutation possibilities in the carboxylate, while $k$ covers the three possibilities in guanidinium. Cartesian coordinates are preserved during the two mutation processes, namely, $\text{wt} \rightarrow m\text{wt}_i$ and $f\text{r}_{\text{pos/neg}} \rightarrow m\text{fr}_{\text{pos/neg},j/k}$. 

c) Distribution of $E_{\text{ip},i}$ for symmetric and asymmetric salt-bridges, $E_{\text{ip}}$ is taken as the mean.
we hydrate our target molecules in a polarizable continuum medium and the geometries are optimized to capture reorganization energies. When a molecule with intramolecular ion-pair interactions is relaxed with a fixed number of water molecules, the resulting fragments will also contain the same number of water molecules with same orientation as in \(\text{wt/mwt}\).

Combining (1) and (2) we arrive at

\[
E_{\text{ip},i} \approx (E_{\text{wt}} - E_{\text{mwt},i}) - (E_{\text{fr}} - E_{\text{mfr},i}) \tag{3}
\]

for the \(i\)-th mutation path. The approximation holds as long as \(E_{\text{al}} \approx \bar{E}_{\text{al}}\), which can be expected when the interacting moieties are appropriately selected to adequately capture polarization effects on the linker atoms. The final value of \(E_{\text{ip}}\) for a given geometry of \(\text{wt}\) is estimated by averaging over all possible alchemical mutations, \(E_{\text{ip}} = \langle E_{\text{ip},i} \rangle\).

The standard deviation with respect to the mean indicates asymmetry of the ion-pair bonding pattern as shown schematically in FIG.1. Hence, an asymmetric bonding pattern will result in distinct \(E_{\text{ip}}\) values whose standard deviation corresponds to the method’s uncertainty. For symmetric bonding patterns, the standard deviation vanishes as all mutation paths are equivalent.

### III. COMPUTATIONAL METHODS

For four large peptides, 1EJG, 1BDK, 1L2Y, and 1SCO, we collected geometries from PDB. For all other systems, we performed geometry optimizations using the \(\omega\)B97X-D3-DFT method with the def2-TZVP basis set. This method was used because of its ability to converge to zwitterionic molecules while preserving the intended bond connectivities. Solvent effects (water medium) were modeled using the conductor-like polarizable continuum medium (CPCM) approach. Explicit solvent interactions were modeled using the microsolvation approach where 2, 4, or 8 explicit water molecules were included in the DFT/CPCM geometry optimizations. In microsolvation investigations, different conformers were generated starting with randomly sampled arrangements of the water molecules. For a given solute, the center of the centroid–centroid distance of the polar moieties involved in non-covalent ion-pair interaction was chosen as a center of a sphere on whose surface the water molecules were randomly dispersed. This implies that in the initial geometries generated, the water molecules were close to the ion-pair containing end of the molecules. Both solutes and solvents were then allowed to undergo geometry relaxation. In vacuum phase geometry optimizations, N-H bond lengths were constrained at the cationic terminals in order to converge to an ion-pair local minimum and prevent converging to a thermodynamically more stable proton-transferred structure. Such constraints were not required with CPCM and microsolvation for most of the systems studied here. All single-point calculations were performed on the \(\omega\)B97X-D3/def2TZVP geometries. For the aforesaid large peptides, such calculations were done using the PDB geometries. DLPNO-CCSD(T) calculations were performed with the aug-cc-pVTZ basis set and TightPNO settings. DFT and DLPNO-CCSD(T) calculations were performed using ORCA 5.0 with the resolution-of-identity (RI) approximation for Coulomb (J) and ‘chain-of-spheres’ (COS) algorithm for exchange integrals. For RI calculations we used the Weigend auxiliary basis sets. DFT calculations were done with the default grid, def2grid2. Intramolecular ion-pair interaction energies were modeled with F/ISAPT0 using aug-cc-pVTZ basis set. This basis set was selected after comparing its accuracy with that of jun-cc-pVDZ for modeling SAPT0-level intramolecular molecular ion-pair interaction energies. All SAPT calculations were performed with the code Psi4.

The basis set effect on ion-pair interaction energies was benchmarked with \(\omega\)B97X-D3 energies calculated using def2-SVP, def2-TZVP, def2-SVPD, and def2-TZVPD basis sets. We have also benchmarked the performance of 10 different DFT approximations from different rungs of Jacob’s ladder using the def2-SVPD basis set. For every functional benchmarked, the recommended semi-empirical dispersion corrections were included. Hence, we included Grimme’s D3 correction with Becke-Johnson damping (D3BJ) and D3Zero corrections for functionals without long-range corrections. For PBE we included Grimme’s D3 correction with Becke-Johnson damping (D3BJ) and D3Zero corrections. For functionals with implicit long-range corrections such as LC-PBE, B3LYP, B3LYP, TPSS, and TPSS we included D3BJ, while for M06-2X we invoked D3Zero. For functionals with implicit long-range corrections such as LC-PBE, LC-BLYP, CAM-B3LYP, and \(\omega\)B97X-D3 no additional dispersion corrections were included. Intramolecular ion-pair interaction energies of peptides reported here are at the \(\omega\)B97X-D3/def2-SVPD level.

### IV. RESULTS AND DISCUSSION

In FIG.1 we discussed the influence of an ion-pair interaction on the net stability of the compact conformer over the long conformer. While the compact conformer is most stable when its polar groups are suitably oriented for maximal interaction, this comes at the cost of internal strain. In large molecules, this balance between strain and ion-pair bonding along with myriad other non-covalent interactions determine the global minimum’s structure. Although studies have quantified the effect of interactions between polar groups from a super-molecule approach, or as relative free energy changes using molecular dynamics simulations, it is of interest to directly estimate such intramolecular interaction energies for a given structure.

In this study, we pursue this goal with F/ISAPT0 and a thermodynamic cycle. To benchmark these methods, we design two sets of model systems. The first
set comprises nine dimers stabilized by intermolecular ion-pair bonding. The interaction energies of these molecules were estimated in a supermolecule fashion with the CCSD(T)/aug-cc-pVTZ method. Using these values as reference, we benchmarked SAPT0/jun-cc-pVDZ and SAPT0/aug-cc-pVTZ methods, and found the aug-cc-pVTZ to be more suitable for SAPT modeling of ion-pair systems. Subsequently, we modeled intramolecular ion-pair interactions in a second set with 26 molecules using the aug-cc-pVTZ basis set in combination with the F/ISAPT0 method. For this set, we also benchmark $E_{ip}^{inter}$ determined with the thermodynamic cycle based on different DFT methods. With the best set up, we report salt-bridge interaction energies of model tripeptides and large proteins in solvent medium.

