Appendix 1

Symbols and their definitions

- x : Instantaneous configuration (positions, box vectors)
- N_{H_20}: Number of water molecules
- N_{Na^+} : Number of cations
- N_{Cl^-} : Number of anions
- N_{NaCl} : Number of salt pairs beyond minimal neutralizing ions; equal to $\min\{N_{Na^+}, N_{Cl^-}\}$
- N : Sum of total number of waters and ions in the system
- θ : Vector species labels with N elements that identifies which molecules are waters and which are ions; $\theta_i = 0$ indicates water, $\theta_i = +1$ indicates monovalent cations, and $\theta_i = -1$ indicates monovalent anions
- z : total charge number of the macromolecules in the simulation
- $n(\theta)$: total charge number of the ions in the simulation

$$n(\theta) = \sum_{i=1}^{N} \theta_i \tag{26}$$

- $U(x, \theta)$: Potential energy for a system with configuration x and water/ion identities θ, units of energy
- p : External pressure, units of energy · length^{-3}
- V : Instantaneous box volume, units of length^3
- T : Absolute temperature, units of temperature
- k_B : Boltzmann constant, units of energy · temperature^{-1}
- β : Inverse temperature ($\equiv 1/k_BT$), units of energy^{-1}
- I : Ionic strength, where instantaneous ionic strength for configuration x is given by

$$I(x, \theta) \equiv \frac{1}{2}\frac{1}{V(x)}\left(z^2 + \sum_{i=1}^{N} \theta_i^2\right) \tag{27}$$

  Note that ionic strength includes minimal neutralizing counterions in the sum.
- $\Delta\mu$: Chemical potential difference for extracting a NaCl molecule from bulk water and depositing two water molecules to bulk water; an abbreviation of $\Delta\mu_{2\cdot H_2O-NaCl}$
- $f(N_{NaCl})$: Free energy to replace $2N_{H_2O}$ water molecules with N_{NaCl} salt pairs in bulk water; an abbreviation of $f(N_{NaCl}, N, p, T)$.
- $\Delta f(N_{NaCl})$: Free energy to add one more salt pair and remove two additional water molecules in a box of water than contains N_{NaCl} salt pairs already; equal to $f(N_{NaCl} + 1) - f(N_{NaCl})$; an abbreviation of $\Delta f(N_{NaCl}, N, p, T)$
- $Z(N_{NaCl}, N, p, T)$: Isothermal-isobaric configurational partition function

$$Z(N_{NaCl}, N, p, T) \equiv \int dx\, e^{-\beta[U(x;N_{NaCl})+pV(x)]} \tag{28}$$

- $\Xi(\Delta\mu, N, p, T)$: Semigrand-isothermal-isobaric configurational partition function expressed as a sum over all θ

$$\Xi(\Delta\mu, N, p, T) = \sum_{\theta} \delta(n(\theta), -z) \int \mathrm{d}x\, e^{-\beta[U(x,\theta)+pV(x)+\Delta\mu N_{NaCl}(\theta)]}, \tag{29}$$

  and expressed as a sum of number of ions and water molecules

$$\Xi(\Delta\mu, N, p, T) \equiv \sum_{N_{\mathrm{NaCl}}=0}^{N/2} \frac{N!}{N_{\mathrm{Na^+}}!N_{\mathrm{Cl^-}}!N_{\mathrm{H_2O}}!} Z(N_{\mathrm{NaCl}}, N, p, T)\, e^{\beta\Delta\mu N_{\mathrm{NaCl}}}, \tag{30}$$

where $N_{\mathrm{NaCl}} = \min\{N_{\mathrm{Na^+}}, N_{\mathrm{Cl^-}}\}$ and $N = N_{\mathrm{Na^+}} + N_{\mathrm{Cl^-}} + N_{\mathrm{H_2O}}$. The upper bound of the summation—valid when $z = 0$ and N is even—is required as two water molecules are removed for every N_{NaCl}.

- $\pi(x, \theta; N, p, T, \mu)$: Semigrand-isothermal-isobaric probability density with charge neutrality constraint

$$\pi(x, \theta; \Delta\mu, N, p, T) = \frac{1}{\Xi(\Delta\mu, N, p, T)}\, \delta(n(\theta), -z)\, e^{-\beta[U(x,\theta)+pV(x)+\Delta\mu N_{\mathrm{NaCl}}(\theta)]}, \tag{31}$$

where the dependence of $\pi(x, \theta; \Delta\mu, N, p, T)$ on z is omitted for brevity

- $\langle A \rangle_{\Delta\mu,N,p,T}$: Expectation of $A(x, \theta)$ in $(\Delta\mu, N, p, T)$ ensemble

$$\langle A \rangle_{\Delta\mu,N,p,T} \equiv \frac{1}{\Xi(\Delta\mu, N, p, T)} \sum_{\theta} \delta(n(\theta), -z) \int dx\, A(x, \theta)\, e^{-\beta[U(x,\theta)+pV(x)+\Delta\mu N_{\mathrm{NaCl}}(\theta)]} \tag{32}$$

- $\langle A \rangle_{N_{\mathrm{NaCl}},N,p,T}$: Expectation of $A(x)$ in $(N_{\mathrm{NaCl}}, N, p, T)$ ensemble

$$\langle A \rangle_{N_{\mathrm{NaCl}},N,p,T} \equiv \frac{1}{Z(N_{\mathrm{NaCl}}, N, p, T)} \int dx\, A(x)\, e^{-\beta[U(x;N_{\mathrm{NaCl}})+pV(x)]} \tag{33}$$

Appendix 2

Salt concentration in the thermodynamic limit

The purpose of this section is to derive an expression that relates the chemical potential to the salt concentration in a macroscopic saline reservoir (equation 19). This relationship is used in the calibration of our osmostat. The derivation will proceed by first, justifying the macroscopic concentration as the thermodynamic limit of the mean concentration, and second, rewriting the resultant expression in a manner that is amenable to computation.

