Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.
This journal is © the Owner Societies 2020

Supplementary Information

Binding of Biologically Relevant Divalent Cations to Aqueous Carboxylates: Molecular Simulations Guided by Raman Spectroscopy

Denilson Mendes de Oliveira[1], Samual R. Zukowski[1], Vladimir Palivec[2], Jérôme Hénin[3,4], Hector Martinez Seara[2], Dor Ben-Amotz[1*], Pavel Jungwirth[2*], Elise Duboué-Dijon[2,3,4*]

[1] *Department of Chemistry, Purdue University, West Lafayette, Indiana 47907*

[2] *Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 16610 Prague 6, Czech Republic*

[3] *CNRS, Université de Paris, UPR 9080, Laboratoire de Biochimie Théorique, 13 rue Pierre et Marie Curie, 75005, Paris, France*

[4] *Institut de Biologie Physico-Chimique-Fondation Edmond de Rothschild, PSL Research University, Paris, France*

*email: *pavel.jungwirth@uochb.cas.cz*; *bendor@purdue.edu*; *duboue-dijon@ibpc.fr*

1 Additional experimental details and results

Bound acetate spectrum. Extracting the bound acetate spectrum for a series of measured MAc_2 (M = Ca, Mg, Zn) solutions is non-trivial because little is known about the bound spectrum band shapes. While MCR does not require information about band profiles, a family of curves representing possible bound components can be obtained.[1,2] Specifically, this involves performing a second round SMCR analysis of the first round SC spectra, including the first round SMCR SC spectra of NaAc and $M(Ac)_2$, performed over a frequency range of 800 cm^{-1} to 1800 cm^{-1}. This second round SMCR decomposition effectively removes the unbound contribution to the first round $M(Ac)_2$ SC spectra. However, the resulting bound spectra may be any one of a family of possible bound spectra, as indicated by the blue regions in the spectra shown in Figure S1. A strict boundary to this spectral family is established by a non-negativity constraint to avoid unphysical spectral components, yielding the blue spectra that differ most from the unbound acetate (dashed black spectrum). The range of possible bound spectra shown in Figure S1 pertain to bound spectra with different spectral weights of the latter limiting bound spectrum and the unbound acetate spectrum.

Our MAc_2 measured spectra indicate that the CC Raman cross section is unchanged upon cation binding and thus all spectra were divided by the CC band area, so as to correspond to the same acetate concentration. The bound component that best describes ion pairing in MAc_2 solutions (dashed red curves in Figure S1) was obtained by determining which of the possible bound spectra yielded the most self-consistent (least concentration dependent) one-to-one binding constant. The resulting variance in the binding constants across the series of possible bound spectra is shown in Figure S2. The location of the minimum in each curve identifies the particular bound spectrum that produces the most self-consistent one-to-one binding constant. The end point of each curve represents the limiting bound spectrum determined by the non-

negativity constraint (corresponds to the blue spectra in Figure S1 that differ most from the unbound spectrum).

