Uncovering Differences in Hydration Free Energies for Model Compound Mimics of Charged Sidechains of Amino Acids

Martin J. Fossat†, Xiangze Zeng†, and Rohit V. Pappu*

Department of Biomedical Engineering and Center for Science & Engineering of Living Systems (CSELS)

Washington University in St. Louis, St. Louis, MO 63130, USA

ABSTRACT

Free energies of hydration are fundamental quantities that are relevant for modeling and understanding sequence-specific conformational and phase equilibria of intrinsically disordered proteins (IDPs). Of particular importance are accurate values for the free energies of hydration of model compounds that mimic the charged sidechains of Arg, Lys, Asp, and Glu. Direct measurements of these quantities are challenging, and accordingly the values used to date rely on decompositions of measurements of whole salts. This has created considerable uncertainty regarding the desired quantities. Here, we adapt and deploy a Thermodynamic Cycle based Proton Dissociation in conjunction with published data from direct measurements along the relevant thermodynamic cycle to obtain new values for the free energies of hydration for model compounds that mimic the sidechains of Arg, Lys, Asp, and Glu. We obtain independent assessments for these inferred values using direct free energy calculations based on the polarizable AMOEBA forcefield and water model. The TCPD and AMOEBA derived values agree with one another. The new numbers suggest that Arg has the least favorable free energy of solvation, followed by Lys. On average, the sidechains of Asp and Glu have free energies of hydration that are ~1.8 times more favorable than Arg and ~1.5 times more favorable than Lys. Calculations used to extract the temperature dependence of free energies of hydration reveal positive heat capacities of hydration for Arg and Lys and negative heat capacities of hydration for Asp and Glu. We analyze the hydration structures around the different solutes to uncover a structural explanation for the observed differences in free energies of hydration. Although the sidechains of Arg and Lys are both amphiphilic, we find that the Arg sidechain has more of a hydrophobic character. In contrast, both Asp and Glu are hydrophilic both in terms of their solvation thermodynamics and hydration structure. Our results provide a physico-chemical explanation for recent accounts documenting the differences of Arg and Lys with respect to one another and with respect to Asp / Glu as determinants of conformational and phase equilibria of IDPs.

1. INTRODUCTION

There is growing interest in uncovering the sequence-specific conformational preferences of intrinsically disordered proteins (IDPs)[1], and in using these insights to quantify sequence-specific contributions to the driving forces for phase separation [2]. Investigations that combine theory,[3] computation,[4] and multi-pronged experiments[5] highlight how the interplay between sequence-specific intra-chain and chain-solvent interactions determines the conformational preferences[6] and functions of IDPs[7] for given solution conditions. Indeed, IDPs can be classified into distinct categories based on their sequence-ensemble relationships [1]. These are being leveraged to describe the form [8], functions [9], and overall phase behaviors [10] of different IDPs.

In a purely additive model [11], the inferred quantitative sequence-ensemble relationships of IDPs can be rationalized using the free energies of hydration of model compound mimics of sidechain and backbone moieties [12]. The temperature and solute dependence of conformational

equilibria can also be rationalized by how free energies of hydration of model compound mimics of sidechain and backbone moieties change with changes to solution conditions such as temperature [13]. Indeed, the molecular transfer model (MTM) of Thirumalai coworkers[14] is a direct illustration of how free energies of solvation can be used to obtain predictive, coarse-grained descriptions of conformational transitions as a function of denaturant concentrations [15].

The free energy of hydration ($\Delta\mu_h$) at a specific temperature and pressure is defined as the change in free energy change associated with transferring the solute of interest from a dilute vapor phase into water [16]. Vapor pressure osmometry with radioactive labeling was an early method adopted by Wolfenden[17] to measure free energies of hydration. While this works for polar solutes, including neutral forms of ionizable species, it cannot be used to measure free energies of hydration of ionizable residues because of the ultra-low vapor pressures and the confounding effects of ion-pairing in the gas phase. Calorimetry [18] is also problematic because of the large magnitudes of free energies of hydration for ionizable residues [19]. And because stable solutions are electroneutral [19b], measurements of free energies of hydration of ionic species have to rely on parsing numbers derived from measurements of whole salts against those of a suitable reference system [20]. Parsing measurements for whole salts also rests on the assumption of minimal ion-pairing or clustering, which need not be true in general, especially for organic ions [21].

The challenges associated with estimating free energies of hydration for model compound mimics of charged sidechains are illustrated in the tabulations of Wuttke et al.,[13]. They adapted previously tabulated values of $\Delta\mu_h$ [22] using three different models. In Model 1 they used measured enthalpies and heat capacities for neutral compounds [18b], i.e., protonated Asp and Glu and deprotonated Lys and Arg. Model 2 uses the enthalpies of hydration estimated by Marcus [23] and the heat capacities of hydration tabulated by Abraham and Marcus [24]. These numbers were derived from measurements of whole salts and then decomposed using different models to arrive at consensus estimates for the enthalpies and heat capacities. Model 3 uses the heat capacities of Model 2; analysis of data for homologous series of model compounds were then used to make empirical choices for the enthalpies. The large uncertainties due to difficulties associated with direct measurements of $\Delta\mu_h$ are illustrated by different estimates obtained using different models. For example, Models 1 and 2 yield values of 50.37 cal $mol^{-1}K^{-1}$ and 5.30 cal $mol^{-1}K^{-1}$, respectively for the heat capacity of hydration of the acetate ion. Likewise, large variations were obtained depending on how one parses experimental data for whole salts or uses measurements for neutralized versions of charged model compounds. The overall conclusion is that quantities of fundamental importance namely the values of $\Delta\mu_h$ for model compound mimics of Arg, Lys, Asp, and Glu sidechains, are plagued by considerable uncertainties.

Here, we adapt a **T**hermodynamic **C**ycle based on **P**roton **D**issociation (TCPD) to obtain improved estimates for free energies of hydration of charged amino acids [25]. This approach uses inputs from: (i) direct measurements of proton dissociation energies in the gas phase; (ii) measured pK_a values in the aqueous phase; (iii) measured free energies of solvation of uncharged variants of charged residues; and (iv) reliable estimates of the free energy of hydration of the proton in the aqueous phase. The TCPD approach was introduced by Pliego and Riveros[26]. They estimated absolute values for $\Delta\mu_h$ for 30 different univalent ions, many of which are organic ions. They relied on the estimates for the free energy of hydration of the proton that was provided by Tissandier et al., [27]. As discussed below, this value is on the high side in terms of magnitude. Further, there were challenges with obtaining $\Delta\mu_h$ for the model compound that mimics the Arg sidechain because of persistent uncertainties regarding its pK_a in the aqueous phase and the absence of data for proton

dissociation in the gas phase. Improved estimates are now available for all the relevant quantities. Accordingly, we adapted these for obtaining TCPD derived values for the free energies of hydration of charged amino acids as illustrated in **Figure 1**. The TCPD derived estimates serve as updated reference values for the free energies of hydration of charged amino acids. Importantly, we go beyond the TCPD derived values and show that these values are reproduced using free energy calculations that use the AMOEBA forcefield for water and model compound mimics of the sidechains of Arg, Lys, Asp, Glu. This gives us confidence in analyzing and interpreting structural differences within first hydration shells and related quantities to explain differences in hydration free energies of model compounds that mimic sidechains of charged residues.

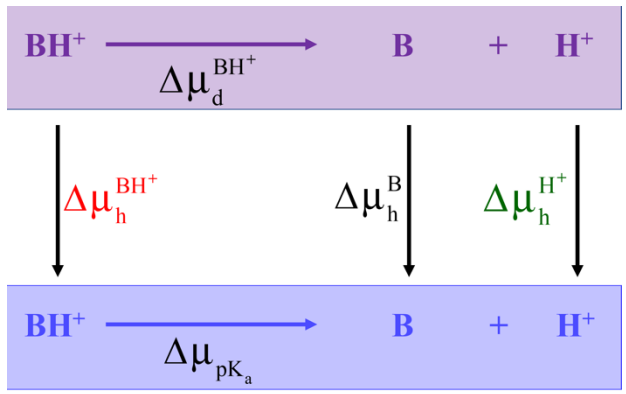

Figure 1: Illustration of the TCPD approach. The schematic shows the transfer of a strong base and the proton dissociation reaction from the gas phase (purple) into the aqueous phase (blue).