The mean concentration in the thermodynamic limit

Following the definition of the concentration given in equation 20, the mean salt concentration in the semigrand ensemble considered here is given by

$$\langle c \rangle_{\Delta\mu,N,p,T} = \left\langle \frac{N_{\mathrm{NaCl}}(\theta)}{V(x)} \right\rangle_{\Delta\mu,N,p,T}. \tag{34}$$

We seek an approximation to this expression that it is appropriate for large, macroscopic amounts of liquid saline. For brevity, all expectation values with respect to the thermodynamic ensemble $(\Delta\mu, N, p, T)$ in this section will henceforth be abbreviated as $\langle\cdot\rangle$.

The concentration is a function of two correlated random variables, the number of salt pairs $N_{\mathrm{NaCl}}(\theta)$ and the total volume $V(x)$. A common way to approximate the expectation value, or mean, of a function of random variables is to perform a Taylor expansion about the mean of the arguments. The Taylor expansion (up to the second-order) of the function $g(a,b)$ about the means $\langle a\rangle$ and $\langle b\rangle$, is

$$\begin{aligned} g(a,b) &= g\big(\langle a\rangle,\langle b\rangle\big) + \frac{\partial g}{\partial a}\Big|_{\langle a\rangle,\langle b\rangle}\big(a-\langle a\rangle\big) + \frac{\partial g}{\partial b}\Big|_{\langle a\rangle,\langle b\rangle}\big(b-\langle b\rangle\big) + \frac{1}{2}\frac{\partial^2 g}{\partial a^2}\Big|_{\langle a\rangle,\langle b\rangle}\big(a-\langle a\rangle\big)^2 \\ &+ \frac{1}{2}\frac{\partial^2 g}{\partial b^2}\Big|_{\langle a\rangle,\langle b\rangle}\big(b-\langle b\rangle\big)^2 + \frac{\partial^2 g}{\partial a\,\partial b}\Big|_{\langle a\rangle,\langle b\rangle}\big(a-\langle a\rangle\big)\big(b-\langle b\rangle\big) + \ldots \end{aligned} \tag{35}$$

This expansion is particularly useful because the first order terms of the expanded mean $\langle g(a,b)\rangle$ are zero i.e. $\langle a-\langle a\rangle\rangle = 0$ and $\langle b-\langle b\rangle\rangle = 0$. Hence, truncating the expansion to the second order leaves us with the approximation

$$\begin{aligned} \big\langle g(a,b)\big\rangle &\approx g(\langle a\rangle,\langle b\rangle) + \frac{1}{2}\frac{\partial^2 g}{\partial a^2}\Big|_{\langle a\rangle,\langle b\rangle}\big\langle (a-\langle a\rangle)^2\big\rangle + \frac{1}{2}\frac{\partial^2 g}{\partial b^2}\Big|_{\langle a\rangle,\langle b\rangle}\big\langle (b-\langle b\rangle)^2\big\rangle \\ &+ \frac{\partial^2 g}{\partial a\,\partial b}\Big|_{\langle a\rangle,\langle b\rangle}\big\langle (a-\langle a\rangle)(b-\langle b\rangle)\big\rangle \\ &= g(\langle a\rangle,\langle b\rangle) + \frac{1}{2}\frac{\partial^2 g}{\partial a^2}\Big|_{\langle a\rangle,\langle b\rangle}\mathrm{Var}(a) + \frac{1}{2}\frac{\partial^2 g}{\partial b^2}\Big|_{\langle a\rangle,\langle b\rangle}\mathrm{Var}(b) + \frac{\partial^2 g}{\partial a\,\partial b}\Big|_{\langle a\rangle,\langle b\rangle}\mathrm{Cov}(a,b), \end{aligned} \tag{36}$$

where $\mathrm{Var}(a)$ and $\mathrm{Cov}(a,b)$ denote the variance and covariance, respectively. Returning to the salt concentration, we relate c to the above with $g(N_{\mathrm{NaCl}}, V) = N_{\mathrm{NaCl}}/V$, and evaluate the partial derivatives to find that

$$\langle c\rangle \approx \frac{\langle N_{\mathrm{NaCl}}\rangle}{\langle V\rangle} + \frac{\langle N_{\mathrm{NaCl}}\rangle}{\langle V\rangle^3}\mathrm{Var}(V) - \frac{1}{\langle V\rangle^2}\mathrm{Cov}(V, N_{\mathrm{NaCl}}). \tag{37}$$

The leading term $\langle N_{\mathrm{NaCl}}\rangle/\langle V\rangle$ is the macroscopic expression that we seek. Thus, we require that the variance and covariance terms vanish in the thermodynamic limit. To show that they indeed do, we exploit the useful correspondence between partial derivatives and covariance in statistical thermodynamics. First, note that

$$\begin{aligned} \mathrm{Var}(V) &= (k_B T)^2 \frac{\partial^2 \ln(\Xi)}{\partial p^2} \\ &= -k_B T \frac{\partial\langle V\rangle}{\partial p}, \end{aligned} \tag{38}$$

where $\Xi \equiv \Xi(\Delta\mu, N, p, T)$ and is defined in equation 7. Also, note that

$$\begin{aligned} \mathrm{Cov}(V, N_{\mathrm{NaCl}}) &= (k_B T)^2 \frac{\partial^2 \ln(\Xi)}{\partial p \, \partial \Delta\mu} \\ &= k_B T \frac{\partial \langle V \rangle}{\partial \Delta\mu}. \end{aligned} \tag{39}$$