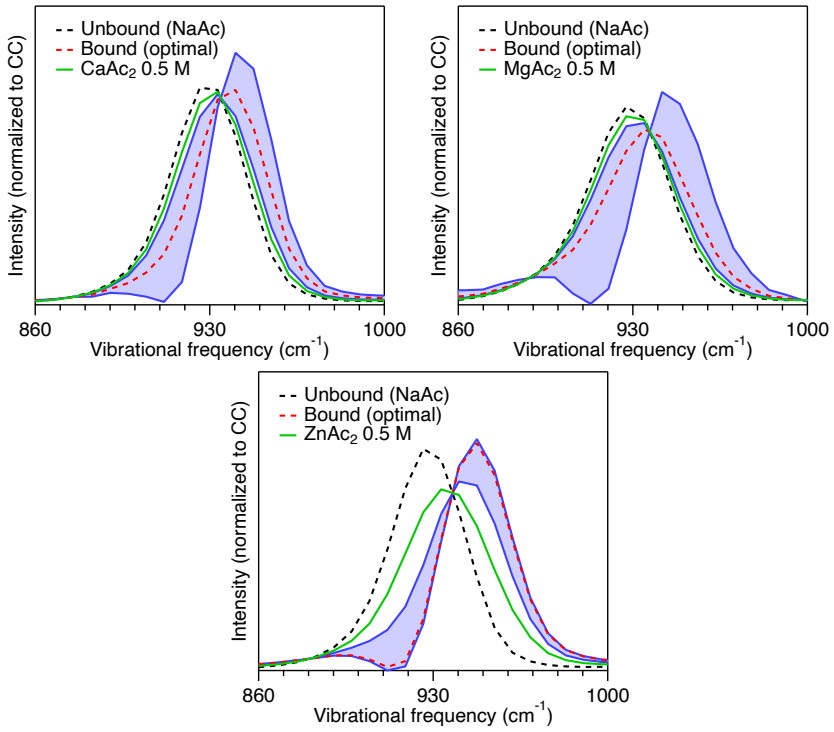

Figure S1. Family of possible bound spectra (blue shade) in the CC region obtained for MAc_2, as well as the unbound (dashed black) and optimally self-consistent bound spectrum (dashed red), and the experimental SC spectra at a total acetate concentration of 1 M. All spectra shown are normalized to unit area over the CC stretch band.

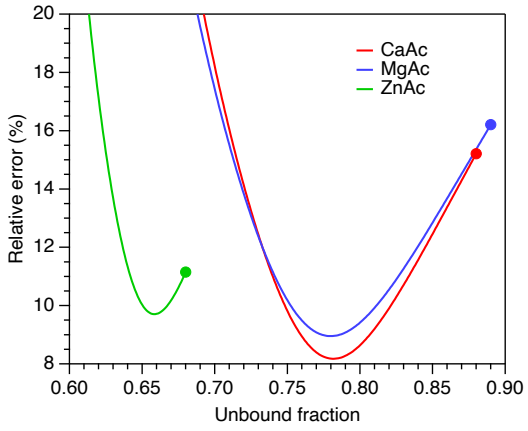

Figure S2. Relative error with respect to the mean value of the binding constants estimated using the range of bound spectrum candidates shown in Figure S1. The horizontal axis represents the equilibrium unbound fraction obtained at a total acetate concentration of 1 M for each member of the family of

bound spectra (terminated by the point obtained from the limiting bound spectrum derived from the non-negativity constraint).

Binding constant determination using TLS regression. The SC spectrum of a MAc_2 solution can be represented by a linear combination of the unbound (free) and bound acetate spectra. Thus, ion-pairing binding constants can be estimated by determining the relative contributions of these two species to the measured spectrum. Total least squares (TLS) regression can be used to obtain these relative contributions as long as the two pure component spectra are known *a priori*. We have shown that the NaAc SC spectrum can be taken as the spectrum of the unbound species and have obtained the bound acetate spectrum using the procedure described above. Therefore, if X is the spectrum of some mixture of Ac^- and M^{2+}, represented by a $1 \times n$ vector (where n is the number of frequency values in the spectrum), TLS can be used to decompose such spectrum into the unbound and bound components. Specifically, let P be a $2 \times n$ matrix of pure component (unbound and bound) spectra and b be a 1×2 vector that contains the spectral weights of the two components in X, such that

$$X = b\,P \tag{1}$$

The coefficients in b are estimated by inverting Eq. 1 as follows

$$b = XP^T(PP^T)^{-1} \tag{2}$$

where P^T represents the transpose of matrix P. Note that if the spectra X and the two component spectra in P are all normalized so as to pertain to the same solute concentration, then the elements of the vector b are the percentages of the total concentration corresponding to the bound and unbound acetate species. In this work, the X spectrum was taken as the SC spectrum obtained from an aqueous MAc_2 solution, and the component spectra were the SC spectrum of the free acetate (obtained from aqueous NaAc) and the bound $M^{2+}Ac^-$ contact-ion pair spectrum obtained as described above. All of the spectra were normalized to CC stretch band area to

obtain concentration-independent spectra given that the CC Raman cross section is unchanged upon ion paring.

