

Thermodynamic Modelling and Experimental Data Reveal that Sugars Stabilize Proteins According to an Excluded Volume Mechanism

Andrea Arsiccio,[†] Tim Sarter,[‡] Ilaria Polidori,[‡] Gerhard Winter,[‡] Roberto Pisano,^{*,¶} and Joan-Emma Shea^{*,†,§}

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

[‡]*Department of Pharmacy, Ludwig-Maximilians-University, 81377 Munich, Germany*

[¶]*Molecular Engineering Laboratory, Department of Applied Science and Technology, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino 10129, Italy*

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

E-mail: roberto.pisano@polito.it; shea@ucsb.edu

Abstract

We present a new thermodynamic model to investigate the relative effects of excluded volume and soft interactions contributions in determining whether a cosolute will either destabilize or stabilize a protein in solution. This model is unique in considering an atomistically detailed model of the protein, and accounting for the preferential accumulation/exclusion of the osmolyte molecules from the protein surface. Importantly, we use molecular dynamics simulations and experiments to validate the model. The experimental approach presents a unique means of decoupling excluded volume

and soft interaction contributions using a linear polymeric series of cosolutes with differing number of glucose subunits, from 1 (glucose) to 8 (maltooctaose), as well as an 8-mer of glucose units in the closed form (γ -CD). By studying the stabilizing effect of cosolutes along this polymeric series using lysozyme as model protein, we validate the thermodynamic model and show that sugars stabilize proteins according to an excluded volume mechanism.

Introduction

The unfolding of proteins is strongly influenced by the surrounding environment, and it is well established that the addition of certain molecules can either stabilize or impair the native fold. Some of these molecules accumulate in living species in response to external stressors,^{1,2} and are generally referred to as cosolutes or osmolytes. Among them, trimethylamine N-oxide (TMAO), sucrose, betaine and proline act as stabilizers, while urea is a denaturant.

The thermodynamic mechanism underlying the role of osmolytes in protein folding has been the subject of intense investigation in the past few years, and has been explained using the concepts of preferential exclusion³⁻⁵ and free energy of transfer.⁶⁻⁸ Both these concepts rely on the observation that the unfolded state exposes more surface area than the native one, especially for the backbone and hydrophobic side chains.

According to the preferential exclusion scenario, stabilizing osmolytes are excluded from the protein surface, and this should increase the surface tension of the solvent, i.e., the work required to enlarge the surface of the peptide. In contrast, denaturing molecules, like urea, accumulate around the protein surface, shifting the equilibrium in favor of the unfolded state. Evidence of preferential exclusion, or interaction, can be obtained by measuring the preferential exclusion coefficient Γ , by means of dialysis/densimetry techniques⁹⁻¹¹ or vapor pressure osmometry.^{5,12-15} More details on the Γ coefficient will be given in the Thermodynamic Model section below.

The free energies of transfer ΔG^{tr} is equally important to consider as it represent the

energetic contributions for moving a solute from one solvent to another. It has been observed that the free energy of transfer of the unfolded state between water and an aqueous solution of a stabilizing osmolyte is larger than the corresponding quantity for the native fold.^{6,16} This means that the unfolding process is energetically unfavorable in the presence of stabilizing cosolutes, because it leads to the exposure of new groups that are less compatible with the osmolyte. The opposite is true in the case of urea; the favorable interaction between urea and the peptide backbone/sidechains stabilizes expanded structures.¹⁶

The stabilizing (or destabilizing) action of osmolytes can further be dissected into two contributions. One is related to excluded volume effects (i.e., the work for the creation of a cavity within the solvent, treated as a hard-particle fluid), and the other is associated with soft interactions (i.e., van der Waals, hydrogen-bonding, and electrostatics).^{17,18} In 2003, Schellman¹⁷ performed a thermodynamic analysis to investigate the contributions of excluded volume and soft interaction effects to protein stabilization. He studied five proteins and four cosolvents, including both denaturants and osmolytes, and came to the conclusion that excluded volume contributes to the stabilization of the native structure, while soft interactions contribute to destabilization. Here, we refine this investigation and develop a novel thermodynamic model to evaluate the two contributions to protein stabilization for different common osmolytes, including TMAO, sucrose and urea. Compared to previous models, we refine the description of the volume change upon unfolding, trying to evaluate this quantity as accurately as possible. We also take into account the presence of density gradients within the system, using the previously introduced preferential exclusion coefficient Γ .

We show that the excluded volume effect drives stabilization in the case of sucrose, while the soft interaction term also plays a major role in the case of TMAO. The opposite is true for urea, that promotes the unfolding process thanks to its favorable soft-interactions with the protein moieties. We then specifically focus our attention on sugars, and combine our thermodynamic modelling with molecular dynamics simulations and experimental data. We

select a polymeric series of cosolutes going from glucose (number of monomeric units $n = 1$) to maltooctaose and γ -cyclodextrin (γ -CD). Maltooctaose and γ -CD have the same degree of polymerization ($n = 8$), but different shape (linear vs. closed). The polymeric series allows us to use experimental data to validate the thermodynamic model herein proposed. We study the stabilizing effect of the cosolutes along the polymeric series on proteins using differential scanning fluorimetry (DSF).¹⁹ The experimental data are compared to the output of the thermodynamic model, and demonstrate that sugars stabilize proteins according to an excluded volume mechanism.

Materials and Methods

Thermodynamic Model

The transfer free energy ΔG^{tr} quantifies the energy cost, or gain, for moving a molecule from solvent A to solvent B,

$$\Delta G^{tr} = G^{solv}(B) - G^{solv}(A) \quad (1)$$

where G^{solv} is the solvation free energy.

As previously observed, the solvation energy can be viewed as the sum of an excluded volume and a 'soft interaction' contribution.^{17,18} The excluded volume term G^{ev} is associated to the creation of a cavity within the solvent, in which to place the solute. In calculating this contribution, the solvent is treated as a hard-particle fluid, with no soft interactions. The soft interactions are instead taken into account in G^{si} ,

$$G^{solv} = G^{ev} + G^{si} \quad (2)$$

Equations 1 and 2 imply that the transfer free energy can also be dissected into an excluded volume and a soft interaction term,

$$\Delta G^{tr} = \Delta G^{ev} + \Delta G^{si} \quad (3)$$

where each contribution is expressed as a difference between solvent B and solvent A ($\Delta G^{ev \text{ or } si} = G^{ev \text{ or } si}(B) - G^{ev \text{ or } si}(A)$).

It is intuitive that the energetic cost for the creation of a cavity depends on the amount of unoccupied volume available for the insertion of the solute. The denser the solvent is, the more difficult it is to find sufficient free volume. Scaled-particle theory (SPT)²⁰⁻²³ is generally used to translate these intuitive considerations into quantitative and rigorous estimates of G^{ev} .

If we assume the solute to be a sphere with radius σ_c , the work for creating the cavity is related to the probability $P(\sigma_c)$ of finding no solvent particles within the spherical region surrounding the cavity. In turn, $P(\sigma_c)$ can be expressed as the ratio between the ensemble average volume available for the solvent after insertion of the solute $\langle V_{avail}(\sigma_c) \rangle$, and the total volume V_{tot} ,²⁴

$$G^{ev} = -RT \ln P(\sigma_c) = -RT \ln \left(\frac{\langle V_{avail}(\sigma_c) \rangle}{V_{tot}} \right) = -RT \ln \left(1 - \frac{\langle V_{unavail}(\sigma_c) \rangle}{V_{tot}} \right) \quad (4)$$

where T is temperature, and R the universal gas constant. The last equality in Equation 4 makes use of the relation $\langle V_{avail}(\sigma_c) \rangle = V_{tot} - \langle V_{unavail}(\sigma_c) \rangle$, where $\langle V_{unavail}(\sigma_c) \rangle$ is the volume that becomes unavailable to the solvent particles upon insertion of the solute.

When the process of protein unfolding is considered, a quantity of interest is the difference in the free energy of transfer between the unfolded (U) and native (N) states, $\Delta G_U^{tr} - \Delta G_N^{tr}$. If this difference is positive, solvent B stabilizes the native state compared to solvent A. In this work, solvent A will consist of pure water, while solvent B will be an aqueous osmolyte solution. In this specific case, the difference $\Delta G_U^{tr} - \Delta G_N^{tr}$ is related to the m -value,

$$m = \frac{\partial \Delta G^{N \rightarrow U}(c_3)}{\partial c_3} \quad (5)$$

which describes the effect of the osmolyte on protein stability.²⁵ We assumed here the m -value to be constant, i.e., independent of c_3 . This is true if the difference in the surface area between the native and unfolded state of the protein is not dependent on c_3 , which is a common assumption:²⁶

$$\Delta G_{c_3}^{N \rightarrow U} = \Delta G_{0M}^{N \rightarrow U} + m c_3 \quad (6)$$

where c_3 is the osmolyte concentration and $\Delta G^{N \rightarrow U}$ the free energy of unfolding. According to the transfer model, schematized in Figure 1, the m -value equals the difference in the transfer free energies of U and N from pure water to a 1M osmolyte solution,²⁷

$$m = \Delta G_{1M}^{N \rightarrow U} - \Delta G_{0M}^{N \rightarrow U} = \Delta G_U^{tr,0 \rightarrow 1M} - \Delta G_N^{tr,0 \rightarrow 1M} \quad (7)$$

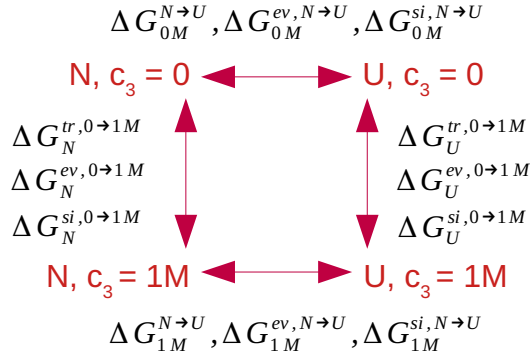


Figure 1: Schematic representation of the transfer model.

Again, the difference $\Delta G_U^{tr,0 \rightarrow 1M} - \Delta G_N^{tr,0 \rightarrow 1M}$ can be dissected into an excluded volume and a soft interaction term,

$$m = \Delta G_U^{tr,0 \rightarrow 1M} - \Delta G_N^{tr,0 \rightarrow 1M} = (\Delta G_U^{ev,0 \rightarrow 1M} - \Delta G_N^{ev,0 \rightarrow 1M}) + (\Delta G_U^{si,0 \rightarrow 1M} - \Delta G_N^{si,0 \rightarrow 1M}) \quad (8)$$

Using again the transfer model schematized in Figure 1, it is also possible to write,

$$m = \Delta G_U^{tr,0 \rightarrow 1M} - \Delta G_N^{tr,0 \rightarrow 1M} = (\Delta G_{1M}^{ev,N \rightarrow U} - \Delta G_{0M}^{ev,N \rightarrow U}) + \Delta \Delta G^{si} \quad (9)$$

where $\Delta \Delta G^{si} = \Delta G_U^{si,0 \rightarrow 1M} - \Delta G_N^{si,0 \rightarrow 1M} = \Delta G_{1M}^{si,N \rightarrow U} - \Delta G_{0M}^{si,N \rightarrow U}$.