2. METHODS

2.1 Details of the TCPD approach: The concepts underlying the TCPD approach are summarized in **Figure 1**. The model compounds used as mimics for these sidechains are n-propylguanidine (Arg), 1-butylamine (Lys), acetate (Asp), and propionate (Glu). For bases, deprotonation reactions in the gas and aqueous phases are written as $BH^+ \xrightleftharpoons{\Delta\mu_d^{BH^+}} B + H^+$ and $BH^+ \xrightleftharpoons{\Delta\mu_{pK_a}} B + H^+$, respectively. Here, $\Delta\mu_d^{BH^+}$ quantifies the change in free energy that accompanies the dissociation of a proton from the base in the gas phase [28] whereas $\Delta\mu_{pK_a}$ is the equivalent quantity in the aqueous phase [29]. The free energy of proton dissociation in the aqueous phase can be obtained from knowledge of the pK_a for the model compound of interest whereby $\Delta\mu_{pK_a} = RT\ln\left(10^{pK_a}\right)$. Here, where $R = 1.98717 \times 10^{-3}$ kcal /mol-K and T is temperature. This approach is the complement of the method used by Jorgensen and Briggs [30] for their *a priori* calculation of pK_a values from knowledge / assumption of free energies of hydration of charged and uncharged variants. Using measured values of $\Delta\mu_d^{BH^+}$ and of $\Delta\mu_h^{B}$ combined with estimation of $\Delta\mu_{pK_a}$ based on measured pK_a values, and knowledge of the free energy of hydration of the proton $\Delta\mu_h^{H^+}$ allows the usage of TCPD analysis to obtain the free energy of hydration $\Delta\mu_h^{BH^+}$ of the protonated base as shown in Equation (1):

$$\Delta\mu_h^{BH^+} = \left(\Delta\mu_h^{B} + \Delta\mu_d^{BH^+} - \Delta\mu_{pK_a} + \Delta\mu_h^{H^+}\right); \quad (1)$$

Likewise, for acids, the free energies of hydration ($\Delta\mu_h^{A^-}$) of the deprotonated forms are calculated using Equation (2):

$$\Delta\mu_h^{A^-} = \left(\Delta\mu_h^{AH} - \Delta\mu_d^{AH} + \Delta\mu_{pK_a} - \Delta\mu_h^{H^+}\right); \quad (2)$$

Here, $\Delta\mu_h^{AH}$ is the free energy of hydration of the protonated form of the acid, which is measured directly [31], and $\Delta\mu_{pK_a}$ is estimated using measured pK_a values of the acid [32]; $\Delta\mu_d^{AH}$ quantifies the change in free energy that accompanies the dissociation of a proton from the acid in the gas phase; the value to be used for $\Delta\mu_h^{H^+}$ is identical to that of Equation (1).

The TCPD approach has been proposed [33] and used in the literature [26], and its usage requires accurate measurements of the relevant parameters [34]. These are now available in the form of accurate proton dissociation / association energies in the gas phase, well-established and reliable values for $\Delta\mu_h^{B}$ and $\Delta\mu_h^{AH}$, improved estimates of the pK_a values, especially for the Arg sidechain, and consensus estimates for $\Delta\mu_h^{H^+}$ that have been obtained using quantum mechanical calculations.

2.2 Set up of simulations using the AMOEBA forcefield: The free energy calculations we report here are an extension of the recent simulations performed by Zeng et al., [12]. The authors developed parameters of the requisite model compounds for the AMOEBA forcefield. The details of the forcefield parameterization, set up of the simulations, and free energy calculations may be found in the work of Zeng et al., [12]. In the interest of completeness, we include a summary of the overall simulation setup and analysis of simulation results. Simulations were performed using the TINKER-OpenMM package [35]. For each model compound, the simulations were performed using a cubic water box with periodic boundary conditions. The initial dimensions of the central cell were set to be 30×30×30 $Å^3$. All molecular dynamics simulations were performed using the RESPA integrator [36] with an inner time step of 0.25 ps and an outer time step of 2.0 fs in isothermal-isobaric ensemble (NPT) ensemble. The target temperature was set to be one of 275 K, 298 K, 323 K, 348 K, or 273 K and the target pressure was set to be 1 bar. The temperature and pressure were controlled using a stochastic velocity rescaling thermostat [37] and a Monte Carlo constant pressure algorithm [38], respectively. The particle mesh Ewald (PME) method [39] with PME-GRID being 36×36×36 $Å^3$, and *B*-spline interpolation [40], with a real space cutoff of 7 Å was used to compute long-range corrections to electrostatic interactions. The cutoff for van der Waals interactions was set to be 12 Å. This combination has been verified [41] in previous work for AMOEBA-based free energy simulations [42].

2.3 Free energy calculations: We used the Bennett Acceptance Ratio (BAR)[43] method to quantify the free energies of solvation for the model compounds of interest. Details of the simulation setup and free energy estimations are identical to those of Zeng et al., [12].The solute is grown in using two different Kirkwood coupling parameters λ_{vdW} and λ_{el} that scale the strengths of

solute-solute and solute-solvent van der Waals and electrostatic interactions. A series of independent molecular dynamics simulations were performed in the NPT ensemble for different combinations of λ_{vdW} and λ_{el}. A soft-core modification of the Buffered-14-7 function was used to scale the vdW interactions as implemented in Tinker-OpenMM [35]. For each pair of λ values, we performed simulations, each of length 6 ns, at the desired temperature and a pressure of 1 bar. We then used the TINKER `bar` program to calculate the free energy difference between neighboring windows defined in terms of the scaling coefficients. For every combination of λ_{vdW} and λ_{el}, we set aside the first 1 ns simulation as part of the equilibration process.

3. RESULTS

3.1 Free energies of hydration calculated using the TCPD approach: Values of free energies of hydration for model compound mimics of the charged forms of Arg, Lys, Asp, and Glu sidechains were calculated at 298 K using the TCPD approach – see Equations (1) and (2) and **Figure 1**. The free energy of hydration of the proton $\Delta\mu_h^{H^+}$ is a crucial parameter that determines the outputs we obtain using the TCPD approach. The difference between the gas phase basicity [44] and $\Delta\mu_h^{H^+}$ quantifies the relative importance of bond energy and the free energy of solvation. Positive values for this difference imply that the favorable free energy of hydration of the proton cannot compensate for the loss of bond energy in the gas phase. In order to achieve the target pK_a value for the ionizable moiety, suitably large magnitudes for $\Delta\mu_h^{BH^+}$ and $\Delta\mu_h^{A^-}$ help offset the loss of bond energy in the gas phase.

Values of $\Delta\mu_{pK_a}$ are derived from measurements of pK_a values for the relevant model compound mimics of Arg, Lys, Asp, and Glu. The gas phase dissociation energies $\Delta\mu_d^{BH^+}$ and $\Delta\mu_d^{AH}$, estimated from gas-phase basicity measurements [44], are available from the literature for three of the four model compounds. For n-propylguanidine, which is the model compound mimic of the Arg sidechain, we set the gas phase basicity to be an average over seven different quantum-mechanical (QM) calculations [7]. The errors in our estimates of the free energy of hydration of Arg account for the standard deviation across the different estimates for the gas phase basicity. This choice is rationalized by the observed congruence between measured gas phase basicity values and QM calculations for the model compound mimics of Lys, Asp, and Glu. This congruence suggests that the QM calculations for the gas phase basicity of Arg are likely to be reliable.

Values for free energies of hydration for uncharged constructs, i.e., $\Delta\mu_h^{B}$ and $\Delta\mu_h^{AH}$ were obtained from the Hydration Free Energy Database curated by Mobley and Guthrie [7]. The free energy of hydration of the proton was set to be -252.5 kcal/mol, which is the value obtained using free energy calculations and the AMOEBA forcefield [20]. This value is very similar to the consensus across multiple, independent quantum mechanical calculations (see footnote to **Table 1**). Estimates for the free energy of hydration of the proton can range from -251 kcal / mol to -264 kcal/mol. Accordingly, the different values we get to choose for $\Delta\mu_h^{H^+}$ are the main source of variation in our estimates of $\Delta\mu_h^{BH^+}$ and $\Delta\mu_h^{A^-}$ using the TCPD approach. The pK_a values of all four model compounds were taken from the Physical/Chemical Property Database (PHYSPROP)[3] database and they reflect updates from the measurements of Fitch et al., [45] and Xu et al., [46]. These

measurements move the consensus estimate for the pK_a of Arg up from 12.6 (which was never independently measured) to 13.6.

Results from application of the TCPD approach are summarized in **Table 1**. In addition to the values of $\Delta\mu_h^{BH^+}$ and $\Delta\mu_h^{A^-}$ we also tabulate the values used for measured and / or calculated gas phase basicity values, measured pK_a values, and free energies of hydration for the uncharged forms of the model compounds. These estimates are for a temperature of 298 K. Errors in estimates of the independently measured / calculated quantities, including the standard deviation across different estimates for the proton hydration free energy, are propagated in our usage of Equations (1) and (2) to arrive at errors in estimates of the desired free energies of hydration. We also quantify the differences between estimates of $\Delta\mu_h^{BH^+}$ and $\Delta\mu_h^{A^-}$ obtained using the TCPD approach and values that are used currently in different simulation paradigms.

