Split the charge difference in two! A rule of thumb for adding proper amounts of ions in MD simulations.

Matías R. Machado[1]*, Sergio Pantano[1]

[1]Biomolecular Simulations Group, Institut Pasteur de Montevideo, Montevideo, Uruguay. Mataojo 2020, CP 11400.

*To whom correspondence should be addressed:

Matías R. Machado, Tel/Fax: +598-2522 0910, e-mail: mmachado@pasteur.edu.uy

RUNNING TITLE

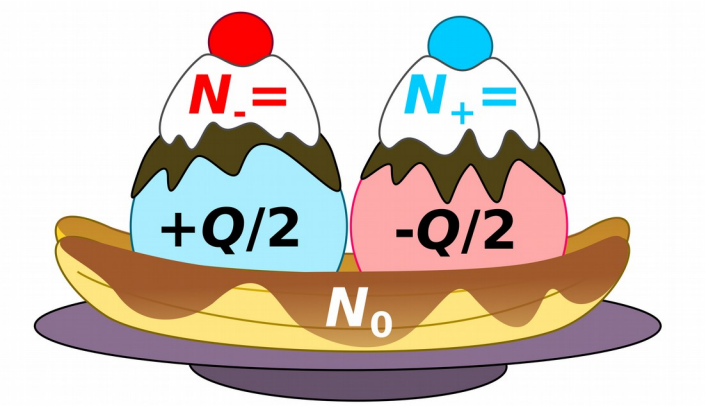
Rule of thumb for adding ions to simulation boxes.

KEYWORDS

Molecular Dynamics, System Setup, Salt, Neutralization.

ABSTRACT

Despite the relevance of properly setting ionic concentrations in Molecular Dynamics (MD) simulations, methods or practical rules to set ionic strength are scarce and rarely documented. Based on a recently proposed thermodynamics method we provide an accurate rule of thumb to define the electrolytic content in simulation boxes. Extending the use of good practices in setting up MD systems is promptly needed to ensure reproducibility and consistency in molecular simulations.



***Q*: Solute charge, N_0: ion pairs at [salt]**

TOC

TEXT BODY

The correct description of molecular features is of major concern to properly study biological systems by Molecular Dynamics (MD) simulations. The aqueous solvent plays a relevant role in the hydrophobic/hydrophilic balance, which determines the aggregation properties leading to protein folding, ligand binding, membrane formation and other large-scale phase separation phenomena in nucleus and cytosol [1,2]. The increasing computational power is permitting realistic simulations in terms of ionic and molecular species, time scales and replicas that increase the robustness of predictions granting direct comparison against state-of-the-art experimental techniques. Concomitantly, it is helping to point out methodological limitations. Clearly, the correct representation of the aqueous solvent is of paramount importance in MD [3]. The very nature of the chosen water model impacts the observed dynamics in intrinsically disordered proteins and protein folding simulations [4,5]. Even the amount of water used in simulation boxes may affect the protein dynamics by preventing the correct representation of bulk properties [6]. Surprisingly, the accurate description of the electrolytic environment has received much less attention in the literature, being poorly documented in most methodological sections of scientific publications. Indeed, there are no clear standards for simple issues as how to calculate the proper amount of ions to add in computational boxes. Roughly, in million atoms simulations of viral particles and huge macromolecular complexes, which are becoming frequent in the literature [7], the physiological salt concentration may exceed 2000 ionic pairs in the box. Precise documentation or standards to correctly and systematically calculate such quantities are clearly not negligible matters. Recently, Schmit et al. showed that the most commonly used method for adding salt to simulation results in a sensible overestimation of the effective salt concentration [8]. Here we explore the limits of this approximation and test it on a comprehensive and non-redundant data set of proteins including over 20% of structures reported in the Protein Data Bank (PDB) to provide a rule of thumb valid for the vast majority of molecular systems.