Second, we make use of the isothermal compressibility

$$\kappa_T \equiv -\frac{1}{\langle V \rangle} \frac{\partial \langle V \rangle}{\partial p}, \tag{40}$$

and introduce the isothermal susceptibility of the volume with respect to the chemical potential

$$\chi_T \equiv \frac{1}{\langle V \rangle} \frac{\partial \langle V \rangle}{\partial \Delta\mu}, \tag{41}$$

The susceptibilities κ_T and χ_T are bulk properties that measure the relative amount the volume of a system responds to changes in pressure and chemical potential, respectively. They are intensive quantities, such that they do not scale with the size of the system. These allow us to re-write the approximation of the mean concentration (equation 37) as

$$\langle c \rangle \approx \frac{\langle N_{\mathrm{NaCl}} \rangle}{\langle V \rangle} - \frac{1}{k_B T} \frac{\langle N_{\mathrm{NaCl}} \rangle}{\langle V \rangle^2} \kappa_p - \frac{1}{k_B T} \frac{1}{\langle V \rangle} \chi_T. \tag{42}$$

To proceed, note that in the second term, both N_{NaCl} and $\langle V \rangle$ are extensive, and rise in proportion to the total number of molecules in the system N. Thus, approximating the mean concentration as $\langle N_{\mathrm{NaCl}} \rangle / \langle V \rangle$ incurs an error that is $\mathcal{O}(\langle V \rangle^{-1})$, which tends to zero in the thermodynamic limit. We therefore define the macroscopic concentration of a saline reservoir as

$$\langle \hat{c} \rangle \equiv \frac{\langle N_{\mathrm{NaCl}} \rangle}{\langle V \rangle}. \tag{43}$$

We require the macroscopic concentration to be amenable to computational analysis

While the expression for the macroscopic concentration above does not appear immediately useful, we now show how $\langle \hat{c} \rangle$ can be calculated for wide range of applied chemical potentials by pre-calculating the free energies to insert salt into a system, $f(N_{\mathrm{NaCl}})$ ($\equiv f(N_{\mathrm{NaCl}}, \Delta\mu, N, p, T)$), and the average volume as a function of the number of salt pairs, $\langle V \rangle_{N_{\mathrm{NaCl}}}$ ($\equiv \langle V \rangle_{N_{\mathrm{NaCl}}, N, p, T}$).

To begin, it is useful to expand the definition of $\langle N_{\mathrm{NaCl}} \rangle$ given by equation 17 into

$$\langle N_{\mathrm{NaCl}} \rangle = \frac{\sum_{N_{\mathrm{NaCl}}=0} N_{\mathrm{NaCl}} \, e^{-f(N_{\mathrm{NaCl}}) + \beta\Delta\mu N_{\mathrm{NaCl}}}}{\sum_{N_{\mathrm{NaCl}}=0} e^{-f(N_{\mathrm{NaCl}}) + \beta\Delta\mu N_{\mathrm{NaCl}}}}. \tag{44}$$

Next, we derive an expression for $\langle V \rangle$ that will cancel with the denominator of equation 44 when evaluating $\langle \hat{c} \rangle$. Using the representation of the semigrand density given by equation 8, the mean

volume is given by

$$\begin{aligned}
\langle V \rangle &= \frac{\sum_{N_{\mathrm{NaCl}}=0} \int dx\, V(x)\, e^{-\beta(U(x;N_{\mathrm{NaCl}})+pV(x)+\Delta\mu N_{\mathrm{NaCl}}(\theta))}}{\sum_{N_{\mathrm{NaCl}}=0} e^{-f(N_{\mathrm{NaCl}})+\beta\Delta\mu N_{\mathrm{NaCl}}}} \\
&= \frac{\sum_{N_{\mathrm{NaCl}}=0} e^{\beta\Delta\mu N_{\mathrm{NaCl}}} \int dx\, V(x)\, e^{-\beta(U(x;N_{\mathrm{NaCl}})+pV(x))}}{\sum_{N_{\mathrm{NaCl}}=0} e^{-f(N_{\mathrm{NaCl}})+\beta\Delta\mu N_{\mathrm{NaCl}}}} \\
&= \frac{\sum_{N_{\mathrm{NaCl}}=0} e^{\beta\Delta\mu N_{\mathrm{NaCl}}} \int dx\, V(x)\, e^{-\beta(U(x;N_{\mathrm{NaCl}})+pV(x))} \cdot \int dx'\, e^{-\beta(U(x';N_{\mathrm{NaCl}})+pV(x'))}}{\sum_{N_{\mathrm{NaCl}}=0} e^{-f(N_{\mathrm{NaCl}})+\beta\Delta\mu N_{\mathrm{NaCl}}} \cdot \int dx''\, e^{-\beta(U(x'';N_{\mathrm{NaCl}})+pV(x''))}} \\
&= \frac{\sum_{N_{\mathrm{NaCl}}=0} e^{\beta\Delta\mu N_{\mathrm{NaCl}}} \langle V \rangle_{N_{\mathrm{NaCl}}} \cdot e^{-f(N_{\mathrm{NaCl}})}}{\sum_{N_{\mathrm{NaCl}}=0} e^{-f(N_{\mathrm{NaCl}})+\beta\Delta\mu N_{\mathrm{NaCl}}}} \\
&= \frac{\sum_{N_{\mathrm{NaCl}}=0} \langle V \rangle_{N_{\mathrm{NaCl}}} e^{-f(N_{\mathrm{NaCl}})+\beta\Delta\mu N_{\mathrm{NaCl}}}}{\sum_{N_{\mathrm{NaCl}}=0} e^{-f(N_{\mathrm{NaCl}})+\beta\Delta\mu N_{\mathrm{NaCl}}}},
\end{aligned} \tag{45}$$