![Figure 2. Constitution of benchmark molecules with intramolecular ion-pair bonding. The set is formed by combining 3 cationic terminals—C: monodentate (ammonium), C: bidentate (formamidinium), C: tridentate (guanidinium) with 3 anionic terminals—A: monodentate (hydroxy), A: bidentate (carboxylate), A: tridentate (sulphate) via three linkers—L: -(CH$_2$)$_n$-, L: -(CH$_2$)$_n$-, L: -(CH$_2$)$_n$-. The various terminal groups used when fragmenting the system into monomers are denoted by X/Y: -CH$_3$, -CH$_2$CH$_3$, & -CH$_2$CH$_2$CH$_3$.](image)

**A. Model ion-pair systems**

In Fig. 2, we describe the composition of model ion-pair systems considered in this study. The polar groups comprises three cationic ($C_n$), three anionic ($A_n$) moieties that are biochemically relevant. The diverse resonance structures of these polar moieties facilitate various bonding patterns as seen, for example, in guanidinium–carboxylate salt-bridges. The cation set includes ammonium ($C_1$), formamidinium ($C_2$), and guanidinium ($C_3$) moieties with 1, 2 and 3 resonance structures, respectively, while the anion set includes oxide ($A_1$), carboxylate ($A_2$), and sulphate ($A_3$) moieties with a similarly increasing number of resonance structures.

The fist set of model systems used for quantifying $E_{ip}^{inter}$ were formed by attaching methyl groups to the polar moieties, directed away from the ion-pair bonding. This set contains 9 dimers formed by combining 3 cations with 3 anions. For each system, we estimated $E_{ip}^{inter}$ for 3 different centroid-centroid distances, resulting in 27 energies. The second set of model systems to quantify $E_{ip}^{intra}$ comprise these same polar moieties by covalently connected via three different linker ($L_n$) fragments, $n = 1, 2, 3$. For linkers, we chose linear alkane chains with 3 ($L_1$), 6 ($L_2$), and 9 ($L_3$) methylene units. These non-polar linkers are not expected to interfere with the salt-bridging interactions. Of the 27 possibilities, C$_2$L$_1$A$_3$ was ignored because during geometry optimization it cyclized to an amide.

**B. Basis set for SAPT modeling of ion-pair interactions**

In Table. I we report SAPT0-level $E_{ip}^{inter}$ using jun-cc-pVDZ and aug-cc-pVTZ basis sets for 9 model ion pairs with three different centroid-centroid distances. In these complexes, the polar moieties faced each other while the methyl groups were pointed outwards. Further, for C$_1$–A$_{1/2/3}$ and C$_{2/3}$–A$_1$ type complexes, we accessed the 3.0 Å, 5.0 Å, and 7.0 Å centroid-centroid distances, while the remaining complexes were placed at 4.0 Å, 5.0 Å, and 7.0 Å to avoid small interatomic distances. We compared the SAPT0 results with those from a supermolecule scheme based on CCSD(T)/aug-cc-pVTZ energies. For short centroid-centroid distances, the interaction energies are high. On an average, we note that the performance of SAPT0/jun-cc-pVDZ to be inferior to SAPT0/aug-cc-pVTZ, with the largest deviations noted at short contacts. Hence, in the remainder of the study, we use the aug-cc-pVTZ basis set for modeling $E_{ip}^{intra}$ at the F/ISAPT0 level.

**C. Intramolecular ion-pair interactions in model systems**

For 26 model systems, we determine $E_{ip}^{intra}$ using the thermodynamic cycle (Eq. 3) and compare the results with F/ISAPT0/aug-cc-pVTZ values ($E_{ip}^{intra}$). The increasing size of the system with linker lengths rendered CCSD(T)/aug-cc-pVTZ level calculations prohibitively expensive. Hence, we resorted to the DLPNO variant of CCSD(T) for the thermodynamic cycle, and pare. Since the inherent limitations of both methods stem from different sources, at the limit of their mutual convergence
Table I. Deviation of intermolecular ion-pair interaction energies of 9 dimer model systems estimated using SAPT0/aug-cc-pVTZ and SAPT0/jun-cc-pVDZ from CCSD(T)/aug-cc-pVTZ reference. The mean absolute deviation of both methods are reported in the last row. Values in parenthesis in the first column the centroid–centroid distances between the monomers in Å. All other values are reported in kcal/mol.

| System    | Reference | aug-cc-pVTZ | jun-cc-pVDZ |
|-----------|-----------|-------------|-------------|
| C₁ – A₁ (3.0) | -103.25   | 3.94        | 3.63        |
| C₁ – A₁ (5.0) | -60.09    | 4.15        | 4.61        |
| C₁ – A₁ (7.0) | -60.09    | -14.16      | -13.84      |
| C₁ – A₂ (3.0) | -121.52   | 1.54        | -1.29       |
| C₁ – A₂ (5.0) | -73.24    | -0.06       | 0.21        |
| C₁ – A₂ (7.0) | -50.34    | 0.25        | 0.54        |
| C₁ – A₃ (3.0) | -110.43   | -1.02       | -4.21       |
| C₁ – A₃ (5.0) | -68.60    | -0.20       | -0.18       |
| C₁ – A₃ (7.0) | -48.63    | 0.14        | 0.35        |
| C₂ – A₁ (3.0) | -131.13   | 0.88        | -2.55       |
| C₂ – A₁ (5.0) | -84.76    | -0.75       | -0.69       |
| C₂ – A₁ (7.0) | -49.49    | -0.07       | 0.22        |
| C₂ – A₂ (4.0) | -122.10   | 0.74        | -2.37       |
| C₂ – A₂ (5.0) | -89.67    | -1.01       | -1.30       |
| C₂ – A₂ (7.0) | -55.88    | -0.10       | 0.13        |
| C₂ – A₃ (4.0) | -103.93   | -2.39       | -5.53       |
| C₂ – A₃ (5.0) | -80.47    | -1.44       | -2.13       |
| C₂ – A₃ (7.0) | -53.36    | -0.21       | -0.10       |
| C₃ – A₁ (3.0) | -126.03   | 2.61        | -1.22       |
| C₃ – A₁ (5.0) | -71.83    | -0.15       | 0.04        |
| C₃ – A₁ (7.0) | -48.06    | 0.10        | 0.34        |
| C₃ – A₂ (4.0) | -106.55   | 0.82        | -2.18       |
| C₃ – A₂ (5.0) | -82.19    | -0.15       | -0.58       |
| C₃ – A₂ (7.0) | -53.01    | 0.29        | 0.48        |
| C₃ – A₃ (4.0) | -99.95    | -0.92       | -4.61       |
| C₃ – A₃ (5.0) | -77.64    | -0.95       | -1.64       |
| C₃ – A₃ (7.0) | -51.40    | 0.11        | 0.22        |

The predictions can be expected to be closer to the reality. In Table I we report on the deviation in interaction energies ($E_{\text{int}}$) determined by F/ISAPT0 and $E_{\text{ip}}$ using Eq. 3 with total energies from DLPNO-CCSD(T) and ωB97X-D3-DFT levels for various choices of terminal groups: -CH₃, -CH₂CH₃, and -CH₂CH₂CH₃.