Previous work often estimated the $\Delta G^{ev,N \rightarrow U}$ term using SPT theory. In SPT, the work of cavity formation is expressed as a third-order polynomial in the radius of the cavity σ_c ,

$$G^{ev} = \alpha + \beta \sigma_c + \gamma \sigma_c^2 + P \frac{4}{3} \pi \sigma_c^3 \quad (10)$$

where the coefficients α , β and γ depend on the size and concentration of the osmolyte, and P is pressure. If G^{ev} is estimated for both the native and unfolded states, $\Delta G^{ev,N \rightarrow U}$ can be computed. Then, knowing $\Delta G^{ev,N \rightarrow U}$ in both water and the aqueous osmolyte solution it is possible to calculate $\Delta \Delta G^{ev} = \Delta G_{1M}^{ev,N \rightarrow U} - \Delta G_{0M}^{ev,N \rightarrow U}$.

In order to do this, the radius of the native and unfolded states must be known. In previous work, the radii of proteins were estimated from their solvent accessible surface areas (SASAs).^{28,29} The SASA of the native state was obtained from the crystal structure, and this SASA was doubled to obtain an approximate estimate for the denatured form. Experimentally measured hydrodynamic radii of native and unfolded states were also used.³⁰ In other cases, the native state was modelled as a sphere and the unfolded state as a prolate spherocylinder.³¹ In this latter case, Equation 10 had to be modified to take into account the spherocylindrical shape of the cavity.^{24,32,33} This approximate description of the solute radius and shape obviously introduces an error in the computation, as the excluded volume term has been shown to strongly depend on the solute size.^{18,34,35}

Here we start from the consideration that the excluded volume is the sum of a 'target' volume and a 'gap' volume (the same terminology employed in Ref.¹⁷ is here used). The target volume is the real van der Waals volume of the solute. The gap volume is, instead, a gap space, with thickness equal to the van der Waals radius of the solvent σ , surrounding

the solute surface area (see Figure 2).

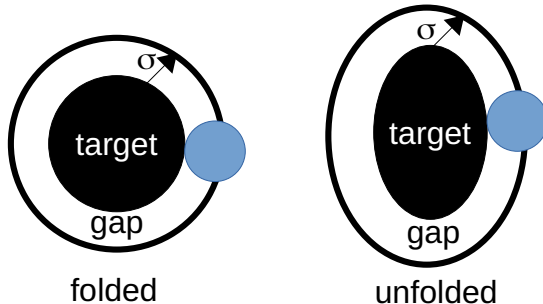


Figure 2: The excluded volume is the sum of the van der Waals volume of the target molecule, plus a gap layer with thickness equal to the solvent radius σ . The larger outer sphere is the SASA of the target molecule.

Denaturation can be viewed as a transition that modifies the SASA, but not the van der Waals volume of the peptide.^{31,36} The van der Waals volume change upon unfolding is indeed a generally small quantity, which can be both positive or negative,^{37,38} and can safely be neglected at ambient pressure. The volume change upon denaturation ΔV can hence be computed solely from the change in SASA of the protein, that is directly related to the 'gap' volume,

$$\Delta V = (V_{target,U} + V_{gap,U}) - (V_{target,N} + V_{gap,N}) \approx (V_{gap,U} - V_{gap,N}) \approx \Delta SASA \sigma \quad (11)$$

The description of the volume change upon denaturation proposed in this work, albeit still approximate, takes into account the real shape of the protein rather than making use of a simplified description as spheres, or spherocylinders, and this should help improve the reliability of the results.

The evaluation of the SASA change upon unfolding $\Delta SASA$ is affected by some uncertainty, as the unfolded state is not a single species with a defined structure, but rather an ensemble of conformations ranging from completely expanded to more collapsed chains. The uncertainty in the SASA estimation is, therefore, a reflection of the uncertainty in the definition of the unfolded state for a protein, and any other approximation for the volume

change upon unfolding would therefore suffer from the same level of unpredictability. The use of the SASA to estimate volumetric changes of the protein upon unfolding has the advantage that the SASA is easy to compute for, at least, the native state. Approximating the native fold as a sphere, as it was for instance done in previous models, would instead introduce an additional source of error, which we avoid in our calculations through the use of the SASA. Moreover, the change in SASA upon unfolding is already used in the framework of the transfer model to estimate the m -value, meaning that using the SASA to also quantify the excluded volume contribution to the m -value introduces no additional sources of uncertainty to the model.

When dealing with an unfolding process, Equation 4 could hence be rewritten as,

$$\Delta G^{ev,N \rightarrow U} = -RT \ln \left(1 - \frac{\langle V_{unavail} \rangle}{V_{tot}} \right) \approx -RT \ln \left(1 - \frac{N_{tot} \Delta SASA \sigma}{V_{tot}} \right) = -RT \ln(1 - \Delta SASA \rho \sigma) \quad (12)$$

where N_{tot} is the total number of molecules in the volume V_{tot} , and ρ the number density of the solvent.

Equation 12 is valid in the small cavity size limit. As it is done in the derivation of the SPT equation,^{21,24} it is then possible to expand the function $\ln(1 - x)$ in a MacLaurin power series, valid for $x = \Delta SASA \rho \sigma \ll 1$, truncated at the linear term,

$$\Delta G^{ev,N \rightarrow U}(\Delta SASA) \approx \Delta G^{ev,N \rightarrow U}(0) + \frac{d\Delta G^{ev,N \rightarrow U}}{dSASA}(0) \Delta SASA \approx RT \Delta SASA \rho \sigma \quad (13)$$

Equation 13 predicts that a high solvent density and large solvent radius, combined with significant conformational changes upon unfolding, increase the energy cost of cavity creation, as would be intuitively expected.

Equations 13 and 9 can finally be combined to obtain,

$$m = \Delta G_U^{tr,0 \rightarrow 1M} - \Delta G_N^{tr,0 \rightarrow 1M} = RT\Delta SASA[\rho_{1M}\sigma_{1M} - \rho_{0M}\sigma_{0M}] + \Delta\Delta G^{si} \quad (14)$$

where,

$$\Delta\Delta G^{ev} = \Delta G_{1M}^{ev,N \rightarrow U} - \Delta G_{0M}^{ev,N \rightarrow U} = RT\Delta SASA[\rho_{1M}\sigma_{1M} - \rho_{0M}\sigma_{0M}] \quad (15)$$

ρ_{0M} and σ_{0M} are the number density and molecular radius of pure water ($0.0333 \text{ \AA}^{-3}$ and $1.4 \text{ \AA}$, respectively), while ρ_{1M} and σ_{1M} are those of a 1M osmolyte solution.

The effective density and radius of the water-osmolyte solutions can be computed as weighted averages,

$$\sigma_{1M} = \left(\frac{n_1}{n_1 + n_3} \sigma_{0M} + \frac{n_3}{n_1 + n_3} \sigma_3 \right) \quad (16)$$

$$\rho_{1M} = \left(\frac{n_1}{n_1 + n_3} \rho_{0M} + \frac{n_3}{n_1 + n_3} \rho_3 \right) \quad (17)$$

Here, n_i is the number of molecules of type i in the region close to the protein surface where the cosolutes concentration is perturbed by the presence of the protein, the subscripts 1 and 3 refer to water and the osmolyte, respectively, and σ_3 , ρ_3 are the molecular radius and number density of the osmolyte.

While representing a simplification, the weighted averages in Equations 16 and 17 still allow us to take into account the effect of concentration gradients within the solution. Indeed, it is to be noted that n_1 and n_3 do not represent the total number of water and osmolyte molecules in the system. In the approach herein proposed, they represent local values, corresponding to the first solvation shell of the protein (here defined as a shell around the protein surface having width equal to the effective diameter of the water-osmolyte solution $2\sigma_{1M}$),

$$v_3 n_3 + v_1 n_1 = 2SASA\sigma_{1M} \quad (18)$$

where v_3 and v_1 are the molecular volumes of the osmolyte and water, respectively.

The choice of focusing on the first solvation shell has been made because it is well known that osmolyte molecules do not have a uniform distribution in a protein solution. They tend to either accumulate, or be excluded from the peptide surface, and this also dictates their stabilizing or denaturing action according to the preferential hydration theory.^{39,40} Specifically, stabilizing osmolytes are excluded from the protein domain, as such promoting conformations with reduced SASA, like the native fold. The opposite is true for denaturing species, like urea, that preferentially solvate the peptide, displacing water molecules. This should be taken into account for a more accurate estimation of the excluded volume contribution upon unfolding. The treatment of density inhomogeneities proposed in this work represents, therefore, an improvement compared to previous models, that neglected any concentration gradient in protein-osmolyte solutions.

Once the total number of water and osmolyte molecules in the system (N_1 and N_3) are known, it is possible to compute n_1 and n_3 thanks to the relation between the free energy of transfer and the preferential exclusion coefficient Γ ,⁴¹

$$\Delta G^{tr} = -\frac{\partial \mu_3}{\partial \ln c_3} \Gamma \quad (19)$$

where $\partial \mu_3 / \partial \ln c_3$ is the derivative of the osmolyte chemical potential with respect to the osmolyte concentration.

Γ then relates N_1 and N_3 to n_1 and n_3 ,

$$n_3 = \Gamma + \frac{N_3}{N_1} n_1 \quad (20)$$

The value of the preferential exclusion coefficient Γ changes between the folded and unfolded states, because the distribution of the osmolytes within the solution depends on the

solvent accessible surface area exposed by the protein. In this work, for the sake of simplicity, an average value of Γ between the folded and unfolded extremes was employed for the calculations. This is an approximation that simplifies our model, but still allows us to consider an inhomogeneous distribution of the osmolytes within the solution, therefore representing an improvement compared to previous approaches, where preferential interaction effects were totally neglected.

Calculation Procedure

Here we address in more detail the procedure used for the calculation of the relevant quantities, and of the final output, of the thermodynamic model outlined above. The input data for the calculation is a Protein Data Bank (pdb)⁴² file of the native state of the protein. Then, the computation proceeds as follows:

1. the SASA of the native state is computed from the input pdb file using the algorithm by Lee and Richards,^{43,44} with a probe size equal to 1.4 Å. The solvent accessible surface area of the denatured state is instead obtained using the method introduced by Creamer et al.⁴⁵ Creamer et al. defined two models, a lower and an upper one, that bracket the SASA of the unfolded state, and correspond to compact and expanded conformations, respectively. Both models are used in the calculation, together with an interpolation between the two limiting extremes, as proposed by Schellman.¹⁷ Knowledge of both the native and unfolded SASAs allows the computation of the $\Delta SASA$ upon unfolding;
2. the m -value is then computed, using the additive approach proposed in Refs.^{6,41,46} More specifically, the difference in transfer free energies between the denatured and native state of a polypeptide containing n residues is obtained as a summation over the side chains ($\Delta g_{sc}^{tr,j}$) and backbone (Δg_{bb}^{tr}) contributions,

$$\Delta G_U^{tr,0 \rightarrow 1M} - \Delta G_N^{tr,0 \rightarrow 1M} = \sum_{j=1}^n \Delta g_{sc}^{tr,j} \Delta \alpha_{sc}^j + \Delta g_{bb}^{tr} \sum_{j=1}^n \Delta \alpha_{bb}^j \quad (21)$$

Each contribution is weighed by the average fractional change in solvent accessible surface area SASA of residue j in going from the native N to the denatured U state,

$$\Delta\alpha^j = \frac{SASA_U^j - SASA_N^j}{SASA_{Gly-j-Gly}^j} \quad (22)$$

where $SASA_{Gly-j-Gly}^j$ is the solvent accessibility of amino acid j in the tripeptide $Gly-j-Gly$. The $\Delta g_{sc}^{tr,j}$ and Δg_{bb}^{tr} values (to be inserted into Eq. 21) for the different osmolytes considered in this work (urea, sucrose and TMAO) were taken from previous experimental works;^{6,8,41,47-50}

