Calcium ions in aqueous solutions: Accurate force field description aided by ab initio molecular dynamics and neutron scattering

Tomas Martinek, Elise Duboué-Dijon, Štěpán Timr, Philip E Mason, Katarina Baxová, Henry Fischer, Burkhard Schmidt, Eva Pluhařová, Pavel Jungwirth

Calcium Ions in Aqueous Solutions: Accurate Force Field Description Aided by *Ab Initio* Molecular Dynamics and Neutron Scattering

Tomas Martinek,[1] Elise Duboué-Dijon,[1] Štěpán Timr,[1] Philip E. Mason,[1] Katarina Baxová,[1] Henry E. Fischer,[2] Burkhard Schmidt[3], Eva Pluhařová,[4,a)] and Pavel Jungwirth[1,a)]

[1]Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Flemingovo nám. 542/2 160 00 Prague, Czech Republic

[2]Institut Laue-Langevin, 71 avenue des Martyrs, CS 20156, 38042 Grenoble cedex 9, France

[3]Institut für Mathematik, Freie Universität Berlin, Arnimallee 6, D-14195 Berlin, Germany

[4]J. Heyrovský Institute of Physical Chemistry, Czech Academy of Sciences, v.v.i., Dolejškova 2155/3, 182 23 Prague 8, Czech Republic

[a)] Authors to whom correspondence should be addressed. Electronic mail: pavel.jungwirth@uochb.cas.cz, eva.pluharova@jh-inst.cas.cz

ABSTRACT

We present a combination of force field and *ab initio* molecular dynamics simulations together with neutron scattering experiments with isotopic substitution that aim at characterizing ion hydration and pairing in aqueous calcium chloride and formate/acetate solutions. Benchmarking against neutron scattering data on concentrated solutions together with ion pairing free energy profiles from *ab initio* molecular dynamics allows us to develop an accurate calcium force field which accounts in a mean-field way for electronic polarization effects via charge rescaling. This refined calcium parameterization is directly usable for standard molecular dynamics simulations of processes involving this key biological signaling ion.

I. INTRODUCTION

The calcium ion is a key signaling species involved in many biological processes including allosteric enzyme activation, neurotransmitter release, muscle contraction, and other biological processes.[1–4] Consequently, numerous recent computer modeling studies have aimed at elucidating the molecular mechanisms of the biological actions of calcium.[5–10] In general, a molecular simulation can only be as accurate as the underlying force field. With the commonly employed non-polarizable force fields, problems may arise to accurately describe aqueous ions of high charge density, such as alkali earth dications, due to important, but neglected, electronic polarization effects.[11] As a consequence , such simulations may not accurately describe without additional parameterization the ion pairing of these ions in aqueous solutions and their binding to negatively charged groups in proteins, nucleic acids, or phospholipids.[12–16]

Recently, a simple but physically well justified way of accounting in a mean-field way for electronic polarization (missing in non-polarizable force fields) via rescaling ionic charges has been suggested.[17,18] The scaling factor of ~ 0.75 is the inverse square root of the electronic (high frequency) part of the dielectric constant of water.[19] This scaling factor is directly applicable to atomic ions with integer charge. For atoms with partial charges, including those forming water molecules, the situation is more complicated, not least because varying extent of charge scaling has already been applied implicitly when fitting the force field against experimental observables.[18] In our previous studies, we have applied this approach to develop a charge scaled model of calcium benchmarked against structural neutron scattering data[11,20] and successfully applied it to quantify its affinity to a common calcium-binding protein calmodulin[13] and to phospholipids in model membranes.[14,15]

Here, we introduce an additional benchmarking approach, namely *ab initio* molecular dynamics (AIMD), which allows (albeit at a significant computational cost) to generate accurate free energy profiles[21] for pairing of calcium with its counter-ion in water. We apply this approach to aqueous calcium chloride and formate solutions, the latter serving as a model for the interaction of calcium with the carboxylic side chain groups of glutamate or aspartate and the protein C-terminus. Together with neutrons scattering experiments on analogous systems, this allows us to refine the calcium force field, such that it is now applicable for accurate simulations of biological processes involving this crucial signaling ion. As a bonus, we obtain a faithful description of these important calcium salt solutions with

quantitative molecular details concerning the hydration structure of the ions, as well as their tendency to form contact and solvent-shared ion pairs.

II. METHODS

A. Experimental details – neutron scattering

Four solutions of calcium acetate were prepared with different isotopic constitutions, which are summarized in TABLE I. All these solutions were prepared using a common procedure as follows. Calcium oxide (anhydrous 99.99%, Sigma-Aldrich) was dissolved in the stoichiometric amount of acetic acid to yield a calcium acetate solution, which consists after neutralization of a ratio water to acetate 2:56.55. For convenience, this is referred to as a 1 m calcium acetate solution in this study. The solution contains 2 m acetate and 1 m calcium concentrations.