Table 1: Summary of inputs and outputs in the TCPD approach used for estimating values free energies of hydration for model compounds that mimic charged versions of Arg, Lys, Asp, and Glu. † Column seven quantifies the discrepancies between estimates for free energies of hydration that are based on the TCPD approach and extant estimates used in the literature.

Sidechain mimicked by the model compound	**Measured pK_a[47]**	**Gas phase basicity (kcal / mol)[28, 48]**	**Measured values of $\Delta\mu_h^{B}, \Delta\mu_h^{AH}$ (kcal / mol)[44]**	**TCPD derived values for $\Delta\mu_h^{BH^+}, \Delta\mu_h^{A^-}$ (kcal / mol) (I)**	**Commonly used values for $\Delta\mu_h^{BH^+}, \Delta\mu_h^{A^-}$ (kcal / mol)[22, 49] II**	**(I) – (II) (kcal / mol)**
Arg	13.6 [46]	234.6 ± 2.0	-10.9 ± 1.9	**-47.3 ± 2.7**	-70.9	*23.6*
Lys	10.7 [50]	211.3 ± 0.5	-4.3 ± 1.9	**-60.1 ± 1.9**	-70.9	*10.8*
Asp	4.76 [51]	341.4 ± 1.2	-6.7 ± 1.9	**-89.1 ± 2.2**	-77.3	*-11.8*
Glu	4.88 [51]	340.4 ± 1.4	-6.5 ± 1.9	**-87.7 ± 2.3**	-77.3	*-10.4*

†Independent estimates for the free of hydration of the proton are as follows: -249.5 kcal / mol [52], -252.5 kcal / mol [20], -252.7 kcal / mol [53], -253.4 kcal / mol [54], -254.3 kcal / mol [34], -262.4 kcal / mol [55], and -264.0 kcal / mol [27]. In our calculations, we use the value of -252.5 kcal / mol, which was reported by Grossfield et al.,[20] based on their free energy calculations that used the AMOEBA forcefield [56]. The error bars for the desired free energies incorporate the contributions, via error propagation, from standard errors in gas phase basicity values (column 3) and free energies of hydration of the uncharged species (column 4). Across the seven estimates, the standard deviation for the free energy of hydration the proton is 5.5 kcal / mol, although there are two distinct clusters of values. By setting the standard error of the free energy of hydration of the proton to the standard deviation while ignoring the fact that five of the seven values are closer to the chosen value of -252.5 kcal / mol, we obtain error bars for the desired free energies that

are as follows: 6.15 kcal / mol, 6.15 kcal / mol, 5.84 kcal / mol, and 5.94 kcal / mol for the model compound mimics of Arg, Lys, Asp, and Glu, respectively. These represent extreme estimates that give equal weight to the outliers in free energies of hydration of the proton namely, -262.4 kcal / mol [55] and -264.0 kcal / mol [27].

As expected, the estimated free energies of hydration are large and negative. However, the TCPD based estimates revealed unexpected trends. Despite being a strong base, the estimated value $\Delta\mu_h^{BH^+}$ is ~12 kcal / mol less favorable for the model compound mimic of Arg when compared to the model compound mimic of Lys. Both compounds have surfactant-like character with hydrocarbon tails and charged headgroups. However, the chemical structures and electron cloud delocalization across the planar guanidinium and tertiary amines are different from one another. Further, the estimates for $\Delta\mu_h^{BH^+}$ (Arg and Lys) are smaller in magnitude than those of $\Delta\mu_h^{A^-}$ (Asp and Glu) implying that the acids are more favorably hydrated than bases – an observation that is concordant with results for free energies of hydration of small anions versus cations [57]. These results suggest that the ~30-40 kcal / mol differences in free energies of hydration of basic versus acidic residues will likely have a profound impact on the conformational and phase equilibria of acid versus base-rich IDPs.

3.2 Independent assessments of TSPD estimates using free energy calculations based on the AMOEBA forcefield: We used free energy calculations [12] based on the AMOEBA forcefield [58] to obtain independent assessments of the estimates from the TSPD approach. For each of the model compounds, the AMOEBA based simulations were used to calculate free energies of hydration at four different temperatures *viz*., 275 K, 298 K, 323 K, 348 K, and 373 K. The results are summarized in **Table 2**. At 298 K, the differences between the free energies of hydration calculated using the AMOEBA forcefield versus those estimated using the TCPD approach range between 0.4 and ~1.6 kcal / mol. The root mean squared deviation (RMSD) between the two estimates, calculated across all four model compounds is 0.96 kcal / mol. Importantly, these differences are considerably smaller than the RMSD of 14.97 kcal / mol that we obtain when comparing the TCPD derived free energies to current estimates derived from decompositions of data for whole salts. The consistency between TCPD derived estimates and direct free energy calculations using the AMOEBA forcefield suggest that the values summarized in **Tables 1** and **2** are useful as new touchstones for free energies of hydration of charged sidechains.

Table 2: Summary of results obtained from free energy calculations based on the AMOEBA forcefield.

Sidechain mimicked by the model compound	$\Delta\mu_h^B$ and $\Delta\mu_h^{AH}$ (kcal / mol) 275 K	$\Delta\mu_h^B$ and $\Delta\mu_h^{AH}$ (kcal / mol) 298 K	$\Delta\mu_h^B$ and $\Delta\mu_h^{AH}$ (kcal / mol) 323 K	$\Delta\mu_h^B$ and $\Delta\mu_h^{AH}$ (kcal / mol) 358 K	Difference between AMOEBA and TCPD values at 298 K (kcal / mol)†
Arg	-47.63 ± 0.11	**-46.72 ±0.11**	-45.67 ±0.11	-45.45 ±0.09	*-0.61 ± 2.7*
Lys	-61.22 ± 0.10	**-60.49 ±0.09**	-59.53 ±0.09	-58.99 ±0.08	*-0.37 ± 1.9*
Asp	-90.39 ± 0.08	**-89.91 ±0.08**	-89.02 ±0.07	-88.15 ±0.07	*-0.81 ± 2.2*
Glu	-86.84 ± 0.12	**-86.16 ±0.12**	-85.24 ±0.12	-84.32 ±0.11	*+1.58 ± 2.3*

†These are differences between values shown in column 3 of **Table 2** and column 5 of **Table 1**. Error bars represent statistical errors [59] in BAR-based estimates for the free energies.

3.3 Insights from analysis of the temperature dependence of calculated free energies of hydration: Table 2 shows how the free energies of hydration vary with temperature for each of the model compounds. The consistent trend is of free energies of hydration becoming less favorable as temperature increases. We fit the temperature dependent data for $\Delta\mu_h^{BH^+}$ (Arg and Lys) and $\Delta\mu_h^{A^-}$ (Asp and Glu) to the integral of the Gibbs-Helmholtz equation in order to estimate the enthalpy of hydration (Δh) and the heat capacity of hydration (Δc_P) at a reference temperature of 298 K. In doing so, we assume that values of Δh and Δc_P are independent of temperature, a conjecture that is supported by the linear increase in the magnitudes of $\Delta\mu_h^{BH^+}$ and $\Delta\mu_h^{A^-}$ with increasing temperature. The integral of the Gibbs-Helmholtz equation is written as:

$$\Delta\mu_h(T) = \frac{\left[\Delta\mu_h(T_0) - \Delta h\right]T}{T_0} + \Delta h + \Delta c_P\left[T\left(1 - \ln\frac{T}{T_0}\right) - T_0\right]; \quad (3)$$

To use Equation (3), we set T_0 = 298 K, substitute the calculated value of $\Delta\mu_h(T_0)$, and estimate Δh and Δc_P using a Levenberg-Marquardt nonlinear least squares algorithm. The values we obtain for Δh and Δc_P are shown in **Table 3.** As a test of the quality of the fit, we compare the values of $\Delta\mu_h(T)$ from free energy calculations to those obtained using Equation (3). For the latter, we use the parameters listed in **Table 3**. The comparisons are shown in **Figure 2**.

Table 3: Parameters for Δh and Δc_P extracted from non-linear least squares analysis of computed temperature dependent free energies and fits to Equation 3.