Because of widely used periodic boundary conditions and algorithms for calculating long-range interactions as the popular Particle Mesh Ewald (PME) approach [9,10], electrically neutral systems are mandatory [11,12]. This can be achieved by adding a uniform background charge density (the default in many MD engines), also known as the uniform neutralizing plasma method [13], or by complementing the solution with explicit counterions. As mentioned by Schmit et al. [8], the two most frequent ways of adding ions are the counterion-neutralization and the add-then-neutralize (AN) methods. While the former strategy does not consider the presence of salt, the later adds an excess of counterions to the system, which may result in an increase of the effective salt concentration. This effect is more marked in highly charged systems like oligomeric complexes (e.g. viral particles [14]), polyelectrolytes, etc. Schmit et al. [8] introduced the method screening layer tally by container average potential (SLTCAP) derived from statistical mechanics as a valid solution to the AN issue. According to SLTCAP, the effective number of positive/negative monovalent ions to be added into the computational box ($N_{\pm}$) is given by:

$$N_{\pm} = v_w c_0 e^{\mp Arcsinh\left(\frac{Q}{2 v_w c_0}\right)} \quad (1)$$

Where v_w is the water volume of the simulation box in reduced units, c_0 is the desired salt concentration and Q is the total charge of the solute in electron units.

Important assumptions behind the SLTCAP model are: i) The solute is at infinite dilution; ii) The system is at equilibrium with a large enough reservoir of ions and solvent in the bulk; iii) The salt is a pair of monovalent ions; iv) Ions do not have specific interactions with each other or with the solute; and v) The solvent volume remains the same upon addition of ions.

Although using finite simulation boxes with a fixed number of particles may limit the reproduction of macroscopic salt concentrations [15], the SLTCAP method still constitutes a major improvement in standard MD simulations, which comes at no computational cost. Moreover, as SLTCAP is a model-independent approach using models of ions and solutes that correctly describe the osmotic pressure may further improve the description of the electrolytic effects [16–19].

Because of mathematical convenience we simplify Eq. 1 by noticing that $N_0 = v_w c_0$ is the number of ions at a given salt concentration and solvent volume in a pure solution (without solute). Using the property $Arcsinh(-z) = - Arcsinh(z)$ and the definition $Arcsinh(z) = \ln(z + (1+z^2)^{1/2})$, we can write:

$$N_{\pm} = N_0 e^{\ln\left(\left(\frac{\mp Q}{2N_0}\right) + \sqrt{1+\left(\frac{\mp Q}{2N_0}\right)^2}\right)} \quad (2)$$

which is equivalent to:

$$N_{\pm} = N_0 \sqrt{1+\left(\frac{Q}{2N_0}\right)^2} \mp \frac{Q}{2} \quad (3)$$

Despite Eq. 1 and 3 are equivalent, the latter is more convenient for further interpretations. The first term is always positive regardless of the charge of the solute and can be interpreted as a correction to N_0, while the second is a constant depending on the solute's charge. The sign of Q determines whether ions are added or subtracted. Assuming a water density (ρ_w) of 1000 g/L at 298 K and a molecular weight (M_w) of 18 g/mol, N_0 can be easily calculated from the number of added water molecules to the simulation box by:

$$N_0 = \frac{N_w M_w c_0}{\rho_w} \sim \frac{N_w c_0}{56} \quad (4)$$

Where N_w is the number of water molecules, c_0 is the salt concentration in Molar units and 1/56 is the conversion factor. Eq. 4 renders the best N_0 estimation against MD simulations using sophisticated osmostats [15].

We then analyze the limits of Eq. 3 in terms of N_0 and Q. If $N_0 >> Q$, because the salt concentration is

high or the system volume (simulation box) is big enough, then Eq. 3 may be approximated by:

$$N_{\pm} \sim N_0 \mp \frac{Q}{2} \qquad (5)$$

We call Eq. 5 the SPLIT method, because it implies equally redistributing the charge excess of the solute between the ionic species. Notice that N_0 does not have to be very large to fulfill such a condition. Indeed, the difference between SLTCAP (Eq. 3) and SPLIT (Eq. 5) for values of N_0 larger than 2*Q is less than 1% (Figure 1A). If $N_0 \sim Q$, then in practical terms SLTCAP may also be approximated by SPLIT with an error of ~ 7%, i.e. comparable with intrinsic rounding effects of adding an integer number of ions. In this regard, we recommend rounding up fractional results of SPLIT. In the limit of application, if $N_0 \sim Q/2$, then the difference between SLTCAP and SPLIT is ~ 17%, compromising the validity of the later approximation. Notice that, if $Q >> N_0$, because of high charged solutes and/or very low salt concentrations, then both SLTCAP and AN approach the counterion-only result.