Binding constants calculated from performing TLS on the CC band are reported in Figure 4 (and Table 2 after extrapolation to zero ionic strength) of the parent manuscript. Those results do not change significantly using the COO^- or the CC and COO^- bands together when performing the TLS fits. Note, however, that the spectral region near the OH stretch band cannot be included in the TLS regression because the free acetate spectrum was obtained with a monovalent cation (Na^+) while the other solutions contained M^{2+} ions which have a more dramatic effect on water structure. Table S1 shows a comparison between the binding constants obtained performing TLS on the CC band or COO^- band.

Table S1. Ion-pairing binding constants* obtained from TLS fits to different vibrational bands.

Vibrational bands	$MgAc_2$	$CaAc_2$	$ZnAc_2$
CC	1.1 ± 0.1	1.1 ± 0.1	4.5 ± 0.4
COO^-	1.2 ± 0.1	1.1 ± 0.1	4.0 ± 0.3
CC and COO^-	1.2 ± 0.2	1.1 ± 0.1	4.2 ± 0.3

*Error bars are the standard deviation of the values of *K* obtained across the 0.2 to 1 M total acetate concentration range.

Influence of ion pairing on acetate hydration shell spectra. The decrease in intensity of the hydration-shell OH band in Figure S3 and Figure 2B of the parent manuscript as function of acetate concentration is due to the interaction of acetate with the counterions. It is likely that the small depletion of SC OH stretch band for NaAc is due to solvent-shared ion-pair formation since there are no significant frequency shifts of the COO^-, CC or CH, bands, thus implying that there are no direct-contact ion pairs between Na^+ and Ac^- over our experimental concentration range. However, our results are also consistent with the possibility that there are some Na^+Ac^- contact-ion pairs, but such contacts do not detectably influence the COO^-, CC or CH band shape, frequency, or intensity, as is the case for $M^{2+}Ac^-$ contact ion pairs (with M = Mg, Ca, or Zn).