Sidechain mimicked by the model compound	Estimated Δh at 298 K (kcal / mol)	Estimated Δc_P (cal / mol-K)
Arg	-57.24	69.39
Lys	-70.37	29.98
Asp	-98.65	-44.97
Glu	-96.62	-8.97

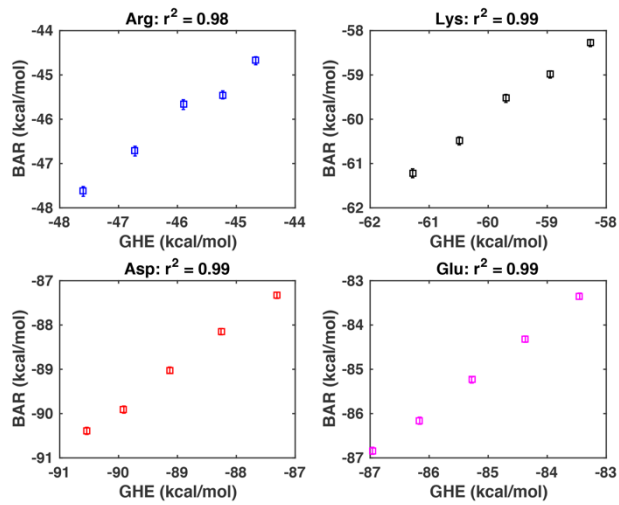

Figure 2: Assessment of the correlation between temperature-dependent free energies of hydration calculated using the integral of the Gibbs-Helmholtz equation (GHE) and the direct calculations from AMOEBA-based simulations that use the BAR estimator.

3.4 Analysis of temperature dependent enthalpy and entropy of hydration: Using the Gibbs-Helmholtz equation and parameters shown in **Table 3**, we estimated the temperature dependence of the enthalpy and entropy of hydration using Equation (4) below.

$$\Delta h(T) = \Delta h(T_0) + \Delta c_P (T - T_0)$$
$$\Delta s(T) = -\frac{\left[\Delta \mu_h(T_0) - \Delta h(T_0)\right]}{T_0} + \Delta c_P \ln\left(\frac{T}{T_0}\right) \quad (4)$$

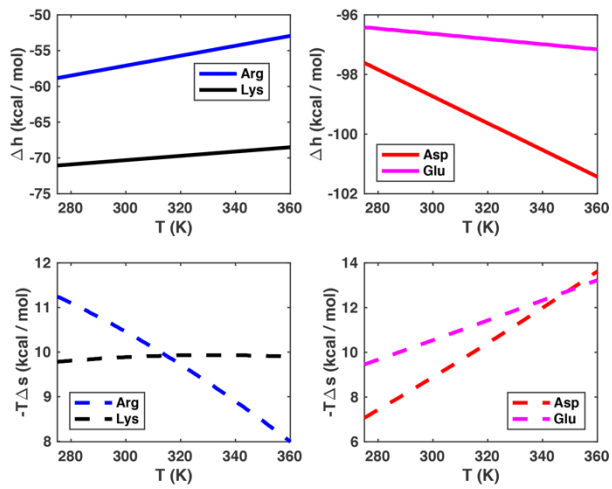

Figure 3: Temperature dependent enthalpies and entropies of hydration for the four model compound mimics of Arg, Lys, Asp, and Glu. These decompositions were calculated using the integral of the Gibbs-Helmholtz equation and parameters from **Table 3**.

The results are shown in **Figure 3**. There is a clear difference in the temperature dependencies for basic versus acidic molecules. With increasing temperature, the enthalpy of hydration becomes less favorable for Arg and Lys while it becomes more favorable for Asp and Glu. The unfavorable entropic contribution to the free energy of hydration stays roughly constant for Lys and decreases with increasing temperature for Arg. In contrast, the unfavorable entropic contribution to the free energy of hydration increases with increasing temperature for Asp and Glu. Each of the results shown in **Figure 3** is a direct consequence of the negative heat capacity of hydration for Asp and Glu, which contrasts with the positive heat capacity of hydration for Arg and Lys and all other model compounds that mimic backbone and sidechain moieties.

To understand the origins of the observations summarized in **Figure 3**, we performed three sets of reference simulations, one for the Cl^- ion and two for alchemic variants of the Cl^- ion. The anionic Cl^- ion is weakly polarizable in the AMOEBA model and carries a net charge of -*e*. We computed free energies of hydration for the following temperatures: 275 K, 298 K, 323 K, and 348 K. The results for $\Delta\mu_h$ as a function of temperature are shown in **Figure 4**. Here, we also show results for two alchemic versions of the Cl^- ion namely, an uncharged version Cl^0 and a cationic version Cl^+, where we flip the sign of the charge. From the temperature dependent values of $\Delta\mu_h$ for Cl^-, Cl^0, and Cl^+ we extract estimates for the enthalpy of hydration Δh at 298 K and the heat capacity of hydration Δc_P. These values are summarized in **Table 4.** Comparison of the parameters in **Tables 3** and **4** reveal the following: Anions that mimic Asp and Glu and the Cl^- ion have negative Δc_P values. The magnitude of Δc_P decreases and approaches zero as the length of the alkyl chain increases – see comparisons of Δc_P values for mimics of Asp versus Glu. The Δc_P values are positive for model compound mimics of Arg, Lys, and the alchemic Cl^+ ion. The magnitude of the positive Δc_P increases with hydrophobicity and surprisingly, n-propylguanidine has a higher Δc_P when compared to the neutral, alchemic Cl^0 solute. These numerical findings prompted detailed comparisons of hydration structures, and these are presented next.

Table 4: Parameters for Δh and Δc_P extracted from non-linear least squares analysis of computed temperature dependent free energies and fits to Equation 3.

Solute	$\Delta\mu_h$ at 298 K	Estimated Δh at 298 K (kcal / mol)	Estimated Δc_P (cal / mol-K)
Cl^-	-86.17 ± 0.06	-87.72	-46.55
Cl^0	2.04 ± 0.04	-2.70	63.31
Cl^+	-65.53 ± 0.05	-68.65	11.35

3.5 Comparative analysis of hydration structures around the different solutes: Molecular theories for hydrophobic and hydrophilic hydration rest on comparative analyses of hydration structures around solutes and the effects of solutes on density inhomogeneities within water. The chemical structures of the mimics of Arg, Lys, Asp, and Glu are subtly or significantly different from another. We computed the spatial density profiles of water molecules around each of the solutes. The results from this analysis are summarized pictorially in **Figure 5**. Here, each panel shows regions around each solute where the density of oxygen and hydrogen atoms from water molecules rises above a prescribed cutoff value – see caption. These calculations emphasize the accumulation of water molecules around the functional groups within each solute.

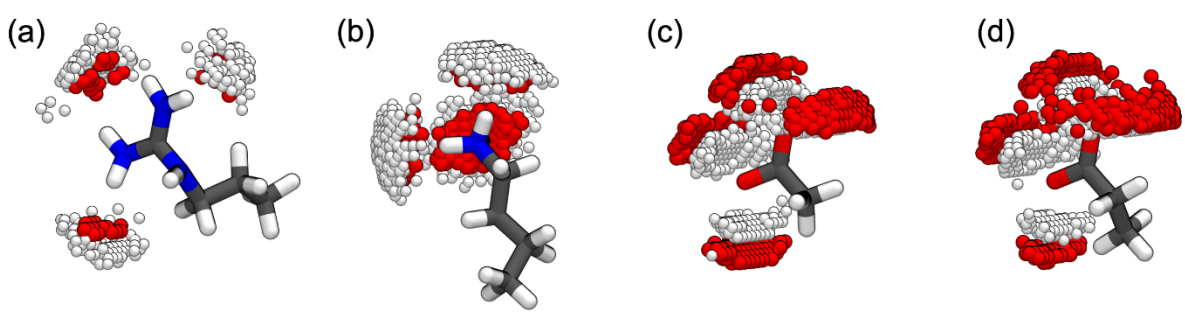

Figure 5: **Hydration structures around the model compounds mimicking the sidechains of (a) Arg, (b) Lys, (c) Asp and (d) Glu.** The red and white spheres around the model compounds denote areas with a time-averaged density of water oxygen number and water hydrogen number larger than 0.2 Å^{-3}. Positions further than 2Å away from the model compound are not shown. To calculate the density, we define two vectors $\vec{r}_y$ and $\vec{r}_{xy}$ for the model compounds to align all the frames in the trajectory. All the coordinates for atoms in the frame are translated and rotated so that the central atom in the model compound is in the origin of the simulation box and $\vec{r}_y$ points to the y direction and $\vec{r}_{xy}$ is in the x-y plane. For (a), the central atom is the carbon atom in the guanidine group, $\vec{r}_y$ is the vector pointing from the central atom to the nitrogen atom bonded with two carbon atoms, $\vec{r}_{xy}$ is the vector pointing from the central atom to one of the nitrogen atoms bonded with two hydrogens. For (b), the central atom is the nitrogen atom, $\vec{r}_y$ is the vector pointing from the central atom to the carbon atom bonded with nitrogen, $\vec{r}_{xy}$ is the vector pointing from the central atom to one of the three hydrogens bonded with it. For (c) and (d), the

central atom is the carbon atom bonded with oxygen atoms, $\vec{r}_y$ is the vector pointing from the central atom to the carbon atoms bonded with the central atom, $\vec{r}_{xy}$ is the vector pointing from the central atom to one of the oxygen atoms. Each panel was made using VMD [60].