In general, for all choices of terminal groups, and across all methods, the deviation between $E_{\text{ISAPT}}$ and $E_{\text{ip}}$ (DLPNO-CCSD(T)) largely follows the order $L_1 > L_2 > L_3$. This trend is reflected better in the standard deviation (RMSD) and mean percentage absolute deviation (MPAD) than the mean absolute deviation (MAD). In $L_1$ systems, the net ion-pair interaction is not only mediated through space but also through polarization of the short linker. This through-bond contribution is not accounted for in $E_{\text{ISAPT}}$. However, in the case of longer linkers ($L_2$ and $L_3$), the through-bond contribution to $E_{\text{ISAPT}}$ becomes negligible justifying the MO localization scheme adopted. Hence, the F/ISAPT0 modeling of ion-pair interaction is most appropriate when the polar groups are separated by sufficiently long linkers.

In the thermodynamic cycle, we do not employ localized orbitals in order to describe the electron density as realistically as possible. Hence, when the spatial extent of the polar group is confined only until the nearest neighbor group (terminal group: -CH₃) and further through-bond polarization is not captured. Therefore, increasing the length of the terminal group from -CH₃ to -CH₂CH₂CH₃ results in improved agreement between $E_{\text{ISAPT}}$ and DLPNO-CCSD(T)-based $E_{\text{ip}}$ with a mean deviation of $2.73 \pm 2.49$ kcal/mol (deviation of $\approx 2\%$) for the 9 $C_2$/$L_3A_y$ molecules. The remaining discrepancies are presumably within the uncertainty of the model arising from averaging over mutation pathways (see FIG. 3).

When replacing the DLPNO-coupled-cluster energies in Eq. 3 with DFT values, we see good agreement for basis sets with a diffuse function: def2-SVPD and def2-TZVPD. For the $L_3$ systems with the longest terminal group explored, the mean deviations of the ωB97X-D3-DFT based estimations of $E_{\text{ip}}$ from the ISAPT values are 3.25 and 2.97 for def2-SVPD and def2-TZVPD basis sets, respectively. Hence in further applications, we use the def2-SVPD basis set due to its favorable cost-accuracy trade-off.

In FIG. 3 we compare the absolute values of $E_{\text{ISAPT}}$ and ωB97X-D3/def2-SVPD-based $E_{\text{ip}}$ for 26 bench-
Table II. Comparison of ISAPT-based intramolecular interaction energies, $E^{\text{int}}$, between the polar groups for 26 benchmark systems (see FIG.2) with intramolecular ion-pair interaction energies, $E_{\text{ip}}$, estimated using the thermodynamic cycle (Eq.3) with DLPNO-CCSD(T) and DFT total energies. ISAPT and DLPNO-CCSD(T) calculations were performed using the aug-cc-pVTZ (AVTZ) basis set while $\omega$B97X-D3 DFT calculations were done with the def2- family of basis sets. Mean absolute deviation (MAD), and root-mean-square deviation (RMSD) in kcal/mol along with mean percentage absolute deviation (MPAD) are reported. Thermodynamic cycle based estimation was performed for terminal groups of varying length; these groups are denoted as X and Y in FIG.2.

| Methods                    | $C_3L_4$-systems (8) | $C_2L_2A_y$-systems (9) | $C_4L_4A_y$-systems (9) | All systems (26) |
|-----------------------------|-----------------------|--------------------------|--------------------------|-------------------|
|                             | MAD       | RMSD       | MPAD     | MAD       | RMSD       | MPAD     | MAD       | RMSD       | MPAD     |
| Terminal group: $-\text{CH}_3$ |           |            |          |           |            |          |           |            |          |
| $\omega$B97X-D3/def2-SVP    | 24.36     | 11.09      | 21.96    | 20.04     | 9.28       | 14.89     | 8.99      | 15.44      | 21.94    | 9.93      | 17.26    |
| $\omega$B97X-D3/def2-TZVP   | 18.16     | 10.50      | 16.63    | 12.38     | 7.34       | 13.36     | 6.77      | 9.46       | 14.50    | 8.63      | 11.56    |
| $\omega$B97X-D3/def2-SVPD   | 15.97     | 10.73      | 14.72    | 9.78      | 6.81       | 7.25      | 5.82      | 7.19       | 11.82    | 8.40      | 9.53     |
| $\omega$B97X-D3/def2-TZVPD  | 16.07     | 10.58      | 14.91    | 9.75      | 6.50       | 7.27      | 5.89      | 7.19       | 11.83    | 8.18      | 9.59     |
| DLPNO-CCSD(T)/AVTZ          | 16.08     | 10.33      | 14.87    | 9.52      | 6.35       | 7.08      | 5.48      | 6.89       | 11.62    | 8.11      | 9.41     |
| Terminal group: $-\text{CH}_2\text{CH}_3$ |           |            |          |           |            |          |           |            |          |
| $\omega$B97X-D3/def2-SVP    | 17.52     | 9.78       | 16.55    | 12.82     | 5.47       | 9.80      | 4.53      | 9.52       | 14.86    | 6.98      | 12.13    |
| $\omega$B97X-D3/def2-TZVP   | 13.04     | 10.12      | 12.58    | 6.83      | 4.92       | 5.29      | 3.02      | 5.32       | 8.92     | 7.05      | 7.54     |
| $\omega$B97X-D3/def2-SVPD   | 12.43     | 10.55      | 11.89    | 5.41      | 5.19       | 4.21      | 3.25      | 3.52       | 7.37     | 7.40      | 6.33     |
| $\omega$B97X-D3/def2-TZVPD  | 12.47     | 10.53      | 12.00    | 5.34      | 5.11       | 4.19      | 2.83      | 3.60       | 7.38     | 7.34      | 6.39     |
| DLPNO-CCSD(T)/AVTZ          | 12.34     | 10.45      | 11.89    | 4.88      | 4.77       | 3.84      | 3.41      | 2.72       | 3.17     | 6.98      | 7.28     |
| Terminal group: $-\text{CH}_2\text{CH}_2\text{CH}_3$ |           |            |          |           |            |          |           |            |          |
| $\omega$B97X-D3/def2-SVP    | 13.09     | 10.02      | 12.79    | 9.48      | 5.08       | 7.30      | 11.41     | 2.91       | 8.29     | 11.26     | 6.70     |
| $\omega$B97X-D3/def2-TZVP   | 11.13     | 10.50      | 10.80    | 4.97      | 4.80       | 3.88      | 4.83      | 2.44       | 3.53     | 6.81      | 6.94     |
| $\omega$B97X-D3/def2-SVPD   | 10.71     | 10.86      | 10.32    | 4.57      | 5.07       | 3.60      | 3.25      | 2.57       | 2.40     | 6.00      | 7.24     |
| $\omega$B97X-D3/def2-TZVPD  | 10.87     | 10.88      | 10.47    | 4.55      | 5.05       | 3.59      | 2.97      | 2.52       | 2.22     | 5.95      | 7.26     |
| DLPNO-CCSD(T)/AVTZ          | 10.93     | 10.87      | 10.51    | 4.21      | 4.76       | 3.32      | 2.73      | 2.49       | 2.07     | 5.77      | 7.25     |