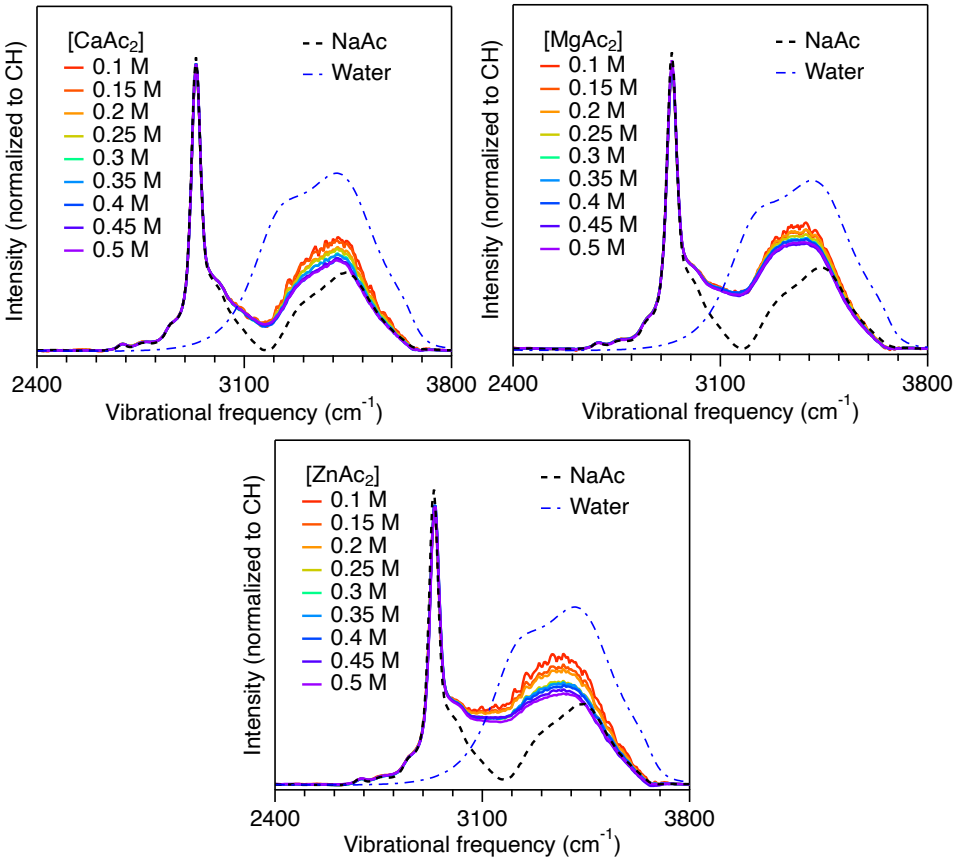

Figure S3. Hydration-shell spectra for magnesium acetate (top), calcium acetate (middle) and zinc(II) acetate (bottom).

2 Additional computational methodological details

Force fields. Several non-polarizable force fields with full charge were tested for each cation, as follows:

- For Ca^{2+}: "Ca_m" corresponds to the calcium developed by the Merz group,[3] "Ca_n" to that developed by the Netz group.[4]
- For Mg^{2+}: "Mg_m" is the magnesium force field by the Merz group,[3] and "Mg_c" that by D. Tobias and coworkers.[5]
- For Zn^{2+}: the zinc force field suggested by Karplus and coworkers, "Zn_k".[6]

We used a previously published Amber-like force field for acetate, "Ac_full".[7]

We also tested different ECC scaled-charge force fields, with both the traditional 0.75 charge scaling factor, and a milder 0.8 charge scaling factor:

- For Ca^{2+}: "Ca_s"[8] and "Ca_2s"[7] are two previously derived ECC force fields with slightly different parameters. "Ca_3s" is a test force field based on "Ca_2s", with a +1.6 charge corresponding to a milder 0.8 scaling.
- For Mg^{2+}: "Mg_ss" and "Mg_sb"[9] are two previously suggested ECC force fields with slightly different parameters. "Mg_s3" is a test force field based on "Mg_ss", with a +1.6 charge corresponding to a milder 0.8 scaling.
- For Zn^{2+}: "Zn_s" is a previously reported ECC force field,[9] and "Zn_s2" is a test force field based on "Zn_s" with a +1.6 charge.

For acetate, we used the previously published scaled-charge ECC version of the Amber-like acetate force field, "Ac_sc".[7] To estimate how sensitive the binding free energy results were to the charge distribution on the acetate anion, we tested a second acetate force field, where the RESP charges are derived from a QM B3LYP/6-31G* calculation using a PCM continuum description of the water solvent, which results in slightly more polarized charges, "Ac_sc2". A

0.8 scaling version of "Ac_sc", "Ac_sc3", was also tested to combine with cation force fields with the same milder scaling.

A parameter file with all the force field definitions is provided as supplementary data,[10] using the same naming convention as above.

Binding free energy calculations. During the alchemical transformation simulations, the binding geometry of the ion pair (contact monodentate, contact bidentate, solvent-shared) was enforced by a flat-bottomed harmonic restraint on the distance between the cation and the carbon of the carboxylate group. The restraining potential is zero for distances between r_{min} and r_{max}, and harmonic beyond with a spring constant k. The details of the employed restraint for each cation are provided in Table S2.

Table S2. Details of the restraining potential used to maintain a given ion binding geometry (M=monodentate, B=bidentate, SShIP= solvent-shared ion pair) during alchemical transformation simulations.