The distinction between hydrophobic and hydrophilic hydration is typically attributed to differences in hydration structure within the first hydration shell [61], to the spatial organization of the first shell with respect to the bulk [61-62], to density inhomogeneities in the vicinity of the solute [63], and their long-range effects [64]. We quantify hydration structures in terms of two-parameter probability distribution functions. Here, we follow the approach of Gallagher and Sharp [62] and compute the joint radial and angular distribution functions $\rho(r,\theta)$ where the definition of r and θ are as shown in **Figure 6**. For neat water, the distribution functions are computed using all pairs of water molecules that are within 7 Å of one another. For solute-solvent systems, the distribution functions are computed for water molecules that are within 7 Å of the first hydration shell around each solute. For each system, $\rho(r,\theta)\delta r\delta\theta$ quantifies the probability that a pair of water molecules will be in a distance interval r and $r+\delta r$ *and* have relative orientations that are between θ and $\theta+\delta\theta$.

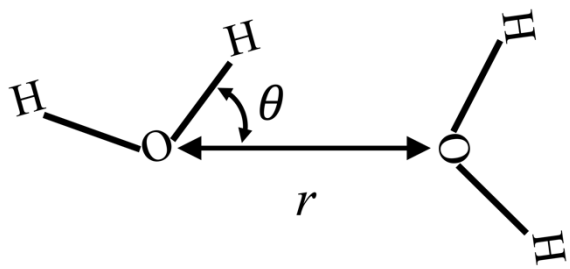

Figure 6: **Definition of *r* and θ for characterizing water structures as defined by Gallagher and Sharp [62].** Here, r is the distance between two water oxygen atoms, θ is the smallest angle in the four O-O_x-H_x angles, where H_x is bonded with O_x.

Optimal hydrogen bonding is realized for short distances and values of θ that are close to zero. These results are shown in **Figure 7**. The basin corresponding to $r \approx 2.8$ Å and values of $\theta < 10^\circ$ are evident in each of the four panels of **Figure 7** and these peaks represent the optimal hydrogen bonded geometries for water molecules. This peak becomes pronounced for Cl^0 and Cl^+, which are the alchemic neutral and cationic solutes, respectively. The density in the interval 3 Å $< r <$ 7 Å *and* $0^\circ < \theta < 60^\circ$ is significantly higher for water in the presence of the anion when compared to neat water or in the presence of the uncharged, non-polar Cl^0 solute. In this region, there are two distinct peaks, and this increased density is consistent with the calculated negative Δc_P values in that it points to distinct structural preferences of water molecules in the presence of anionic solutes. In contrast, for water around the cationic Cl^+ solute, the density is dispersed across the interval 3 Å $< r <$ 7 Å and $40^\circ < \theta < 120^\circ$ i.e., reflected about an axis that intersects the ordinate at $\approx 40^\circ$. As with the uncharged Cl^0 solute, there is a sharp density for water in the regime corresponding to optimal hydrogen bonding and this is weakened as r increases for the Cl^+ solute.

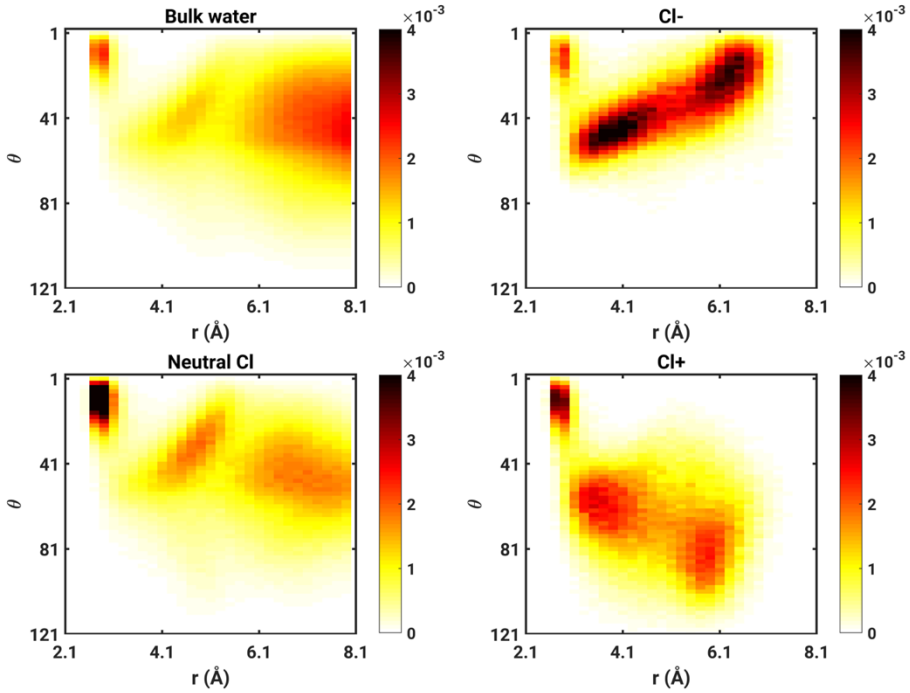

Figure 7: Joint distributions *r* and θ (see Figure 6) for bulk water and the waters in the first solvation shell of Cl^-, Cl^0 (neutral Cl), and Cl^+ at 298K. For Cl-, Cl+ and the neutral Cl, the water is considered in the first solvation shell if the distance from the solute to the water oxygen atom is smaller than the radius of the 1st solvation shell, which are 4.0 Å, 3.9 Å and 5.4 Å for Cl-, Cl+ and neutral Cl, respectively. For each system, the histogram is calculated from a 6 ns long trajectory with a saving interval of 1 ps. The bin size is 0.2 Å for r and 2° for θ.

Figure 8 shows difference density distributions $\Delta\rho(r,\theta)$, where for each solute X, the difference distribution is calculated as $\rho_X(r,\theta) - \rho_w(r,\theta)$, where $\rho_X(r,\theta)$ and $\rho_w(r,\theta)$ are the joint distributions for the solute-solvent system with solute X and neat water, respectively. In these difference distribution plots, regions where there is an enhancement of density vis-à-vis neat water are in hot colors, whereas the regions where there is a depletion of density compared to neat water are shown in cool colors. The difference distributions highlight the fundamental differences between hydrophobic hydration seen for Cl^0 and the anion versus cation systems.

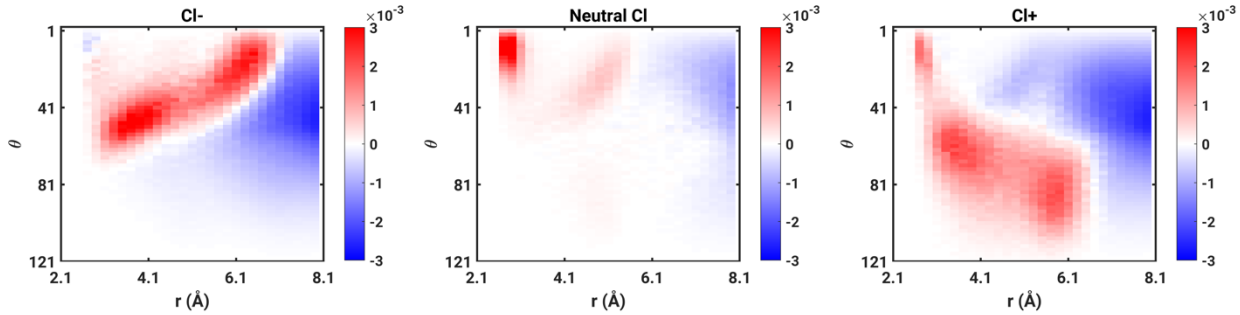

Figure 8: The difference density distributions for the three reference systems.

Next, we analyzed the difference density distributions for model compounds that mimic Arg, Lys, Asp, and Glu, respectively. The results, shown in **Figure 9**, reveal several striking features. The difference density distributions for Arg and Lys are clearly very different from one another. When compared to neat water, there is a significant increase of density in the basin corresponding to $r \approx 2.8$ Å and $\theta < 10^{\circ}$. This increase is similar to that of the model hydrophobic solute Cl^0. For Lys, the density in the basin corresponding to $r \approx 2.8$ Å and values of $\theta < 10^{\circ}$is considerably lower than that of neat water or the model hydrophobic solute Cl^0. Instead, there is a pronounced increase in density in the region corresponding to 3 Å $< r <$ 7 Å and $40^{\circ} < \theta < 120^{\circ}$, which is concordant with the observations for Cl^+, although the distribution is considerably more uniform for Lys. The difference density distributions for Asp and Glu are qualitatively similar to that observed for Cl^-, showing increased preference for the interval 3 Å $< r <$ 7 Å *and* $0^{\circ} < \theta < 60^{\circ}$ and a clear weakening, vis-à-vis neat water, for the basin corresponding to $r \approx 2.8$ Å and values of $\theta < 10^{\circ}$. Taken together, these features indicate that Arg behaves more like a hydrophobic solute when compared and Lys, and the impact of the anionic moieties on water structure is mutually consistent, being qualitatively similar to that of Cl^- while also providing a rationalization for the negative heat capacities reported for these solutes.