Importantly, previous observations are valid regardless of the actual volume of the system. We next analyze the application range of SPLIT in common setups for MD simulation of biological systems. To do that, a representative subset of structures from the PDB is chosen according to the following criteria: the main set of selected structures contain proteins but not nucleic acids, have a sequence length between 20 and 1000 residues, do not have modified residues, share less than 90% of sequence identity and are solved by X-Ray diffraction. Additionally, non-redundant protein-DNA/RNA complexes were taken from the manually curated Protein-DNA Interface Database (PDIdb, http://melolab.org/pdidb) [20] and the Nucleic Acid Database (NDB, http://ndbserver.rutgers.edu) [21]. Biological assemblies from the selected set of ~34000 entries are first processed to add missing atoms and remove alternative conformations, ligands, ions and crystallographic waters. Curated structures are then solvated in standard octahedral boxes that extend up to 1.2 nm from the solute. All the procedure is done by the LEAP utility of AMBER Tools [22]. Only Lysine, Arginine, Aspartic and Glutamic acids are considered in their charged state. For nucleic acids, only their phosphates are used to assign the charge. Raw data and scripts to generate it are available as supporting information.

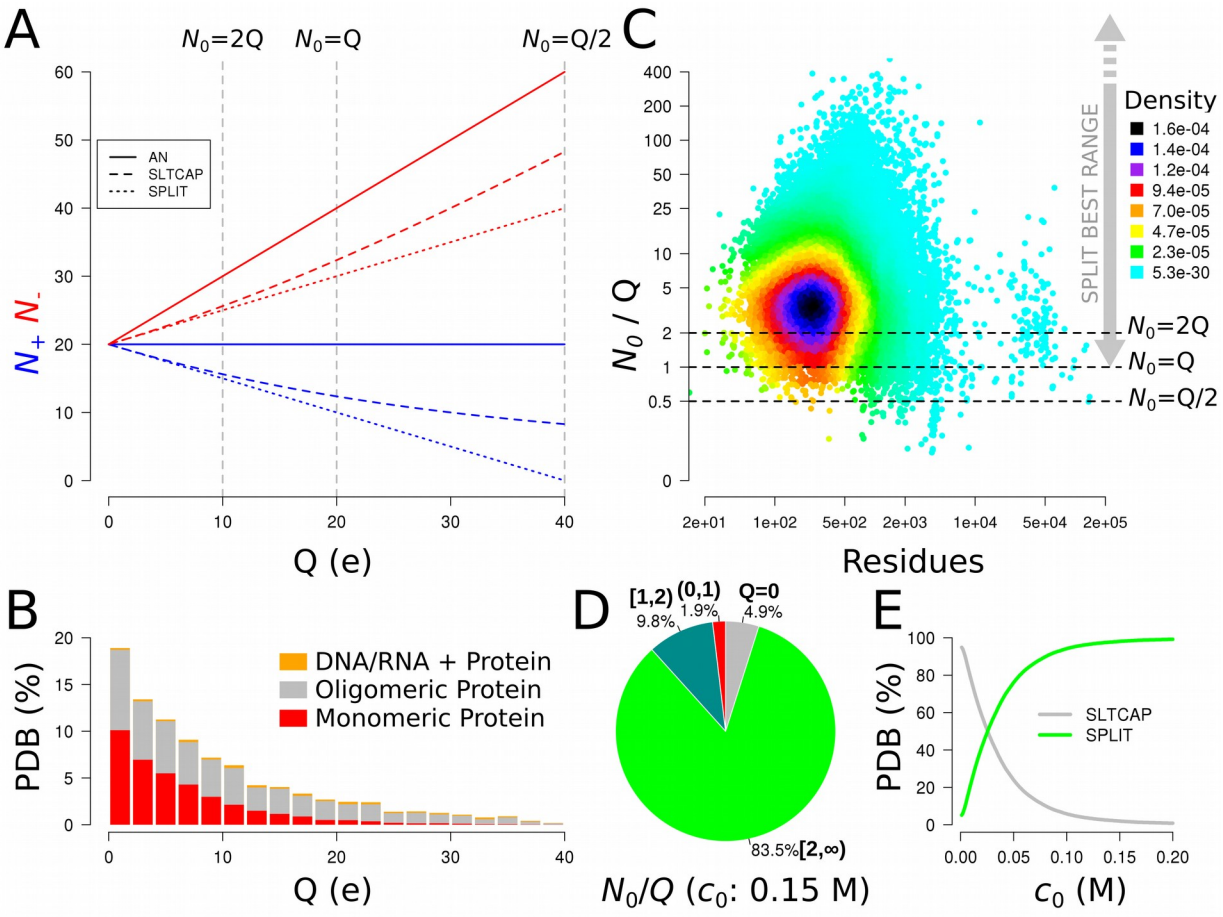


Figure 1. A) Calculated number of salt ions needed to neutralize the system and achieve a bulk concentration of 0.15 M in a solvent volume of 221 nm^3 (~ 7400 water molecules at room temperature) according to the solute charge. B) Distribution of absolute macromolecule charges in a representative PDB data set of ~34000 structures. C) Estimated ions and charge ratio from a PDB data set at a salt concentration of 0.15 M. D) Proportion of N_0/Q ratios from panel C. Intervals are expressed in mathematical notation for simplicity. Neutral systems (Q = 0) are included. E) Percentage of correctly described systems from the PDB data set by SPLIT ($N_0/Q \geq 1$) against those requiring SLTCAP ($N_0/Q < 1$) as a function of salt concentration.