3. a value of Γ is computed, using Equation 19. The value of ΔG^{tr} used for this purpose is an average between the native $\Delta G_N^{tr,0 \rightarrow 1M}$ and unfolded $\Delta G_U^{tr,0 \rightarrow 1M}$ extremes. The values of $\partial\mu_3/\partial \ln c_3$ for each osmolyte are taken from Table I of Ref.⁴¹ ;
4. n_1 and n_3 are then computed by solving Equations 18 and 20. The value of σ_{1M} in Equation 18 is calculated through Equation 16. Also in this case, the SASA to be inserted into Equation 18 is an average between those of the native and unfolded states. The molecular volumes v_3 and v_1 , that appear in Equation 18, are obtained from Table I of Ref.⁴¹ The osmolytes radii σ_3 , necessary to compute σ_{1M} , are the same already used in previous work: 4.05 Å for sucrose, 2.32 Å for urea and 2.7 Å for TMAO.³¹ The N_3/N_1 values to be inserted into Equation 20 are derived from the experimental values of c_1 (molar concentration of water in the osmolytes solution) listed in Table I of reference,⁴¹ as: $N_3/N_1 = c_3/c_1 = 1M/c_1$;
5. the values of σ_{1M} and ρ_{1M} can be computed once n_1 and n_3 are known. Values of ρ_3 are computed as the inverse of the experimental v_3 listed in Table I of Ref.⁴¹ The excluded volume contribution is then obtained as detailed in Equation 15;
6. finally, knowing $m = \Delta G_U^{tr,0 \rightarrow 1M} - \Delta G_N^{tr,0 \rightarrow 1M}$ and $\Delta\Delta G^{ev} = \Delta G_{1M}^{ev,N \rightarrow U} - \Delta G_{0M}^{ev,N \rightarrow U}$, the soft-interaction contribution $\Delta\Delta G^{si}$ is extracted from Equation 9.

Molecular Dynamics Simulations

Hen egg white lysozyme (PDB 1LKS⁵¹) was simulated at 298 K and pH 3 in presence of a 150 mM concentration of the following cosolutes (Figure 3): γ -cyclodextrin (γ -CD), maltooctaose, maltoheptaose, maltohexaose, maltopentaose, maltotetraose, maltotriose, maltose and glucose. They represent a polymeric series, with glucose as the constituent subunit. Maltooctaose and γ -CD have the same degree of polymerization ($n = 8$), but different shape (linear vs. closed).

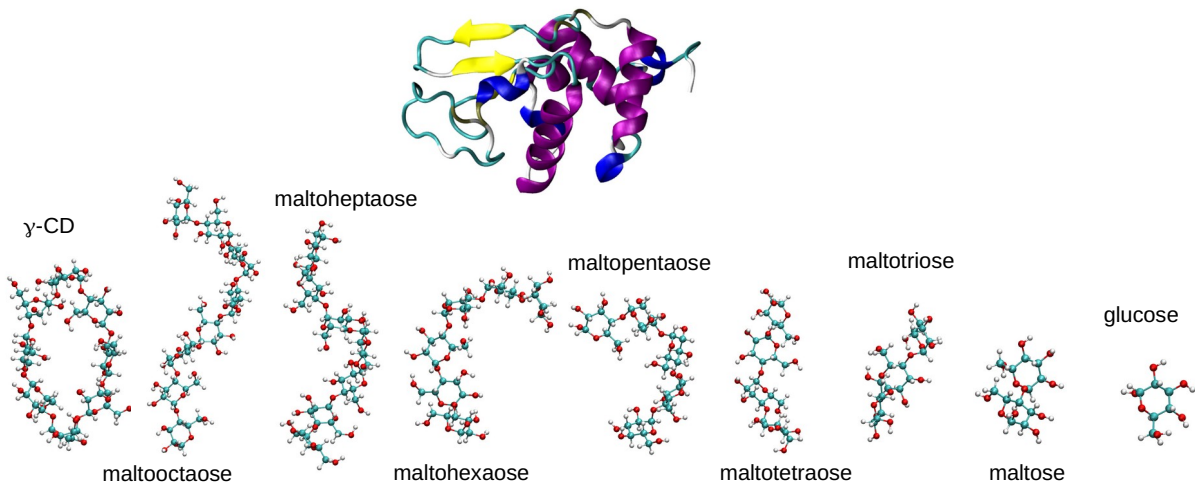


Figure 3: Cartoon representation of lysozyme, and ball-and-stick representation of the polymeric series of glucose (from glucose to maltooctaose and γ -CD) employed in this work.

The protonation state of lysozyme was adjusted using the H++ server (<http://newbiophysics.cs.vt.edu/H++/>⁵²), and simulations were performed starting from the native configuration.

One protein molecule was introduced into a cubic simulation box, with 7.64 nm side length, and the overall charge was neutralized using Cl^- molecules. Protein-free systems were also simulated (i.e., only the selected cosolute along the polymeric series was introduced within the simulation box at 150 mM concentration, and solvated in water), and in this case the box size was set to 8 x 8 x 8 nm.

Each system was first energy minimized using the steepest descent algorithm, and then equilibrated for 1 ns at 1 bar and 298 K, using Berendsen pressure (3 ps relaxation time) and

temperature (0.5 ps relaxation time) coupling.⁵³ The production run was then performed at 1 bar and 298 K for 100 ns using the Nosé-Hoover thermostat^{54,55} (0.5 ps relaxation time) and Parrinello-Rahman barostat⁵⁶ (3 ps relaxation time). For analyses, the first 50 ns were deemed as an equilibration, and discarded from the computation of equilibrium values.

In all simulations, the proteins were described with the CHARMM36m force field,⁵⁷ while the CHARMM TIP3P model was used for water.⁵⁸ The ADD force field was used for the sugar molecules.⁷

For all simulations, Gromacs 2018⁵⁹ was used. Periodic boundary conditions were used, and the cut-off radius for both Coulombic (calculated using the PME method⁶⁰) and Lennard-Jones interactions was set to 1.2 nm. A 2 fs timestep was used, and the Lincs algorithm was employed for constraining all bonds,⁶¹ while the SETTLE algorithm kept the water molecules rigid.⁶²

Experimental Procedure

Materials

For thermal unfolding studies, hen egg white lysozyme (Sigma Aldrich, Steinheim, Germany) in 10 mM citrate buffer pH 3.0 in the presence of different sugars was used. Buffer ingredients were dissolved in highly purified water from a Sartorius Arium[®] pro system (Sartorius Corporate Administration GmbH, Göttingen, Germany) and pH was adjusted either with hydrochloric acid (VWR, Darmstadt, Germany) or sodium hydroxide (Bernd Kraft GmbH, Duisburg, Germany). Hen egg white lysozyme as lyophilized powder was added and the formulations were sterile filtered with 0.20 μ m polyvinylidene fluoride filters (VWR International, Darmstadt, Germany). The protein concentration was measured spectrophotometrically using a NanoDrop[™]One (Thermo Fisher Scientific, Madison, USA) and an absorbance of 26.4 for a 1% solution measured in a 1 cm pathlength.^{63,64} The pH after preparation was ± 0.1 of the target value. Chemicals were obtained as follows: citric acid, D-glucose monohydrate, D-maltose monohydrate and maltotriose hydrate from Sigma

Aldrich (Steinheim, Germany), maltohexaose and maltooctaose from CycloLab (Budapest, Hungary), and γ -cyclodextrin from Wacker Chemie (Munich, Germany).

Differential Scanning Fluorimetry (nanoDSF)

The thermal unfolding of hen egg white lysozyme in the presence of different sugars in varying concentrations was studied by nanoDSF. Standard nanoDSF grade capillaries were filled with solutions containing 1 g/L of protein. All measurements were performed in triplicates. Using the Prometheus NT.48 (NanoTemper Technologies, Munich, Germany) system, a temperature ramp of 1 °C/min from 20 to 90 °C was applied and the intrinsic protein fluorescence intensity at 330 nm and 350 nm after excitation at 280 nm was measured. The change in the fluorescence intensity ratio (F_{350}/F_{330}) was plotted and the apparent protein melting temperatures (T_m) were determined from the maximum of the first derivative of each curve.

Dynamic Light Scattering (DLS)

Solutions with a sugar concentration of 50 mM were prepared and centrifuged for 10 min at 10000 rpm. 50 μ L of the solutions were pipetted in triplicates into a 384 well plate (Corning, Kennebunk, USA), and afterwards the plate was centrifuged at 2000 rpm for 2 minutes using a Heraeus Megafuge 40 centrifuge equipped with an M-20 well plate rotor (Thermo Fisher Scientific, Wilmington, USA). Each well was sealed with 10 μ L of silicon oil and centrifuged again at 2000 rpm for 2 minutes. The well plate was placed in a DynaPro® DLS plate reader III (Wyatt Technology Europe, Dernbach, Germany) and 10 acquisitions per well with an acquisition time of 5 seconds at a temperature of 25 °C were collected. The apparent hydrodynamic radii were calculated by the Dynamics V7.10 software (Wyatt Technology, Santa Barbara, USA) using a viscosity of 0.89 mPa·s.

Results and Discussion

The Excluded Volume Term Dominates for Sucrose, while Urea and TMAO Influence Protein Behavior Through Soft Interactions

Three common osmolytes were studied in the present work, at a reference concentration of 1 M. Two of them are stabilizing (sucrose and TMAO), while one is a denaturant (urea). The m -values, excluded volume ($\Delta\Delta G^{ev}$) and soft interaction ($\Delta\Delta G^{si}$) contributions were calculated, for the case of five well-studied proteins: ribonuclease-T (RNT, pdb 9RNT⁶⁵), ribonuclease-A (RN, pdb 1RND⁶⁶), hen egg white lysozyme (LZ, pdb 1LKS⁵¹), staphylococcus nuclease (SN, pdb 1SNO⁶⁷), and T4 lysozyme (T4L, pdb 2LZM⁶⁸). This same set of proteins was already used in previous work.¹⁷

The results of this analysis are shown in Figure 4. The sign convention used in the present work implies that positive values of m , $\Delta\Delta G^{ev}$ and $\Delta\Delta G^{si}$ indicates stabilization of the native fold, while negative values are associated with increased tendency of unfolding.

As mentioned in the Thermodynamic Model section, three models of the denatured state were used,^{17,45} corresponding to expanded, compact or intermediate conformations. The bars shown in Figure 4 correspond to the intermediate model of the unfolded state. In contrast, the positive and negative error bars bracket the results obtained using the upper or lower models, respectively.

The following general observations can be drawn from the results displayed in Figure 4. First, the excluded volume contribution is always positive, as such stabilizing the native fold. Unfolding results in an increased SASA, and the creation of a cavity to accommodate this expanded structure is more difficult in presence of osmolytes.

In the case of sucrose, the excluded volume contribution is very positive, and dominates over the negative soft interaction term. The large size of sucrose molecules is, therefore, what mostly drives their stabilizing effect.