Figure 4. Effect of increasing the terminal group length, from ethyl to propyl, on $\omega$B97X-D3/def2-SVPD-based $E_{\text{ip}}$ values in vacuum and with the CPCM implicit solvation model.
Table III. Deviation of ion-pair interaction energies of 9 $C_xL_3A_y$ benchmark systems estimated using F/ISAPT0/aug-cc-pVTZ and the thermodynamic cycle (Eq.3) based on various density functional approximations with the def2-SVPD basis set. For comparison, DLPNO-CCSD(T)/aug-cc-pVTZ and $\omega$B97X-D3 values from Table II are also provided. Mean absolute deviation (MAD), root-mean-square deviation (RMSD), mean percentage absolute deviation (MPAD) are reported in kcal/mol.

| Methods             | MAD  | RMSD | MPAD |
|---------------------|------|------|------|
| PBE (D3BJ)          | 3.32 | 2.29 | 2.51 |
| BLYP (D3BJ)         | 2.78 | 2.39 | 2.11 |
| TPSS (D3BJ)         | 2.93 | 2.44 | 2.20 |
| LC-PBE              | 4.54 | 2.14 | 3.32 |
| LC-BLYP             | 3.24 | 2.23 | 2.43 |
| PBE0 (D3BJ)         | 3.34 | 2.33 | 2.47 |
| B3LYP (D3BJ)        | 3.01 | 2.65 | 2.26 |
| M06-2X (D3Zero)     | 3.49 | 3.18 | 2.59 |
| CAM-B3LYP           | 3.01 | 2.63 | 2.25 |
| $\omega$B97X-D3     | 3.25 | 2.57 | 2.40 |
| DLPNO-CCSD(T)       | 2.73 | 2.49 | 2.07 |

mark ion-pairs. $E_{ip}$ values were estimated using the $-\text{CH}_2\text{CH}_2\text{CH}_3$ terminal group. Interaction energies computed using both methods lie in the range of -160 to -80 kcal/mol. Such large values are expected in the vacuum phase as noted previously in the EDA calculations of model ion-pair systems.[93]. As noted in Table I with increasing linker length, the agreement improves as reflected by the Pearson correlation coefficients. The correlation is 0.98 when using a sufficiently long linker. In general, the magnitude of interaction energy is large when the polar groups are connected by a long and flexible linker resulting in a more compact structure.

The best agreement between F/ISAPT0 and the thermodynamic cycle is seen for the 9 $C_xL_3A_y$ systems with the X/Y=$\text{CH}_2\text{CH}_2\text{CH}_3$ terminal groups, in Table III. Hence, we benchmark the performance of 10 DFT approximations only for these systems. Several of the range-separated and long-range corrected hybrid-DFT methods show good agreement with the reference F/ISAPT0 values. All methods (barring LC-PBE) differ amongst each other within 1 kcal/mol suggesting the uncertainty due to the choice of the DFT method to be smaller than the inherent uncertainty in the thermodynamic cycle. Since $\omega$B97X-D3 was used for geometry optimization, we continue with the same method in combination with the def2-SVPD basis set for the rest of the study.

Motivated by the good agreement with the DFT-based thermodynamic cycle and F/ISAPT0 gas-phase interaction energies, we now move on to see the effect of solvent on $E_{ip}$. This is relevant because salt-bridging are mostly found in proteins, which in turn are often experimentally isolated in the aqueous phase. Alas, the ISAPT-formalism has not been extended to solvent phase. However, there is no limitation in applying the thermodynamic cycle in the solvent phase. The importance of accurately modeling the aqueous phase to study non-covalent interactions[94, 95], reactions[96], and ligand–metal interactions[97] has been well established. The magnitudes of $E_{ip}$ is expected to decrease in the aqueous phase where the ion-pair interaction is screened. To this end, using the vacuum phase geometries, we performed CPCM total energy calculations for all 26 benchmark systems.

FIG. 4 presents $E_{ip}$ calculated at $\omega$B97X-D3/def2-SVPD and $\omega$B97X-D3(CPCM; water)/def2-SVPD levels. While the vacuum phase values are $< -80$ kcal/mol, in the CPCM phase even the strongest interaction has an $E_{ip}$ of about -40 kcal/mol. In general, the magnitude of vacuum phase $E_{ip}$ decreases by a factor of 3.6–9.3 in the CPCM phase. Further, the effect of CPCM on $E_{ip}$ preserves the qualitative trends in ion-pair strengths as noted in the vacuum phase. In both media, we note the interactions due to monodentate polar moieties (ammonium or oxide) resulting in the strongest ion-pair interaction. We explored the ion-pair interactions with longer terminal groups ($-\text{CH}_2\text{CH}_3$ and $-\text{CH}_2\text{CH}_2\text{CH}_3$), and noted their influence on $E_{ip}$ to be minimal in the CPCM phase. FIG. 4 ascertains that our estimation of $E_{ip}$ is efficient in the continuum solvation paradigm capturing commonly expected physical trends. This motivated us to investigate $E_{ip}$ for some biologically relevant model systems with CPCM and with microsolvation modeling to account for explicit solute-solvent interactions.

D. Microsolvation effects on model tripeptides

Upon realizing the quenching effect of CPCM on $E_{ip}$ in the 26 model systems we expand our study towards biologically relevant molecules with implicit and microsolvation modeling. The functionality of most macro-molecules is strongly influenced by their interaction with water molecules in the medium. For example, explicit solute-solvent interactions influence pharmacodynamic activities of ligand-protein complexes.[98]. It has been shown that quantitatively accurate modeling of salt-bridges realistic an adequate description of the solvent phase.[99]. Since previous studies have shown that the zwitterionic form of amino acids is best described in the presence of water molecules[99, 100], we modeled the four tripeptides—lysine-glycine-aspartic acid (KGD), lysine-glycine-glutamic acid (KGE), arginine-glycine-aspartic acid (RGD), and arginine-glycine-glutamic acid (RGE)—with various degrees of microsolvation. Their C–terminals were amidated with methylamine to confine salt-bridging interactions to only between side-chains and to prevent the participation of main chain carboxylic groups.[2].

In FIG. 5, we show the definitions of the polar terminal groups of for the four tripeptides studied here. We fragmented the positive amino acids by taking all atoms of the amino acid up to the C$_{\alpha}$ atom of the adjacent residue, i.e. glycine in this case. Hence, the NH part of the K-
Figure 5. Definitions of terminal groups for charged amino acid residues: lysine (K), arginine (R), aspartic acid (D), and glutamic acid (E). For both N/C-terminal polar residues, all atoms until the C$_{\alpha}$ atom of the adjacent residue are included in the terminal groups by saturating the dangling bond with H. For the tripeptides, KGD, KGE, RGE, and RGD, the carboxylic acid group in the main chain is amidated by methyl amine.

