ADD force field for sugars and polyols: predicting the additivity of protein-osmolyte interaction

Arsiccio et al. (2020). The ADD force field for sugars and polyols: predicting the addictivity of protein-osmolyte interaction. Journal of Physical Chemistry B 124(36):7779-7790.

Corresponding author: Roberto Pisano, e-mail: roberto.pisano@polito.it

The ADD Force Field for Sugars and Polyols: Predicting the Additivity of Protein-Osmolyte Interaction

Andrea Arsiccio,[†] Pritam Ganguly,[†] Lorenzo La Cortiglia,[‡] Joan-Emma Shea,[†,¶] and Roberto Pisano[*,‡]

[†]*Department of Chemistry and Biochemistry, University of California, Santa Barbara, California 93106, United States*

[‡]*Department of Applied Science and Technology, Politecnico di Torino, 24 corso Duca degli Abruzzi, Torino 10129, Italy*

[¶]*Department of Physics, University of California, Santa Barbara, California 93106, United States*

E-mail: roberto.pisano@polito.it

Abstract

The protein-osmolyte interaction has been shown experimentally to follow an additive construct, where the individual osmolyte-backbone and osmolyte-sidechains interactions contribute to the overall conformational stability of proteins. Here, we computationally reconstruct this additive relation using molecular dynamics simulations, focusing on sugars and polyols, including sucrose and sorbitol, as model osmolytes. A new set of parameters (ADD) is developed for this purpose, using the individual Kirkwood-Buff integrals for sugar-backbone and sugar-sidechain interactions as target experimental data. We show that the ADD parameters can reproduce the additivity of protein-sugar interactions, and correctly predict sucrose and sorbitol self-association, as well as their interaction with water. The accurate description of the separate osmolyte-backbone and osmolyte-sidechain contributions also automatically translates into a good prediction of preferential exclusion from the surface of ribonuclease A and α-chymotrypsinogen A. The description of sugar polarity is improved compared to previous force fields, resulting in closer agreement with the experimental data and better compatibility with charged groups, such as the guanidinium moiety. The ADD parameters are developed in combination with the CHARMM36m force field for proteins, but good compatibility is also observed with the AMBER 99SB-ILDN and the OPLS-AA force fields. Overall, exploiting the additivity of protein-osmolyte interactions is a promising approach for the development of new force fields.

Introduction

Carbohydrates, including sugars and polyols, are a common class of osmolytes, and play a crucial role in biological systems. They are commonly accumulated by organisms that undergo temperature stress, such as freezing, or anhydrobiosis. For this reason, they are the dominant solutes encountered in many organisms, such as terrestrial plants, insects, amphibians and some polar fishes.[1] Carbohydrates are also often used in protein formulations to preserve their biological activity against external stress. They are generally believed

to stabilize a protein by altering the structural and dynamic properties of water,[2,3] or by being excluded from the peptide surface.[4–6] This exclusion creates a thermodynamically unfavorable situation, that favors the more compact native conformation against the expanded unfolded ones.[7–9]

Significant effort has recently been invested in developing approaches to understand the molecular mechanisms that drive protein-sugar interactions. In particular, molecular dynamics simulations have emerged as a powerful tool for an in-depth investigation of molecular-level phenomena. Several force fields were proposed in the literature for carbohydrates,[10–14] and Cloutier et al.[15] recently developed new parameters for these molecules (which they named KBPs), by modifying the original CHARMM36 force field.[16–18] They used the Kirkwood-Buff (KB) integrals[19–22] as target experimental data, and showed that the new force field could accurately predict both self-association and exclusion of carbohydrates from the protein surface. This represents a crucial step forward in our understanding of protein-sugar interactions, especially considering that previous parameters often overestimated sugar self-interaction[23] and resulted in sugar molecules that were preferentially interacting with the protein.[18]

Here we take one step further in this direction, studying the individual contributions of the backbone and amino acid sidechains to the overall protein-sugar KB integrals. A similar additive approach developed using transfer free energies (which are related to KB integrals) was proposed by Auton et al.[24–26] to explain the thermodynamics of osmolyte-induced protein transitions, and was investigated using molecular dynamics simulations for the case of urea.[27] In this work, we sought to reproduce the experimentally observed additivity of protein-carbohydrate interaction using the KBP parameters, and found that they led to an incorrect description of carbohydrate polarity, overestimating the interaction with aromatic and hydrophobic moieties. We also observed poor compatibility of KBP sugars with charged groups.

A new set of parameters is hence developed to better reproduce experimental values of the

KB integrals for sugar-backbone and sugar-sidechain interaction, using sucrose and sorbitol as model osmolytes. By properly adjusting charges and Lennard-Jones coefficients for the hydroxyl group, we obtain a good description of the solution properties, that does not require separate parameters for water and non-water interaction, as instead occurs with the KBPs. The new force field (in the following referred to as ADD) improves the description of sugars polarity, resulting in a better prediction of their interaction with aromatic and hydrophobic groups. Results also improve for the charged groups, despite quantitative deviation from the experimental data. We further show that the KBP model predicts too unfavorable sugar-guanidinium interactions, leading to phase-separation in guanidinium chloride (GdmCl) solutions. The ADD force field overcomes this issue and stabilizes the solution phase of GdmCl. The ADD parameters also accurately predict carbohydrate self-association, as well as their interaction with water. We show that the additivity of protein-osmolyte interaction can be reproduced using ADD sugars, and this automatically translates into a prediction of preferential exclusion from ribonuclease A (RNase A) and α-chymotrypsinogen A (α-Cgn A) that is in good agreement with the experimental data.

The ADD force field is developed in conjunction with CHARMM36m for proteins,[28] but we observe good compatibility also with other popular force fields, such as AMBER 99SB-ILDN[29] and OPLS-AA.[30] Overall, our simulation results show that the additive behavior that was experimentally observed for osmolyte-protein interactions can be recovered in molecular dynamics simulations, allowing for detailed atomistic insights into the effects of sugars on proteins stability, and providing a promising strategy for the development and validation of force field parameters.

Materials and Methods

Theoretical Background

In this section, some basic concepts of the Kirkwood-Buff theory[19] and transfer model[24] will be introduced. We will refer to a three-component system containing water (component 1), a protein (component 2) and an osmolyte (component 3). We will also consider the case of dilute component 2.

The free energy of unfolding (native N to unfolded U conversion) $\Delta G^{N\to U}$ of the protein in such a system is generally assumed to be a linear function of the osmolyte concentration c_3,[31]

$$\Delta G^{N\to U}_{c_3} = \Delta G^{N\to U}_{0M} + mc_3 \tag{1}$$

It was shown that the slope m of this equation may be predicted using the concept of transfer free energy ΔG_{tr}.[25] In summary, a thermodynamic cycle was imagined, where the difference between the free energies of unfolding in presence and absence of 1 M osmolyte equals the difference in the transfer free energies of N and U from pure water to a 1 M osmolyte solution,[32]

$$m = \Delta G^{N\to U}_{1M} - \Delta G^{N\to U}_{0M} = \Delta G^{0\to 1M}_{tr,U} - \Delta G^{0\to 1M}_{tr,N} \tag{2}$$

More specifically, an additivity construct was found to be valid, where the difference in transfer free energies between the denatured and native state of a polypeptide containing n residues could be obtained by summing the contributions given by the amino acid side chains ($\Delta g^{sc}_{tr,j}$) and by the peptide backbone (Δg^{bb}_{tr}),[24]

$$\Delta G^{0\to 1M}_{tr,U} - \Delta G^{0\to 1M}_{tr,N} = \sum_{j=1}^{n} \Delta g^{sc}_{tr,j} \Delta \alpha^{sc}_{j} + \Delta g^{bb}_{tr} \sum_{j=1}^{n} \Delta \alpha^{bb}_{j} \tag{3}$$

Each contribution is weighted by the average fractional change in solvent accessible sur-

face area $SASA$ of residue j in going from the native N to the denatured U state,

$$\Delta\alpha_j = \frac{SASA_{j,U} - SASA_{j,N}}{SASA_{j,Gly-X-Gly}} \tag{4}$$

where $SASA_{j,Gly-X-Gly}$ is the solvent accessibility of amino acid X in the tripeptide Gly-X-Gly, and X is the amino acid residue type j.

The m value can also be related to the Kirkwood-Buff integrals (KBIs) G_{ij}, which are used to describe the solvation behavior around a reference particle i in the framework of the KB theory,

$$G_{ij} = 4\pi \int_0^\infty (g_{ij}(r) - 1) r^2 dr \tag{5}$$

$g_{ij}(r)$ is the radial distribution function, which describes the variations in component j density as a function of the distance r from component i. A value of $G_{ij} < 1$ indicates exclusion, while $G_{ij} > 1$ indicates accumulation of component j around the reference i.