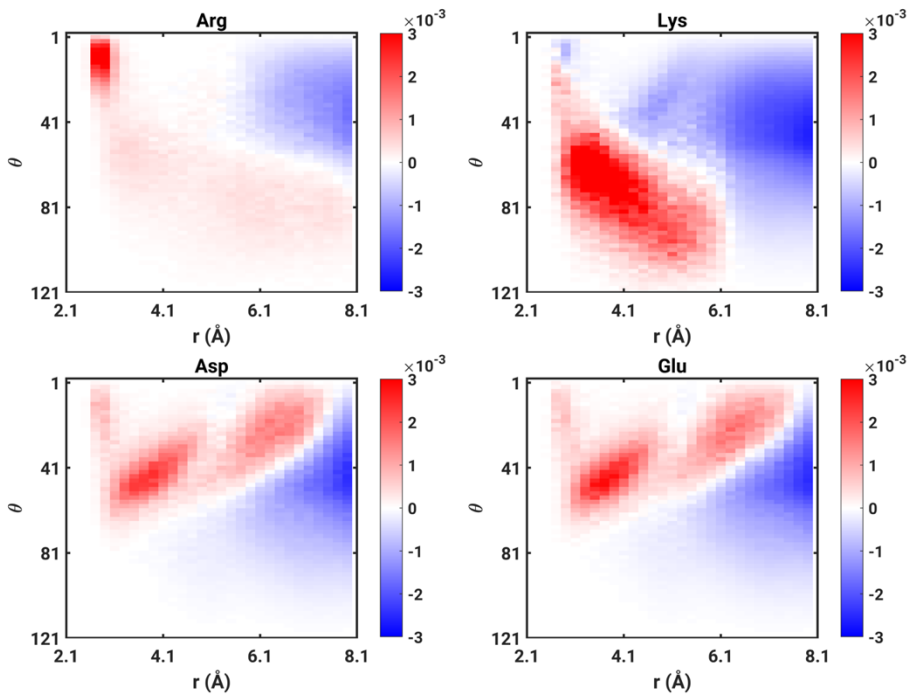

Figure 9: Differences density distributions for the model compound mimicking the sidechains of the charged amino acids referenced to that neat water. The comparisons are shown for 298K.

Finally, we computed the numbers of water molecules that make up the first hydration shells around each of the four solutes. On average, there are 15 water molecules in the first shell around Arg, ≈ 9 water molecules in the first shell around Asp and Glu, and ≈ 5 water molecules in the first shell around Lys. The larger numbers of water molecules around Arg are concordant with signatures of hydrophobic hydration when compared to Lys, Asp, or Glu. Based on the estimates for the average numbers of water molecules in the first hydration shells, it follows that the free

energy of hydration per water molecule is ≈ -3 kcal / mol for Arg, ≈ -12 kcal / mol for Lys, ≈ -10 kcal / mol for Asp, and ≈ -9.6 kcal / mol for Glu. These estimates suggest that free energy cost for displacing individual water molecules from the first hydration shell will be smallest for Arg, largest for Lys, and ≈ 10 kcal / mol per molecule for Asp / Glu.

4. DISCUSSION

4.1 Summary of main findings: We introduced our adaptation of the TCPD approach to estimate free energies of hydration from direct measurements of accessible quantities. The measured values are taken from the literature and adapted for use with a consensus estimate for the free energy of hydration of the proton. The results shown in **Table 1** provide a new, experimentally anchored touchstone for free energies of hydration of model compounds that mimic charged sidechains of amino acid residues. We used free energy calculations based on the AMOEBA forcefield to compute free energies of hydration at different temperatures for the model compounds that mimic sidechains of Arg, Lys, Asp, and Glu. Values obtained from free energy calculations are very similar to those obtained from TCPD-based estimations. This consistency provides confidence regarding the robustness of the new numbers.

Using free energies of hydration that were calculated at different temperatures, we obtained estimates for the enthalpy of hydration at a reference temperature of 298 K and the heat capacity of hydration. The latter was estimated by assuming that the heat capacity of hydration is independent of temperature. This is a reasonable assumption, and its validity is assessed by the quality of the agreement between the direct calculations of $\Delta\mu_h$ as a function of temperature and the values we obtain using the integral of the Gibbs-Helmholtz equation.

Importantly, we find three surprising results: (1) Arg and Lys are non-equivalent in terms of their hydration preferences. Contrary to expectations based on partition coefficients between water and octanol [65], the hydration free energy we obtain for Arg is consistently less favorable than that of Lys. Additionally, the heat capacity of hydration, which is positive for both species, is approximately 2.3 times larger for Arg when compared to Lys. This large heat capacity of hydration is indicative of a more hydrophobic character for Arg. (2) The free energies of hydration for acidic residues are considerably more favorable than for basic residues. This preference is well established for small, spherical ions [57]. Here, we find that it is retained for the compounds that mimic charged sidechains of amino acids. (3) The heat capacity of hydration is negative for Asp and Glu. Negative heat capacities have been attributed to differences in hydration structure and the propagation of these effects beyond the first hydration shell. This is evident in the difference density distributions that we compute for the negatively charged solutes.

Measurements have reported negative heat capacities of hydration for whole salts, and this been taken to imply that hydrophilic hydration, i.e., the hydration of anions as well as cations is associated with negative values for Δc_P [66]. This has been questioned by Sedlmeier and Netz [64]. They showed that while the sum of heat capacities for anions and cations in whole salts can be negative, the negative values arise strictly from the anions. Inasmuch as their work uses the SPC/E water model and non-polarizable forcefields, it appears that the negative heat capacity of hydration for anions is a generic and robust attribute that does not depend on polarizability. Indeed, we find that the negative heat capacity is evident for Cl^- ions whereas the heat capacity is positive for alchemically transformed Cl^+ and Cl^0 solutes.

4.2 Relevance of our findings to recent studies on IDPs: Ongoing investigations of the determinants of driving forces of phase separation in IDPs and in RNA binding domains have

revealed striking differences in the contributions of charged sidechains to the driving forces for phase separation [67]. The most salient observation is that Arg and Lys are fundamentally different [68] as drivers of phase separation [67, 69]. Replacing Arg residues with Lys significantly weakens the driving forces for phase separation in a variety of systems [70]. In a similar vein, Sørensen and Kjaergaard recently showed that Arg-rich polyampholytic IDPs prefer considerably more compact conformations when compared to Lys-rich counterparts [71]. Increasing the fraction of acidic residues within IDPs tends to have a solubilizing effect on IDPs [72].

Various conjectures have been offered to explain the differences between Arg and Lys as drivers of phase separation and their differences on conformational equilibria of IDPs. The "Y-aromaticity" of Arg, its high quadrupole moment, favorable interactions with pi systems, and the apparent ability to engage in water-mediated attractive interactions have been implicated as features that distinguish Arg from Lys residues as more potent drivers of phase separation [2, 70]. The destabilizing effects of increased negative charge on phase separation driven exclusively by homotypic interactions among IDPs can be attributed to the favorable free energies of solvation of the carboxylate moieties in acidic residues. However, a definitive rationale for explaining the differences between Arg and Lys and the manifest differences between the contributions of Asp / Glu versus Arg / Lys has been lacking. Our results suggest a simple explanation, pointing to the differences in free energies of hydration and hydration structure as being the prime candidates for the differences in charged sidechains as determinants of conformational and phase equilibria of IDPs. This suggests that changes to sequences that alter the solvation thermodynamics of charged residues can have manifest impacts on specificity and energetics of molecular recognition, order-disorder transitions, and phase behavior.

AUTHOR INFORMATION

[†]Co-first authors, *corresponding author, E-mail: pappu@wustl.edu

The authors declare that there are no competing financial interests

ACKNOWLEDGMENTS

We are grateful for the privilege of contributing this work as part of the festschrift celebrating Professor Dev Thirumalai's numerous and influential contributions to physical chemistry, soft matter physics, and biophysics. Professor Thirumalai has been an inspiration and a generous mentor over the years. We are grateful to Chengwen Liu and Pengyu Ren for their collaboration on the original free energy calculations [12], developing parameters for the AMOEBA forcefield, and technical assistance. Our work is supported by grants from the US National Science Foundation (DMR1729783), the US National Institutes of Health (5R01NS056114), and the St. Jude Research Collaborative on Membraneless Organelles.

References

1. Das, R. K.; Ruff, K. M.; Pappu, R. V., Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr Opin Struct Biol* **2015,** *32*, 102-12.

2. Wang, J.; Choi, J.-M.; Holehouse, A. S.; Zhang, X.; Jahnel, M.; Lemaitre, R.; Maharana, S.; Pozniakovsky, A.; Drechsel, D.; Poser, I.; Pappu, R. V.; Alberti, S.; Hyman, A. A., A molecular grammar underlying the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **2018,** *174*, 688-699.