Cation	Ca^{2+}			Mg^{2+}			Zn^{2+}		
Binding geometry	M	B	SShIP	M	B	SShIP	M	B	SShIP
r_{min} (Å)	3.1	2	4	2.6	2	3.8	2.8	2	3.8
r_{max} (Å)	4	3.1	6.5	3.8	2.6	5.8	3.8	2.8	5.8
k (10^3 kJ/mol/nm^2)	50	50	20	100	100	100	50	100	20

In the main text (Methods section), we detail how different corrections need to be taking into account and added to the raw alchemical transformation result to yield the desired standard binding free energy. $\Delta G_{bind}° = \Delta G_1 - \Delta G_2 + \Delta G_{PBC} + \Delta G_{restr} + \Delta G_{unrestr}$. We provide below (**Table S3**) the detail of the different free energy terms for one specific case (Ca^{2+}-acetate, ECC

force fields Ca_2s + Acetate_sc) in a bidentate geometry to illustrate the magnitude of the different terms, and the importance of applying all the correction terms.

Table S3. Detailed values of the different free energy terms estimated to obtain the standard binding free energy for calcium-acetate (Ca_2s + Ac_sc) in a bidentate ion pair geometry.

Free energy Term	ΔG_1	ΔG_2	ΔG_{PBC}	ΔG_{restr}	$\Delta G_{unrestr}$	**Total $\Delta G_{bind}°$**
kJ/mol	806.95	807.01	−1.95	6.85	0.53	**+5.4**

Symmetry in monodentate pairs and convergence of free energy calculations. When decoupling from a monodentate ion pair, two symmetric binding modes exist, each one involving either carboxylate oxygen atom. These two modes are metastable, and hence sometimes difficult to sample evenly in the fully coupled state, depending on the exchange kinetics. Hence, the decoupling simulation may be seen as sampling the transition from one half of the space of bound configurations to the full unbound space, and therefore the decoupling free energy would be underestimated by RT ln(2). However intuitively satisfying, we show that this reasoning is incorrect here. In practice, we estimate the free-energy difference based on exponential averages of energy differences (Zwanzig's exponential formula,[11] or Bennett's acceptance ratio[12]). Thus, the free energy is determined entirely by the statistical distribution of energy differences between the end-states (in practice, intermediate states, as the transformation is stratified). For symmetry reasons, this distribution is the same for both binding modes. Sampling only one of the two modes well, and the other not at all, would yield exactly the same free energy as sampling both modes. Therefore, the question is: does partial sampling of one binding mode lead to a biased free energy estimate?

To assess this, we reanalyzed the decoupling data for a window with incomplete sampling of one binding mode. Configuration space was split into two parts according to the sign of Δ = dO1-dO2, where dOi is the distance between the metal ion and carboxylate oxygen i. This

partitions configuration space into two symmetry-equivalent halves, each containing one of the monodentate binding modes, as well as half of all unbound configurations. The simulated ensemble as sampled by the MD trajectories was split into two separate ensembles according to the sign of Δ, and the free energy was estimated separately based on each of the sub-ensembles. In all cases, both sub-ensembles yielded the same free energy as the global estimate, well within the error bars. This is explained by the data in Figure S4, which shows that even in the case of partial sampling of one of the modes, both modes yield very similar energy distributions. This indicates that once a transition from one mode to the other occurs, relaxation within that mode is fast, yielding the correct ensemble within the simulated time. In other words, a key factor that determines possible detrimental effects of symmetry-related binding modes in such calculations is not the low transition rate between the modes, but a possible slow relaxation rate *within a mode*. Provided that internal relaxation is fast enough, then metastability between the modes is irrelevant for the purpose of free energy estimation.

