Supporting Information for Understanding Glucagon Aggregation: *In-silico* Insights and Experimental Validation

Roberto Pisano,[*,†] Andrea Arsiccio,[†] Valerie Collins,[‡] Patrick King,[‡] Marco Macis,[¶] Walter Cabri,[¶,§] and Antonio Ricci[*,¶]

†*Department of Applied Science and Technology, Politecnico di Torino, 24 corso Duca degli Abruzzi, IT-10129, Torino, Italy*

‡*RedShift BioAnalytics, 80 Central Street, Boxborough, MA-01719, USA*

¶*Fresenius Kabi iPSUM, 108 via Roma, IT-20081, Cassina de Pecchi, Italy*

§*Center for Chemical Catalysis Department of Chemistry "Giacomo Ciamician", University of Bologna, 40126 Bologna, Italy*

E-mail: roberto.pisano@polito.it; antonio.ricci@fresenius-kabi.com

Explicit Solvent Simulations of Lactose

Explicit solvent molecular dynamics simulations of lactose in presence of the different amino acids were performed with the objective to extract the free energy of transfer values to then use in the implicit solvent simulations. For this purpose, the software Gromacs 2018.6 [1] was used.

Lactose was modelled using the ADD force field for carbohydrates,[2] in combination with the CHARMM TIP3P water model,[3] and the CHARMM36m force field[4] for amino acids, whenever present.

A scheme of the explicit solvent simulations performed for lactose in this work, with the corresponding box size, temperature and duration, is listed in Table S1.

Table S1: List of the simulations performed in this work.

Sim. type #	solute	lactose conc. mol/L	box size nm	T K	duration ns
1	-	1	8 x 8 x 8	300	60
2	capped amino acids	1	8 x 8 x 8	300	60
3	NAG_xA	1	8 x 8 x 8	300	60

In sim. type 1, lactose at 1 mol/L was simulated in absence of cosolutes. For sim. type 2, all the 20 naturally occurring amino acids, in their capped form (i.e., acetylated N-terminus and amidated C-terminus), were simulated in 1 M lactose. Also the N-acetyl glycinamide series (NAG_xA) was simulated (sim. type 3 in Table S1). NAG_xA corresponds to a series of molecules with a varying number x of glycine residues linked by a peptide bond and whose termini are blocked by an acetyl (N-terminus) and an amide (C-terminus) moiety. The NAG_xA series considered comprised the variants with 3 (NAG_3A), 4 (NAG_4A), 5 (NAG_5A) or 8 (NAG_8A) glycine units. In all cases, the box was cubic with ≈ 8 nm side length. For simulations type 2 and 3 the box included 25 amino acid/NAG_xA molecules. For charged residues, Na^+ or Cl^- ions were added to reach neutrality.

In all cases, the cut-off radius for both Coulombic (calculated using the PME method[5])

and Lennard-Jones interactions was set to 1.2 nm, and periodic boundary conditions were used. Each box was first energy minimized with the steepest descent algorithm, and then equilibrated for 1 ns at 1 bar and 300 K in the NPT ensemble, using Berendsen pressure (3 ps relaxation time) and temperature (0.5 ps relaxation time) coupling.[6] The simulations were then run at 300 K and 1 bar in the NPT ensemble, controlling temperature and pressure with the Nosé-Hoover thermostat[7,8] (0.5 ps relaxation time) and Parrinello-Rahman barostat[9] (3 ps relaxation time), respectively. A 2 fs time-step was used, and configurations were saved every 2 ps. The Lincs algorithm was employed for constraining all bonds,[10] while the SETTLE algorithm was used to keep the water molecules rigid.[11] The last 40 ns were used for the analyses.

Kirkwood-Buff integrals

Kirkwood-Buff integrals (KBIs)[12,13] G_{ij} are used to describe the solvation behavior of component j around a reference particle i, and are defined as spherical integrals of the radial distribution function $g_{ij}(r)$ over the distance r between i and j.

The Kirkwood-Buff integrals were calculated in this work as follows,

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

where R is a value at which convergence is reached (generally, the running KBIs converged for values of R comprised between 1.0 and 1.4 nm). To correct for finite size effects, the correction suggested in Ref.[14] was applied.

A value of $G_{ij} < 1$ indicates exclusion, while $G_{ij} > 1$ indicates accumulation of component j around the reference i.

Here, we computed the KBIs for amino acid-lactose (G_{23}) and amino acid-water (G_{12}) interaction. We then obtained the values $\gamma_i = G_{23} - G_{12}$ for each amino acid i, and computed the side chain contributions by subtracting the value for glycine γ_{gly},

$$\gamma_i^{sc} = \gamma_i - \gamma_{gly} \tag{2}$$

We also obtained a value γ^{bb} for the backbone by simulating the N-acetyl glycinamide series $\mathrm{NAG}_x\mathrm{A}$ in presence of lactose, as described in Ref.[2] The $\gamma^{bb \text{ or } sc}$ values can then be converted to $g_{bb \text{ or } sc}^{tr}(c)$ values using the following equation,

$$g_{bb \text{ or } sc}^{tr}(c) = \frac{-\mathrm{R}Tc\gamma^{bb \text{ or } sc}}{1 - c(G_{13} - G_{33})} \tag{3}$$

where R is the universal gas constant, and G_{13} and G_{33} are the water-lactose and lactose-lactose Kirkwood-Buff integrals, respectively.

The values of $g_{bb \text{ or } sc}^{tr}(c)$ obtained in this way for lactose, and those computed for HPβCD in previous work,[15] are listed in Table S2.

Table S2: Free energy of transfer values $g^{tr}_{k,sc}(c)$ and $g^{tr}_{bb}(c)$ used in the implicit solvent simulations.

$g^{tr}(c)$ kJ mol^{-1}	10.7 % w/w lactose	10.7 % w/w HPβCD[1]
Ala	0.098	-0.268
Arg	0.265	-0.294
Asn	-0.016	-0.180
Asp	0.475	0.381
Cys	-0.180	-0.453
Gln	0.142	-0.374
Glu	0.470	0.275
Gly	0	0
His	0.127	-0.278
Ile	0.072	-0.898
Leu	0.005	-0.769
Lys	0.355	0.091
Met	-0.036	-1.054
Phe	-0.017	-1.117
Pro	0.027	-0.426
Ser	0.028	-0.096
Thr	-0.025	-0.198
Trp	-0.516	-1.440
Tyr	-0.313	-1.305
Val	0.081	-0.374
backbone	0.016	-0.146

[1] obtained from our computational work in explicit solvent[15]

References

(1) 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.

(2) 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.

(3) 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.

(4) 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.

(5) 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.

(6) 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.

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

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

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

(10) 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.

(11) 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.

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

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

(14) 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.

(15) 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.