We first notice from the data set distribution that 80% of the systems have charge lower than +/-20 e (Figure 1B). Monomeric proteins of charge greater than +/-20 e are rare in the PDB (~5%), such high charge magnitudes are more frequent in oligomeric proteins (~26%) and protein-DNA/RNA complexes (~43%). However, we expect oligomeric complexes to be larger than monomeric proteins, requiring bigger computational boxes and hence more electrolytes and co-ions which ends up compensating for high solute charges. Taking together the data reported in Figures 1A and B, we speculate that most structures should belong to systems with an N_0/Q ratio larger than 1. This hypothesis is confirmed by computing Eq. 4 at c_0 = 0.15 M from the actual solvation box of each system (Figure 1C). Indeed, a detailed analysis of the N_0/Q distribution shows that ratios above 2 correspond to the vast majority, while ratios below 1 are quite rare (Figure 1D). Moreover, only ~5% of the protein-DNA/RNA complexes were outside the range of application of SPLIT. Notice that there is also ~5% of neutral structures, which don't require neutralization. As the most populated systems gather around 200

residues and N_0/Q ratios of ~3 (Figure 1C), it is important to evaluate the dependence of these results as a function of the salt concentration. Figure 1E shows the percentage of correctly described PDB structures from the data set by SPLIT ($N_0/Q \geq 1$) against those requiring SLTCAP ($N_0/Q < 1$) for different salt concentrations keeping the same simulation box. In this scenario, up to 50% of the systems at ~ 0.03 M of salt can be described by the simple SPLIT method, proving a wide range of applications. Although this analysis includes nearly 1700 protein-DNA/RNA complexes, naked nucleic acids (or other strong polyelectrolytes), present higher charge densities than proteins. This feature may place them in a more susceptible region of the N_0/Q ratio. For example, solvating the famous 12-mer B-form DNA of Drew-Dickerson (PDB id. 1BNA [23]) under the previous condition, renders $N_0/Q = 0.8$ for a salt concentration of 0.15 M. However, increasing the box side to 0.15 nm shifts the ratio to 1.1, which is within the application range of SPLIT. It is important to keep in mind that nucleic acids are most often bound to divalent (metallic) ions that are not to be confused with the monovalent electrolytes considered here. Metallic ions often work as cofactors of nucleic acids shaping the structure/function and decreasing the effective charge density of these molecules. Hence, although box sizes and degree of packing of solutes may condition the applicability of SPLIT, these systems should be considered with special care.