where the third and fourth line exploit the definition of the ensemble average for a fixed N_{NaCl}. Inserting the expressions for the average number of salt pairs (equation 44) and the average volume (equation 45) into the macroscopic concentration (equation 43), we arrive at

$$\langle \hat{c} \rangle = \frac{\sum_{N_{\mathrm{NaCl}}=0} N_{\mathrm{NaCl}}\, e^{-f(N_{\mathrm{NaCl}})+\beta\Delta\mu N_{\mathrm{NaCl}}}}{\sum_{N_{\mathrm{NaCl}}=0} \langle V \rangle_{N_{\mathrm{NaCl}}}\, e^{-f(N_{\mathrm{NaCl}})+\beta\Delta\mu N_{\mathrm{NaCl}}}},$$

which is the same as equation 19 from the main text. Pertinently, the denominators in equations 44 and 45 have canceled, which greatly simplifies the evaluation of the macroscopic concentration for a given $\Delta\mu$.

The magnitude of salt fluctuations

The concentration of salt fluctuates in osmostat simulations. This section briefly outlines how one would expect the magnitude of salt fluctuations to vary with the size of the system based on statistical mechanical principles. By differentiating equation 17, one can show that the variance of the number of salt pairs N_{NaCl} is proportional to the gradient of $\langle N_{\mathrm{NaCl}} \rangle$ with respect to the chemical potential $\Delta\mu$, specifically

$$\mathrm{Var}(N_{\mathrm{NaCl}}) = k_B T \frac{\partial \langle N_{\mathrm{NaCl}} \rangle}{\partial \Delta\mu}. \tag{46}$$

By dividing both sides by $\langle N_{\mathrm{NaCl}} \rangle$, i.e.

$$\frac{1}{\langle N_{\mathrm{NaCl}} \rangle} \mathrm{Var}(N_{\mathrm{NaCl}}) = \frac{1}{\langle N_{\mathrm{NaCl}} \rangle} k_B T \frac{\partial \langle N_{\mathrm{NaCl}} \rangle}{\partial \Delta\mu}, \tag{47}$$

reveals that $\frac{1}{\langle N_{\mathrm{NaCl}} \rangle} \mathrm{Var}(N_{\mathrm{NaCl}})$ is proportional to the *relative* change in the mean of N_{NaCl} in response to altering the chemical potential. As the right-hand-side of the above equation is an intensive quantity, $\frac{1}{\langle N_{\mathrm{NaCl}} \rangle} \mathrm{Var}(N_{\mathrm{NaCl}})$ is also an intensive, implying that

$$\mathrm{Var}(N_{\mathrm{NaCl}}) \propto N_{\mathrm{NaCl}}. \tag{48}$$

Therefore, the scale of the fluctuations in salt amount, as measured by the standard deviation, grows as $\langle N_{\mathrm{NaCl}} \rangle^{1/2}$.

In contrast to the amount of salt, the size of the fluctuations of salt concentration *decreases* with the size of aqueous systems. Water is a highly incompressible fluid, such that small changes in pressure have a very small effect on the volume of aqueous systems. From equations 38 and 40, a low isothermal compressibility implies that the variance of the volume is small with respect to the mean volume (i.e. the relative variance). Assuming that the relative variance of the volume is smaller

than the relative variance of the number of salt pairs, one can use the same approach as that of equation 35 to show that

$$\mathrm{Var}(c) = \mathrm{Var}\left(\frac{N_{\mathrm{NaCl}}}{V}\right) \tag{49}$$

$$\approx \frac{1}{\langle V\rangle^2}\mathrm{Var}(N_{\mathrm{NaCl}}) \tag{50}$$

Using the fact that, for bulk-like water, $\langle V\rangle \propto \langle N_{\mathrm{H_2O}}\rangle \propto \langle N_{\mathrm{NaCl}}\rangle$ along with equation 48, we arrive at $\mathrm{Var}(c) \sim \langle N_{\mathrm{NaCl}}\rangle^{-1}$ for systems with large amounts of water. Thus, the standard deviation of the salt concentration scales like $\langle N_{\mathrm{H_2O}}\rangle^{-1/2}$ or $\langle N_{\mathrm{NaCl}}\rangle^{-1/2}$ for a fixed chemical potential.

Appendix 3

Algorithmic implementation of the osmostat

This section describes the Metropolis-Hastings procedure from Saltswap [0.52] used to insert and delete salt. Insertion and deletion moves were enhanced with NCMC[32]. To describe its implementation of NCMC within SaltSwap, a more compressed notation is used compared to the original publication. For a more general and detailed exposition on NCMC, we refer readers to the original manuscript.