The difference $\gamma = G_{23} - G_{12}$ is related to the prefential exclusion of the osmolyte from the protein. A negative value of γ indicates preferential exclusion, and vice versa.

Assuming linearity of $\Delta G_{tr}^{0\to c_3}$ with c_3, it is possible to write the following,[25]

$$\frac{\Delta G_{tr}^{0\to 1M}}{1M} = \frac{\Delta G_{tr}^{0\to c_3}}{c_3} = \left(\frac{\partial \mu_2}{\partial c_3}\right)_{T,p,c_2} \tag{6}$$

where μ_2 is the chemical potential of component 2. If this component is infinitely diluted, the following approximation applies,

$$\left(\frac{\partial \mu_2}{\partial c_3}\right)_{T,p,c_2} = -\gamma \left(\frac{\partial \mu_3}{\partial \ln c_3}\right)_{T,p,c_2} \tag{7}$$

The derivative of the osmolyte chemical potential can further be expressed as,[33]

$$\frac{1}{\mathrm{R}T}\left(\frac{\partial \mu_3}{\partial \ln c_3}\right)_{T,p,c_2} = \frac{1}{1 - c_3(G_{13} - G_{33})} \tag{8}$$

Combining Equations 6- 8 it is finally possible to write the following relation between transfer free energies and KBIs,

$$\Delta G_{tr}^{0\rightarrow 1M} = -\frac{\mathrm{R}T\gamma}{1 - c_3(G_{13} - G_{33})} \tag{9}$$

The additivity construct in Equation 3 can therefore also be rewritten in terms of Kirkwood-Buff integrals,

$$m = -\frac{\mathrm{R}T\Delta_N^U\gamma}{1 - c_3(G_{13} - G_{33})} = -\frac{\mathrm{R}T(\sum_{j=1}^{n}\gamma_j^{sc}\Delta\alpha_j^{sc} + \gamma^{bb}\sum_{j=1}^{n}\Delta\alpha_j^{bb})}{1 - c_3(G_{13} - G_{33})} \tag{10}$$

where γ_j^{sc} and γ^{bb} are the sidechain and backbone contributions, respectively.

Simulation details

Molecular dynamics simulations were carried out using Gromacs 5.0.7.[34] The objective was to define a suitable combination of parameters for carbohydrates, that respected the additive behavior discussed in the previous section, using sucrose and sorbitol as model molecules. The KBP variant of the CHARMM36 force field for sugars[15] was used as a starting point for parameterization, in combination with the CHARMM TIP3P water model.[35]

A scheme of all simulations performed, with the corresponding box size and duration, is listed in Table 1.

All the 20 naturally occurring amino acids, in their zwitterionic form, were simulated in 1 M sucrose or sorbitol (sim. type 1 and 2 in Table 1). For these simulations the box was cubic with $\approx$ 8 nm side length, and included 25 amino acids molecules. For charged residues, Na^+ or Cl^- ions were added to reach neutrality. Some capped amino acids were also considered. In this latter case, the N-terminus was acetylated and the C-terminus was blocked with an amide group.

The N-acetyl glycinamide series (NAG_xA) was also simulated (sim. type 3 in Table 1). NAG_xA corresponds to a series of molecules with a varying number x of glycine residues

Table 1: List of the simulations performed in this work.

Sim. type #	Solute (component 2)	Osmolyte (component 3)	box size nm	duration ns
1	zwitterionic/capped AA	1 M sucrose	8 x 8 x 8	60
2	zwitterionic/capped AA	1 M sorbitol	8 x 8 x 8	60
3	NAG_xA	1 M sucrose/sorbitol	8 x 8 x 8	60
4	-	1 M sucrose/sorbitol	8 x 8 x 8	60
5	RNase A pH 3	0.1M, 0.7M, 1M sucrose	7.7 x 7.7 x 7.7	60
6	RNase A pH 7	0.7M sucrose	7.7 x 7.7 x 7.7	60
7	RNase A pH 2	2.2M sorbitol	7.7 x 7.7 x 7.7	60
8	RNase A pH 5.5	0.55M, 2.2M sorbitol	7.7 x 7.7 x 7.7	60
9	α-Cgn A pH 3	0.7M sucrose	8.1 x 8.1 x 8.1	60
10	Trpzip1 pH 7	1 M sucrose	6.1 x 6.1 x 6.1	150

linked by a peptide bond and whose termini are blocked by an acetyl and an amide moiety. In particular, the number of internal glycine units has been varied from 3 (NAG_3A) to 6 (NAG_6A). Also in this case the box length was 8 nm, and the carbohydrate concentration adjusted to 1 M. The CHARMM36m[28] force field was used for the amino acids and the NAG_xA series in simulations 1-3.

For the determination of the sugar-sugar G_{33} and water-sugar G_{13} KB integrals a simulation box was used that contained only 1 M sucrose or sorbitol in water (sim. type 4 in Table 1). Also in this case the simulation box was about 8 x 8 x 8 nm.

The preferential exclusion of sucrose from ribonuclease A (RNase A) at pH 3 or 7, or from α-chymotrypsinogen A (α-Cgn A) at pH 3, and of sorbitol from RNase A at pH 2 or 5.5 was also assessed (sim. types 5-9 in Table 1). The RNase A and α-Cgn A configuration files were obtained from the RCSB Protein Data Bank (PDB: 1KF5[36] for RNase A and 2CGA[37] for α-Cgn A). The side length of the box in this case was approximately 7.7 nm for RNase A and 8.1 nm for α-Cgn A, and the pH was set by adjusting the protonation state of the amino acids using the H++ server, version 3.2 (http://biophysics.cs.vt.edu/H++[38]). Neutrality of the systems was guaranteed by the addition of Cl^- ions. The simulations with RNase A and α-Cgn A were perfomed using the CHARMM36m[28] force field for the protein.

Simulations were also performed for Trpzip1 (SWTWEGNKWTWK) (PDB 1LE0[39]) in the presence of 1 M sucrose. These simulations were carried out using the CHARMM36m,[28] AMBER 99SB-ILDN[29] or OPLS-AA[30] force fields for the peptide. CHARMM TIP3P[35] was used in combination with CHARMM36m, while the original TIP3P model[40] was used with AMBER and OPLS. More details about these simulations can be found in the Supporting Information.

In all cases, periodic boundary conditions were used, and the cut-off radius for both Coulombic (calculated using the PME method[41]) and Lennard-Jones interactions was set to 1.2 nm. Each box was first energy minimized with the steepest descent algorithm, and then equilibrated for 1 ns at 1 bar and 300 K (sim. 1-4 and 10) or 293 K (sim. 5-9) in the NPT ensemble, using Berendsen pressure and temperature coupling[42] at 1 ps relaxation time. The simulations were then run at the same temperature used for equilibration and at 1 bar in the NPT ensemble, controlling temperature and pressure with the Nosé-Hoover thermostat[43,44] (0.5 ps relaxation time) and Parrinello-Rahman barostat[45] (3 ps relaxation time), respectively. For simulations 5-9, the temperature was equilibrated at 293 K to allow a direct comparison with experimental data.[5,46] A 2 fs time-step was used, and configurations were saved every 2 ps. The Lincs algorithm was employed for constraining all bonds,[47] while the SETTLE algorithm kept the water molecules rigid.[48]

Analyses of Simulation Results

The last 40 ns (for simulations 1-9), or 100 ns (for simulation 10) were used for the analyses.

In simulations 1-4, the KB integrals were calculated by taking the average of the running KB integrals,

$$G_{ij} = 4\pi \int_0^R (g_{ij}(r) - 1) r^2 dr \tag{11}$$

at values of R where convergence is reached, which often occurs between 1.0 and 1.4 nm.[49]

In simulations 5-10, the preferential exclusion of excipient molecules from the protein surface was quantified using the preferential interaction coefficient $\Gamma(r)$, defined as

$$\Gamma(r) = n_3(r) - n_1(r)\left(\frac{n_3 - n_3(r)}{n_1 - n_1(r)}\right) \tag{12}$$

where $n_i(r)$ is the number of molecules of class i that on average are within a distance r from the protein surface. In contrast, n_i is the total number of molecules i in the simulation box. Γ was extracted from the trajectory, by computing the average value of the running $\Gamma(r)$ between 0.7 and 1.0 nm, and will be used in the following.