As a final analysis, we compare the use of AN, SLTCAP or SPLIT to simulate two small proteins of high charge. Granulysin (PDB id. 1L9L [24]) is a protein of 74 residues with a net charge of +11 e and a radius of ~12 nm, which has antimicrobial effects by creating holes in target cell membranes. The capsid protein of the Japanese Encephalitis virus (JEV capsid, PDB id. 5OW2 [25]) is a dimer of 146 total residues with a net charge of +20 e and a radius of ~1.5 nm, which helps to wrap and stabilize the viral genome. Systems are prepared as described before and resulting simulation boxes are detailed in Table 1. The AMBER 14SB force field [26] is used in combination with the TIP3P water model [27] and the Joung and Cheatham parameters for Na+ and Cl- ions [28]. Systems are minimized and then equilibrated in NVT by first restraining protein's heavy atoms for 1 ns and then Cα atoms for 4 ns with force constants of 2092 kJ mol^{-1} nm^{-2}. Five replicates of 100 ns in NPT are simulated with the AMBER GPU code [29]. While three replicates of 30 ns in semigrand canonical ensemble applying an osmostat are simulated with the OpenMM GPU code [30] by following protocols described in ref. [15]. All simulations are performed at 300 K and 1 bar by using the Langevin Thermostat [31] and a Monte Carlo Barostat. A 2 fs time step is used by applying constraints to Hydrogen atoms [32,33]. A cut-off of 0.1 nm is used for non-bonded interactions, while long-range electrostatics are considered by PME. The detailed protocol including input flags and scripts is available as supporting information.

At macroscopic salt concentrations of 0.10 M and 0.15 M, the N_0/Q ratio equals 1 for Granulysin and JEV capsid, respectively (Table 1). In these scenarios, the number of estimated ions by SPLIT and SLTCAP are still very similar and close to the average value obtained by simulations using an osmostat (Figure 2A). As the salt concentration increases, the similarity between results of SLTCAP and SPLIT does it too. On the other hand, the AN approach overestimates the ion amounts, as it was already observed by Schmit et al. [8]. This issue is more pronounced in JEV capsid due to its higher charge,

leading to values away from the osmostat distribution. The same general behavior is observed within the explored salt ranges for both systems. Overall, AN estimates an effective salt concentration of about 0.05 M and 0.08 M larger than SPLIT/SLTCAP for Granulysin and JEV capsid, respectively. This effect has a consequence on the local electrolytic environment around the solutes, which is observed by the increased concentration of both ionic species within 1 nm from the protein (Figure 2B). The same is observed in the occupancy of ions nearby charged residues of the proteins (Figure 2C).

Table 1. Details of simulated system.

System	Q (e)	N_w	[NaCl] (M)	[a]N_0	AN	[b]SPLIT	SLTCAP
Granulysin	+11	6071	0	11 (Cl-)	-	-	-
			0.10	11	11 (Na+), 22 (Cl-)	6 (Na+), 17 (Cl-)	7 (Na+), 18 (Cl-)
			0.15	16	16 (Na+), 27 (Cl-)	11 (Na+), 22 (Cl-)	11 (Na+), 22 (Cl-)
			0.20	22	22 (Na+), 33 (Cl-)	17 (Na+), 28 (Cl-)	17 (Na+), 28 (Cl-)
JEV capsid	+20	7419	0	20 (Cl-)	-	-	-
			0.15	20	20 (Na+), 40 (Cl-)	10 (Na+), 30 (Cl-)	12 (Na+), 32 (Cl-)
			0.20	26	26 (Na+), 46 (Cl-)	16 (Na+), 36 (Cl-)	18 (Na+), 38 (Cl-)
			0.25	33	33 (Na+), 53 (Cl-)	23 (Na+), 43 (Cl-)	24 (Na+), 44 (Cl-)

[a] Calculated from Eq. 4.
[b] Calculated from Eq. 5. Fractional numbers are rounded up.

In conclusion, for most standard setups in proteins and other biological systems, the simple strategy of splitting the charge excess of the solute in equal parts between both electrolytic species is a very good and trivial approximation to SLTCAP. In that respect, either AN or SPLIT can be considered as particular cases of SLTCAP formulation. In practice, SPLIT can be safely applied to any system whenever $N_0/Q \geq 1$. Interestingly, the application of SPLIT or SLTCAP may not be limited to all-atom simulation but extended to coarse-grained models, even though they lump several water molecules into one or few effective interaction beads. In the particular case of the SIRAH force field developed by us [34], each water model (named WT4) represents 11 atomistic water molecules at the same density [35]. To apply either method, it is just enough to correct Eq. 4 by the actual mass of WT4 to give $N_0 \sim N_{WT4}*c_0/5$.

Finally, we would like to stress the fact that adopting this kind of simple and practical rules by the Biocomputing community should redound in increasing consistency and reproducibility of MD simulations. This is in line with initiatives oriented to maximize the effectiveness of the computational efforts carried out worldwide, aimed at facilitating the analysis and re-use data for research (https://www.eosc-portal.eu) [36,37].