3. Lin, Y.-H.; Forman-Kay, J. D.; Chan, H. S., Theories for Sequence-Dependent Phase Behaviors of Biomolecular Condensates. *Biochemistry* **2018,** *57* (17), 2499-2508.

4. Shea, J.-E.; Best, R. B.; Mittal, J., Physics-based computational and theoretical approaches to intrinsically disordered proteins. *Current Opinion in Structural Biology* **2021,** *67*, 219-225.

5. Borgia, A.; Borgia, M. B.; Bugge, K.; Kissling, V. M.; Heidarsson, P. O.; Fernandes, C. B.; Sottini, A.; Soranno, A.; Buholzer, K. J.; Nettels, D.; Kragelund, B. B.; Best, R. B.; Schuler, B., Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **2018,** *555* (7694), 61-66.

6. Martin, E. W.; Holehouse, A. S.; Grace, C. R.; Hughes, A.; Pappu, R. V.; Mittag, T., Sequence Determinants of the Conformational Properties of an Intrinsically Disordered Protein Prior to and upon Multisite Phosphorylation. *Journal of the American Chemical Society* **2016,** *138* (47), 15323-15335.

7. Das, R. K.; Huang, Y.; Phillips, A. H.; Kriwacki, R. W.; Pappu, R. V., Cryptic sequence features within the disordered protein p27 Kip1 regulate cell cycle signaling. *Proceedings of the National Academy of Sciences USA* **2016**.

8. (a) Sawle, L.; Ghosh, K., A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins. *The Journal of Chemical Physics* **2015,** *143* (8), 085101; (b) Sawle, L.; Huihui, J.; Ghosh, K., All-Atom Simulations Reveal Protein Charge Decoration in the Folded and Unfolded Ensemble Is Key in Thermophilic Adaptation. *Journal of Chemical Theory and Computation* **2017,** *13* (10), 5065-5075.

9. Sherry, K. P.; Das, R. K.; Pappu, R. V.; Barrick, D., Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor. *Proceedings of the National Academy of Sciences* **2017,** *114* (44), E9243-E9252.

10. (a) Monahan, Z.; Ryan, V. H.; Janke, A. M.; Burke, K. A.; Rhoads, S. N.; Zerze, G. H.; O'Meally, R.; Dignon, G. L.; Conicella, A. E.; Zheng, W.; Best, R. B.; Cole, R. N.; Mittal, J.; Shewmaker, F.; Fawzi, N. L., Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. *EMBO J* **2017,** *36* (20), 2951-2967; (b) Dignon, G. L.; Zheng, W.; Kim, Y. C.; Mittal, J., Temperature-Controlled Liquid–Liquid Phase Separation of Disordered Proteins. *ACS Central Science* **2019,** *5* (5), 821-830.

11. Auton, M.; Bolen, D. W., Additive Transfer Free Energies of the Peptide Backbone Unit That Are Independent of the Model Compound and the Choice of Concentration Scale. *Biochemistry* **2004,** *43* (5), 1329-1342.

12. Zeng, X.; Liu, C.; Fossat, M. J.; Ren, P.; Chilkoti, A.; Pappu, R. V., Design of intrinsically disordered proteins that undergo phase transitions with lower critical solution temperatures. *APL Materials* **2021,** *In press*, bioRxiv: 2020.11.13.381897.

13. Wuttke, R.; Hofmann, H.; Nettels, D.; Borgia, M. B.; Mittal, J.; Best, R. B.; Schuler, B., Temperature-dependent solvation modulates the dimensions of disordered proteins. *Proc Natl Acad Sci U S A* **2014,** *111* (14), 5213-8.

14. Thirumalai, D.; Samanta, H. S.; Maity, H.; Reddy, G., Universal Nature of Collapsibility in the Context of Protein Folding and Evolution. *Trends in Biochemical Sciences* **2019,** *44* (8), 675-687.

15. O'Brien, E. P.; Ziv, G.; Haran, G.; Brooks, B. R.; Thirumalai, D., Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model. *Proceedings of the National Academy of Sciences* **2008,** *105* (36), 13403.

16. Ben-Naim, A.; Ting, K.-L.; Jernigan, R. L., Solvation thermodynamics of biopolymers. I. Separation of the volume and surface interactions with estimates for proteins. *Biopolymers* **1989,** *28* (7), 1309-1325.

17. Wolfenden, R., Interaction of the peptide bond with solvent water: a vapor phase analysis. *Biochemistry* **1978,** *17* (1), 201-204.

18. (a) Makhatadze, G. I.; Lopez, M. M.; Privalov, P. L., Heat capacities of protein functional groups. *Biophysical Chemistry* **1997,** *64* (1), 93-101; (b) Makhatadze, G. I.; Privalov, P. L., Contribution of Hydration to Protein Folding Thermodynamics: I. The Enthalpy of Hydration. *Journal of Molecular Biology* **1993,** *232* (2), 639-659.

19. (a) Marcus, Y., The hydration entropies of ions and their effects on the structure of water. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases* **1986,** *82* (1), 233-242; (b) Marcus, Y., Electrostriction, Ion Solvation, and Solvent Release on Ion Pairing. *The Journal of Physical Chemistry B* **2005,** *109* (39), 18541-18549.

20. Grossfield, A.; Ren, P.; Ponder, J. W., Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field. *Journal of the American Chemical Society* **2003,** *125* (50), 15671-15682.

21. Marcus, Y.; Hefter, G., Ion Pairing. *Chemical Reviews* **2006,** *106* (11), 4585-4621.

22. Vitalis, A.; Pappu, R. V., ABSINTH: a new continuum solvation model for simulations of polypeptides in aqueous solutions. *J Comput Chem* **2009,** *30* (5), 673-99.

23. Marcus, Y., The thermodynamics of solvation of ions. Part 2.—The enthalpy of hydration at 298.15 K. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases* **1987,** *83* (2), 339-349.

24. Abraham, M. H.; Marcus, Y., The thermodynamics of solvation of ions. Part 1.—The heat capacity of hydration at 298.15 K. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases* **1986,** *82* (10), 3255-3274.

25. Pearson, R. G., Ionization potentials and electron affinities in aqueous solution. *Journal of the American Chemical Society* **1986,** *108* (20), 6109-6114.

26. Pliego, J. R.; Riveros, J. M., New values for the absolute solvation free energy of univalent ions in aqueous solution. *Chemical Physics Letters* **2000,** *332* (5), 597-602.

27. Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Tuttle, T. R.; Coe, J. V., The Proton's Absolute Aqueous Enthalpy and Gibbs Free Energy of Solvation from Cluster Ion Solvation Data. *The Journal of Physical Chemistry A* **1998,** *102* (46), 9308-9308.

28. Hunter, E. P. L.; Lias, S. G., Evaluated Gas Phase Basicities and Proton Affinities of Molecules: An Update. *Journal of Physical and Chemical Reference Data* **1998,** *27* (3), 413-656.

29. Bodnarchuk, M. S.; Heyes, D. M.; Dini, D.; Chahine, S.; Edwards, S., Role of Deprotonation Free Energies in pKa Prediction and Molecule Ranking. *Journal of Chemical Theory and Computation* **2014,** *10* (6), 2537-2545.

30. Jorgensen, W. L.; Briggs, J. M., A priori pKa calculations and the hydration of organic anions. *Journal of the American Chemical Society* **1989,** *111* (12), 4190-4197.

31. Stull, D. R., Vapor Pressure of Pure Substances. Organic and Inorganic Compounds. *Industrial & Engineering Chemistry* **1947,** *39* (4), 517-540.

32. Stover, M. L.; Jackson, V. E.; Matus, M. H.; Adams, M. A.; Cassady, C. J.; Dixon, D. A., Fundamental Thermochemical Properties of Amino Acids: Gas-Phase and Aqueous Acidities and Gas-Phase Heats of Formation. *The Journal of Physical Chemistry B* **2012,** *116* (9), 2905-2916.

33. Lee, S.; Cho, K.-H.; Lee, C. J.; Kim, G. E.; Na, C. H.; In, Y.; No, K. T., Calculation of the Solvation Free Energy of Neutral and Ionic Molecules in Diverse Solvents. *Journal of Chemical Information and Modeling* **2011,** *51* (1), 105-114.

34. Zhang, H.; Jiang, Y.; Yan, H.; Yin, C.; Tan, T.; van der Spoel, D., Free-Energy Calculations of Ionic Hydration Consistent with the Experimental Hydration Free Energy of the Proton. *The Journal of Physical Chemistry Letters* **2017,** *8* (12), 2705-2712.