The number of amino acid-osmolyte hydrogen bonds was also measured. In particular, we evaluated the parameter χ, which represents the relative contribution of hydrogen bonding between the amino acid and the osmolyte with respect to the total number of intermolecular hydrogen bonds formed by the amino acid,

$$\chi = \frac{\text{number of amino acid-osmolyte hydrogen bonds}}{\text{total number of amino acid-osmolyte and amino acid-water hydrogen bonds}} \tag{13}$$

To determine the presence of a hydrogen bond, a geometrical criterion was used, requiring that the distance between donor and acceptor was less than 0.35 nm, and that the angle formed between the hydrogen atom and the line joining the COMs of donor and acceptor was smaller than 30°.

Results and Discussion

An Additivity Investigation for the KBP Force Field

The KBP force field[15] is a variant of CHARMM36,[16–18] which was developed using the KB integrals as target experimental data. It was observed that by changing the atomic partial charges and the Lennard-Jones parameter ε of the alcohol groups, it was possible to improve

the prediction of protein-carbohydrate, carbohydrate-carbohydrate and carbohydrate-water interactions. In the original CHARMM36 force field, the partial charges of alcohol O and H are -0.65 and 0.42, respectively. The magnitude of these charges is reduced in the KBP variant to -0.50 and 0.18, respectively. The partial charges for the other atoms in the sucrose and sorbitol molecules considered in this work are shown in Figure S1.

The Lennard-Jones parameters ε for the alcohol O and H atoms are 0.804 kJ mol^{-1} and 0.192 kJ mol^{-1}, respectively, in the original CHARMM36. The KBP force field, in contrast, defines two ε values for each hydroxyl atom, one specific for water interactions (0.900 kJ mol^{-1} and 0.300 kJ mol^{-1} for O and H, respectively), and the other applicable to all non-water interactions (0.450 kJ mol^{-1} for O and 0.120 kJ mol^{-1} for H). In contrast, there is no change in the σ parameter for the alcohol atoms in the KBP force field compared to the orginal CHARMM36. Four separate terms are defined in the KBPs for the water-specific interactions, one for each combination of the two hydroxyl atoms and the two types of water atoms. In each of these terms, the overall Lennard-Jones parameters of the carbohydrate-water interaction are defined as the geometric (for ε) or arithmetic (for σ) mean of the single atomic values. These modifications to the original CHARMM36 force field made self-association less favorable, and promoted carbohydrate-water interaction.

Cloutier et al.[15] observed that the prediction of sugar-sugar, sugar-protein and sugar-water KB integrals remarkably improved for carbohydrates with the KBP. This resulted in a better prediction of the experimental preferential exclusion of sugars from the protein surface.[4,6] The first objective of the present work was to investigate whether the KBP force field for sugars also respected the additive properties previously described (Equation 10). We therefore verified if the KBP parameters could predict the experimental values of $\gamma = G_{23} - G_{12}$ for the individual amino acid-sugar interaction, as reported in Auton et al.[25]

The backbone (γ^{bb}) and sidechain (γ^{sc}) contributions are considered separately in the additive approach outlined in the Materials and Methods section. Following the same approach described by Auton et al.,[24,25] it is assumed that the additivity exists also within a

single amino acid. This implies that the sidechain contribution γ^{sc}, defined as the effect of substituting the side chain for a hydrogen atom, can be calculated by subtracting the KB integral γ_{GLY} for glycine (where the sidechain is just a hydrogen atom) to the KB integral of the amino acid i being considered γ_i,

$$\gamma_i^{sc} = \gamma_i - \gamma_{GLY} \tag{14}$$

Here, attention will be focused on zwitterionic residues, in line with what was done experimentally.[50]

The backbone contribution γ^{bb} can be computed using, for instance, the constant increment method[26] for the NAG_xA series. According to this construct, the contribution of the peptide unit can be obtained from the slope of a plot of the γ values versus the number of internal glycine units x (Figure 1a),

$$\gamma = \gamma^{eg} + \gamma^{bb}x \tag{15}$$

where γ^{eg} is the contribution of the end groups.

We can observe that a good linearity is observed for the NAG_xA series in 1 M sucrose in Figure 1a ($R^2 \approx 0.95$), obtaining a value of $\gamma^{bb} \approx -0.128$ nm^3. This linearity further confirms the existence of additivity for the osmolyte-polypeptide interaction, as was also experimentally observed.[26]

Figure 2a shows a comparison between the γ^{sc} and γ^{bb} values for the case of 1 M sucrose as a model osmolyte, obtained using either the KBP force field (red bars), or the experimental data by Auton et al.[25] (dashed bars).

As can be observed, the KBP force field gives results that are not always in perfect accordance with experiments. For instance, it overestimates the interaction between sucrose and aromatic residues (Phe, Tyr and Trp), and also tends to predict preferential interaction (positive γ^{sc}) of sucrose with some non-polar groups (Ala, Val, Leu, Ile), while preferential

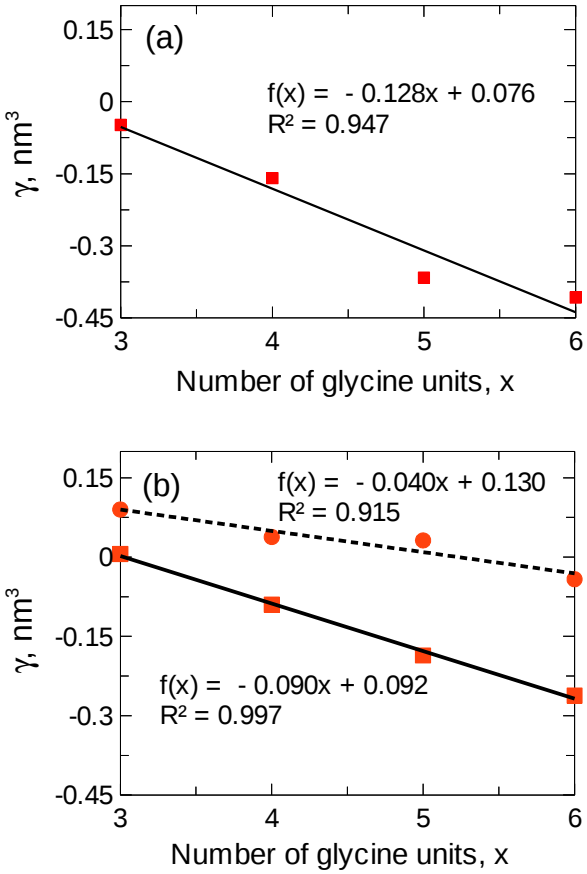

Figure 1: KB integrals ($\gamma = G_{23} - G_{12}$) for the N-acetyl glycinamide series NAG_xA, as function of the number of internal glycine units x. The results shown in panel (a) were obtained for 1 M sucrose, and using the KBP force field.[15] Panel (b) displays results for the ADD force field. Here, the solid line with red squares is for 1 M sucrose, while the dashed one with red circles is for 1 M sorbitol.