The opposite is true for urea. This denaturing osmolyte is small, hence resulting in a

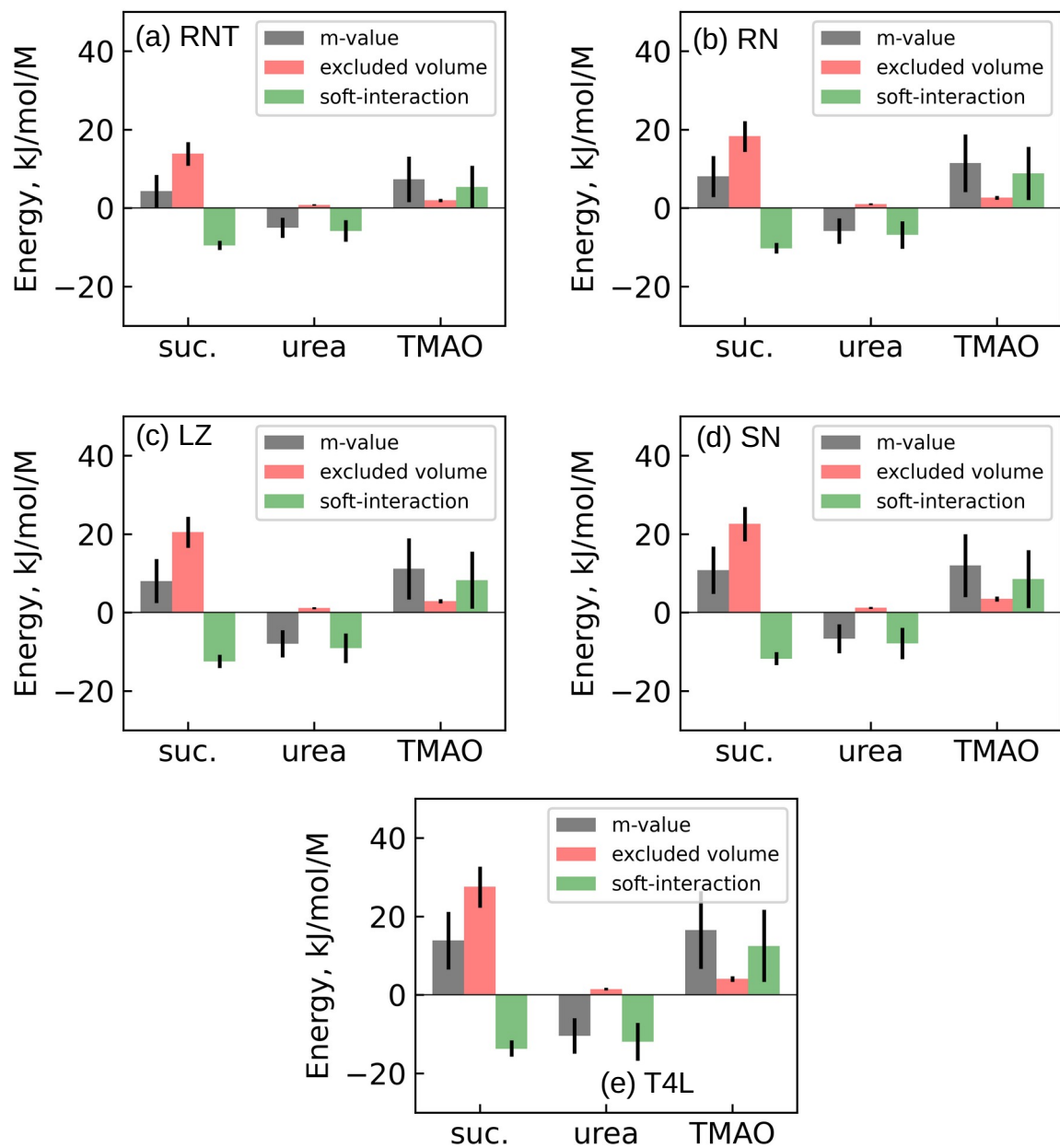


Figure 4: m -values, excluded volume ($\Delta\Delta G^{ev}$) and soft interaction ($\Delta\Delta G^{si}$) contributions to the unfolding of five well-known proteins: ribonuclease-T (RNT), ribonuclease-A (RN), hen egg white lysozyme (LZ), staphylococcus nuclease (SN), and T4 lysozyme (T4L). A positive value indicates stabilization of the native fold, and vice versa.

small excluded volume contribution, which is offset by a negative (i.e., destabilizing) soft interaction term.

The results obtained so far agree, at least qualitatively, with previous results by Schellman.¹⁷ A discrepancy between our results and Schellman’s calculations is instead observed for the case of TMAO. Schellman found that TMAO had a destabilizing soft interaction contribution, which was, however, compensated for by the excluded volume term, leading to overall stabilization. This is what occurs in our study for sucrose, but not for TMAO. Figure 4 shows that both the excluded volume and soft interaction terms are positive (i.e., stabilizing) for TMAO according to our calculations, and the sum of the two results in a pronounced stabilizing effect. Our results therefore suggest a different mechanism of protein stabilization by TMAO, compared to other common osmolytes such as sucrose. It is important to note that Schellman used a larger value of TMAO radius in his calculations (2.95 Å) compared to our work (2.7 Å). However, we repeated our study also with a larger value $\sigma_3 = 2.95$ Å for TMAO (data not shown), and found that the soft-interaction contribution decreased, but still remained positive (i.e., favorable).

Effect of Protein SASA, Preferential Exclusion and Osmolyte Concentration on the Excluded Volume and Soft Interaction Terms

The model proteins used in this work are characterized by different SASAs, ranging from 5413.1 Å² for RNT to 8573.0 Å² for T4L (in their native states). It is interesting to see how the energetic terms involved in the protein-osmolyte interaction change with the SASA of the protein considered, and such a plot is displayed in Figure 5.

The m -value increases with SASA for stabilizing osmolytes, and decreases for urea, but the trend is not perfectly linear, as evident from the values of the coefficient of determination R^2 shown in Figure 5a. On the contrary, the excluded volume term increases linearly with the protein SASA for all osmolytes, as would be expected considering its strong relation with the system geometry. The R^2 values in Figure 5b are all larger than 0.90, and the figure also

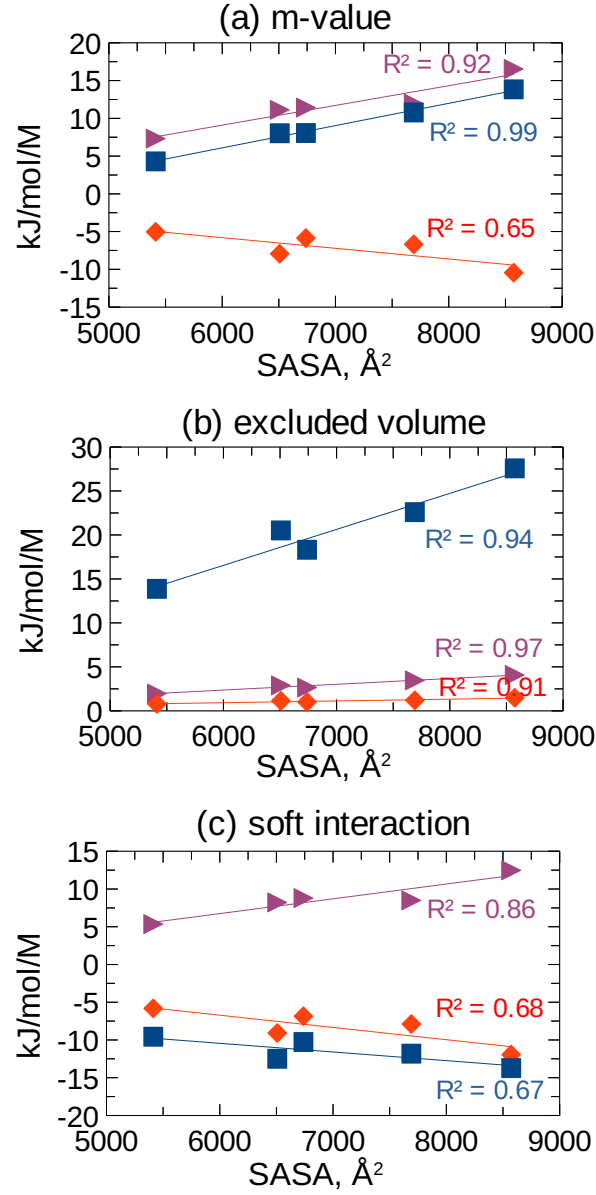


Figure 5: Evolution of the (a) m -value, (b) excluded volume ($\Delta\Delta G^{ev}$), and (c) soft interaction ($\Delta\Delta G^{si}$) contributions as function of the protein SASA (as measured for the native fold). Blue: sucrose, red: urea, purple: TMAO.

clearly shows the net separation between sucrose (blue line), whose energetics is dominated by the excluded volume term, and the other smaller osmolytes. It is also interesting to observe that the slope of the excluded volume vs. SASA curves is directly related to the molecular size, with urea having the smallest slope and sucrose the highest.

Finally, the soft interaction term (Figure 5c) is only poorly related with SASA. TMAO (purple curve) is the only osmolyte whose soft interaction term positively contributes to stabilization, and the degree of stabilization increases with the protein SASA in this case ($R^2 = 0.86$). The $\Delta\Delta G^{si}$ term of sucrose and urea becomes more negative (i.e., more destabilizing) when increasing the SASA.

The approach for the dissection of the excluded volume and soft interaction contributions to protein stabilization proposed in the present work takes into account the presence of concentration gradients around the protein surface. This result is achieved using the preferential exclusion coefficient Γ , that quantifies the accumulation ($\Gamma > 0$) or depletion ($\Gamma < 0$) of osmolytes around the protein surface, and can be computed from the free energies of transfer (Equation 19). Figure 6 shows the effect of Γ on the excluded volume term ($\Delta\Delta G^{ev}$), for the specific case of SN as model protein.

Neglecting the presence of concentration gradients within the solution ($\Gamma = 0$, Figure 6b) is equal to assuming that $\sigma_{1M} = [(N_1\sigma_{0M})/(N_1 + N_3) + (N_3\sigma_3)/(N_1 + N_3)]$ and $\rho_{1M} = [(N_1\rho_{0M})/(N_1 + N_3) + (N_3\rho_3)/(N_1 + N_3)]$. If we compare this situation to the 'real' Γ case (Figure 6a, where Γ has been computed using Equation 19), we observe that the excluded volume term of stabilizing osmolytes is generally overestimated ($\Delta\Delta G^{ev} = 24.80$ or 3.81 kJ/mol/M for sucrose and TMAO, respectively, at $\Gamma = 0$, compared to 22.60 or 3.46 kJ/mol/M when concentration gradients are considered). In contrast, the $\Delta\Delta G^{ev}$ term is underestimated for urea (1.08 kJ/mol/M at $\Gamma = 0$ against 1.20 kJ/mol/M for 'real' values of Γ).