G and R-G amide bonds were included as a part of the cationic fragments, K and R. Similarly, the CO part of the DG and EG amide bonds were included as a part of the anionic fragments, D and E. The dangling bonds were saturated by capping with an H atom. Hence, the terminal groups are now different for the two non-interacting components of $fr$ (in FIG. 1b) and they depend on the exact sequence of amino acids running from the N to C-terminals. This implies that the terminal groups for fragments of KGD and DGK will be different. The coordinates of the atoms in these fragments were kept frozen as in the compact structures and the mutation paths were followed as discussed before in Sec.[1]

In order to study the influence of water molecules on $E_{ip}$, we inspected the 10 conformers of RGE in the presence of 4 water molecules in a background CPCM environment. While the influence of conformers on intramolecular H-bonding interactions has been studied before[101], here we wanted to study the conformer influence on intramolecular salt-bridging interactions. We started with RGE relaxed in CPCM without any explicit water molecule as the initial structure to which we added water molecules. Hence, when the initial structure showed a bidentate side-on[32] interaction between guanidinium and carboxylate groups, it is surprising to observe in FIG. 6 a monodentate structure as the global minima. Previous studies have noted a better stabilization of solvent-separated ion-pair in lysine-glutamate dipeptide with an increasing number of explicit water molecules[99]. The remaining 9 complexes showed side-on interactions. We calculated the relative $E_{ip}$ values with reference to the most stable conformer with a side-on interaction.

While for all side-on conformers, the $\Delta E_{ip}$ is very similar, for the monodentate complex it is found to be destabilizing. Inspecting this structure revealed that this is due to an unfavorable orientation that diminishes the salt-bridging interaction. The net stability of the global minimum is due to solvation energy dominating over the salt-bridging contribution. It may be realized that for a highly stable complex, the water molecules and the polar groups must involve in maximal bonding (including H−bond network among the water molecules). The role of solvation in stabilizing a system decreases when the H−bonds are formed with weakly polar groups or when the water network is disconnected. In FIG. 6 we note this effect on the conformers on the right side, where the complexes destabilize when all solvent molecules do not form extended networks with the polar groups. Therefore, it may be concluded that, while favorable orientations with water molecules may benefit the overall stability of the complex, it comes at the cost of quenched salt-bridging interactions.

Varying the number of water molecules in the microsolvation of tripeptide will provide insights into the effect of solvation in the bulk water scenario. In FIG. 7 we plotted $E_{ip}$ for all 4 tripeptides in the presence of 0, 2, 4, and 8 water molecules. The zero-water case corresponds to the CPCM limit with one minimum energy structure. For the case of 2, 4, and 8 water complexes, we sampled 10 conformers starting with different arrangements of water molecules, of which we report only the global minimum's $E_{ip}$ regardless of the bonding denticity. Thus, for all 4 peptides, increasing the number of water molecules mostly results in quenched salt-bridging interactions. This is a physically expected trend that has been captured by the model. However, the degree to which the $E_{ip}$ changes is different across the peptides. This can be attributed to the relatively small conformer space sampled and due to the finite degree of hydration.

A closer look at the structures sheds more light on the magnitude of the interaction energies reported in FIG. 7. For KGD, with an increasing number of water molecules, the interactions between the polar fragments diminish. While this trend is seemingly similar in KGE, in the case of 8 water molecules KGE adopts a solvent separated ion-pair form, while that of KGD is solvent-shared with a
Figure 6. Relative energy, $\Delta E$, and relative ion-pair interaction energy, $\Delta E_{ip}$, of 10 low-energy conformers of microsolvated RGE (arginine-glycine-glutamic acid) estimated using thermodynamic cycles based on $\omega$B97X-D3(PCM; water)/def2-SVP total energies. All energies are reported by subtracting the values of the most stable conformer with a side-on guanidinium (arginine) - carboxylate (glutamic acid) salt-bridge. Minimum energy structures of all 10 conformers of RGE(H$_2$O)$_4$ are shown. Conformers are labelled from 0-9 (0 being the global minimum). For clarity, structures have been arranged above and below the horizontal axis. The leftmost structure is with a monodentate salt-bridge bonding where only one of the two O atoms of the carboxylate moiety binds to both NH atoms (see inset) while the structure right to it shows a bidentate interaction (see inset).

stronger ion-pair interaction. For RGD, all solvated complexes prefer the monodentate arrangement. Hence with increasing degree of hydration, $E_{ip}$ of RGD approaches a limit of $> -10$ kcal/mol. In the case of RGE, 0, 2, and 8 water complexes form a side-on bidentate arrangement, while with 4 water molecules, the most stable conformer has a monodentate arrangement with an elevated $E_{ip}$. It may be noted that the linker length is longer in RGE compared to RGD favoring a more stable side-on salt-bridging. Hence, by comparing the bonding pattern of the minima we can qualitatively interpret $E_{ip}$ values estimated using the thermodynamic cycle.

E. Application to large peptides

We collected the NMR resolved structures of crambin (1EJG)[102], bradykinin (1BDK)[103], tryptophan zipper I (1L2Y)[104], and a scorpion toxin OSK1 (1SCO)[105]. Crambin (1EJG) is a neutral plant protein with 46 amino-acid residues and 642 atoms stabilized by the salt-bridging interaction between a carboxylate group on the main chain of Asn46 and a guanidinium group of Arg10. 1BDK is 11 residues long with 187 atoms with a net charge of +3. 1L2Y has a salt-bridging interaction between Arg16 and Asp9, containing 304 atoms and a net charge of +1. 1SCO is a scorpion toxin with an overall charge of +8 and 595 atoms, presenting a $i$-to-$i+4$ salt-bridge between Arg12 and Glu16. These 4 peptides...
Figure 7. Effect of microsolvation on intramolecular ion-pair interaction energy in four tripeptides: KGD, KGE, RGD, and RGE. In all cases, 0-water corresponds to CPCM; water implicit solvation model. With 2, 4, and 8 explicit water molecules, 10 conformers were identified for which structures and $E_{ip}$ of the lowest energy conformer are shown. Guanidinium-carboxylate salt-bridges are classified as monodentate or side-on. Green lines denote ion-pair/H-bonding.

provide an opportunity to understand the significant role orientation plays in deciding the strength of a salt-bridge. Their structures are on display in FIG. [5]. In these proteins, centroid-centroid distances of ion-pairs are < 5.0Å. 1EJG shows a side-on salt-bridge (Asn46-Arg10) with a short centroid-centroid distance leading to a significantly large magnitude of $E_{ip}$, ~24.7 kcal/mol. 1BDK is stabilized in a monodentate fashion with a relatively weaker interaction energy than 1EJG. This is followed by 1L2Y presenting an even more diminished interaction. Finally, in 1SCO we note a small positive $E_{ip}$. A closer inspection of the individual $E_{ip}$ values obtained through various mutation paths provide us with further insight. As noted before in FIG. [1], a salt-bridge with a symmetric bonding pattern should show a smaller variance across mutation paths. Hence, for 1EJG, where the salt-bridge is symmetric, all $E_{ip}$ obtained are of similar magnitude with a small standard deviation. However, for monodentate arrangements in 1BDK and 1L2Y, the symmetry is lost and we note a larger variation in the $E_{ip}$. For 1SCO, while most mutation paths yield small positive values, there is one path that yields a small negative value. Hence, the salt-bridging in this system is essentially non-interacting, if not mildly unfavorable. This could be speculated to be due to the unfavorable orientation of the interacting moieties. Despite the relatively short centroid-centroid distance, the polar groups are oriented away from each other, resulting in a net repulsion, unlike all the other cases explored here. The small positive values arise when we explore mutation paths that involve these unfavorably oriented atoms. This example demonstrates how a salt-bridging with a reasonably small centroid-centroid distance can be overall irrelevant.