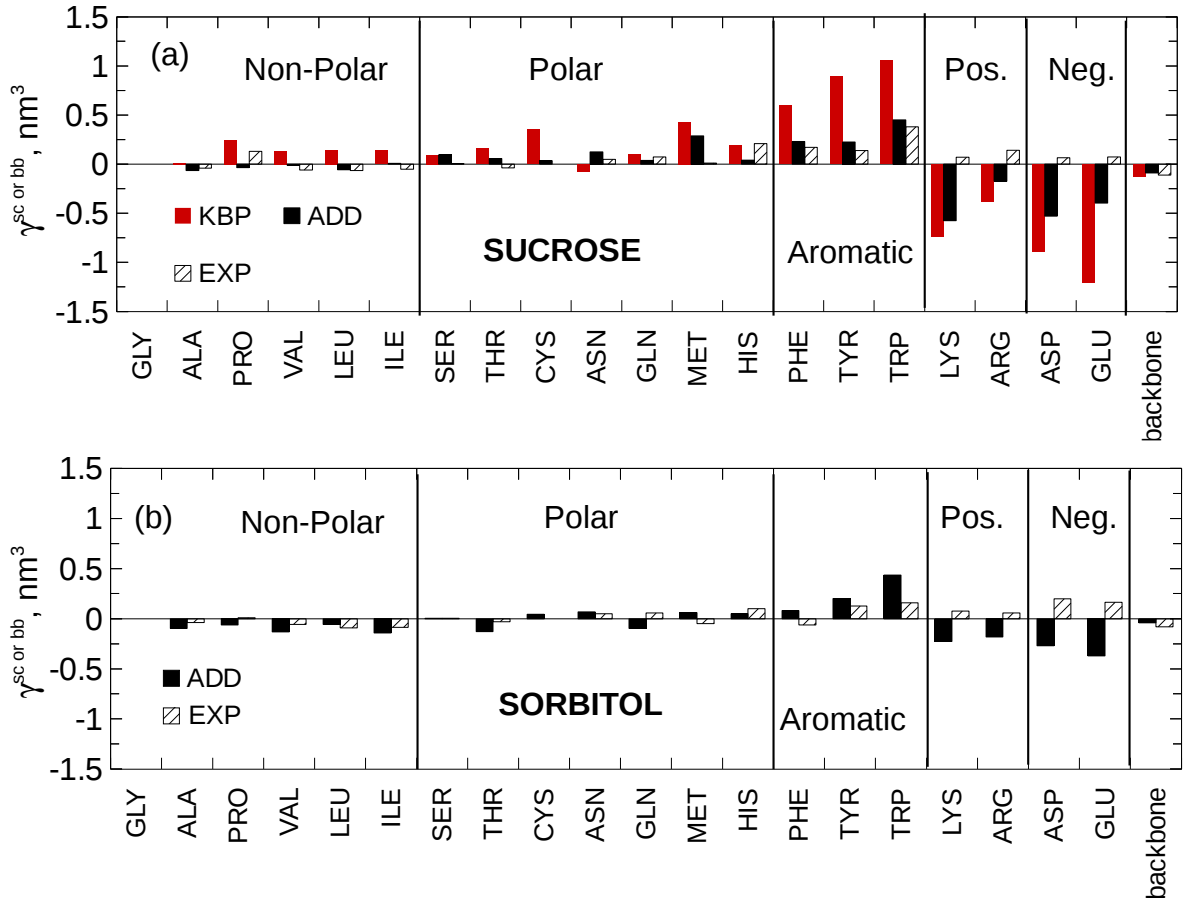

Figure 2: KB integrals γ^{sc} and γ^{bb} for the different amino acids side chains and the backbone, and for (a) 1 M sucrose or (b) 1 M sorbitol as model osmolytes. Red, black and dashed bars correspond to the KBP, the ADD and the experimental data,[24,50] respectively. In both panels, the amino acids are divided into groups, depending on their side chain properties (non-polar, polar, aromatic, positively and negatively charged, moving from left to right), and ordered according to their molecular weight (increasing from left to right) in each group. Zwitterionic amino acids were used in the simulations.

exclusion (negative γ^{sc}) was experimentally observed. This result is not surprising considering that the magnitude of partial charges for the alcohol atoms is reduced in the KBP force field, as such decreasing the polar behavior of this group. In line with this, simulations of the KBP carbohydrates with lysozyme predicted that sugars were interacting more strongly with hydrophobic residues, like Ala, Ile and Phe,[15] than with hydrophilic patches.

This increased interaction of KBP sucrose with non-polar and aromatic residues is compensated for by a significant exclusion of simulated sucrose from charged amino acids. This also is not in line with the experimental data, where a slight interaction was observed (Figure 2a). However, we found that it is not possible to reproduce this experimental behavior, i.e., preferential interaction with charged side chains, with any choice of force field parameters for the hydroxyl group. We hypothesized that this may be due to the very high ionic strength in simulations of the charged residues, which does not reproduce real experimental conditions. For computational reasons, we are limited to small simulation boxes, and the presence of a huge number of ions in a small system may lead to the observed discrepancy. To verify this hypothesis, we tried two different simulation conditions for lysine, arginine, aspartic and glutamic acid. In the first one, 25 amino acid molecules were introduced into the simulation box (8 x 8 x 8 nm) as described in the Materials and Methods section, and neutralized with Na^+ or Cl^- ions, resulting in a high ionic strength. In the second case, only one residue was inserted into the box, with a corresponding 25-fold reduction in ionic strength. We observed no significant change in the computed KB integrals for these two configurations, suggesting that the influence of the ionic strength is not dramatic and cannot therefore be the reason for the observed disagreement between simulations and experimental data. We note that the experiments were carried out with the neutral salts of the amino acids. The discrepancies for the charged amino acids may also arise from how the amino acids interact with their counterions. Further analysis would be needed in this direction. Finally, another possible explanation for this discrepancy may be related to the presence of charge effects in the setup used to obtain the experimental values, where solubility data were used as inputs,

the activity coefficients were ignored and therefore some non-ideality of charged systems may have been neglected.[50]

Developing the ADD Force Field

We tried to further improve the prediction of experimental γ values between carbohydrates and amino acids, by modifying the non-bonded parameters for sugars. The KBP set was used as the starting point for this investigation. Cloutier et al.[15] showed that modifying the charges and the Lennard-Jones parameter ε of the alcohol groups should be enough to tune the solution properties. The final objective was to obtain a force field that could accurately predict not only the experimental amino acid-sugar KB integrals, but also sugar-sugar (G_{33}), sugar-water (G_{13}) and sugar-protein (Γ) interactions. For this purpose, we made the following assumptions: i) changing the partial charges of O and H alcohol atoms affects the polar behavior of sugars. This will be useful to adjust their polarity, making the interaction with the non-polar groups less favorable; ii) the Lennard-Jones parameter ε for interaction with water affects G_{33} and G_{13}.

We started by modifying the partial charges of O and H alcohol atoms, using sucrose (at 1 M concentration) as model osmolyte (see Table 2). The Lennard-Jones parameters for this first tuning step were the same as for the KBP force field. The γ^{sc} values for two model amino acids, one polar (Asn) and the other apolar (Ala), were used as reference to adjust sucrose polarity. The simulations results for these two residues, listed on each row of Table 2 for the different combinations of O and H charges here investigated, were compared to the experimental values, shown on the last row of the table. As evident, the KBP values (-0.50 for O and 0.18 for H, first row of Table 2) result in a reversal of sign compared to the experimental data. That is, KBP sucrose interacts with Ala and is excluded from Asn, while the opposite is observed in experiments. This confirms that the apolar behavior of sugars is enhanced by the KBP parameters.

Other reference values we used to adjust the force field parameters were the KB integrals

Table 2: Effect of O and H charges on sucrose polarity (ALA and ASN γ^{sc}), self-association (G_{33}), and interaction with water (G_{13}). The Lennard-Jones parameters were the same as for the KBP force field. The last column corresponds to the root mean square error (RMSE) between the simulated and experimental values (listed in the last row), for each combination of partial charges. A 1 M concentration was selected for sucrose.

O charge	H charge	ALA γ^{sc}, nm^3	ASN γ^{sc}, nm^3	G_{33}, nm^3	G_{13}, nm^3	RMSE, nm^3
-0.50	0.18	0.011	-0.073	-0.843	-0.251	0.069 [1]
-0.50	0.25	-0.048	-0.088	-0.760	-0.250	0.077 [1]
-0.60	0.28	-0.158	-0.113	-1.070	-0.130	0.167
-0.62	0.30	-0.032	-0.151	-1.070	-0.130	0.167
-0.64	0.35	-0.206	0.328	-0.940	-0.170	0.174
-0.64	0.37	-0.081	0.055	-0.910	-0.176	0.055[2]
-0.64	0.38	-0.430	-0.261	-0.824	-0.206	0.249
-0.64	0.39	-0.218	-0.365	-0.786	-0.218	0.226
-0.65	0.25	-0.257	0.089	-1.170	-0.090	0.217
-0.65	0.33	-0.045	0.244	-1.060	-0.130	0.161[3]
-0.65	0.40	-0.384	0.007	-0.620	-0.260	0.201
	Exp.[4]	-0.039	0.050	-0.819	-0.220	

[1]Sugars too apolar

[2]Not enough excluded from protein

[3]Bad prediction of sugar-sugar interaction: possibility to modify Lennard-Jones parameters

[4]Experimental values from[50–52]

for sucrose-sucrose (G_{33}) and sucrose-water (G_{13}) interaction. As already shown by Cloutier et al.,[15] the KBP force field results in an extremely good prediction of these values.

The magnitude of the partial charges of O and H atoms needs to be increased to achieve the correct polarity. For this reason, we focused our attention on the values of the partial charges closer to that of the original CHARMM36 force field. For instance, the combination -0.64 for O and 0.37 for H results in a correct prediction of the experimental sign of γ^{sc} for both Ala and Asn, and also describes fairly well the interaction of sucrose with both water and other sucrose molecules. However, when we tried to apply these force field parameters to the case of RNase A at pH 3 (sim. 5 in Table 1), we observed that they led to an underestimation of the exclusion of sucrose from the protein.

The root mean square error (RMSE) reported in the last column of Table 2 was computed considering all four parameters G_{33}, G_{13} and γ^{sc} for both Ala and Asn. Instead, Figure 3a shows the RMSE calculated considering as reference values only the γ^{sc} for Ala and Asn. This 3D plot, obtained varying the O and H charges, shows that the minimum RMSE is observed for the combination -0.65 for O and 0.33 for H (also highlighted with an arrow in Figure 3a). This possible choice of force field parameters, hence, correctly describes sucrose polarity, resulting in a good prediction of γ^{sc} for both Ala and Asn. As shown in Table 2, the combination -0.65 for O, 0.33 for H underestimates sucrose-sucrose interaction (too negative G_{33}). However, as mentioned in our working hypotheses, we expect that it would be possible to adjust the values of G_{33} and G_{13} by selecting the proper Lennard-Jones parameters ε for water interaction.