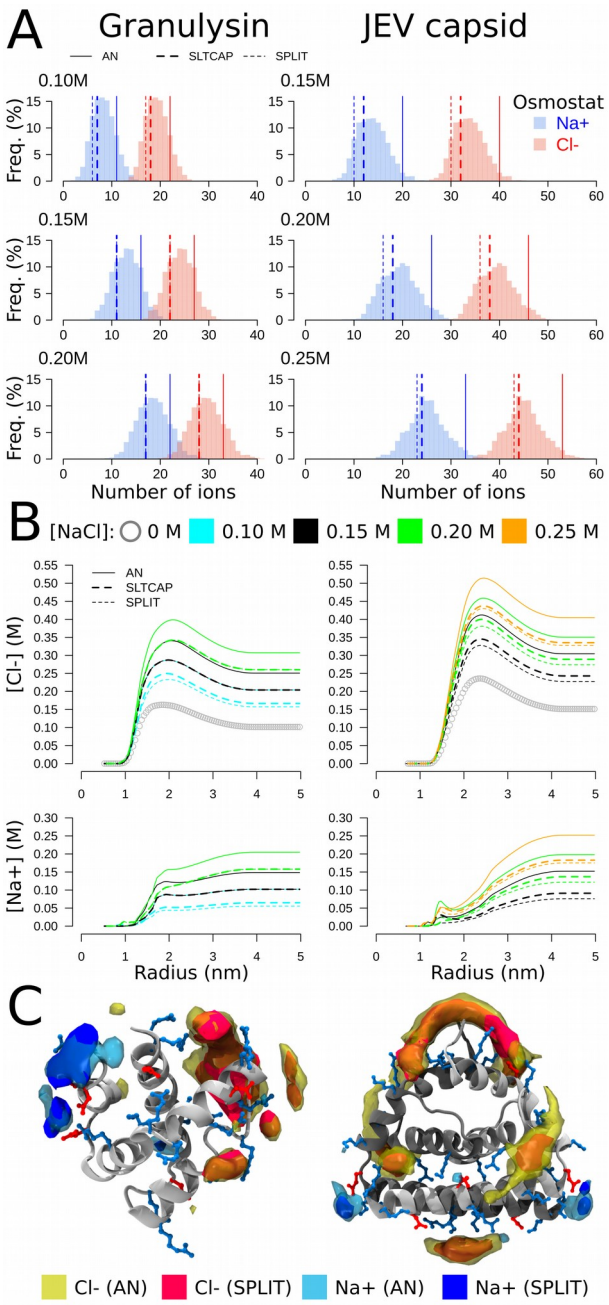


Figure 2. Representation of macroscopic salt concentrations with different neutralizing strategies. Results are shown for Granulysin (left panels) and JEV capsid (right panels). A) Number of ions in the computational box during MD simulations with an osmostat. Expected values for AN, SLTCAP and SPLIT are shown. B) Radial ion molarity from the protein's center. C) Volumetric occupancy of ions at 0.15 M during MD simulations using AN (transparent surfaces) or SPLIT (solid surfaces) setups as calculated with the VolMap plugin of VMD [38]. Surfaces are plotted at isovalues of 0.007. Basic and acidic residues are represented as blue or red balls and sticks respectively.

ACKNOWLEDGMENT

M.R.M. and S.P. belong to the SNI program of ANII. We would like to thank L. Darré, P. Dans and E. Barrera for comments on the manuscript.

FOUNDING

This work was partially funded by FOCEM (MERCOSUR Structural Convergence Fund), COF 03/11.

REFERENCES

(1) Ball, P. Water Is an Active Matrix of Life for Cell and Molecular Biology. *Proc. Natl. Acad. Sci.* **2017**, *114* (51), 13327–13335. https://doi.org/10.1073/pnas.1703781114.

(2) Alberti, S. Phase Separation in Biology. *Curr. Biol.* **2017**, *27* (20), R1097–R1102. https://doi.org/10.1016/j.cub.2017.08.069.

(3) Bhattacharjee, N.; Biswas, P. Structure of Hydration Water in Proteins: A Comparison of Molecular Dynamics Simulations and Database Analysis. *Biophys. Chem.* **2011**, *158* (1), 73–80. https://doi.org/10.1016/j.bpc.2011.05.009.