The osmostat move begins with the random choice of whether to insert or delete salt. The protocol is denoted $\Lambda \in \{ \Lambda_{\text{insert}}, \Lambda_{\text{delete}} \}$, and the time reversed protocol is denoted $\tilde{\Lambda}$, where $\tilde{\Lambda}_{\text{insert}} = \Lambda_{\text{delete}}$ and $\tilde{\Lambda}_{\text{delete}} = \Lambda_{\text{insert}}$. The probability to insert or delete a salt pair, $P(\Lambda|N_{\text{NaCl}})$, depends on the number of salt molecules, N_{NaCl}, in the system in the following way:

$$P(\Lambda_{\text{insert}}|N_{\text{NaCl}}) = \begin{cases} 1 & \text{if } N_{\text{NaCl}} = 0; \\ 1/2 & \text{if } 0 < N_{\text{NaCl}} < N_{\text{NaCl,max}}, \\ 0 & \text{if } N_{\text{NaCl}} = N_{\text{NaCl,max}}; \end{cases} \tag{51}$$

$$P(\Lambda_{\text{delete}}|N_{\text{NaCl}}) = \begin{cases} 0 & \text{if } N_{\text{NaCl}} = 0; \\ 1/2 & \text{if } 0 < N_{\text{NaCl}} < N_{\text{NaCl,max}}, \\ 1 & \text{if } N_{\text{NaCl}} = N_{\text{NaCl,max}}; \end{cases} \tag{52}$$

where for all simulations except the SAMS calibration simulations, $N_{\text{NaCl,max}} = \frac{1}{2}(N - (N \mod 2))$ was chosen as two water molecules are required for the insertion of a Na^+ and Cl^- pair. In the SAMS calibration simulations, $N_{\text{NaCl,max}}$ was set to twenty. The particular choices of $P(\Lambda_{\text{delete}}|N_{\text{NaCl}})$ and $P(\Lambda_{\text{insert}}|N_{\text{NaCl}})$ ensure that insertions are always attempted when there is no salt in the system, and deletions are always attempted when the number of salt pairs has reached maximum capacity.

For the insertion of salt, any two water molecules could be selected for transformation into Na^+ and Cl^-. Similarly, for the removal of salt, any Na^+ ion and Cl^- ion could be selected for transformation into two water molecules. Formally, let $S(N)$ denote the set $\{1, 2, ..., N\}$, i.e. the set of indices for all water molecules and ions. For salt insertion, the index of candidate Na^+ ion was a random uniform sample from the set $\{i \in S(N) : \theta_i = 0\}$ and the index of the Cl^- ion was a random uniform sample from the set $\{j \in S(N) : \theta_j = 0, i \neq j\}$. For salt removal, indices were selected randomly and uniformally from the sets $\{i \in S(N) : \theta_i = +1\}$ and $\{j \in S(N) : \theta_j = -1\}$. As indices were chosen with equal probability within each set of possible candidates, the ratio of selection probabilities for molecule indices for forward and reverse protocols are given by

$$\frac{P(i, j|\Lambda_{\text{insert}})}{P(i, j|\Lambda_{\text{delete}})} = \frac{N_{\text{H}_2\text{O}}(N_{\text{H}_2\text{O}} - 1)}{(N_{\text{Na}^+} + 1)(N_{\text{Cl}^-} + 1)}, \tag{53}$$

and

$$\frac{P(i, j|\Lambda_{\text{delete}})}{P(i, j|\Lambda_{\text{insert}})} = \frac{N_{\text{Na}^+} N_{\text{Cl}^-}}{(N_{\text{H}_2\text{O}} + 1)(N_{\text{H}_2\text{O}} + 2)} \tag{54}$$

Following the choice of protocol and pair of molecules that would be transmuted, NCMC was used to enhance the efficiency of the insertion or deletion attempt. This implementation of NCMC consists of a fixed series of *perturbation* and *propagation* kernels over a fixed alchemical path. For both insertion and deletion moves, the alchemical path is a linear interpolation the nonbonded parameters of the water model and the ions. This particular alchemical path ensured that charge neutrality was maintained throughout the NCMC procedure.

The alchemical path is broken up into T segments that are uniformally spaced with respect to the nonbonded parameters. At state t, the configuration of the system will be denoted as x_t and the values of the nonbonded parameters for molecules i and j will be denoted as λ_t^{ij}. A single NCMC *step* corresponds to the application of the perturbation kernel followed by a the propagation kernel. When in state t, the perturbation kernel updates the nonbonded parameters $(x_t, \lambda_t^{ij}) \to (x_t, \lambda_{t+1}^{ij})$, and the propagation kernel updates the configuration $(x_t, \lambda_{t+1}^{ij}) \to (x_{t+1}, \lambda_{t+1}^{ij})$. Each propagation kernel consists of K steps of Langevin dynamics using the parameters described in Simulation Details. A propagation kernel is also applied to the system before the first perturbation kernel to ensure the time symmetry of the protocol. The instantaneous change in the potential energy that results from the application of the perturbation kernel is recorded for each NCMC step and summed to produce the total work performed on the system by the protocol:

$$W^{ij}(X_T, \Lambda) = \sum_{t=1}^{T} U(x_t, \lambda_{t+1}^{ij}) - U(x_t, \lambda_t^{ij}), \tag{55}$$

where the nonequilibrium trajectory $X_T \equiv (x_0, x_1, ..., x_T)$. The difference between the protocol work and applied chemical potential $\Delta\mu$, along with the move proposal probabilities, determines whether a move is accepted or rejected. For the insertion of salt $\Delta\mu(\Lambda_{\text{insert}}) = 2\mu_{\text{H}_2\text{O}} - \mu_{\text{NaCl}}$, and for the deletion of salt $\Delta\mu(\Lambda_{\text{delete}}) = 2\mu_{\text{NaCl}} - \mu_{\text{H}_2\text{O}}$. Attempts are accepted with the following probability

$$A^{ij}(X_T, \Lambda) = \min\left\{1, \frac{P(i,j|\tilde{\Lambda})P(\tilde{\Lambda}|\tilde{N}_{\text{NaCl}})}{P(i,j|\Lambda)P(\Lambda|N_{\text{NaCl}})} \exp\left(-\beta W^{ij}(X_T, \Lambda) + \beta\Delta\mu(\Lambda)\right)\right\}. \tag{56}$$

To preserve pathwise detailed balance, velocities were reversed upon acceptance. If a move is accepted, θ_i and θ_j are updated to reflect the new molecule identities.