For this reason, the combination of partial charges -0.65 for O and 0.33 for H was further explored, and the ε parameters (for interaction with water) modified with the objective to improve the prediction of sucrose-sucrose and sucrose-water interaction (Table 3). We started from the ε values of the KBP force field (0.90 kJ mol^{-1} for O and 0.30 kJ mol^{-1} for H) and progressively reduced them to improve the matching with the experimental data (shown in the last row of Table 3). The choice that seemed to result in the best prediction

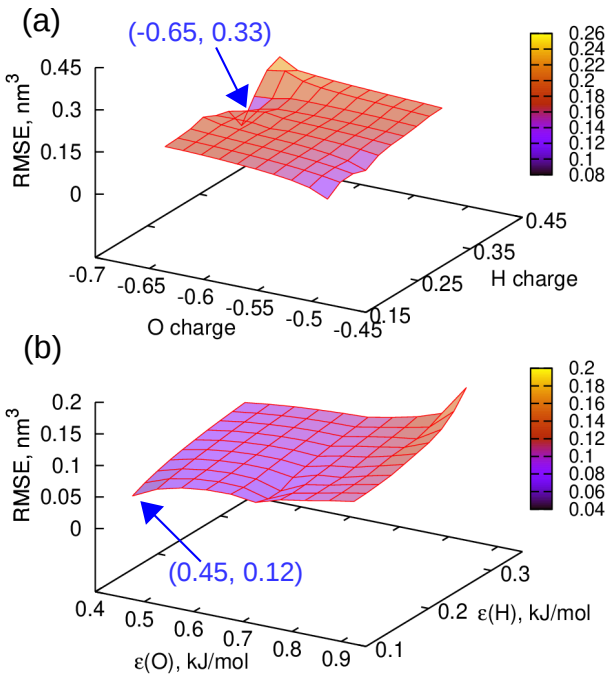

Figure 3: (a) Root mean square error (RMSE) between experimental and simulated values of γ^{sc} for Ala and Asn, as a function of the O and H charges. The Lennard-Jones parameters were the same as for the KBP force field. (b) RMSE between experimental and simulated values of KB integrals G_{33} and G_{13} as function of the Lennard-Jones parameters ε for water interaction. The O and H partial charges were set to -0.65 and 0.33, respectively. The data shown in both panels are for 1 M sucrose as model osmolyte.

of G_{33} and G_{13} values was 0.45 kJ mol^{-1} for O and 0.12 kJ mol^{-1} for H. This combination minimizes the root mean square error (RMSE, shown in the last colum of Table 3 and in Figure 3b) between simulated and experimental data, and also has the advantage of requiring no separate Lennard-Jones parameters for water and non-water interactions. In the following we will refer to this optimized combination of parameters as to the ADD force field, to underline the fact that it was obtained from an additivity analysis.

Table 3: Effect of Lennard-Jones parameters ε for water interaction on the KB integral between 1 M sucrose and water (G_{13}) and on sucrose self-association (G_{33}). The O and H partial charges were -0.65 and 0.33, respectively. The last column corresponds to the root mean square error (RMSE) between the simulated and experimental values (listed in the last row), for each combination of Lennard-Jones parameters.

ε(O), kJ/mol	ε(H), kJ/mol	G_{33}, nm^3	G_{13}, nm^3	RMSE, nm^3
0.90	0.30	-1.060	-0.130	0.182
0.85	0.25	-0.967	-0.155	0.115
0.78	0.18	-0.937	-0.165	0.092
0.70	0.15	-0.897	-0.173	0.064
0.45	0.12	-0.758	-0.193	0.047[1]
	Exp.[2]	-0.819	-0.220	

[1]No separate Lennard-Jones parameters for water and non-water interactions
[2]Experimental values from [51,52]

Table 4 summarizes the parameters for the original CHARMM36, KBP and ADD force fields. The non-alcohol atoms partial charges for the KBP and ADD force fields are the same, and are shown in Figure S1 for the two molecules (sucrose and sorbitol) that will be used to verify the performances of the ADD parameters in the next section.

The ADD Force Field Improves the Prediction of Sugars and Polyols Polarity

The sidechain and backbone γ contributions were recomputed for the ADD force field. To verify that the sidechain contribution would not vary depending on the terminal capping

Table 4: Summary of O and H partial charges and Lennard-Jones parameters ε (for interaction with all, non-water or water only atoms) for the original CHARMM36, KBP and ADD force fields.

force field	O charge	H charge	interaction type	ε(O), kJ/mol	ε(H), kJ/mol
original CHARMM36	-0.65	0.42	all	0.804	0.192
KBP	-0.50	0.18	non-water	0.450	0.120
			water	0.900	0.300
ADD	-0.65	0.33	all	0.450	0.120

conditions, both capped (acetylated and amidated) and zwitterionic amino acids were considered. In Figures S2a-c we show γ^{sc} values computed using either capped or zwitterionic amino acids in 1 M sucrose or sorbitol, and obtained using the ADD or KBP force fields (sim. 1 and 2 in Table 1). The trend was essentially the same for both the capped and the zwitterionic form. These results are in line with the work of Nozaki and Tanford,[53] where the interaction of the sidechain of a branched organic compound with the solvent was found to be approximately independent of the interaction of the backbone to which the side chain was attached. The simulation results, therefore, seem to confirm the existence of additivity within the individual amino acid (i.e., amino acid KB integral = backbone contribution + side chain contribution). This also means that, in our specific case, both zwitterionic and capped amino acids may be considered to compute the side chain contribution, without dramatically affecting the final results. Similarly to what was done before, we will therefore focus our attention on zwitterionic amino acids.

Figure 1b further confirms the existence of a good linearity for the NAG$_x$A series, also for the ADD force field. This holds true for both 1 M sucrose (solid line with red squares) and 1 M sorbitol (dashed line with red circles). The values of γ^{bb} are ≈ -0.090 nm^3 and ≈ -0.040 nm^3 for sucrose and sorbitol, respectively.

Figure 2 shows a comparison between the γ^{sc} and γ^{bb} values obtained either experimentally (dashed bars) or with the ADD (black bars) force field, for either 1 M sucrose (Figure 2a) or 1 M sorbitol (Figure 2b). Overall, the ADD force field resulted in an improved pre-

diction of experimental data, compared to the KBP parameters. This is also evident from the RMSE values listed in Table 5. The ADD force field resulted in a considerably lower error, compared to the KBPs, especially for non-polar and aromatic sidechains. The observed improvement is mostly related to the enhanced polar behavior of the ADD sugars compared to the KBP force field. As can be observed, the ADD parameters (black bars) correctly predict exclusion of sugars from non-polar residues, and do not overestimate sucrose-aromatic residues interaction, as instead occurred with the KBP force field (red bars, Figure 2a).

In the case of the ADD force field, similarly to the KBPs, we observed a discrepancy between simulated and experimental values for charged amino acids. As already mentioned, no possible choice of force field parameters for the hydroxyl group could solve this problem. To investigate this further, simulations of 1 M sucrose, described with either the KBP or ADD force fields, were perfomed in presence of 4 M guanidinium chloride (GdmCl). GdmCl is a strong denaturant, often used in experimental investigations of protein folding. The guanidinium group is also present in the side chain of arginine, and is responsible for the polarity of this amino acid. Details on the simulations of 1 M sucrose in presence of GdmCl can be found in the Supporting Information. The KB integrals extracted from the obtained trajectories (Table S1) indicated a tendency of KBP sucrose to self-interact in presence of GdmCl. This self-interaction, coupled to an extremely unfavorable sucrose-ions interaction, translated into demixing of the solution (Figure S3). The ADD force field could ameliorate this situation, by decreasing sucrose self interaction, and preventing demixing of the solution. This suggests that the interaction of ADD sucrose with charged groups, such as the guanidinium moiety also present in the sidechain of arginine, is improved compared to the KBPs.

Table 6 further shows a comparison between the original CHARMM36, KBP and ADD force fields for what concerns the prediction of sugars self-association (G_{33}), interaction with water (G_{13}) and preferential exclusion from proteins (Γ, for RNase A and α-Cgn A as model systems). The original CHARMM36 force field resulted in a clear overestimation of self-

Table 5: RMSE between experimental and simulated values of γ^{sc} and γ^{bb} for 1 M sucrose. For the γ^{sc} values, the RMSE has also been calculated separately for the different sidechain groups.