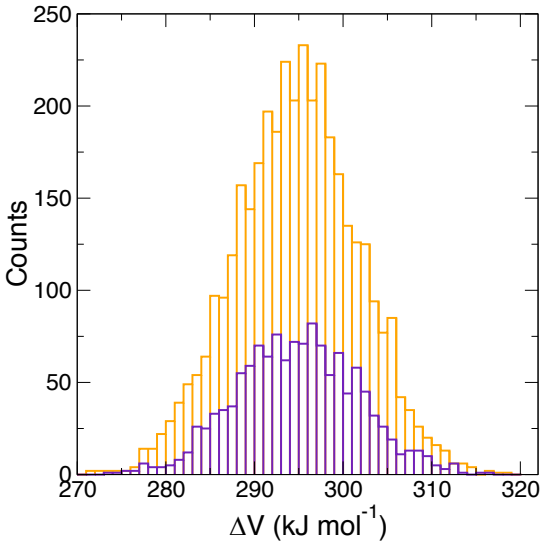

Figure S4. Distribution of the potential energy difference ΔV between the two neighboring intermediate states (which enters in the free energy calculation) in a chosen window with nonsymmetrical sampling of the two binding modes, for the subsensemble of configurations where the cation is closer to O1 (Δ<0) (orange) or closer to O2 (Δ>0) (purple).

3 Additional computational results

Free energy calculations. We provide below the final free energy calculation results, after applying all the corrections detailed above, for each full and scaled charge tested force field (the results for the AMOEBA force field are already provided in the main text Table 1). Table S4 summarizes the calcium-acetate binding free energy for each calcium acetate force field combination in the different ion pair geometries; Table S5 summarizes the magnesium results and Table S6 those for zinc.

Table S4. Standard binding free energy of calcium with acetate as calculated for different ion pair geometries using either a full charge force field, an ECC force field with the usual 0.75 scaling factor or with a 0.8 scaling factor. The details of the employed force field are provided in the above. The error bar on all calculated values is about 0.5 kJ mol^{-1}.

$\Delta G_{bind}°$ (kJ mol^{-1})	Ca_m + Ac_full	Ca_n + Ac_full	Ca_2s + Ac_sc	Ca_s + Ac_sc	Ca_2s + Ac_sc2	Ca_3s + Ac_sc3
Contact bidentate	−12.3	−21.7	+5.4	+6.0	+6.3	+2.5
Contact monodentate	−10.1	−18.4	+2.0	+5.7	−0.6	0.0
Total contact	***−13.2***	***−22.3***	***+1.5***	***+1.0***	***−0.7***	***−0.8***
Solvent-shared	−6.2	−6.7	−2.0	−2.0	−1.9	−2.7

Table S5. Standard binding free energy of magnesium with acetate as calculated for different ion pair geometries using either a full charge force field, an ECC force field with the usual 0.75 scaling factor or with a 0.8 scaling factor. The details of the employed force field are provided in the above. The error bar on all calculated values is about 0.5 kJ mol^{-1}.

$\Delta G_{bind}°$ (kJ mol^{-1})	Mg_c + Ac_full	Mg_m + Ac_full	Mg_ss + Ac_sc	Mg_sb + Ac_sc	Mg_s3 + Ac_sc3
Contact bidentate	−29.9	−13.0	+18.1	+19.7	+15.6
Contact monodentate	−34.1	−37.4	−0.6	−0.1	−5.3
Total contact	***−34.5***	***−37.4***	***−0.6***	***−0.1***	***−5.3***
Solvent-shared	−6.6	−6.8	−3.4	−1.1	−1.3

Table S6. Standard binding free energy of zinc with acetate as calculated for different ion pair geometries using either a full charge force field, an ECC force field with the usual 0.75 scaling factor or with a 0.8 scaling factor. The details of the employed force field are provided in the above. The error bar on all calculated values is about 0.5 kJ mol^{-1}.

$\Delta G_{bind}°$ (kJ mol^{-1})	Zn_k + Ac_full	Zn_s + Ac_full	Zn_s2 + Ac_sc
Contact bidentate	−14.9	+21.4	+19.4
Contact monodentate	−36.8	+0.6	−3.7
Total contact	***−36.8***	***+0.6***	***−3.7***
Solvent-shared	−6.9	−2.0	−2.4

Additional simulations were performed on sodium acetate, using the much employed Joung-Cheatham force field[13] as a full charge reference, a previously published ECC force field,[14] and the AMOEBA force field.[15] Binding free energy results are summarized in **Table S7**.