The extreme cases of huge accumulation ($\Gamma = 10$, Figure 6c) and depletion ($\Gamma = -10$, Figure 6d) have also been considered. For very positive values of Γ the solution density

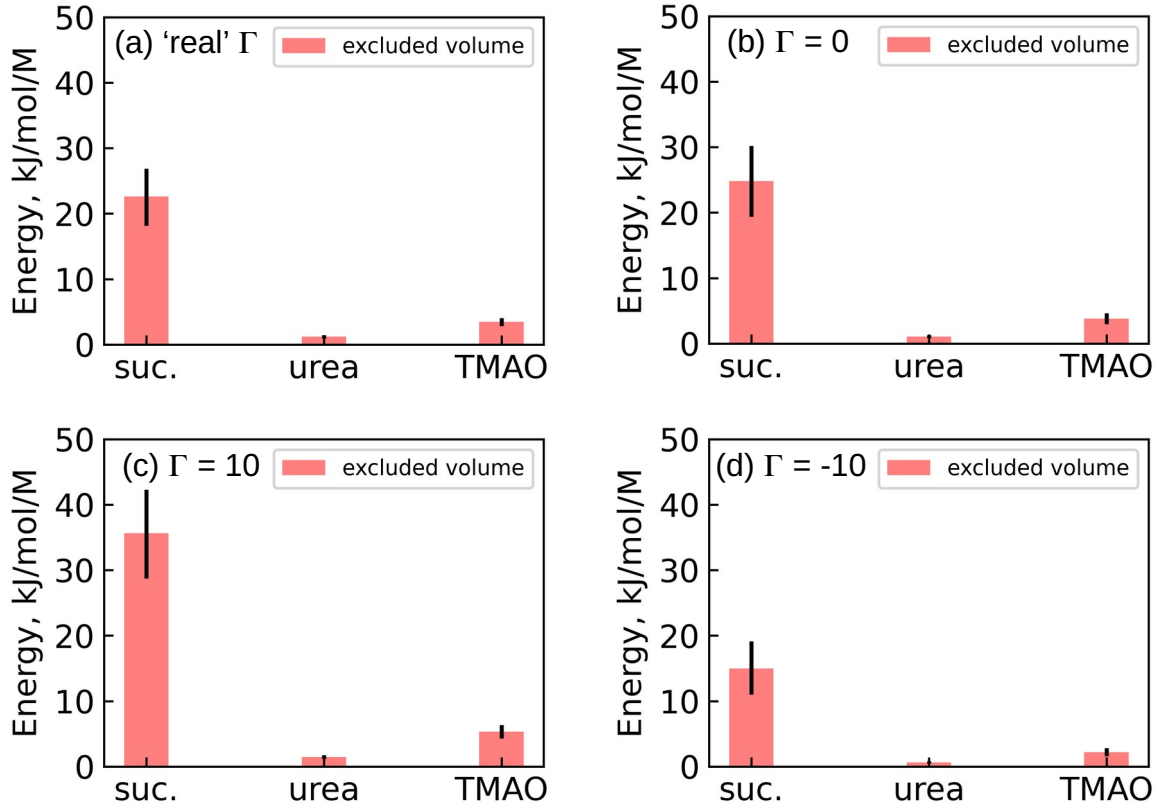


Figure 6: Effect of the preferential exclusion coefficient Γ on the excluded volume contribution ($\Delta\Delta G^{ev}$), for the specific case of staphylococcus nuclease. (a) 'real' Γ as computed from Equation 19, (b) $\Gamma = 0$, (c) $\Gamma = 10$, (d) $\Gamma = -10$.

close to the protein is higher compared to pure water, and the excluded volume contribution is consequently increased. The opposite occurs in the case of a very negative Γ . The protein is mostly surrounded by water in this case, and the excluded volume contribution to stabilization is hence reduced.

Finally, the effect of the osmolyte concentration was also addressed, for the specific case of sucrose and urea. Sucrose was studied at both 0.5 M and 1 M concentration, and 1 M and 2 M urea solutions were also investigated. The same model proteins previously investigated were considered for this analysis, as shown in Figure 7.

As expected, the excluded volume contribution becomes more positive when increasing the concentration, while the soft interaction term becomes more negative (i.e., more destabilizing) for both sucrose and urea. However, the increase in the excluded volume contribution prevails for sucrose, leading to an increase in the overall m -value with concentration, while the opposite is true for urea.

As outlined in the Calculation Procedure section, an additive approach was employed for the computation of m -values, using experimental values of transfer free energy contributions. For urea, two different sets of transfer free energy contributions exist,⁶⁹ as proposed by either Auton, Holthauzen and Bolen¹⁶ (urea - B) or by Moeser and Horinek⁸ (urea - H). Auton, Holthauzen and Bolen¹⁶ took the activity coefficient of glycine in water and urea into account when deriving their transfer free energies for the side chains. However, they made a mistake in the conversion of activity coefficient data between concentration scales. Moeser and Horinek⁸ corrected for this mistake in their new set of transfer free energy values. The Bolen set predicts that urea interactions with the protein backbone contribute the most to unfolding, while the Horinek set gives equal importance to both backbone and sidechains interactions. Despite this difference, the m -value predictions for the two sets of transfer free energies remain similar because of compensation effects, and the two sets lead to almost equal results in the framework of the thermodynamic model herein proposed (Figure S1).

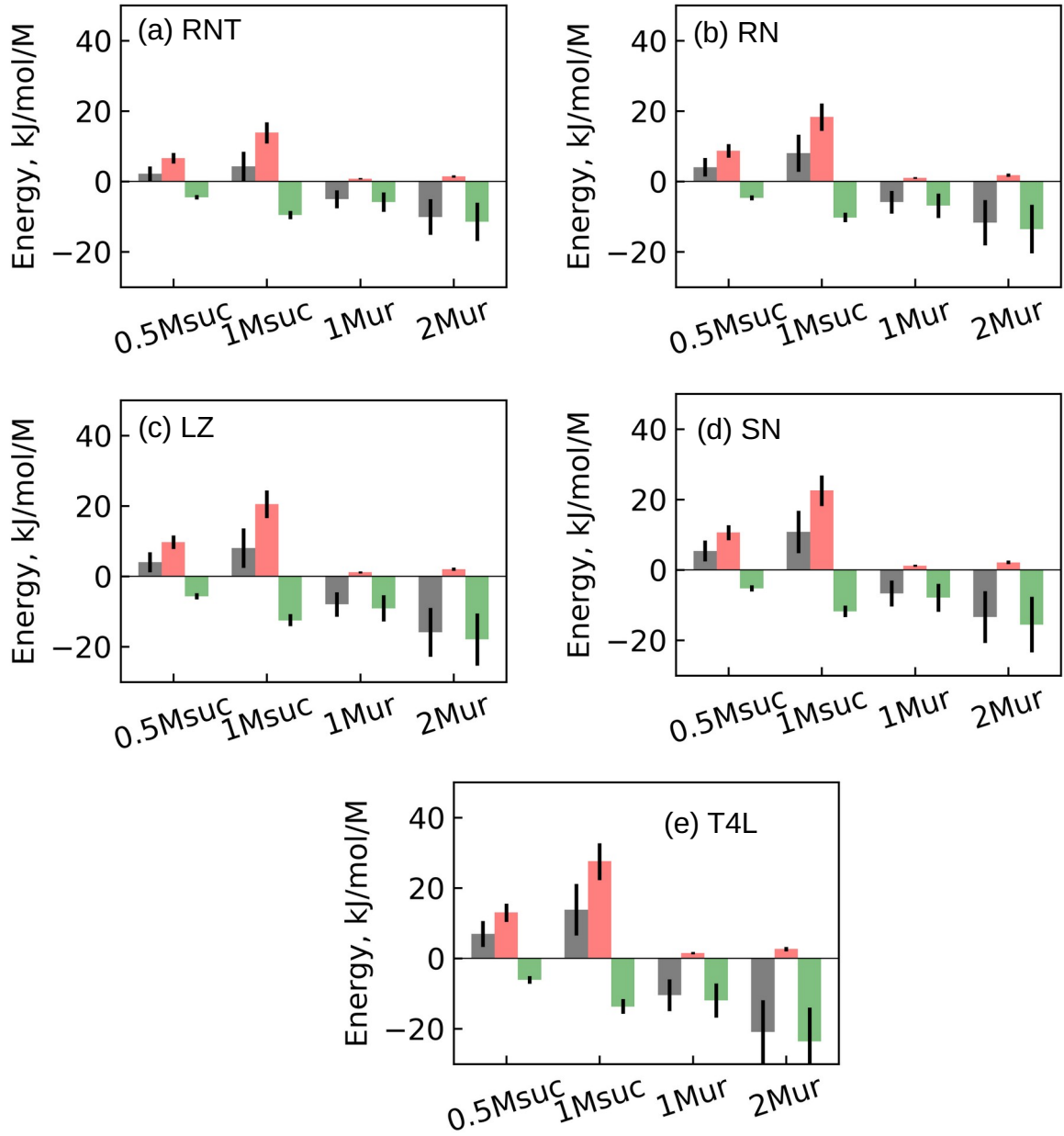


Figure 7: Effect of sucrose and urea concentration on protein stability. Grey bars: total energetic contribution $\Delta\Delta G^{N \rightarrow U} = \Delta G_{c_3}^{N \rightarrow U} - \Delta G_{0M}^{N \rightarrow U}$, orange bars: excluded volume contribution ($\Delta\Delta G^{ev}$), green bars: soft interaction contribution ($\Delta\Delta G^{si}$). A positive value indicates stabilization of the native fold, and vice versa.

Sugars Stabilize Proteins According to An Excluded Volume Mechanism

We have so far discussed the predictions of a thermodynamic model that aims at dissecting the contributions of excluded volume and soft interactions to protein stabilization by osmolytes. Validating such predictions is not straightforward, as experimental approaches can provide m -values, but cannot distinguish between the excluded volume and the soft interaction contributions.

We here propose an experimental approach to validate the results of the thermodynamic model. This experimental approach makes use of polymeric cosolutes containing a different number n of glucose monomers: γ -cyclodextrin (γ -CD, $n = 8$), maltooctaoase ($n = 8$), maltoheptaoase ($n = 7$), maltohexaoase ($n = 6$), maltopentaoase ($n = 5$), maltotetraose ($n = 4$), maltotriose ($n = 3$), maltose ($n = 2$) and glucose ($n = 1$). They represent a polymeric series, with glucose as the constituent subunit.

For this system, we assume that the soft interaction term scales with the number of monomeric (glucose) units, while the excluded volume contribution should depend on the size of the polymeric molecule. The size of a linear polymer containing n subunits should be, on average, larger than that of a closed polymer having the same number of subunits. For this reason, the excluded volume contribution should be larger for the linear polymer maltooctaoase ($n = 8$), than for its closed counterpart γ -CD. In contrast, the soft interaction term should be unaffected by the (linear or closed) shape. For this reason, the overall stabilizing effect of a n -mer should be lower if the n -mer is closed rather than linear (Figure 8). We are implicitly assuming here that the circular shape of γ -CD does not affect the soft interactions term, i.e., we are hypothesizing that changes in both the protein and cyclodextrin shape upon interaction tend to maximize the soft interactions between the glucose units of γ -CD and the protein. Because of these rearrangements in shape, the cyclodextrin does not interact with the protein only on one side (the side of its first contact with the protein) but with all of its glucose subunits. If this is true, the difference in stabilizing effect between the

closed and linear form could be attributed to the excluded volume term only, providing a way to experimentally decouple the two contributions involved in protein stabilization.

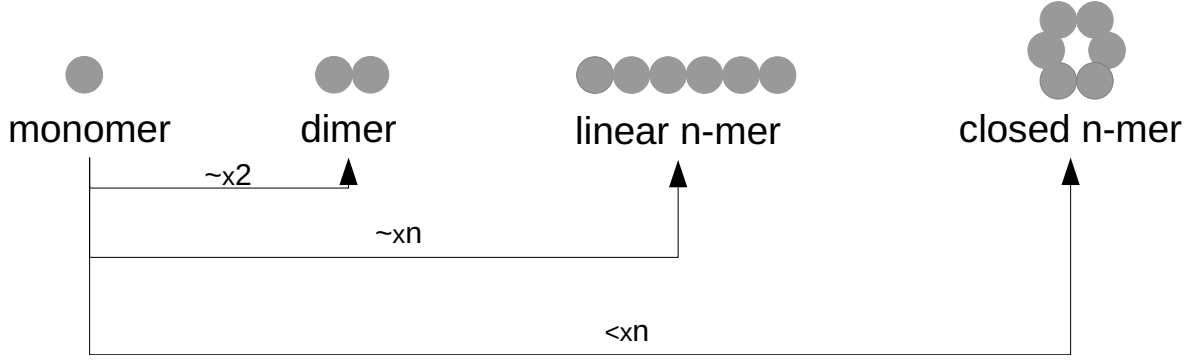


Figure 8: If the cosolute concentration is kept constant, a linear dimer should have a stabilizing effect which is about twice that of the monomer. Similarly, a linear n -mer should stabilize the protein about n times more than the monomer. In contrast, the excluded volume contribution of a closed n -mer should be reduced, leading to an overall level of stabilization that is $< n$ times that of the monomer.