V. CONCLUSIONS

In this study, we present a thermodynamic cycle to quantify intramolecular ion-pair interactions between two polar moieties of a molecule, without complex bifurcating interactions. We benchmark the efficiency of this approach by comparing intramolecular interaction energies obtained for a set of 26 zwitterions stabilized by ion-pair interactions against predictions from ISAPT. We found the deviations between the two approaches to be minimal for molecules containing long linkers and when the
Figure 8. Strength of guanidinium–carboxylate salt-bridge in the peptides 1EJG, 1BDK, 1L2Y, and 1SCO estimated using a thermodynamic cycle with ωB97X-D3(CPCM; water)/def2-SVPD total energies. The centroid–centroid distance between the interacting moieties are also given. In each case, 6 possible alchemical mutations result in different values of $E_{ip}$ shown in histograms. The net salt-bridge interaction energy corresponds to the average marked by horizontal line and the standard deviation is denoted by an envelope.

terminal groups are adequately capped. Going from the vacuum phase to the aqueous phase, the ion-pair interactions are quenched as expected.

We modeled microsolvation of four biologically relevant tripeptides and found increasing degree of solvation to quench their $E_{ip}$. In the presence of solvent molecules, the weakest interaction was found for a solvent-separated salt-bridge where the polar groups are surrounded by a cage of water molecules. The salt-bridge interaction is stronger for conformers with a bidentate side-on bonding. On the other hand, in conformers with a monodentate bonding, the salt-bridge is more exposed to solvent molecules resulting in better hydration of the biomolecule. In the aqueous environment both hydration and salt-bridging are competing stabilizing factors, the latter favoring a folded structure. We applied the thermodynamic cycle on experimental structures of 4 large proteins and found the salt-bridging interaction to be strongest for 1EJG with a side-on bonding, while 1BDK and 1L2Y with a monodentate bonding showed weak interactions. In 1SCO, even though the distance between polar groups suggests a possible salt-bridge bonding, due to the unfavorable orientation of the corresponding moieties, its $E_{ip}$ turned out to be insignificant. The favorable computational overhead of the proposed method promises explorations of salt-bridging interaction in other biological systems in the solvent phase. Since the present model can only account for simple, non-bifurcated salt-bridges, future endeavors may concentrate on quantifying interactions for complex salt-bridges.

ACKNOWLEDGMENTS

We acknowledge support of the Department of Atomic Energy, Government of India, under Project Identification No. RTI 4007. All calculations have been performed using the Helios computer cluster, which is an integral part of the MolDis Big Data facility, TIFR Hyderabad [http://moldis.tifrh.res.in].
VI. AUTHOR DECLARATIONS

A. Conflicts of Interest

The authors have no conflicts of interest to disclose.

B. Data Availability

The data that support the findings of this study are openly available in Github, Refs. [106]

[1] Renato Ferreira de Freitas and Matthieu Schapira, “A systematic analysis of atomic protein–ligand interactions in the pdb,” MedChemComm 8, 1970–1981 (2017).
[2] Sandeep Kumar and Ruth Nussinov, “Close-range electrostatic interactions in proteins,” ChemBioChem 3, 604–617 (2002).
[3] Paola Gilli, Loretta Pretto, and Gastone Gilli, “Pa/pka equalization and the prediction of the hydrogen-bond strength: a synergism of classical thermodynamics and structural crystallography,” J. Mol. Struct. 844, 328–339 (2007).
[4] Paola Gilli, Loretta Pretto, Valerio Bertolasi, and Gastone Gilli, “Predicting hydrogen-bond strengths from acid-base molecular properties. the p k a slide rule: Toward the solution of a long-lasting problem,” Acc. Chem. Res. 42, 33–44 (2009).
[5] Paola Gilli and Gastone Gilli, “Hydrogen bond models and theories: The dual hydrogen bond model and its consequences,” J. Mol. Struct. 972, 2–10 (2010).
[6] Jason E Donald, Daniel W Kulp, and William F DeGrado, “Salt bridges: geometrically specific, designable interactions,” Proteins: Struct., Funct.,Bioinf. 79, 898–915 (2011).
[7] Eric V Anslyn and Dennis A Dougherty, Modern physical organic chemistry (University science books, 2006).
[8] Martin S King, Matthew Kerr, Paul G Crichton, Roger Springett, and Edmund RS Kunji, “Formation of a cytoplasmic salt bridge network in the matrix state is a fundamental step in the transport mechanism of the mitochondrial adp/atp carrier,” Biochim. Biophys. Acta. Bioenerg. 1857, 14–22 (2016).
[9] Esperanza Rivera-de Torre, Juan Palacios-Ortega, Sara Garcia-Linares, José G Gavilanes, and Álvaro Martínez-del Pozo, “One single salt bridge explains the different cytolytic activities shown by actinoporins sticholyisin i and ii from the venom of stichodactyla helianthus,” Arch. Biochem. Biophys 636, 79–89 (2017).
[10] Sabino Pacheco, Isabel Gómez, Jorge Sánchez, Blanca-Ines García-Gómez, Mario Soberón, andAlejandra Bravo, “An intramolecular salt bridge in bacillus thuringiensis cry4ba toxin is involved in the stability of helix α-3, which is needed for oligomerization and insecticidal activity,” Appl. Environ. Microbiol. 83, e01515–17 (2017).
[11] Mason S Smith, Wendy M Billings, Frank G Whitby, McKenzie B Miller, and Joshua L Price, “Enhancing a long-range salt bridge with intermediate aromatic and nonpolar amino acids,” Org. Biomol. Chem. 15, 5882–5886 (2017).
[12] Sven Heiles, Giel Berden, Jos Oomens, and Evan R Williams, “Competition between salt bridge and non-zwitterionic structures in deprotonated amino acid dimers,” Phys. Chem. Chem. Phys. 20, 15641–15652 (2018).
[13] Vahid Dianati, Azar Shamloo, Anna Kwiatkowska, Roxane Desjardins, Armand Soldera, Robert Day, and Yves L Dory, “Rational design of a highly potent and selective peptide inhibitor of pace4 by salt bridge interaction with d160 at position p3,” ChemMedChem 12, 1169–1172 (2017).
[14] Caroline M Gorvin, Valerie N Babinsky, Tomas Malinauskas, Peter H Nissen, Anders J Schou, Aylin C Hanyaloglu, Christian Siebold, E Yvonne Jones, Fadil M Hannan, and Rajesh V Thakker, “A calcium-sensing receptor mutation causing hypocalcemia disrupts a transmembrane salt bridge to activate β-arrestin–biased signaling,” Sci. Signal. 11, eaan3714 (2018).
[15] Xiao Zhang, Linda Li, Ning Li, Xinyu Shi, Liwen Zhou, Shouqin Lü, Shenbao Chen, Debin Mao, and Mian Long, “Salt bridge interactions within the β3 integrin α7 helix mediate force-induced binding and shear resistance ability,” PFEBS J. 285, 261–274 (2018).
[16] Ariel Ben-Bassat, Moshe Giladi, and Yoni Haitin, “Structure of kcnh2 cyclic nucleotide-binding homology domain reveals a functionally vital salt-bridge,” J. Gen. Physiol. 152 (2020).
[17] TJ McManus, SA Wells, and AB Walker, “Salt bridge impact on global rigidity and thermostability in thermophilic citrate synthase,” Phys. Biol. 17, 016002 (2019).
[18] Wen-Shan Zhao, Meng-Yang Sun, Liang-Fei Sun, Yan Liu, Yang Yang, Li-Dong Huang, Ying-Zhe Fan, Xiao-Yang Cheng, Peng Cao, You-Min Hu, et al., “A highly conserved salt bridge stabilizes the kinked conformation of α2, 3-sheet essential for channel function of p2x4 receptors,” J. Biol. Chem. 291, 7990–8003 (2016).
[19] Nina B Hentzen, Valdrin Islami, Martin Kohler, Renato Zenobi, and Helma Wennemers, “A lateral salt bridge for the specific assembly of an abc-type collagen heterotrimer,” J. Am. Chem. Soc. 142, 2208–2212 (2020).
[20] Erik J Spek, An H Bui, Min Lu, and Neville R Kallenbach, “Surface salt bridges stabilize the gcna4 leucine zipper,” Protein Sci. 7, 2331–2337 (1998).
[21] Costantino Vetriani, Dennis L Maeder, Nicola Tolliday, Kitty S-P Yip, Timothy J Stillman, K Linda Britton, David W Rice, Horst H Klumpp, and Frank T Robb, “Protein thermostability above 100 °C: a key role for ionic interactions,” Proc. Natl. Acad. Sci. USA 95, 12300–12305 (1998).