(4) Shabane, P. S.; Izadi, S.; Onufriev, A. V. General Purpose Water Model Can Improve Atomistic Simulations of Intrinsically Disordered Proteins. *J. Chem. Theory Comput.* **2019**, *15* (4), 2620–2634. https://doi.org/10.1021/acs.jctc.8b01123.

(5) Anandakrishnan, R.; Izadi, S.; Onufriev, A. V. Why Computed Protein Folding Landscapes Are Sensitive to the Water Model. *J. Chem. Theory Comput.* **2019**, *15* (1), 625–636. https://doi.org/10.1021/acs.jctc.8b00485.

(6) El Hage, K.; Hédin, F.; Gupta, P. K.; Meuwly, M.; Karplus, M. Valid Molecular Dynamics Simulations of Human Hemoglobin Require a Surprisingly Large Box Size. *eLife* **2018**, *7*, e35560. https://doi.org/10.7554/eLife.35560.

(7) Perilla, J. R.; Goh, B. C.; Cassidy, C. K.; Liu, B.; Bernardi, R. C.; Rudack, T.; Yu, H.; Wu, Z.; Schulten, K. Molecular Dynamics Simulations of Large Macromolecular Complexes. *Curr. Opin. Struct. Biol.* **2015**, *31*, 64–74. https://doi.org/10.1016/j.sbi.2015.03.007.

(8) Schmit, J. D.; Kariyawasam, N. L.; Needham, V.; Smith, P. E. SLTCAP: A Simple Method for Calculating the Number of Ions Needed for MD Simulation. *J. Chem. Theory Comput.* **2018**, *14* (4), 1823–1827. https://doi.org/10.1021/acs.jctc.7b01254.

(9) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092. https://doi.org/10.1063/1.464397.

(10) 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* (19), 8577–8593. https://doi.org/10.1063/1.470117.

(11) Hub, J. S.; de Groot, B. L.; Grubmüller, H.; Groenhof, G. Quantifying Artifacts in Ewald Simulations of Inhomogeneous Systems with a Net Charge. *J. Chem. Theory Comput.* **2014**, *10* (1), 381–390. https://doi.org/10.1021/ct400626b.

(12) Chen, W.; Shen, J. K. Effects of System Net Charge and Electrostatic Truncation on All-Atom Constant PH Molecular Dynamics. *J. Comput. Chem.* **2014**, *35* (27), 1986–1996. https://doi.org/10.1002/jcc.23713.

(13) Darden, T.; Pearlman, D.; Pedersen, L. G. Ionic Charging Free Energies: Spherical versus Periodic Boundary Conditions. *J. Chem. Phys.* **1998**, *109* (24), 10921–10935. https://doi.org/10.1063/1.477788.

(14) Freire, J. M.; Santos, N. C.; Veiga, A. S.; Poian, A. T. D.; Castanho, M. A. R. B. Rethinking the capsid proteins of enveloped viruses: multifunctionality from genome packaging to genome transfection https://febs.onlinelibrary.wiley.com/doi/abs/10.1111/febs.13274 (accessed Aug 19, 2019). https://doi.org/10.1111/febs.13274.

(15) Ross, G. A.; Rustenburg, A. S.; Grinaway, P. B.; Fass, J.; Chodera, J. D. Biomolecular Simulations under Realistic Macroscopic Salt Conditions. *J. Phys. Chem. B* **2018**, *122* (21), 5466–5486. https://doi.org/10.1021/acs.jpcb.7b11734.

(16) Luo, Y.; Roux, B. Simulation of Osmotic Pressure in Concentrated Aqueous Salt Solutions. *J. Phys. Chem. Lett.* **2010**, *1* (1), 183–189. https://doi.org/10.1021/jz900079w.

(17) Yoo, J.; Aksimentiev, A. New Tricks for Old Dogs: Improving the Accuracy of Biomolecular Force Fields by Pair-Specific Corrections to Non-Bonded Interactions. *Phys. Chem. Chem. Phys.* **2018**, *20* (13), 8432–8449. https://doi.org/10.1039/C7CP08185E.