With this in mind, we performed differential scanning fluorimetry (DSF) experiments of lysozyme. We collected thermal unfolding profiles of lysozyme in presence of 0 mM, 50 mM, 100 mM and 150 mM concentration of different cosolutes along the polymeric series (glucose, maltose, maltotriose, maltohexaose, maltooctaose and γ -CD). We fitted the thermal unfolding profiles to a two-state model,

$$\Delta G_{c_3}^{N \rightarrow U} = [(1 - T/T_m)\Delta H(T_m)] + (T - T_m)\Delta C_p - \ln(T/T_m)T\Delta C_p \quad (23)$$

where T_m is the melting temperature and $\Delta H(T_m)$ and ΔC_p are the enthalpy change and change in heat capacity upon unfolding. We then extracted the free energy change upon unfolding at a reference temperature (72.5 °C, which is close to the melting temperature of lysozyme in water) from each unfolding profile, and fitted the obtain values as function of concentration to obtain the m -value for each cosolute according to,

$$\Delta G_{c_3}^{N \rightarrow U}(72.5^\circ C) = \Delta G_{0M}^{N \rightarrow U}(72.5^\circ C) + mc_3 \quad (24)$$

The results of this analysis are displayed in Figure 9. The melting temperature T_m increased with concentration for glucose, maltose, maltotriose, maltohexaose and maltooctaose, but it decreased when γ -CD was added to the lysozyme solution (Figure 9a). The same trend was noted for the free energy change upon unfolding at 72.5 °C (Figure 9b). The unfolding free energy was linear with concentration, in line with Equation 24, as shown by the high values of the coefficient of determination R^2 for the linear fitting. The slopes of such linear fittings correspond to the m -values, which are plotted in Figure 9c as function of the number of glucose units. The m -value linearly ($R^2 = 0.97$) increases with the number of monomeric units from glucose to maltooctaose. However, the m -value is negative (indicating a denaturing behavior) for the closed 8-mer γ -CD.

Despite having the same number of glucose subunits, maltooctaose and γ -CD behave very differently, with maltooctaose being a potent stabilizer for lysozyme, and γ -CD acting instead as a denaturant. This behavior can be rationalized according to the theory proposed in Figure 8: the excluded volume contribution of a closed n -mer (such as γ -CD) should be reduced, leading to an overall level of stabilization that is lower compared to its open counterpart (such as maltooctaose).

This experimental result qualitatively confirms our model prediction that sugars stabilize proteins according to an excluded volume effect, while the soft interaction contribution is instead destabilizing. However, we set out to perform a more quantitative validation of the thermodynamic model herein proposed. For this purpose, we employed the thermodynamic model to predict the energetics contributions involved in the stabilization of lysozyme by glucose, maltose, maltotriose, maltotetraose, maltopentaose, maltohexaose, maltoheptaose, maltooctaose and γ -CD. For this purpose, we applied the following calculation procedure:

1. we computed the SASA of the native state of lysozyme from the input pdb file 1LKS using the algorithm by Lee and Richards,^{43,44} with a probe size equal to 1.4 Å. The solvent accessible surface area of the denatured state was instead obtained using the

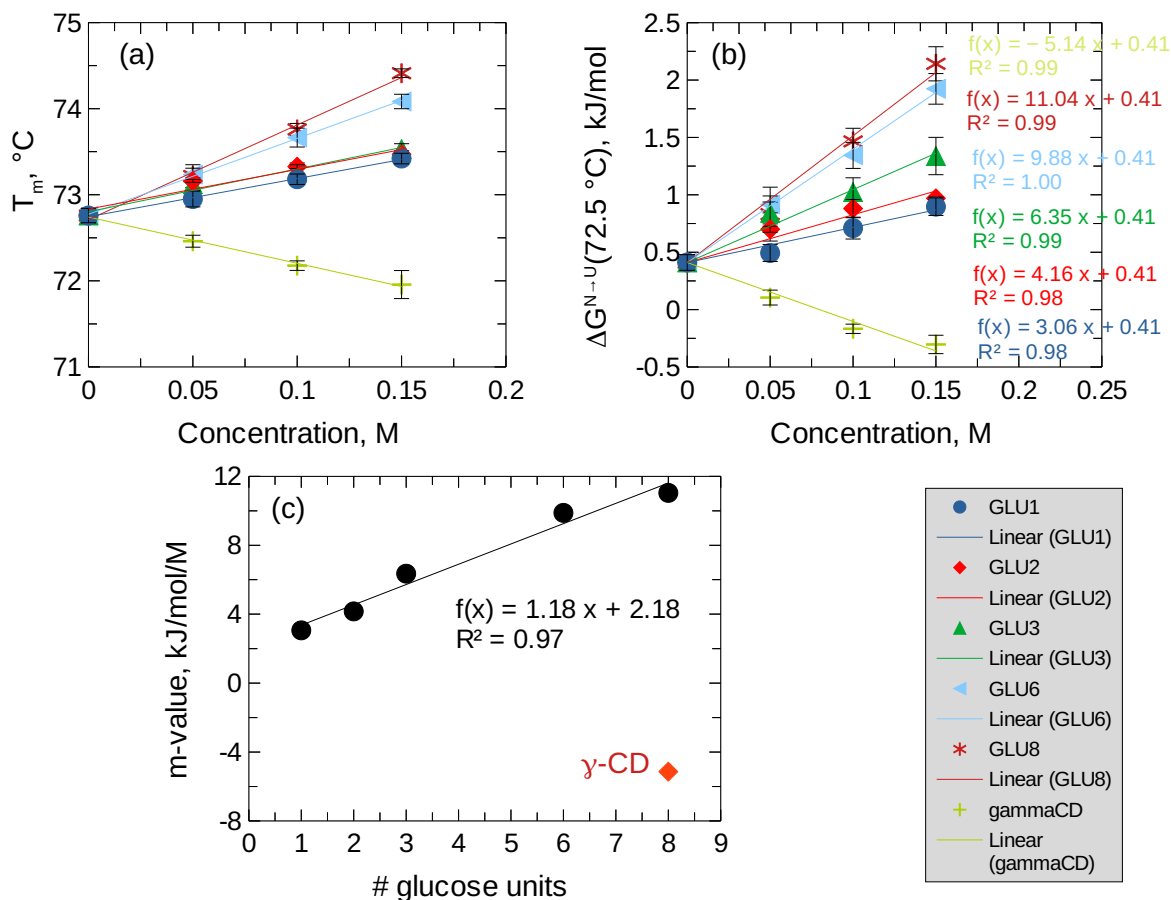


Figure 9: (a) Melting temperature T_m for lysozyme as function of concentration of glucose (GLU1), maltose (GLU2), maltotriose (GLU3), maltohexaose (GLU6), maltooctaose (GLU8) and γ -CD. (b) Free energy change upon unfolding at 72.5 °C for lysozyme as function of cosolute concentration. (c) m -value as function of the number of monomeric (glucose) units.

method introduced by Creamer et al.⁴⁵ Knowledge of both the native and unfolded SASAs allows the computation of the $\Delta SASA$ upon unfolding;

2. we extracted the radius (σ_3) of the cosolutes (glucose, maltose, maltotriose, maltotetraose, maltopentaose, maltohexaose, maltoheptaose, maltooctaose and γ -CD) from protein-free molecular dynamics simulations (i.e., simulations of the cosolutes in pure water). The radii of gyration extracted from such simulations are displayed in Figure S2a. The radius of gyration increases linearly with the number of glucose units from glucose to maltooctaose, but is lower for γ -CD than for maltooctaose due to the closed shape of the cyclodextrin. The radii of gyration predicted by simulations (shown as squares in Figure S2a) display the same trend of the hydrodynamic radii extracted from experimental DLS measurements (triangles in Figure S2a), with the only exception of γ -CD. This discrepancy between simulations and experiments for the cyclodextrin may be due to the tendency of γ -CD to form clusters in solutions, which leads to a higher average hydrodynamic radius measured experimentally;
3. we extracted the $n_3/(n_1 + n_3)$ values to be inserted into Equations 17 and 16 from molecular dynamics simulations. For this purpose, the preferential exclusion coefficient Γ was extracted from the molecular dynamics simulations as,

$$\Gamma(r) = n_3(r) - \frac{N_3}{N_1}n_1(r) \quad (25)$$

where $n_i(r)$ is the number of molecules of type i that are within a distance r from the protein surface, and N_i is the total number of molecules i in the simulation box. The evolution of Γ as function of distance from the protein surface is displayed in Figure S2b. The ratio $n_3/(n_1 + n_3)$ was then evaluated for each cosolute at the distance r at which Γ displays a minimum;

4. we set the m -value for glucose to be equal to the measured experimental value shown

in Figure 9 (3.06 kJ/mol/M). Knowing σ_3 and $n_3/(n_1 + n_3)$ we could compute the excluded volume contribution $\Delta\Delta G^{ev}$ for glucose. From the m -value and $\Delta\Delta G^{ev}$, the soft-interaction contribution $\Delta\Delta G^{si}$ for glucose was then extracted from Equation 9. This soft-interaction contribution is the difference between the soft-interaction term of glucose and that of water, $\Delta\Delta G^{si} = \Delta G_{glu1}^{si,N \rightarrow U} - \Delta G_{water}^{si,N \rightarrow U}$. The soft-interaction term of glucose can further be described as the soft interaction contribution of the repetitive unit in the polymeric series, plus the contribution due to the end groups (i.e., a hydroxyl group and a hydrogen), $\Delta G_{glu1}^{si,N \rightarrow U} = \Delta G_{rep}^{si,N \rightarrow U} + \Delta G_{end}^{si,N \rightarrow U}$;

5. finally, for all the remaining cosolutes along the polymeric series, the corresponding excluded volume contribution was computed from σ_3 and $n_3/(n_1 + n_3)$. The soft-interaction term was calculated as $\Delta\Delta G^{si} = n\Delta G_{rep}^{si,N \rightarrow U} + \Delta G_{end}^{si,N \rightarrow U} - \Delta G_{water}^{si,N \rightarrow U}$, where n is the degree of polymerization. The end groups contribution $\Delta G_{end}^{si,N \rightarrow U}$ was omitted for the cyclodextrin. This last equation implies that we assumed that the soft interaction term scales with the number of monomeric units. The m -value for each cosolute larger than glucose was eventually computed according to Equation 9.

The results of the calculation procedure just outlined are summarized in Figure 10. The thermodynamic model predicts a positive (i.e., stabilizing) excluded volume contribution, and a negative (i.e., destabilizing) soft-interaction contribution for all cosolutes along the polymeric series (Figure 10a). The excluded volume term dominates above the soft-interaction one, leading to overall stabilization, for all cosolutes, with the only exception of γ -CD. In this latter case, the smaller radius of γ -CD, due to its closed shape, leads to a decreased excluded volume contribution. The destabilizing soft-interaction term, therefore, dominates for γ -CD, resulting in a negative m -value.