Pseudo-code for the NCMC osmostat with molecular dynamics

This section contains the pseudo-code of the production osmostat simulations.

Begin algorithm

Choose a macroscopic salt concentration $\hat{c}$.
Infer the chemical potential $\Delta\mu$ by inverting equation 19.
Initialize position and velocity (x_0, v_0), state vector θ_0, and maximum number of iterations M.
for $i \in \{1, 2, ..., M\}$ **do**
    Sample conformations
    Perform 4 ps of Langevin integration with a fixed amount of salt:
    $(x_i^*, v_i^*) \leftarrow \text{Integrate}\big((x_{i-1}, v_{i-1}), 4\text{ ps}\big)$.
    Sample salt concentration
    Randomly select whether to add or remove salt as well as which molecules will be transmuted.
    Define the trial state vector as θ^*.
    Define initial and final nonbonded parameters: $(q_{\text{initial}}, \sigma_{\text{initial}}, \epsilon_{\text{initial}})$ and $(q_{\text{final}}, \sigma_{\text{final}}, \epsilon_{\text{final}})$.
    procedure NCMC($(q_{\text{initial}}, \sigma_{\text{initial}}, \epsilon_{\text{initial}})$,$(q_{\text{initial}}, \sigma_{\text{initial}}, \epsilon_{\text{initial}})$, (x_i^*, v_i^*), θ^*)
        Initialize variables, including protocol work W:
        $W^0 \leftarrow 0$
        $(q^0, \sigma^0, \epsilon^0) \leftarrow (q_{\text{initial}}, \sigma_{\text{initial}}, \epsilon_{\text{initial}})$
        $(x_i^0, v_i^0) \leftarrow \text{Integrate}\big((x_i^*, v_i^*), 20\text{ fs}\big)$
        for $k \in \{1, 2, ..., 1000\}$ **do**
            Linear interpolation of the nonbonded parameters:
            $f^k = k/1000$
            for all atoms in the molecule **do**

$q^k \leftarrow (1 - f^k)q_{\text{initial}} + f^k q_{\text{final}}$
$\sigma^k \leftarrow (1 - f^k)\sigma_{\text{initial}} + f^k \sigma_{\text{final}}$
$\epsilon^k \leftarrow (1 - f^k)\epsilon_{\text{initial}} + f^k \epsilon_{\text{final}}$
end for
Update the protocol work:
$W^k \leftarrow W^{k-1} + U(x_i^{k-1}; q^k, \sigma^k, \epsilon^k) - U(x_i^{k-1}; q^{k-1}, \sigma^{k-1}, \epsilon^{k-1})$
Propagate the system:
$(x_i^k, v_i^k) \leftarrow \text{Integrate}\left((x_i^{k-1}, v_i^{k-1}), 20\ \text{fs}\right)$
end for
Accept or reject using acceptance criterion $A(W^k, \Delta\mu, \theta^*)$
if Accept move **then**
Keep final positions and state vector but reverse velocities:
$(x_i, v_i) \leftarrow (x_i^k, -v_i^k)$
$\theta_i \leftarrow \theta^*$
else
Return positions, velocities and the state vector to after equilibrium sampling:
$(x_i, v_i) \leftarrow (x_i^*, v_i^*)$
$\theta_i \leftarrow \theta_{i-1}$
end if
end procedure
end for
End algorithm

Appendix 4

Validation: Ideal Mixing with the osmostat

In the Results section, Figure 4 *top left* indicates that the chemical potential has been properly calibrated, and Figure 6 shows that the osmostat produces samples that are concordant with physical-chemical intuition. In this section, we apply our osmostat to sample ideal mixing to provide further validation of the SaltSwap code base. Ideal mixing can be simulated with our osmostat by ensuring that salt insertion and deletion accrue no protocol work. This is implemented by using the same forcefield parameters for Na^+ and Cl^- as the water model. As our osmostat also gives the ions the same mass as water, the "ions" sampled over in this section are identical to water except for their labeling.

To validate the sampling of the osmostat, we require an analytical relationship between the chemical potential $\Delta\mu$ and the numbers of salt N_{NaCl} and water molecules $N_{\mathrm{H_2O}}$. The chemical potential used in our osmostat is the difference between the chemical potential of water multiplied by two and Na^+ and Cl^-:

$$\Delta\mu = 2\mu_{\mathrm{H_2O}} - \mu_{\mathrm{Na^+}} - \mu_{\mathrm{Cl^-}}. \tag{57}$$