Sucrose		
RMSE, nm^3	**KBP**	**ADD**
all sidechains + backbone	0.503	0.252
non-polar sidechains	0.149	0.075
polar sidechains	0.199	0.147
aromatic sidechains	0.633	0.073
positive sidechains	0.680	0.507
negative sidechains	1.131	0.536

interaction for both sucrose and sorbitol. The sugar molecules are strongly attracted to one another when using CHARMM36, and tend to repel water, in contrast to experimental data.

The KBP parameters and the ADD force field strongly improve the agreement with experimental values. If we compare more closely the KBP and ADD force fields, we can observe that they provide a similar description of G_{33} and G_{13} values for sucrose, while the KB integrals predicted by the ADD parameters showed the best agreement with experiments for sorbitol, as confirmed by the small RMSE. These values of KB integrals were obtained using the Lorentz-Berthelot rule for the van der Waals cross-interactions. This means that the Lennard-Jones parameters of a cross-interaction are defined as the arithmetic (for σ) or geometric (for ε) mean of the individual atomic values. Sim. 4 in Table 1 were also performed using the geometric combination rule (i.e., geometric mean for both σ and ε), for the case of sucrose as a model osmolyte. The solution KBIs (G_{33} and G_{13}) were computed, and the results for both combination rules and for both KBP and ADD are shown in Table S2. It was observed that the behavior of ADD sucrose was not influenced by the combination rule. In contrast, the KBP sucrose self-interaction became less unfavorable when the geometric mean was used instead of the Lorentz-Berthelot rule, leading to a significant deviation from experimental values. It is important to note that the KBP force field was developed using the Lorentz-Berthelot rule, and uses a separate set of nonbonded parameters for sugar OH-

water interaction. This probably causes imbalance between sugar-sugar and sugar-water interactions when the geometric rule is used.

For what concerns preferential exclusion, the ADD force field seems to provide an accurate estimation of Γ for sucrose, both at low (0.1 M) and high (1.0 M) concentration, performing better than the KBPs at high concentration. The opposite is true for sorbitol and Rnase A. In this case, the KBP parameters seem to reproduce experimental values better, while the ADD force field slightly underestimates preferential exclusion. However, both force fields perform much better than the original CHARMM36, which was shown to result in preferential interaction (i.e., positive Γ) of sugars and polyols, especially at high concentration of cosolutes.[15] What is interesting and important to note is that this improved description of preferential exclusion from proteins derives from the additivity of amino acids/carbohydrate interaction in the case of the ADD force field.

The diffusion coefficients of ADD sucrose and sorbitol were also measured, and compared to the values obtained experimentally,[54,55] or with the KBP and original CHARMM36 force fields. The results of this analysis are illustrated in Figure S4, and show that the KBP parameters lead to higher diffusion coefficients compared to the original CHARMM36 force field. ADD carbohydrates generally display intermediate values, albeit closer to those observed for the KBPs. For sorbitol (Figure S4b), the ADD and KBP parameters improve the description of the experimental trend, while the opposite is true for sucrose (Figure S4a).

Some simulations of Trpzip1 in 1 M ADD sucrose were also performed, using three different force fields (CHARMM36m,[28] AMBER 99SB-ILDN[29] or OPLS-AA[30]) for the peptide. The preferential exclusion coefficient Γ for these simulations is shown in Figure S5. The results with the different protein force fields did not vary significantly (maximum deviation $<=1$), indicating compatibility of the ADD force field with these widely used protein force fields.

As a further step, we would like to illustrate how the γ^{sc} and γ^{bb} values in Figure 2 could be used in the framework of an additive approach. As shown in Eq. 10, an additive

Table 6: Comparison between the sugar-sugar (G_{33}), sugar-water (G_{13}) and sugar-protein (Γ, for the specific case of RNase A and α-Cgn A) interaction as obtained experimentally, or as predicted by the original CHARMM36, KBP and ADD force fields. The RMSE between simulations and experimetal values is also shown.

Sucrose [1]			
KBIs	**original CHARMM36**	**KBP**	**ADD**	**Exp.**
G_{33} 1M suc., nm^3	0.346	-0.843	-0.758	-0.819
G_{13} 1M suc., nm^3	-0.633	-0.251	-0.193	-0.220
RMSE, nm^3	0.87	0.03	0.05	
Pref. Excl.		**KBP**	**ADD**	**Exp.**
Γ (RNase A pH 3 in 0.1M suc.)		-1.12	-0.74	-0.641
Γ (RNase A pH 3 in 0.7M suc.)		-3.5	-3.0	n/a
Γ (RNase A pH 3 in 1.0M suc.)		-4.9	-7.5	-7.611
Γ (RNase A pH 7 in 0.7M suc.)		-5.9	-6.9	n/a
Γ (α-Cgn A pH 3 in 0.7M suc.)		-6.1	-5.6	-7.135
RMSE, -		1.70	0.89	
Sorbitol [2]			
KBIs	**original CHARMM36**	**KBP**	**ADD**	**Exp.**
G_{33} 1M sor., nm^3	0.231	-0.812	-0.417	-0.400
G_{13} 1M sor., nm^3	-0.306	-0.156	-0.171	-0.173
RMSE, nm^3	0.46	0.29	0.01	
Pref. Excl.		**KBP**	**ADD**	**Exp.**
Γ (RNase A pH 2 in 2.2M sor.)		-11.9	-8.9	-12.91
Γ (RNase A pH 5.5 in 0.55M sor.)		-4.7	-2.2	-4.76
Γ (RNase A pH 5.5 in 2.2M sor.)		-15.0	-11.5	-13.26
RMSE, -		1.16	2.93	

[1]Experimental values from [5,51,52]

[2]Experimental values from [46,56–58]

relation exists between Kirkwood-Buff integrals, and this result could be exploited to predict the m value for protein-osmolyte interaction. More specifically, the values of γ^{sc} and γ^{bb} shown in Figure 2, and the G_{13} and G_{33} listed in Table 6 for the ADD force field could be inserted into Eq. 10 to predict the m value. We tried this approach for a protein that was also experimentally investigated, i.e., RCAM-T1 (a reduced and carboxyamidated variant of RNase T1).[59] The side-chain and backbone accessible surface areas of residues in the native fold were calculated from a Protein Data Bank file of RNase T1 (PDB code 2BU4[60]), exploiting the algorithm developed by Lee and Richards[61,62] as modified by Lesser and Rose.[63] A probe size equal to 0.14 nm was used. The solvent accessible surface areas in the denatured state were instead obtained by interpolating[64] between the lower and upper limits representing compact and expanded models of the denatured state, as defined by Creamer et al.[65] The fractional exposure of each group was then computed as described in Eq. 4.

The experimental m value for RCAM-T1 in sucrose is 1.550 kcal mol^{-1} M^{-1}.[59] Using the additive approach by Auton and Bolen, with experimentally measured transfer free energies, a value of 1.060 kcal mol^{-1} M^{-1} was obtained for the m value,[24] in line with experiments. Using Eq. 10, with KB integrals from the ADD force field as inputs, the m value obtained is 1.554 kcal mol^{-1} M^{-1}, which is also in very good agreement with the experimental result. In contrast, if the results obtained with the KBP parameters are substituted into Equation 10, an m value of -0.291 kcal mol^{-1} M^{-1} is calculated, which does not compare well with experiments.

An m value of 1.380 kcal mol^{-1} M^{-1} for RCAM-T1 in sorbitol was also experimentally obtained,[24] and fairly well predicted by Auton et al. using the additive approach based on experimental transfer free energies, which gave a result of 0.816 kcal mol^{-1} M^{-1}. Using the KB integrals as predicted by the ADD force field and Eq. 10, an m value of 1.116 kcal mol^{-1} M^{-1} is obtained, in good accordance with experiments.