Table S7. Standard binding free energy for sodium acetate in different ion pair geometries using either a full charge force field, an ECC force field with the usual 0.75 scaling factor, and the fully polarizable AMOEBA force field. The error bar on all calculated values is about 0.5 kJ mol^{-1}.

$\Delta G_{bind}°$ (kJ mol^{-1})	Full charge FF	ECC FF (scaling 0.75)	AMOEBA
Contact bidentate	+0.4	+5.7	+9.0
Contact monodentate	+1.0	+3.8	+3.0
Total contact	*−1.0*	*+2.9*	*+2.8*
Solvent-shared	−0.9	+1.2	+0.3

***Ab initio* MD free energy profiles.** We computed the *ab initio* free energy profiles along the C(acetate)-cation distance with two different dispersion corrections (Figure S5). Comparison of the different profiles shows that their qualitative features are robust with respect to the details of the calculation setup.

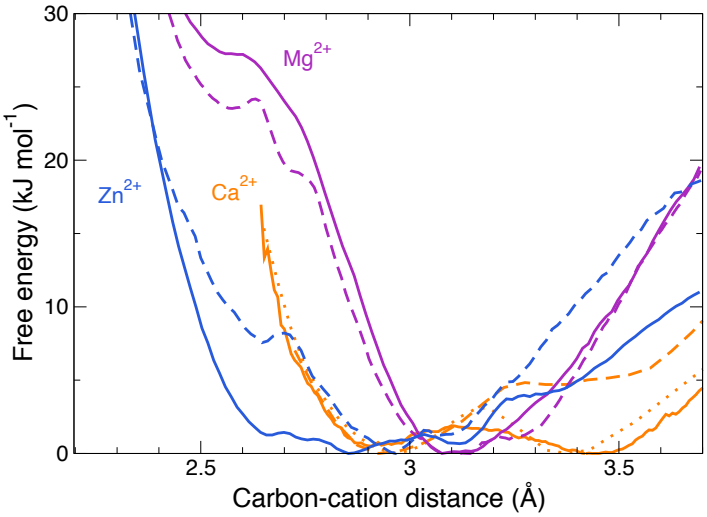

Figure S5. Free energy profiles along the carboxylate carbon-cation distance for calcium (orange), magnesium (purple) and zinc (blue), obtained using either a D3M(BJ) dispersion correction (solid lines), or a D2 dispersion correction (dashes). For Ca^{2+}, the free energy profiles with the D2 correction are taken from previous literature, by the Jungwirth group[7] (dashes) and the Mundy group[16] (dots).

Simulated Raman spectra. The simulated Raman spectra for acetate in different ion pairing geometries with a sodium cation are provided Figure S6, clearly showing that, irrespective of the ion pairing, no significant change is visible on the Raman spectrum, neither in terms of frequency shifts nor intensities.