In order to relate $\Delta\mu$ to N_{NaCl} and $N_{\mathrm{H_2O}}$, we will first consider a solution of water and ions in the (N, p, T) ensemble with fixed particle identities, and then relate the result to the $(\Delta\mu, N, p, T)$ ensemble. For this fixed identity solution, let $N = N_{\mathrm{H_2O}} + N_{\mathrm{Na^+}} + N_{\mathrm{Cl^-}}$ and $N_{\mathrm{Na^+}} = N_{\mathrm{Cl^-}}$. In the (N, p, T) ensemble, the chemical potential for a species s can be expressed as

$$\mu(N, p, T) = \mu_s^o - kT\ln(x_s\gamma_s(N, p, T)), \tag{58}$$

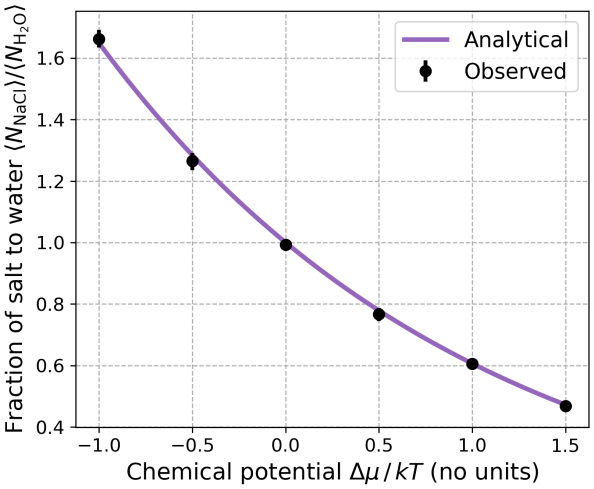
where μ_s^o is the chemical potential of s in some reference state, x_s is the mole fraction of s, and $\gamma_s(N, p, T)$ is the activity coefficient of s. In general, the chemical potential is also dependent on the composition of the system. When Na^+ and Cl^- have the same forcefield parameters and mass as water (i.e they are physically identical), the reference state and activity coefficients must be the same. So using equation 58 and 57 we have

$$\begin{aligned}\Delta\mu(N, p, T) \quad &= 2kT\ln(x_{\mathrm{H_2O}}) - kT\ln(x_{\mathrm{Na^+}}) - kT\ln(x_{\mathrm{Cl^-}}).\\ &= 2kT\ln(x_{\mathrm{H_2O}}) - 2kT\ln(x_{\mathrm{NaCl}})\\ &= 2kT\ln\left(\frac{N_{\mathrm{H_2O}}}{N_{\mathrm{NaCl}}}\right)\end{aligned} \tag{59}$$

where the second line follows from the fact that there are equal numbers of Na^+ and Cl^- ions. In the semigrand canonical $(\Delta\mu, N, p, T)$ ensemble that is sampled by our osmostat, the chemical potential $\Delta\mu$ is a controlled by the user. As this conjugate to the number of salt pairs, equation 59 will apply to the averages $\langle N_{\mathrm{NaCl}}\rangle_{\Delta\mu,\mathrm{N,p,T}}$ and $\langle N_{\mathrm{H_2O}}\rangle_{\Delta\mu,\mathrm{N,p,T}}$, so that we have

$$\frac{\langle N_{\mathrm{NaCl}}\rangle_{\Delta\mu,\mathrm{N,p,T}}}{\langle N_{\mathrm{H_2O}}\rangle_{\Delta\mu,\mathrm{N,p,T}}} = e^{-\frac{1}{2}\beta\Delta\mu}. \tag{60}$$

To test whether our osmostat correctly samples the average salt to water ratio given in equation 60, ideal mixing simulations were performed using SaltSwap on a small box of TIP3P water containing five hundred molecules for a range of chemical potentials. Ten thousand insertion and deletion attempts were made for salt pairs that had the same forcefield parameters as water. Only one perturbation step was used for the ideal NCMC insertion and deletion and the configuration of the system was not propagated during attempts. Figure 1 shows that there is excellent agreement between the relationship predicted by equation 60 and the simulation data.

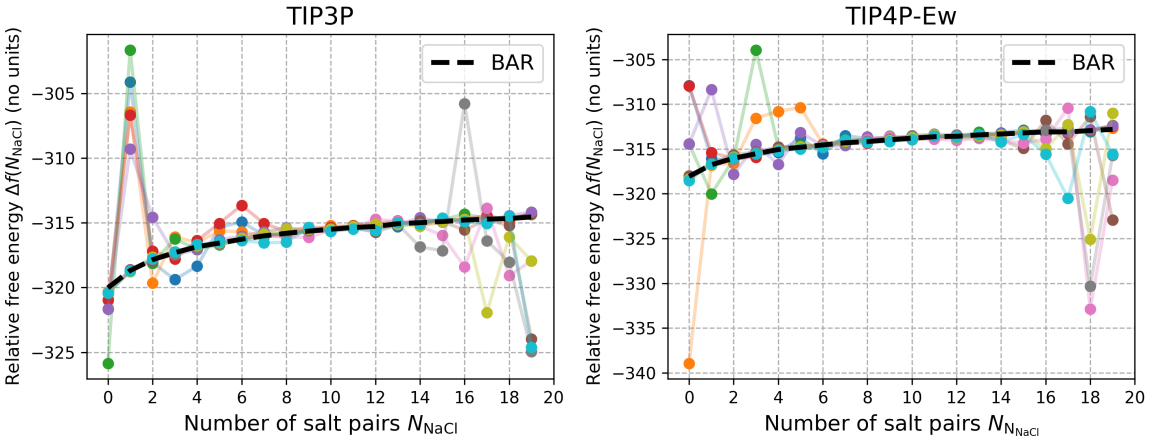


Appendix 4 Figure 1. Validating the osmostat by comparing the observed average salt-water fractions to analytical values for ideal mixing. The relationship between the chemical potential and fraction of average number of salt pairs to water molecules is known exactly for ideal mixing, and is given by equation 60. Ideal mixing was implemented for the osmostat by giving the ions the same forcefield parameters as water. For each simulation at a chemical potential, the equilibration time and statistical inefficiency for the average number of salt pairs $\langle N_{\mathrm{NaCl}} \rangle_{\Delta\mu,\mathrm{N,p,T}}$ and water molecules $\langle N_{\mathrm{H_2O}} \rangle_{\Delta\mu,\mathrm{N,p,T}}$ was determined using the timeseries module of pymbar[75]. The automatically determined equilibration times ranged from 361 and 723 insertion or deletion attempts. Effectively independent samples were extracted using the statistical inefficiency, and the means and 95% confidence intervals were estimated using bootstrap analysis.