As a last step, we show in Figure 4 the χ parameter, as computed with the KBP or ADD force fields, for sucrose and sorbitol. Focusing on Figure 4a, we can observe that the ADD

force field predicts the formation of a larger number of sucrose-amino acid hydrogen bonds compared to the KBP. This is again related to the higher polarity of ADD sugars. We also observe that sucrose (Fig. 4a) forms a larger number of hydrogen bonds with amino acids than sorbitol (Fig. 4b). As already hypothesized, this may be related to the higher number of accessible hydrogen-bonding sites in sucrose compared to sorbitol.[66,67]

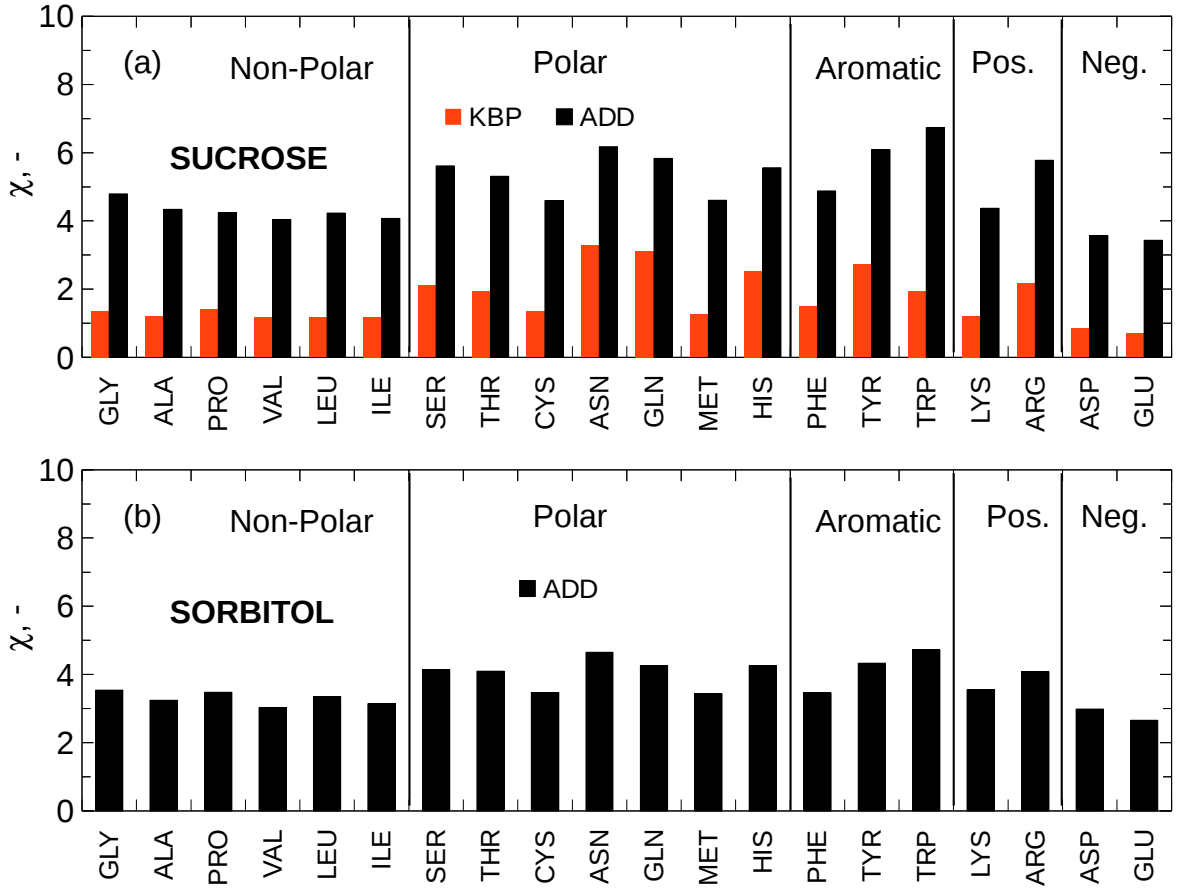

Figure 4: χ parameter for the different amino acids in (a) 1 M sucrose or (b) 1 M sorbitol, as obtained with the KBP (red bars) or ADD (black bars) force fields.

Conclusions

In this work, a new set of parameters for carbohydrates (ADD) has been developed, with the objective to improve the prediction of individual carbohydrate-backbone and carbohydrate-side chains KB integrals. We have shown that molecular dynamics simulations can reproduce

the experimentally observed additive behavior of sugar-protein interaction. The new set of parameters also correctly predicts sucrose and sorbitol self-association, as well as their interaction with water. The ADD force field improves the description of sugars polarity compared to the KBPs, and shows comparable performance in predicting the degree of preferential exclusion from proteins. By properly adjusting charges and Lennard-Jones parameters a good description of the solution properties is obtained, without requiring separate parameters for water or non-water interactions, as instead occurs for the KBPs. The ADD parameters were developed in combination with the CHARMM36m force field for proteins, but good compatibility was observed for the case of the Trpzip1 peptide also with AMBER 99SB-ILDN and OPLS-AA.

The main discrepancy between simulations and experiments regards charged amino acids. No possible choice of non-bonded parameters for the hydroxyl group made it possible to solve this problem, and the origin of this disagreement may lie in the experimental or simulation setup. A high ionic strength in simulations, due to small box sizes, seems not to be the problem, as the results did not change when reducing the ions concentration within the simulated systems. Also, the experiments were carried out by neutralizing the charge groups, and in a way the results depended on the counterions pairing. The presence of neglected deviations from ideality in experiments may also account for the observed discrepancy. It must furthermore be considered that the parameters of the hydroxyl moiety only were tuned in the present work. This choice was made to reduce the number of variables, and because the alcohol atoms represent the functional group of sugars and polyols, thus being responsible for many of their properties. However, the non-bonded parameters of other atoms may need to be adjusted to correctly reproduce the carbohydrates-charged amino acids interactions. For instance, the partial charges of other atom types may need to be modified, or polarizable models[68] may be an option to address this problem. The use of separate Lennard-Jones parameters between alcohols and charged amino acids may also be considered, but, as we observed for the KBP force field in Table S2, this may make the force field not fully compatible

with different combination rules. The hope of the authors is that this work may stimulate further improvement in this direction. A first step has already been made in the present manuscript, where we observed that the ADD parameters showed better compatibility with charged groups than the KBPs, for instance avoiding demixing in sucrose-GdmCl mixtures.

Overall, a new approach has been herein proposed for the development of a force field to describe protein-osmolyte interactions, where the individual backbone-osmolyte and sidechain-osmolyte contributions are used as target data. This new strategy seems to give promising results, and will further be investigated.

Supporting Information

Partial charges of simulated carbohydrates, comparison between zwitterionic and capped amino acids, simulations of sucrose-guanidinium chloride mixtures, ADD and KBP behavior in the case of different combination rules, diffusion coefficients of sucrose and sorbitol, compatibility of ADD with the AMBER 99SB-ILDN and OPLS-AA force fields.

Acknowledgement

The authors acknowledge support from the hpc@polito team (http://www.hpc.polito.it), from the Center for Scientific Computing at the California Nanosystems Institute (CNSI, NSF grant CNS-1725797), and from CINECA under the ISCRA initiative (DisCyc-HP10CQROS1), for the availability of high performance computing resources and support. This work used the Extreme Science and Engineering Discovery Environment, which is supported by National Science Foundation grant number ACI-1548562 (MCA05S027). The authors aknowledge support from the NSF (MCB-1716956) and the NIH (R01-GM118560-01A).

References

(1) Yancey, P. H. Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. *J. Exp. Biol.* **2005**, *208*, 2819–2830.

(2) Magazú, S.; Migliardo, F.; Mondelli, C.; Vadalà, M. Correlation between bioprotective effectiveness and dynamic properties of trehalose–water, maltose–water and sucrose–water mixtures. *Carbohydr. Res.* **2005**, *340*, 2796 – 2801.

(3) Magazú, S.; Migliardo, F.; Telling, M. T. F. Study of the dynamical properties of water in disaccharide solutions. *Eur. Biophys. J.* **2007**, *36*, 163–171.

(4) Timasheff, S. The control of protein stability and association by weak interactions with water: How do solvents affect these processes? *Ann. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 67–97.

(5) Lee, J. C.; Timasheff, S. N. The stabilization of proteins by sucrose. *J. Biol. Chem.* **1981**, *256*, 7193–7201.

(6) Arakawa, T.; Timasheff, S. N. Stabilization of protein structure by sugars. *Biochemistry* **1982**, *21*, 6536–6544.

(7) Wyman, J. In *Linked Functions and Reciprocal Effects in Hemoglobin: A Second Look*; Anfinsen, C., Anson, M., Edsall, J. T., Richards, F. M., Eds.; Adv. Protein Chem.; Academic Press, 1964; Vol. 19; pp 223 – 286.

(8) Tanford, C. Extension of the theory of linked functions to incorporate the effects of protein hydration. *J. Mol. Biol.* **1969**, *39*, 539 – 544.

(9) Timasheff, S. N. Protein-solvent Preferential Interactions, Protein Hydration, and the Modulation of Biochemical Reactions by Solvent Components. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9721 – 9726.

(10) Lins, R. D.; Hünenberger, P. H. A new GROMOS force field for hexopyranose-based carbohydrates. *J. Comput. Chem.* **2005**, *26*, 1400–1412.