35. Harger, M.; Li, D.; Wang, Z.; Dalby, K.; Lagardère, L.; Piquemal, J.-P.; Ponder, J.; Ren, P., Tinker-OpenMM: Absolute and relative alchemical free energies using AMOEBA on GPUs. *Journal of Computational Chemistry* **2017,** *38* (23), 2047-2055.

36. Tuckerman, M. E.; Berne, B. J.; Martyna, G. J., Molecular dynamics algorithm for multiple time scales: Systems with long range forces. *The Journal of Chemical Physics* **1991,** *94* (10), 6811-6815.

37. Bussi, G.; Zykova-Timan, T.; Parrinello, M., Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *The Journal of Chemical Physics* **2009,** *130* (7), 074101.

38. Åqvist, J.; Wennerström, P.; Nervall, M.; Bjelic, S.; Brandsdal, B. O., Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chemical Physics Letters* **2004,** *384* (4), 288-294.

39. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993,** *98* (12), 10089-10092.

40. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995,** *103* (19), 8577-8593.

41. Jing, Z.; Liu, C.; Qi, R.; Ren, P., Many-body effect determines the selectivity for Ca2+ and Mg2+ in proteins. *Proceedings of the National Academy of Sciences* **2018,** *115* (32), E7495.

42. Jiao, D.; Golubkov, P. A.; Darden, T. A.; Ren, P., Calculation of protein–ligand binding free energy by using a polarizable potential. *Proceedings of the National Academy of Sciences* **2008,** *105* (17), 6290.

43. Bennett, C. H., Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics* **1976,** *22* (2), 245-268.

44. Moser, A.; Range, K.; York, D. M., Accurate Proton Affinity and Gas-Phase Basicity Values for Molecules Important in Biocatalysis. *The Journal of Physical Chemistry B* **2010,** *114* (43), 13911-13921.

45. Fitch, C. A.; Platzer, G.; Okon, M.; Garcia-Moreno E., B.; McIntosh, L. P., Arginine: Its pKa value revisited. *Protein Science* **2015,** *24* (5), 752-761.

46. Xu, B.; Jacobs, M. I.; Kostko, O.; Ahmed, M., Guanidinium Group Remains Protonated in a Strongly Basic Arginine Solution. *ChemPhysChem* **2017,** *18* (12), 1503-1506.

47. Rosenberg, S. A.; Hueber, A. E.; Aronson, D.; Gouchie, S.; Howard, P. H.; Meylan, W. M.; Tunkel, J. L., Syracuse Research Corporation's Chemical Information Databases. *Science & Technology Libraries* **2003,** *23* (4), 73-87.

48. Linstrom, P. J.; Mallard, W. G., The NIST Chemistry WebBook: A Chemical Data Resource on the Internet. *Journal of Chemical & Engineering Data* **2001,** *46* (5), 1059-1063.

49. Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L., Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous solution. *Journal of Solution Chemistry* **1981,** *10* (8), 563-595.

50. Tehan, B. G.; Lloyd, E. J.; Wong, M. G.; Pitt, W. R.; Gancia, E.; Manallack, D. T., Estimation of pKa Using Semiempirical Molecular Orbital Methods. Part 2: Application to Amines, Anilines and Various Nitrogen Containing Heterocyclic Compounds. *Quantitative Structure-Activity Relationships* **2002,** *21* (5), 473-485.

51. Tehan, B. G.; Lloyd, E. J.; Wong, M. G.; Pitt, W. R.; Montana, J. G.; Manallack, D. T.; Gancia, E., Estimation of pKa Using Semiempirical Molecular Orbital Methods. Part 1: Application to Phenols and Carboxylic Acids. *Quantitative Structure-Activity Relationships* **2002,** *21* (5), 457-472.

52. Schmid, R.; Miah, A. M.; Sapunov, V. N., A new table of the thermodynamic quantities of ionic hydration: values and some applications (enthalpy–entropy compensation and Born radii). *Physical Chemistry Chemical Physics* **2000,** *2* (1), 97-102.

53. Asthagiri, D.; Pratt, L. R.; Ashbaugh, H. S., Absolute hydration free energies of ions, ion–water clusters, and quasichemical theory. *The Journal of Chemical Physics* **2003,** *119* (5), 2702-2708.

54. Latimer, W. M.; Pitzer, K. S.; Slansky, C. M., The Free Energy of Hydration of Gaseous Ions, and the Absolute Potential of the Normal Calomel Electrode. *The Journal of Chemical Physics* **1939,** *7* (2), 108-111.

55. Zhan, C.-G.; Dixon, D. A., First-Principles Determination of the Absolute Hydration Free Energy of the Hydroxide Ion. *The Journal of Physical Chemistry A* **2002,** *106* (42), 9737-9744.

56. (a) Ren, P.; Ponder, J. W., Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *The Journal of Physical Chemistry B* **2003,** *107* (24), 5933-5947; (b) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T., Current Status of the AMOEBA Polarizable Force Field. *The Journal of Physical Chemistry B* **2010,** *114* (8), 2549-2564.

57. Grossfield, A., Dependence of ion hydration on the sign of the ion's charge. *The Journal of Chemical Physics* **2005,** *122* (2), 024506.

58. Shi, Y.; Wu, C.; Ponder, J. W.; Ren, P., Multipole electrostatics in hydration free energy calculations. *Journal of Computational Chemistry* **2011,** *32* (5), 967-977.

59. Wyczalkowski, M. A.; Vitalis, A.; Pappu, R. V., New Estimators for Calculating Solvation Entropy and Enthalpy and Comparative Assessments of Their Accuracy and Precision. *The Journal of Physical Chemistry B* **2010,** *114* (24), 8166-8180.

60. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **1996,** *14* (1), 33-38.

61. Prabhu, N. V.; Sharp, K. A., Heat capacity in proteins. *Annual Review of Physical Chemistry* **2005,** *56* (1), 521-548.

62. Gallagher, K. R.; Sharp, K. A., A New Angle on Heat Capacity Changes in Hydrophobic Solvation. *Journal of the American Chemical Society* **2003,** *125* (32), 9853-9860.

63. Godawat, R.; Jamadagni, S. N.; Garde, S., Characterizing hydrophobicity of interfaces by using cavity formation, solute binding, and water correlations. *Proceedings of the National Academy of Sciences* **2009,** *106* (36), 15119.

64. Sedlmeier, F.; Netz, R. R., Solvation thermodynamics and heat capacity of polar and charged solutes in water. *The Journal of Chemical Physics* **2013,** *138* (11), 115101.

65. Radzicka, A.; Wolfenden, R., Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry* **1988,** *27* (5), 1664-1670.

66. Kinoshita, M.; Yoshidome, T., Molecular origin of the negative heat capacity of hydrophilic hydration. *The Journal of Chemical Physics* **2009,** *130* (14), 144705.

67. Nott, T. J.; Petsalaki, E.; Farber, P.; Jervis, D.; Fussner, E.; Plochowietz, A.; Craggs, T. D.; Bazett-Jones, D. P.; Pawson, T.; Forman-Kay, J. D.; Baldwin, A. J., Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol Cell* **2015,** *57* (5), 936-47.

68. Fisher, R. S.; Elbaum-Garfinkle, S., Tunable multiphase dynamics of arginine and lysine liquid condensates. *Nature Communications* **2020,** *11* (1), 4628.

69. Brady, J. P.; Farber, P. J.; Sekhar, A.; Lin, Y. H.; Huang, R.; Bah, A.; Nott, T. J.; Chan, H. S.; Baldwin, A. J.; Forman-Kay, J. D.; Kay, L. E., Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proceedings of the National Academy of Sciences USA* **2017,** *114* (39), E8194-E8203.

70. Greig, J. A.; Nguyen, T. A.; Lee, M.; Holehouse, A. S.; Posey, A. E.; Pappu, R. V.; Jedd, G., Arginine-Enriched Mixed-Charge Domains Provide Cohesion for Nuclear Speckle Condensation. *Molecular Cell* **2020,** *77* (6), 1237-1250.e4.

71. Sørensen, C. S.; Kjaergaard, M., Effective concentrations enforced by intrinsically disordered linkers are governed by polymer physics. *Proceedings of the National Academy of Sciences* **2019,** *116* (46), 23124.

72. Guillén-Boixet, J.; Kopach, A.; Holehouse, A. S.; Wittmann, S.; Jahnel, M.; Schlüßler, R.; Kim, K.; Trussina, I. R. E. A.; Wang, J.; Mateju, D.; Poser, I.; Maharana, S.; Ruer-Gruß, M.; Richter, D.; Zhang, X.; Chang, Y.-T.; Guck, J.; Honigmann, A.; Mahamid, J.; Hyman, A. A.; Pappu, R. V.; Alberti, S.; Franzmann, T. M., RNA-Induced Conformational Switching and Clustering of G3BP Drive Stress Granule Assembly by Condensation. *Cell* **2020,** *181* (2), 346-361.e17.