[22] Li Xiao and Barry Honig, “Electrostatic contributions to the stability of hyperthermophilic proteins,” J. Mol. Biol. 289, 1435–1444 (1999).

[23] Sandeep Kumar, Chung-Jung Tsai, and Ruth Nussinov, “Factors enhancing protein thermostability,” Protein Eng. 13, 179–191 (2000).

[24] Sandeep Kumar, Buyong Ma, Chung-Jung Tsai, and Ruth Nussinov, “Electrostatic strengths of salt bridges in thermophilic and mesophilic glutamate dehydrogenase monomers,” Proteins: Struct., Funct., Bioinf. 38, 368–383 (2000).

[25] S Kumar and R Nussinov, “How do thermophilic proteins deal with heat?,” Cell. Mol. Life Sci. 58, 1216–1233 (2001).

[26] Amnon Horovitz, Luis Serrano, Boaz Avron, Mark Bycroft, and Alan R Fersht, “Strength and co-operativity of contributions of surface salt bridges to protein stability,” J. Mol. Biol. 216, 1031–1044 (1990).

[27] Luis Serrano, Amnon Horovitz, Boaz Avron, Mark Bycroft, and Alan R Fersht, “Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles,” Biochemistry 39, 9433–9439 (1990).

[28] Dao Pin Sun, Uwe Sauer, Hale Nicholson, and Brian W Matthews, “Contributions of engineered surface salt bridges to the stability of tet lysozyme determined by directed mutagenesis,” Biochemistry 30, 7142–7153 (1991).

[29] Mark Bycroft, Alan R Fersht, et al., “Surface electrostatic interactions contribute little to stability of barnase,” J. Mol. Biol. 220, 779–788 (1991).

[30] Zachary S Hendisch and Bruce Tidor, “Do salt bridges stabilize proteins? a continuum electrostatic analysis,” Protein Sci. 3, 211–226 (1994).

[31] Pavel Strop and Stephen L Mayo, “Contribution of surface salt bridges to protein stability,” Biochemistry 39, 1251–1255 (2000).

[32] Adriana Huerta-Viga, Saeed Amirjalayer, Sérgio R Domingos, Heline Meunzebra, Alisa Rupenyan, and Sander Woutersen, “The structure of salt bridges between arg-+ and glu- in peptides investigated with 2d-ir spectroscopy: Evidence for two distinct hydrogen-bond geometries,” J. Chem. Phys. 142, 21244 (2015).

[33] S Lotze and HJ Bakker, “Structure and dynamics of a salt-bridge model system in water and dmso,” J. Chem. Phys. 142, 212436 (2015).

[34] Meng Ge, Xia-Yu Xia, and Xian-Ming Pan, “Salt bridges in the hyperthermophilic protein ssh10b are resilient to temperature increases,” J. Biol. Chem. 283, 31690–31696 (2008).

[35] Chi-Ho Chan, Tsz-Ha Yu, and Kam-Bo Wong, “Stabilizing salt-bridge enhances protein thermostability by reducing the heat capacity change of unfolding,” PloS one 6, e21624 (2011).

[36] George I Makhatadze, Vakhtang V Loladze, Dmitri N Ermolenko, XiaoFen Chen, and Susan T Thomas, “Contribution of surface salt bridges to protein stability: guidelines for protein engineering,” J. Mol. Biol. 327, 1135–1148 (2003).

[37] AC Tissot, S Vuilleumier, and AR Fersht, “Importance of two buried salt bridges in the stability and folding pathway of barnase,” Biochemistry 35, 6786–6794 (1996).

[38] Andrew D White, Andrew J Keefe, Jean-Rene Ellmenye, Ann K Nowinski, Qing Shao, Jim Pfendtner, and Shaoyi Jiang, “Free energy of solvated salt bridges: A simulation and experimental study,” J. Phys. Chem. B 117, 7254–7259 (2013).

[39] Mrityunjay K Tiwari and Rajesh K Murarka, “Interaction strength of osmolytes with the anion of a salt-bridge determines its stability,” Phys. Chem. Chem. Phys. 23, 5527–5539 (2021).

[40] Kazufumi Takano, Kimiko Tsuchimori, Yuriko Yamagata, and Katsuhide Yutani, “Contribution of salt bridges near the surface of a protein to the conformational stability,” Biochemistry 39, 12375–12381 (2000).

[41] Anne-Marie Saspe, Robert Rothchild, Duli C Jain, and Cecilia G Unson, “The role of salt bridge formation in glucagon: an experimental and theoretical study of glucagon analogs and peptide fragments of glucagon,” Mol. Med. 8, 251–262 (2002).

[42] Tom Ziegler and Arvi Ranke, “On the calculation of bonding energies by the hartree fock Slater method,” Theor. Chem. Acta 46, 1–10 (1975).

[43] Keiji Morokuma, “Molecular orbital studies of hydrogen bonds. iii. c==o…h-o hydrogen bond in h2co…h2o and h2co… h2o,” J. Chem. Phys. 55, 1236–1244 (1971).

[44] Peifeng Su and Hui Li, “Energy decomposition analysis of covalent bonds and intermolecular interactions,” J. Chem. Phys. 131, 014102 (2009).

[45] Moritz von Hopfgarten and Gernot Frenking, “Energy decomposition analysis,” Wiley Interdiscip. Rev. Comput. Mol. Sci. 2, 43–62 (2012).

[46] Lili Zhao, Moritz von Hopfgarten, Diego M Andrade, and Gernot Frenking, “Energy decomposition analysis,” Wiley Interdiscip. Rev. Comput. Mol. Sci. 8, e1345 (2018).