(11) Damm, W.; Frontera, A.; Tirado–Rives, J.; Jorgensen, W. L. OPLS all-atom force field for carbohydrates. *J. Comput. Chem.* **1997**, *18*, 1955–1970.

(12) Nester, K.; Gaweda, K.; Plazinski, W. A GROMOS Force Field for Furanose-Based Carbohydrates. *J. Chem. Theory Comput.* **2019**, *15*, 1168–1186.

(13) Glennon, T. M.; Zheng, Y.-J.; Le Grand, S. M.; Shutzberg, B. A.; Merz Jr., K. M. A force field for monosaccharides and $(1 \rightarrow 4)$ linked polysaccharides. *J. Comput. Chem.* **1994**, *15*, 1019–1040.

(14) Foley, B. L.; Tessier, M. B.; Woods, R. J. Carbohydrate force fields. *WIREs Comput. Mol. Sci.* **2012**, *2*, 652–697.

(15) Cloutier, T.; Sudrik, C.; Sathish, H. A.; Trout, B. L. Kirkwood–Buff-derived alcohol parameters for aqueous carbohydrates and their application to preferential interaction coefficient calculations of proteins. *J. Phys. Chem. B* **2018**, *122*, 9350–9360.

(16) Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; Mackerell Jr, A. D. Additive empirical force field for hexopyranose monosaccharides. *J. Comput. Chem.* **2008**, *29*, 2543–2564.

(17) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. *J. Chem. Theory Comput.* **2009**, *5*, 2353–2370.

(18) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ_1 and χ_2 dihedral angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.

(19) Kirkwood, J. G.; Buff, F. P. The statistical mechanical theory of solutions. I. *J. Chem. Phys.* **1951**, *19*, 774–777.

(20) Ben-Naim, A. *Molecular Theory of Solutions*; Molecular Theory of Solutions; OUP Oxford, 2006.

(21) Ben-Naim, A. Inversion of the Kirkwood–Buff theory of solutions: Application to the water–ethanol system. *J. Chem. Phys.* **1977**, *67*, 4884–4890.

(22) Smith, P. E. On the Kirkwood–Buff inversion procedure. *J. Chem. Phys.* **2008**, *129*, 124509.

(23) Sapir, L.; Harries, D. Linking trehalose self-association with binary aqueous solution equation of state. *J. Phys. Chem. B* **2011**, *115*, 624–634.

(24) Auton, M.; Bolen, D. W. Predicting the energetics of osmolyte-induced protein folding/unfolding. *Proc. Natl. Acad. Sci.* **2005**, *102*, 15065–15068.

(25) Auton, M.; Bolen, D. W.; Rösgen, J. Structural thermodynamics of protein preferential solvation: Osmolyte solvation of proteins, aminoacids, and peptides. *Proteins* **2008**, *73*, 802–813.

(26) 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*, 1329–1342.

(27) Horinek, D.; Netz, R. R. Can Simulations Quantitatively Predict Peptide Transfer Free Energies to Urea Solutions? Thermodynamic Concepts and Force Field Limitations. *J. Phys. Chem. A* **2011**, *115*, 6125–6136.

(28) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71 – 73.

(29) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78*, 1950–1958.

(30) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(31) Pace, C. N.; Hermans, J. The stability of globular protein. *CRC Critical Reviews in Biochemistry* **1975**, *3*, 1–43.

(32) Tanford, C. Isothermal Unfolding of Globular Proteins in Aqueous Urea Solutions. *J. Am. Chem. Soc.* **1964**, *86*, 2050–2059.

(33) Rösgen, J.; Pettitt, B. M.; Bolen, D. W. Protein Folding, Stability, and Solvation Structure in Osmolyte Solutions. *Biophys. J.* **2005**, *89*, 2988 – 2997.

(34) Abraham, M. J.; Murtola, T.; Schulz, R.; Pall, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19 – 25.

(35) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S. et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(36) Berisio, R.; Sica, F.; Lamzin, V. S.; Wilson, K. S.; Zagari, A.; Mazzarella, L. Atomic resolution structures of ribonuclease A at six pH values. *Acta Cryst.* **2002**, *D58*, 441–450.

(37) Wang, D.; Bode, W.; Huber, R. Bovine chymotrypsinogen A: X-ray crystal structure

analysis and refinement of a new crystal form at 1.8 Å resolution. *J. Mol. Biol.* **1985**, *185*, 595 – 624.

(38) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **2012**, *40*, W537–W541.

(39) Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. Tryptophan zippers: Stable, monomeric β-hairpins. *Proc. Natl. Acad. Sci.* **2001**, *98*, 5578–5583.

(40) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(41) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(42) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(43) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255–268.

(44) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(45) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(46) Xie, G.; Timasheff, S. N. Mechanism of the stabilization of ribonuclease A by sorbitol: preferential hydration is greater for the denatured then for the native protein. *Protein Sci.* **1997**, *6*, 211–221.

(47) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

(48) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952–962.

(49) Ganguly, P.; van der Vegt, N. F. A. Convergence of Sampling Kirkwood–Buff Integrals of Aqueous Solutions with Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 1347–1355.

(50) Liu, Y.; Bolen, D. W. The peptide backbone plays a dominant role in protein stabilization by naturally occurring osmolytes. *Biochemistry* **1995**, *34*, 12884–12891.

(51) Sangster, J.; Teng, T.-T.; Lenzi, F. Molal volumes of sucrose in aqueous solutions of NaCl, KCl, or urea at 25°C. *J. Solution Chem.* **1976**, *5*, 575–585.

(52) Robinson, R. A.; Stokes, R. H. Activity coefficients in aqueous solutions of sucrose, mannitol and their mixtures at 25°C. *J. Phys. Chem.* **1961**, *65*, 1954–1958.

(53) Nozaki, Y.; Tanford, C. The solubility of amino acids and related compounds in aqueous urea solutions. *J. Biol. Chem.* **1963**, *238*, 4074–4081.

(54) Sartorio, R.; Wurzburger, S.; Guarino, G.; Borriello, G. Diffusion of polyols in aqueous solution. *J. Solution. Chem.* **1986**, *15*, 1041–1049.

(55) Ekdawi-Sever, N.; de Pablo, J. J.; Feick, E.; von Meerwall, E. Diffusion of Sucrose and α,α-Trehalose in Aqueous Solutions. *J. Phys. Chem. A* **2003**, *107*, 936–943.

(56) Bonner, O. Osmotic and activity coefficients of sodium chloride-sorbitol and potassium chloride-sorbitol solutions at 25°C. *J. Solution. Chem.* **1982**, *11*, 315–324.

(57) DiPaola, G.; Belleau, B. Polyol-water interactions. Apparent molal heat capacities andvolumes of aqueous polyol solutions. *Can. J. Chem.* **1977**, *55*, 3825–3830.

(58) Banipal, T.; Sharma, S.; Lark, B.; Banipal, P. Thermodynamic and transport properties of sorbitol and mannitol in water and in mixed aqueous solutions. *Indian J. Chem.* **1999**, *38A*, 1106–1115.

(59) Bolen, D.; Baskakov, I. V. The osmophobic effect: Natural selection of a thermodynamic force in protein folding. *J. Mol. Biol.* **2001**, *310*, 955 – 963.

(60) Loris, R.; Langhorst, U.; De Vos, S.; Decanniere, K.; Bouckaert, J.; Maes, D.; Transue, T. R.; Steyaert, J. Conserved water molecules in a large family of microbial ribonucleases. *Proteins* **1999**, *36*, 117–134.

(61) Lee, B.; Richards, F. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* **1971**, *55*, 379 – IN4.

(62) Richards, F. M. Areas, volumes, packing, and protein structure. *Annu. Rev. Biophy. Bioeng.* **1977**, *6*, 151–176.

(63) Lesser, G. J.; Rose, G. D. Hydrophobicity of amino acid subgroups in proteins. *Proteins* **1990**, *8*, 6–13.

(64) Schellman, J. A. Protein stability in mixed solvents: A balance of contact interaction and excluded volume. *Biophys. J.* **2003**, *85*, 108 – 125.

(65) Creamer, T. P.; Srinivasan, R.; Rose, G. D. Modeling unfolded states of proteins and peptides. II. Backbone solvent accessibility. *Biochemistry* **1997**, *36*, 2832–2835.

(66) Arsiccio, A.; Pisano, R. Stability of proteins in carbohydrates and other additives during freezing: The human growth hormone as a case study. *J. Phys. Chem. B* **2017**, *121*, 8652–8660.

(67) Arsiccio, A.; Pisano, R. Water entrapment and structure ordering as protection mechanisms for protein structural preservation. *J. Chem. Phys.* **2018**, *148*, 055102.

(68) Halgren, T. A.; Damm, W. Polarizable force fields. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236 – 242.

Graphical TOC Entry