The predictions of the thermodynamic model for the polymeric series are in agreement with the experimental values (Figures 9 and 10b). This proves that the thermodynamic model successfully identifies the different contributions of excluded volume and soft-interaction effects to protein stabilization by excipients, providing insight into the mecha-

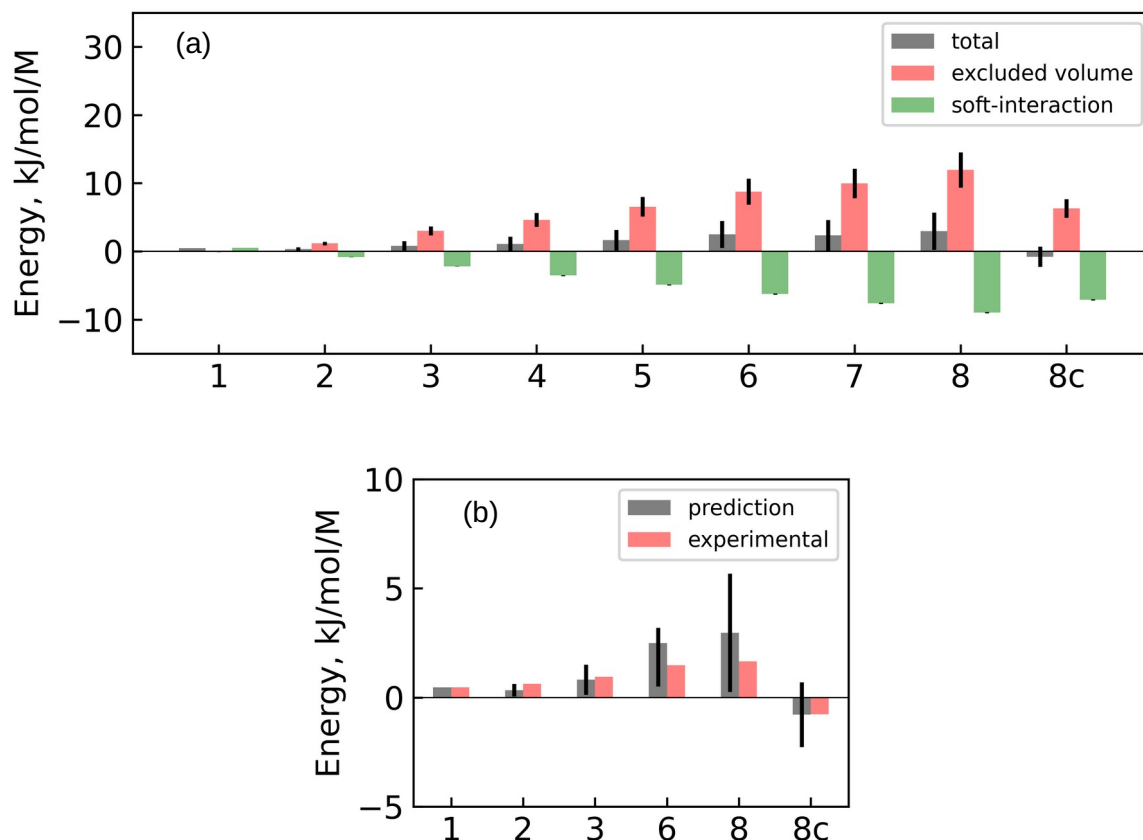


Figure 10: (a) Energetics contributions of 150 mM glucose polymers to lysozyme stabilization. The degree of polymerization (x-axis) ranges from 1 (glucose) to 8 (maltooctaose, linear shape) and 8c (γ -CD, closed shape). The total energetic contribution is dissected into the excluded volume and soft interaction terms. (b) Comparison of the energetics contributions to lysozyme stabilization by 150 mM glucose polymers as predicted by the thermodynamic model, and as determined experimentally. The degree of polymerization (x-axis) ranges from 1 (glucose) to 8 (maltooctaose, linear shape) and 8c (γ -CD, closed shape). In both panels, a positive value indicates stabilization of the native fold, and vice versa.

nisms of protein stabilization. Moreover, an interesting finding from both the model predictions and the experiments is that sugars stabilize proteins according to an excluded volume mechanism, while soft-interactions are instead destabilizing.

Cyclodextrin-Induced Protein Destabilization May Be Driven by a Global Soft-Interaction Mechanism rather than by Specific Direct Interactions

The negative effect of cyclodextrins on the thermodynamic stability of proteins is not new in the literature.⁷⁰ However, the thermodynamic model and experimental data shown in this work may provide some mechanistic insight about this destabilizing effect.

Cyclodextrins are characterized by the presence of a lipophilic cavity and a hydrophilic outer surface, and it is well known in the literature that cyclodextrins can include apolar and aromatic groups within their cavity.^{71–75} If it were true that cyclodextrins could at least partly embed lipophilic amino acid side chains from the protein into their ring cavity, the unfolded state could be stabilized by these complex formations because it exposes more apolar/aromatic residues. The shift of the equilibrium towards the unfolded state could therefore be explained by specific, direct interactions between the protein lipophilic residues and the cyclodextrin.⁷⁶

However, we monitored the formation of inclusions between lysozyme and γ -CD in our molecular dynamics simulations (Figure S3) and we found that the overall level of inclusion was quite small, and the most included residues were generally those most surface-exposed, with no real preference for apolar residues. For instance, the most included side chain belonged to a charged residue (Arg21), and was included for only 4.2 % of the equilibrated trajectory. This is in line with a previous simulation work,⁷⁷ where the interaction of cyclodextrins with proteins was found to be nonspecific, mostly due to attraction between cyclodextrins and the protein backbone, and where cyclodextrins were observed to bind to

strongly exposed residues, independently of their polarity. In addition, the fact that the thermodynamic model herein presented (which considers only a global soft-interaction term and does not account for specific, direct interactions) is enough to reflect the protein destabilization observed experimentally, is indirect evidence that specific inclusion effects are not the dominant mechanism. Our results suggest, therefore, that the cyclodextrin-induced destabilization of proteins may not be driven by specific inclusion of lipophilic residues, but rather by a global soft-interaction mechanism. It is to be noted, however, that some assumptions were made in the derivation of the thermodynamic model herein proposed, and other mechanisms not taken into account by the model and leading to fortuitous compensations may also play a role.

Conclusions

We have proposed a simple model to compute the two fundamental contributions to the free energies of transfer, namely, the excluded volume and soft interaction terms. The equations used here apply to the specific case of protein unfolding, and have been used to study the behavior of common osmolytes, including urea, sucrose and TMAO. Different protein-osmolyte combinations have been studied, and a general trend has been extracted for the various osmolytes.

Our model presents a number of unique features compared to existing approaches. In particular, decoupling the excluded volume and soft interactions contributions is experimentally difficult because only the sum of both these contributions is accessible by experimental techniques. Our work represents a novel approach to decoupling these terms by using a series of glucose polymers and taking advantage of the fact that these polysaccharides can present linear and cyclical forms. Using this polymeric series allows us not only to validate our thermodynamic model, but importantly to experimentally assess (for the first time) the relative importance of the soft interactions and excluded volume contributions to protein

stabilization by sugars. In addition, our model expresses the volume change upon unfolding as a function of the change in solvent accessibility and considers the presence of density gradients within the solution through the preferential exclusion coefficient. These two factors make the thermodynamic approach proposed in the present work more realistic than existing models in the literature. Finally, the conclusion of this study, i.e., that sugars stabilize proteins according to an excluded volume mechanism, also represents a novel and significant finding, not reported in previously published work.

We have found that sucrose stabilizes proteins because of excluded volume effects that overcome the destabilizing soft interaction term. The mechanism of TMAO is different, as in this case stabilization occurs also through soft interactions, that penalize the unfolded state against the native fold. In the case of urea, the excluded volume term is, instead, too small to counterbalance the strongly denaturing soft interaction contribution. For all osmolytes, the excluded volume term scales almost linearly with the dimension (SASA) of the protein molecule considered. This is not true for the soft interaction term, that is only marginally influenced by the peptide SASA.

Overall, these results show that the thermodynamic mechanism underlying the osmolytes effect on protein stability is not universal, and depends on the specific cosolute being considered. For the case of sugars, for instance, only the excluded volume term is stabilizing, while the soft-interaction contributions tend to promote protein unfolding. As an additional finding of this work, our results suggest that the cyclodextrin-induced protein destabilization may be driven by a global soft-interaction mechanism, rather than by specific direct interactions.

Supporting Information

Comparison of urea models, radii of gyration and preferential exclusion coefficients for the glucose polymeric series, inclusion of lysozyme residues within the cavity of γ -cyclodextrin.

Acknowledgement

The authors acknowledge support from the Center for Scientific Computing at the California Nanosystems Institute (CNSI, NSF grant CNS-1725797) and from the hpc@polito team (<http://www.hpc.polito.it>) for the availability of high performance computing resources and support. This work used the Extreme Science and Engineering Discovery Environment, which is supported by the National Science Foundation grant ACI-1548562 (MCA05S027). The authors acknowledge support from the NSF (MCB-1716956 and CHE-1800352) and the NIH (R01-GM118560-01A).

References

- (1) Yancey, P.; Clark, M.; Hand, S.; Bowlus, R.; Somero, G. Living with water stress: evolution of osmolyte systems. *Science* **1982**, *217*, 1214–1222.
- (2) Yancey, P. H. Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. *J. Exp. Biol.* **2005**, *208*, 2819–2830.
- (3) Arakawa, T.; Timasheff, S. N. Stabilization of protein structure by sugars. *Biochemistry* **1982**, *21*, 6536–6544.
- (4) Lee, J. C.; Timasheff, S. N. The stabilization of proteins by sucrose. *J. Biol. Chem.* **1981**, *256*, 7193–7201.
- (5) Ganguly, P.; Polák, J.; van der Vegt, N. F. A.; Heyda, J.; Shea, J.-E. Protein Stability in TMAO and Mixed Urea–TMAO Solutions. *J. Phys. Chem. B* **2020**, *124*, 6181–6197.
- (6) Auton, M.; Bolen, D. W. Predicting the energetics of osmolyte-induced protein folding/unfolding. *Proc. Natl. Acad. Sci.* **2005**, *102*, 15065–15068.
- (7) Arsiccio, A.; Ganguly, P.; La Cortiglia, L.; Shea, J.-E.; Pisano, R. ADD Force Field

- for Sugars and Polyols: Predicting the Additivity of Protein-Osmolyte Interaction. *J. Phys. Chem. B* **2020**, *124*, 7779–7790.
- (8) Moeser, B.; Horinek, D. Unified Description of Urea Denaturation: Backbone and Side Chains Contribute Equally in the Transfer Model. *J. Phys. Chem. B* **2014**, *118*, 107–114.
 - (9) Gekko, K.; Timasheff, S. N. Mechanism of protein stabilization by glycerol: preferential hydration in glycerol-water mixtures. *Biochemistry* **1981**, *20*, 4667–4676.
 - (10) Arakawa, T.; Timasheff, S. N. Protein stabilization and destabilization by guanidinium salts. *Biochemistry* **1984**, *23*, 5924–5929.
 - (11) Timasheff, S. N.; Xie, G. Preferential interactions of urea with lysozyme and their linkage to protein denaturation. *Biophys. Chem.* **2003**, *105*, 421 – 448.
 - (12) Courtenay, E. S.; Capp, M. W.; Anderson, C. F.; Record, M. T. Vapor Pressure Osmometry Studies of Osmolyte-Protein Interactions: Implications for the Action of Osmoprotectants in Vivo and for the Interpretation of “Osmotic Stress” Experiments in Vitro. *Biochemistry* **2000**, *39*, 4455–4471.
 - (13) Anderson, C. F.; Courtenay, E. S.; Record, M. T. Thermodynamic Expressions Relating Different Types of Preferential Interaction Coefficients in Solutions Containing Two Solute Components. *J. Phys. Chem. B* **2002**, *106*, 418–433.
 - (14) Schneider, C. P.; Trout, B. L. Investigation of Cosolute-Protein Preferential Interaction Coefficients: New Insight into the Mechanism by Which Arginine Inhibits Aggregation. *J. Phys. Chem. B* **2009**, *113*, 2050–2058.
 - (15) Courtenay, E. S.; Capp, M. W.; Record Jr., M. T. Thermodynamics of interactions of urea and guanidinium salts with protein surface: Relationship between solute effects