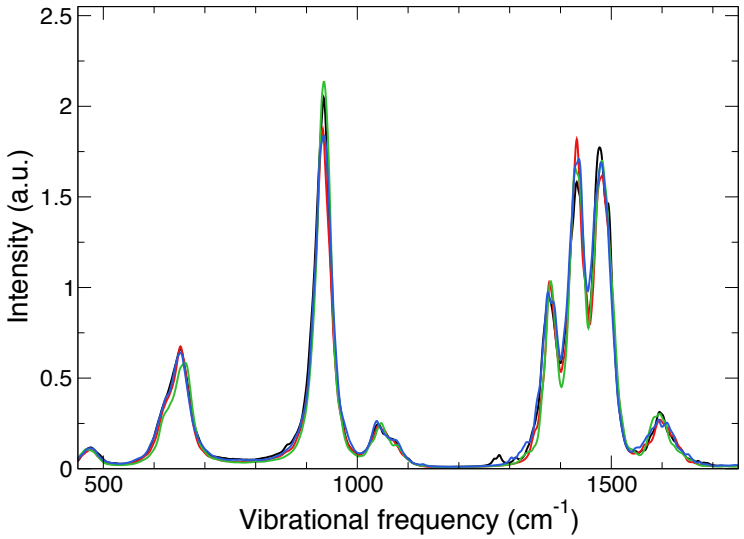

Figure S6. Simulated Raman spectra for acetate in pure water (black) or involved in a contact monodentate (red), contact bidentate (green) or solvent-shared (blue) ion pair with a sodium cation.

References

1. Fega, K. R., Wilcox, D. S. & Ben-Amotz, D. Application of Raman multivariate curve resolution to solvation-shell spectroscopy. *Appl. Spectrosc.*, 2012, **66**, 282–288.
2. de Juan, A. & Tauler, R. Multivariate Curve Resolution (MCR) from 2000: Progress in concepts and applications. *Crit. Rev. Anal. Chem.*, 2006, **36**, 163–176.
3. Li, P., Roberts, B. P., Chakravorty, D. K. & Merz, K. M. Rational design of particle mesh ewald compatible lennard-jones parameters for +2 metal cations in explicit solvent. *J.Chem. Theory Comput.*, 2013, **9**, 2733–2748.
4. Mamatkulov, S., Fyta, M. & Netz, R. R. Force fields for divalent cations based on single-ion and ion-pair properties. *J. Chem. Phys.*, 2013, **138**, 024505.
5. Callahan, K. M., Casillas-Ituarte, N. N., Roeselová, M., Allen, H. C. & Tobias, D. J. Solvation of magnesium dication: molecular dynamics simulation and vibrational spectroscopic study of magnesium chloride in aqueous solutions. *J. Phys. Chem. A*, 2010, **114**, 5141–5148.
6. Stote, R. H. & Karplus, M. Zinc binding in proteins and solution: a simple but accurate nonbonded representation. *Proteins*, 1995, **23**, 12–31.
7. Martinek, T. *et al.* Calcium ions in aqueous solutions: Accurate force field description

aided by ab initio molecular dynamics and neutron scattering. *J. Chem. Phys.*, 2018, **148**, 222813.

8. Kohagen, M., Mason, P. E. & Jungwirth, P. Accurate description of calcium solvation in concentrated aqueous solutions. *J. Phys. Chem. B*, 2014, **118**, 7902–7909.

9. Duboué-Dijon, E., Mason, P. E., Fischer, H. E. & Jungwirth, P. Hydration and Ion Pairing in Aqueous Mg2+ and Zn2+ Solutions: Force Field Description Aided by Neutron Scattering Experiments and Ab Initio Molecular Dynamics Simulations. *J. Phys. Chem. B*, 2017, **122**, 3296–3306.

10. Duboué-Dijon, E. Supporting data files for 'Binding of Biologically Relevant Divalent Cations to Aqueous Carboxylates: Molecular Simulations Guided by Raman Spectroscopy'. 2020, doi: 10.5281/zenodo.3827165.

11. Zwanzig, R. W. High Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.*, 1954, **22**, 1420–1426.

12. Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.*, 1976, **22**, 245–268.

13. 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**, 9020–9041.

14. Kohagen, M., Mason, P. E. & Jungwirth, P. Accounting for Electronic Polarization Effects in Aqueous Sodium Chloride via Molecular Dynamics Aided by Neutron Scattering. *J. Phys. Chem. B*, 2016, **120**, 1454–1460.

15. Ponder, J. W. *et al.* Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B*, 2010, **114**, 2549–2564.

16. Daily, M. D., Baer, M. D. & Mundy, C. J. Divalent Ion Parameterization Strongly Affects Conformation and Interactions of an Anionic Biomimetic Polymer. *J. Phys. Chem. B*, 2016, **120**, 2198–2208.