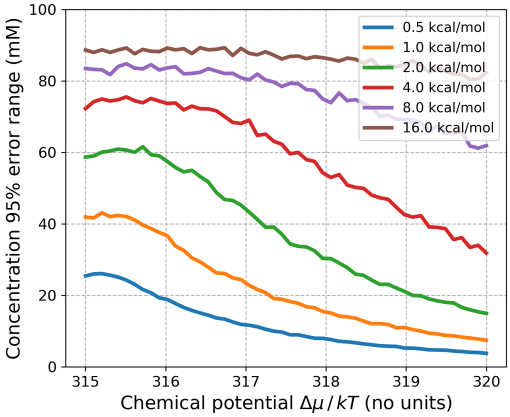
It was also verified that the protocol work was effectively zero for the ideal NCMC transformations. While the protocol work should be exactly zero, the numerical imprecision of our implementation meant this could not always be achieved. The average protocol work for the transformations shown in Figure 1 (which were performed on a CPU Intel Core i7 with one perturbation step) was 1×10^{-7} kT with a maximum absolute value of 8×10^{-5} kT. The NCMC protocol used throughout this study has one thousand perturbation steps and ten propagation steps per perturbation. With this protocol, the average protocol work was estimated using one thousand attempts on a GTX1080 GPU to be 2×10^{-8} kT with a maximum absolute value of 5×10^{-4} kT.

Appendix 5

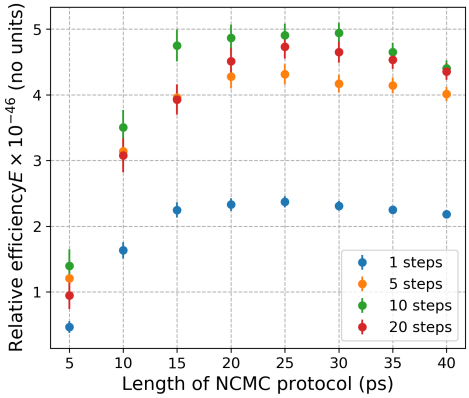
Supplementary figures



Appendix 5 Figure 1. A Comparison of the salt insertion free energies as estimated by SAMS and BAR. The individual SAMS estimates from ten repeats of the relative free energy $\Delta f(N_{\mathrm{NaCl}})$ to insert an Na^+ and Cl^- and remove two water molecules in boxes of TIP3P (left) and TIP4P-Ew (right) for each SAMS simulations. Each color represents an estimate of $\Delta f(N_{\mathrm{NaCl}})$ from each repeat. The relative free energy as calculated by BAR using all the SAMS simulation data is shown for reference (dotted black line). Five of the SAMS repeats were started with the maximum of 20 salt pairs in the system, and the other five started with none. The significant variation between the individual SAMS repeats is due to the rapid accumulation of the biasing potential in the initial stages of the algorithm. This biased the sampling away from the initial states of the simulations and prevented the uniform sampling over the salt numbers.



Appendix 5 Figure 2. The statistical uncertainty of the predicted macroscopic concentration as a function of the chemical potential for different standard errors of the free energies $f(N_{\mathrm{NaCl}})$ in a box of 887 TIP3P water molecules. Using the data from the SAMS calibration simulations, Gaussian noise, with a mean of zero, was added to each estimated free energy $f(N_{\mathrm{NaCl}})$ $N \in \{0, 1, \ldots, 20\}$, for a fixed values of $\langle V \rangle_{N_{\mathrm{NaCl}}}$. Three thousand noisy sample of $f(N_{\mathrm{NaCl}})$ $N \in \{0, 1, \ldots, 20\}$, equation 19 were used to predict the macroscopic concentration for a range of chemical potentials. This figure shows the 95% confidence range of the resultant ensemble of concentrations for different standard deviations of the Gaussian noise about the free energies. One needs to evaluate the free energies $f(N_{\mathrm{NaCl}})$ to within 4 kcal/mol to achieve an error in the concentration that is no larger than roughly 80 mM. The tapering of the statistical error in the concentration at lower values of the chemical potential is due to maximum number of salt pairs used in the calibration (20), which limits that maximum concentration that can be predicted.



Appendix 5 Figure 3. The relative efficiency of salt insertions/deletions in TIP3P water for different numbers of NCMC propagation steps between each perturbation step. Due to the manner in which the nonbonded parameters are updated in the SaltSwap code, it is faster—for a fixed protocol time-length—to perform multiple propagation steps for each perturbation (i.e. update of the nonbonded parameters) during an NCMC insertion/deletion attempt. More propagation steps limit the amount of communication between the CPU and GPU. However, for a fixed total protocol time-length, fewer perturbations increases the thermodynamic length each perturbation must traverse, which decreases the mean acceptance rate of the attempts. Thus, there is a (code-dependent) trade-off in the sampling efficiency between the number of perturbations and propagations steps. This figure shows the efficiency, defined by equation 25, for different numbers of propagation steps at different protocol time-lengths relative to the efficiency of instantaneous insertions and deletions. Ten propagation steps per perturbation step achieve the highest efficiencies, and so were used in all production osmostat simulations.