- on protein processes and changes in water-accessible surface area. *Protein Sci.* **2001**, *10*, 2485–2497.
- (16) Auton, M.; Holthauzen, L. M. F.; Bolen, D. W. Anatomy of energetic changes accompanying urea-induced protein denaturation. *Proc. Natl. Acad. Sci.* **2007**, *104*, 15317–15322.
- (17) Schellman, J. A. Protein stability in mixed solvents: A balance of contact interaction and excluded volume. *Biophys. J.* **2003**, *85*, 108 – 125.
- (18) Tang, K. E.; Bloomfield, V. A. Excluded Volume in Solvation: Sensitivity of Scaled-Particle Theory to Solvent Size and Density. *Biophys. J.* **2000**, *79*, 2222 – 2234.
- (19) Niesen, F. H.; Berglund, H.; Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc* **2007**, *2*, 2212–2221.
- (20) Reiss, H.; Frisch, H. L.; Lebowitz, J. L. Statistical Mechanics of Rigid Spheres. *J. Chem. Phys.* **1959**, *31*, 369–380.
- (21) Lebowitz, J. L.; Helfand, E.; Praestgaard, E. Scaled Particle Theory of Fluid Mixtures. *J. Chem. Phys.* **1965**, *43*, 774–779.
- (22) Mandell, M. J.; Reiss, H. Scaled particle theory: Solution to the complete set of scaled particle theory conditions: Applications to surface structure and dilute mixtures. *J. Stat. Phys.* **1975**, *13*, 113–128.
- (23) Heying, M.; Corti, D. S. Scaled Particle Theory Revisited: New Conditions and Improved Predictions of the Properties of the Hard Sphere Fluid. *J. Phys. Chem. B* **2004**, *108*, 19756–19768.
- (24) Graziano, G. A purely geometric derivation of the scaled particle theory formula for the work of cavity creation in a liquid. *Chem. Phys. Lett.* **2007**, *440*, 221 – 223.

- (25) Pace, C. N.; Hermans, J. The stability of globular protein. *CRC Critical Reviews in Biochemistry* **1975**, *3*, 1–43.
- (26) O'Brien, E. P.; Brooks, B. R.; Thirumalai, D. Molecular Origin of Constant m-Values, Denatured State Collapse, and Residue-Dependent Transition Midpoints in Globular Proteins. *Biochemistry* **2009**, *48*, 3743–3754.
- (27) Tanford, C. Isothermal Unfolding of Globular Proteins in Aqueous Urea Solutions. *J. Am. Chem. Soc.* **1964**, *86*, 2050–2059.
- (28) Davis-Searles, P. R.; Saunders, A. J.; Erie, D. A.; Winzor, D. J.; Pielak, G. J. Interpreting the Effects of Small Uncharged Solutes on Protein-Folding Equilibria. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 271–306.
- (29) Saunders, A. J.; Davis-Searles, P. R.; Allen, D. L.; Pielak, G. J.; Erie, D. A. Osmolyte-induced changes in protein conformational equilibria. *Biopolymers* **2000**, *53*, 293–307.
- (30) O'Connor, T. F.; Debenedetti, P. G.; Carbeck, J. D. Simultaneous Determination of Structural and Thermodynamic Effects of Carbohydrate Solutes on the Thermal Stability of Ribonuclease A. *J. Am. Chem. Soc.* **2004**, *126*, 11794–11795.
- (31) Cozzolino, S.; Oliva, R.; Graziano, G.; Del Vecchio, P. Counteraction of denaturant-induced protein unfolding is a general property of stabilizing agents. *Phys. Chem. Chem. Phys.* **2018**, *20*, 29389–29398.
- (32) Graziano, G. Dimerization Thermodynamics of Large Hydrophobic Plates: A Scaled Particle Theory Study. *J. Phys. Chem. B* **2009**, *113*, 11232–11239.
- (33) Graziano, G. The Gibbs energy cost of cavity creation depends on geometry. *J. Mol. Liq.* **2015**, *211*, 1047 – 1051.
- (34) Pohorille, A.; Pratt, L. R. Cavities in molecular liquids and the theory of hydrophobic solubilities. *J. Am. Chem. Soc.* **1990**, *112*, 5066–5074.

- (35) Floris, F. M.; Selmi, M.; Tani, A.; Tomasi, J. Free energy and entropy for inserting cavities in water: Comparison of Monte Carlo simulation and scaled particle theory results. *J. Chem. Phys.* **1997**, *107*, 6353–6365.
- (36) Graziano, G. How does trimethylamine N-oxide counteract the denaturing activity of urea? *Phys. Chem. Chem. Phys.* **2011**, *13*, 17689–17695.
- (37) Chen, C. R.; Makhatadze, G. I. Molecular determinant of the effects of hydrostatic pressure on protein folding stability. *Nat. Commun.* **2017**, *8*, 14561.
- (38) Chalikian, T. V. Volumetric Properties of Proteins. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 207–235.
- (39) 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.
- (40) 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.
- (41) Auton, M.; Bolen, D. W.; Rös gen, J. Structural thermodynamics of protein preferential solvation: Osmolyte solvation of proteins, aminoacids, and peptides. *Proteins* **2008**, *73*, 802–813.
- (42) Zardecki, C.; Dutta, S.; Goodsell, D. S.; Voigt, M.; Burley, S. K. RCSB Protein Data Bank: A Resource for Chemical, Biochemical, and Structural Explorations of Large and Small Biomolecules. *J. Chem. Educ.* **2016**, *93*, 569–575.
- (43) Lee, B.; Richards, F. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* **1971**, *55*, 379 – IN4.

- (44) Richards, F. M. Areas, volumes, packing, and protein structure. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151–176.
- (45) Creamer, T. P.; Srinivasan, R.; Rose, G. D. Modeling unfolded states of proteins and peptides. II. Backbone solvent accessibility. *Biochemistry* **1997**, *36*, 2832–2835.
- (46) Auton, M.; Bolen, D. W. In *Osmosensing and Osmosignaling*; Häussinger, D., Sies, H., Eds.; Methods Enzymol.; Academic Press, 2007; Vol. 428; pp 397 – 418.
- (47) Liu, Y.; Bolen, D. W. The peptide backbone plays a dominant role in protein stabilization by naturally occurring osmolytes. *Biochemistry* **1995**, *34*, 12884–12891.
- (48) Qu, Y.; Bolen, C. L.; Bolen, D. W. Osmolyte-driven contraction of a random coil protein. *Proc. Natl. Acad. Sci.* **1998**, *95*, 9268–9273.
- (49) Wang, A.; Bolen, D. W. A Naturally Occurring Protective System in Urea-Rich Cells: Mechanism of Osmolyte Protection of Proteins against Urea Denaturation. *Biochemistry* **1997**, *36*, 9101–9108.
- (50) 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.
- (51) Steinrauf, L. K. Structures of Monoclinic Lysozyme Iodide at 1.6 Å and of Triclinic Lysozyme Nitrate at 1.1 Å. *Acta Cryst. D* **1998**, *54*, 767–779.
- (52) 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.
- (53) 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.

- (54) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255–268.
- (55) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- (56) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (57) 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.
- (58) MacKerell, A. D. et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (59) 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.
- (60) 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.
- (61) 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.
- (62) 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.
- (63) Svergun, D. I.; Richard, S.; Koch, M. H. J.; Sayers, Z.; Kuprin, S.; Zaccai, G. Protein hydration in solution: Experimental observation by x-ray and neutron scattering. *Proc. Natl. Acad. Sci.* **1998**, *95*, 2267–2272.

- (64) Kumagai, I.; Kojima, S.; Tamaki, E.; Miura, K.-i. Conversion of Trp 62 of Hen Egg-White Lysozyme to Tyr by Site-Directed Mutagenesis. *J. Biochem.* **1987**, *102*, 733–740.
- (65) Martinez-Oyanedel, J.; Choe, H.-W.; Heinemann, U.; Saenger, W. Ribonuclease T1 with free recognition and catalytic site: Crystal structure analysis at 1.5 Å resolution. *J. Mol. Biol.* **1991**, *222*, 335 – 352.
- (66) Aguilar, C.; Thomas, P.; Mills, A.; Moss, D.; Palmer, R. Newly observed binding mode in pancreatic ribonuclease. *J. Mol. Biol.* **1992**, *224*, 265 – 267.
- (67) Truckses, D. M.; Prehoda, K. E.; Miller, S. C.; Markley, J. L.; Somoza, J. R. Coupling between trans/cis proline isomerization and protein stability in staphylococcal nuclease. *Protein Sci.* **1996**, *5*, 1907–1916.
- (68) Weaver, L.; Matthews, B. Structure of bacteriophage T4 lysozyme refined at 1.7 Å resolution. *J. Mol. Biol.* **1987**, *193*, 189 – 199.
- (69) Arsiccio, A.; Ganguly, P.; Shea, J.-E. A Transfer Free Energy Based Implicit Solvent Model for Protein Simulations in Solvent Mixtures: Urea-Induced Denaturation as a Case Study. *J. Phys. Chem. B* **2022**, *126*, 4472–4482.
- (70) Hartl, E.; Winter, G.; Besheer, A. Influence of Hydroxypropyl-Beta-Cyclodextrin on the Stability of Dilute and Highly Concentrated Immunoglobulin G Formulations. *J. Pharm. Sci.* **2013**, *102*, 4121–4131.
- (71) Achmann, F.; Otzen, D.; Larsen, K.; Wimmer, R. Structural background of cyclodextrin–protein interactions. *Protein Eng.* **2003**, *16*, 905–912.
- (72) Otzen, D. E.; Knudsen, B. R.; Achmann, F.; Larsen, K. L.; Wimmer, R. Structural basis for cyclodextrins’ suppression of human growth hormone aggregation. *Protein Sci.* **2002**, *11*, 1779–1787.

- (73) Koushik, K. N.; Bandi, N.; Kompella, U. B. Interaction of [d-Trp6, Des-Gly10] LHRH Ethylamide and Hydroxy Propyl β -Cyclodextrin (HP β CD): Thermodynamics of Interaction and Protection from Degradation by α -Chymotrypsin. *Pharm. Dev. Technol.* **2001**, *6*, 595–606.
- (74) Arsiccio, A.; Rospiccio, M.; Shea, J.-E.; Pisano, R. Force Field Parameterization for the Description of the Interactions between Hydroxypropyl- β -Cyclodextrin and Proteins. *J. Phys. Chem. B* **2021**, *125*, 7397–7405.
- (75) Yi, Z.; Qasim, M.; Qasim, S.; Warrington, T.; Laskowski, M. Ring-Toss: Capping highly exposed tyrosyl or tryptophyl residues in proteins with β -cyclodextrin. *Biochim. Biophys. Acta* **2006**, *1760*, 372–379.
- (76) Serno, T.; Geidobler, R.; Winter, G. Protein stabilization by cyclodextrins in the liquid and dried state. *Adv. Drug Deliv. Rev.* **2011**, *63*, 1086–1106.
- (77) Rospiccio, M.; Arsiccio, A.; Winter, G.; Pisano, R. The Role of Cyclodextrins against Interface-Induced Denaturation in Pharmaceutical Formulations: A Molecular Dynamics Approach. *Mol. Pharm.* **2021**, *18*, 2322–2333.

Graphical TOC Entry