(18) Tolmachev, D. A.; Boyko, O. S.; Lukasheva, N. V.; Martinez-Seara, H.; Karttunen, M. Overbinding and Qualitative and Quantitative Changes Caused by Simple Na+ and K+ Ions in Polyelectrolyte Simulations: Comparison of Force Fields with and without NBFIX and ECC Corrections. *J. Chem. Theory Comput.* **2019**. https://doi.org/10.1021/acs.jctc.9b00813.

(19) Ahmed, M. C.; Papaleo, E.; Lindorff-Larsen, K. How Well Do Force Fields Capture the Strength of Salt Bridges in Proteins? *PeerJ* **2018**, *6*, e4967. https://doi.org/10.7717/peerj.4967.

(20) Norambuena, T.; Melo, F. The Protein-DNA Interface Database. *BMC Bioinformatics* **2010**, *11*, 262. https://doi.org/10.1186/1471-2105-11-262.
(21) Coimbatore Narayanan, B.; Westbrook, J.; Ghosh, S.; Petrov, A. I.; Sweeney, B.; Zirbel, C. L.; Leontis, N. B.; Berman, H. M. The Nucleic Acid Database: New Features and Capabilities. *Nucleic Acids Res.* **2014**, *42* (Database issue), D114-122. https://doi.org/10.1093/nar/gkt980.
(22) Salomon Ferrer, R.; Case, D. A.; Walker, R. C. An Overview of the Amber Biomolecular Simulation Package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3* (2), 198–210. https://doi.org/10.1002/wcms.1121.
(23) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. Structure of a B-DNA Dodecamer: Conformation and Dynamics. *Proc. Natl. Acad. Sci.* **1981**, *78* (4), 2179–2183. https://doi.org/10.1073/pnas.78.4.2179.
(24) Anderson, D. H.; Sawaya, M. R.; Cascio, D.; Ernst, W.; Modlin, R.; Krensky, A.; Eisenberg, D. Granulysin Crystal Structure and a Structure-Derived Lytic Mechanism. *J. Mol. Biol.* **2003**, *325* (2), 355–365. https://doi.org/10.1016/s0022-2836(02)01234-2.
(25) Poonsiri, T.; Wright, G. S. A.; Solomon, T.; Antonyuk, S. V. Crystal Structure of the Japanese Encephalitis Virus Capsid Protein. *Viruses* **2019**, *11* (7), 623. https://doi.org/10.3390/v11070623.
(26) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.
(27) 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* (2), 926–935. https://doi.org/10.1063/1.445869.
(28) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112* (30), 9020–9041. https://doi.org/10.1021/jp8001614.
(29) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9* (9), 3878–3888. https://doi.org/10.1021/ct400314y.
(30) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; et al. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13* (7), e1005659. https://doi.org/10.1371/journal.pcbi.1005659.
(31) Wu, X.; Brooks, B. R. Self-Guided Langevin Dynamics Simulation Method. *Chem. Phys. Lett.* **2003**, *381* (3), 512–518. https://doi.org/10.1016/j.cplett.2003.10.013.
(32) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327–341. https://doi.org/10.1016/0021-9991(77)90098-5.
(33) Miyamoto, S.; Kollman, P. A. Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13* (8), 952–962. https://doi.org/10.1002/jcc.540130805.
(34) Machado, M. R.; Barrera, E. E.; Klein, F.; Sóñora, M.; Silva, S.; Pantano, S. The SIRAH 2.0 Force Field: Altius, Fortius, Citius. *J. Chem. Theory Comput.* **2019**, *15* (4), 2719–2733. https://doi.org/10.1021/acs.jctc.9b00006.
(35) Darré, L.; Machado, M. R.; Dans, P. D.; Herrera, F. E.; Pantano, S. Another Coarse Grain Model for Aqueous Solvation: WAT FOUR? *J. Chem. Theory Comput.* **2010**, *6* (12), 3793–3807. https://doi.org/10.1021/ct100379f.
(36) Riccardi, E.; Pantano, S.; Potestio, R. Envisioning Data Sharing for the Biocomputing Community. *Interface Focus* **2019**, *9* (3), 20190005. https://doi.org/10.1098/rsfs.2019.0005.
(37) Wilkinson, M. D.; Dumontier, M.; Aalbersberg, Ij. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L. B.; Bourne, P. E.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3* (1), 1–9. https://doi.org/10.1038/sdata.2016.18.
(38) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38, 27–28. https://doi.org/10.1016/0263-7855(96)00018-5.