[47] Bogumil Jeziorski, Robert Moszynski, and Krzysztof Szalewicz, “Perturbation theory approach to intermolecular potential energy surfaces of van der waals complexes,” Chem. Rev. 94, 1887–1930 (1994).

[48] Krzysztof Szalewicz, “Symmetry-adapted perturbation theory of intermolecular forces,” Wiley Interdiscip. Rev. Comput. Mol. Sci. 2, 254–272 (2012).

[49] Edward G Hohenstein and C David Sherrill, “Wavefunction methods for noncovalent interactions,” Wiley Interdiscip. Rev. Comput. Mol. Sci. 2, 304–326 (2012).

[50] Jerome F Gonthier and Clemente Cornimbouef, “Quantification and analysis of intramolecular interactions,” Chimia 68, 221–226 (2014).

[51] Clémence Cornimbouef, “Quantifying intra-and intermolecular phenomena: challenging yet exciting territory for quantum chemistry,” Chimia 68, 512–515 (2014).

[52] Konrad Patkowski, “Recent developments in symmetry-adapted perturbation theory,” Wiley Interdiscip. Rev. Comput. Mol. Sci. 10, e1452 (2020).

[53] Robert M Parrish, Jérôme F Gonthier, Clémence Corminboeuf, and C David Sherrill, “Communication: Practical intramolecular symmetry adapted perturbation theory via hartree-fock embedding,” J. Chem. Phys. 143, 051103 (2015).
12 other functionals,” Theor. Chem. Acc. 120, 215–241 (2008).

[86] Hisayoshi Ikura, Takao Tsuneda, Takeshi Yanai, and Kimihiko Hirao, “A long-range correction scheme for generalized-gradient-approximation exchange functionals,” J. Chem. Phys. 115, 3540–3544 (2001).

[87] Asim Najibi, Marcos Casanova-Páez, and Lars Gogerik, “Analysis of recent blyp-and pbe-based range-separated double-hybrid density functional approximations for main-group thermochemistry, kinetics, and noncovalent interactions,” J. Phys. Chem. A 125, 4026–4035 (2021).

[88] Yoshihiro Tawada, Takao Tsuneda, Susumu Yanagisawa, Takeshi Yanai, and Kimihiko Hirao, “A long-range-corrected time-dependent density functional theory,” J. Chem. Phys. 120, 8425–8433 (2004).

[89] Takeshi Yanai, David P Tew, and Nicholas C Handy, “A new hybrid exchange–correlation functional using the coulomb-attenuating method (cam-b3lyp),” Chem. Phys. Lett. 393, 51–57 (2004).

[90] Rafal Kurczab, Pawel Sliwa, Krzysztof Rataj, and Andrzej J Bojarski, “Salt bridge in ligand–protein complexes—systematic theoretical and statistical investigations,” J. Chem. Inf. Model. 58, 2224–2238 (2018).

[91] Mustapha Carab Ahmed, Elena Papaleo, and Kresten Lindorff-Larsen, “How well do force fields capture the strength of salt bridges in proteins?” PeerJ 6, e4967 (2018).

[92] Yandong Huang, Robert C Harris, and Jana Shen, “Generalized born based continuous constant ph molecular dynamics in amber: Implementation, benchmarking and analysis,” J. Chem. Inf. Model. 58, 1372–1383 (2018).

[93] MJS Phipps, T Fox, CS Tautermann, and C-K Skylar-Isaac, “Intuitive density functional theory-based energy decomposition analysis for protein–ligand interactions,” J. Chem. Theory Comput. 13, 1837–1850 (2017).

[94] Andrea N Bootsma and Steven E Wheeler, “Tuning stacking interactions between asp-arg salt-bridges and heterocyclic drug fragments,” J. Chem. Inf. Model. (2018).

[95] Andrea N Bootsma, Analyse C Doney, and Steven E Wheeler, “Predicting the strength of stacking interactions between heterocycles and aromatic amino acid side chains,” J. Am. Chem. Soc. 141, 11027–11035 (2019).

[96] Alex M Maldonado, Satoshi Hagiwara, Tae Hoon Choi, Frank Eckert, Kathleen Schwarz, Ravishankar Sundaraman, Minoru Otani, and John A Keith, “Quantifying uncertainties in solvation procedures for modeling aqueous phase reaction mechanisms,” J. Phys. Chem. A 125, 154–164 (2021).

[97] Brian M Gentry, Tae Hoon Choi, William S Belfield, and John A Keith, “Computational predictions of metal–macrocycle stability constants require accurate treatments of local solvent and ph effects,” Phys. Chem. Chem. Phys. 23, 9189–9197 (2021).

[98] Alfonso T García-Sosa, “Hydration properties of ligands and drugs in protein binding sites: tightly-bound, bridging water molecules and their effects and consequences on molecular design strategies,” J. Chem. Inf. Model. 53, 1388–1405 (2013).

[99] Eva Pluliahová, Ondrej Marsalek, Burkhard Schmidt, and Pavel Jungwirth, “Peptide salt bridge stability: From gas phase via microhydration to bulk water simulations,” J. Chem. Phys. 137, 185101 (2012).

[100] Gerhard Schwaab, Ricardo Pérez de Tudela, Devendra Mani, Nitish Pal, Tarun Kumar Roy, Fabio Gabas, Ricardo Conte, Laura Durán Caballero, Michele Ceotto, Dominik Marx, et al., “Zwitter ionization of glycine at outer space conditions due to microhydration by six water molecules,” Phys. Rev. Lett. 128, 033001 (2022).

[101] Lucas J Karas, Patrick R Batista, Renan V Viesser, Cláudio F Tormena, Roberto Rittner, and Paulo R de Oliveira, “Trends of intramolecular hydrogen bonding in substituted alcohols: A deeper investigation,” Phys. Chem. Chem. Phys. 19, 16904–16913 (2017).

[102] Kalyaneswar Mandal, Brad L Pentelute, Duhee Bang, Zachary P Gates, Vladimir Yu Torbeev, and Stephen BH Kent, “Design, total chemical synthesis, and x-ray structure of a protein having a novel linear-loop polypeptide chain topology,” Angew. Chem. Int. Ed. 51, 1481–1486 (2012).

[103] Jan Sejbal, John R Cann, John M Stewart, Lajos Gera, and George Kotovych, “An nmr, cd, molecular dynamics, and fluorometric study of the conformation of the bradykinin antagonist b-9340 in water and in aqueous micellar solutions,” J. Med. Chem. 39, 1281–1292 (1996).

[104] Jonathan W Neidigh, R Matthew Fesmiremeyer, and Niels H Andersen, “Designing a 20-residue protein,” Nat. Struct. Biol. 9, 425–430 (2002).

[105] Victor A Jaravine, Dmitry E Nolde, Michail J Reibarkh, Yulia V Korolkova, Sergey A Kozlov, Kirill A Pluzhnikov, Eugene V Grishin, and Alexander S Arseniev, “Three-dimensional structure of toxin osk1 from orthocirrus scrobiculosus scorpion venom,” Biochemistry 36, 1223–1232 (1997).

[106] moldis-group, “https://github.com/moldis-group/si_intraionpair,” (2022).