TABLE I. Isotopic composition of calcium acetate solutions used in neutron scattering measurements

Sample name	Acid used for preparation	Solvent
$H_3CCOOCa$ in H_2O	H_3CCOOH	H_2O
$D_3CCOOCa$ in H_2O	D_3CCOOH	H_2O
$H_3CCOOCa$ in D_2O	H_3CCOOD	D_2O
$D_3CCOOCa$ in D_2O	D_3CCOOD	D_2O

Total neutron scattering patterns were measured for the four calcium acetate solutions (TABLE I) on the D4C diffractometer[22] at the ILL in Grenoble, France, with neutrons with a wavelength of $\lambda = 0.5$ Å. Data[23] were recorded for 3 h for each D_2O sample and for 6 h for each H_2O sample. The raw scattering data were then corrected for multiple scattering and absorption,[24] before being normalized versus a standard vanadium rod. This provided for each sample the total scattering pattern, $S(Q)$. First order differences, $\Delta S(Q)$, were then obtained by subtracting the total scattering patterns of CD_3COO^- (D_3-acetate) and CH_3COO^- (H_3-acetate) solutions. These first order differences were obtained

both in H_2O and D_2O. Each of them can be expressed as a sum of pair-wise structure factors, which contains only structural data from the substituted non-exchangeable hydrogen "*Hsub*" to any other atom in solution, "*X*", with all the other terms canceling out. The expressions for the first order differences in Q-space are provided below, "*Hex*" referring to the exchangeable hydrogen atoms on water and the prefactors (expressed in millibarns) being calculated from the concentrations and coherent scattering length of each nucleus.[25] The offset is subtracted so that the scattering patterns oscillate around zero at long Q.

$$\Delta S^{D_2O}_{\mathrm{HsubX}}(Q) = 27.66\, S_{\mathrm{HsubHex}}(Q) + 12.03\, S_{\mathrm{HsubOw}}(Q) + 0.85\, S_{\mathrm{HsubOc}}(Q) + \\ +0.97\, S_{\mathrm{HsubC}}(Q) - 0.17\, S_{\mathrm{HsubCa}}(Q) + 0.32\, S_{\mathrm{HsubHsub}}(Q) - 42.02 \qquad (1)$$

$$\Delta S^{H_2O}_{\mathrm{HsubX}}(Q) = -15.51\, S_{\mathrm{HsubHex}}(Q) + 12.03\, S_{\mathrm{HsubOw}}(Q) + 0.85\, S_{\mathrm{HsubOc}}(Q) + \\ +0.97\, S_{\mathrm{HsubC}}(Q) - 0.17\, S_{\mathrm{HsubCa}}(Q) + 0.32\, S_{\mathrm{HsubHsub}}(Q) + 1.15 \qquad (2)$$

The first order difference in H_2O was further subtracted from the first order difference in D_2O to yield the second order difference, $\Delta\Delta S(Q)$, which contains only *HsubHex* correlations.

$$\Delta\Delta S(Q) = \Delta S^{D_2O}_{\mathrm{HsubX}}(Q) - \Delta S^{H_2O}_{\mathrm{HsubX}}(Q) = 43.17\, S_{\mathrm{HsubHex}}(Q) - 43.17 \qquad (3)$$

which can be Fourier-transformed to yield the corresponding second-order difference in r-space:

$$\Delta\Delta G(r) = \Delta G^{D_2O}_{\mathrm{HsubX}}(r) - \Delta G^{H_2O}_{\mathrm{HsubX}}(r) = 43.17\, g_{\mathrm{HsubHex}}(r) - 43.17 \qquad (4)$$

B. Computational details

1. Force field molecular dynamics

A set of aqueous solutions was simulated to investigate the association of the calcium cation with its most common counterions – chloride anion as an example of the simplest spherical counterion, and formate or acetate (Ac) anions to mimic the association with carboxylic groups of biological molecules. Systems containing one ion pair (a calcium cation with a single chloride, formate or acetate counterion) in explicit water were simulated to obtain the free energy profiles for the corresponding ion associations and compare the behavior of different force fields with *ab initio* (DFT) simulations. In addition, force field simulations of concentrated solutions of $CaCl_2$ and $CaAc_2$ were performed to directly compare the results computed with different force fields with experimental neutron scattering data. TABLE

II summarizes the compositions of the different simulated systems and provides in each case additional details about the simulation setup.

TABLE II. Overview of simulated systems with relevant details. For simulations in the N*p*T ensemble, the cell size is the average value obtained from simulation runs, which slightly depends on the employed force field.

System	Composition	Ensemble	Cell size	cutoffs
Calcium chloride ion pair	1 Ca^{2+}, 1 Cl^-, 64 H_2O	NVT	12.5 Å	6 Å
4 m $CaCl_2$ solution	52 Ca^{2+}, 104 Cl^-, 719 H_2O	N*p*T	~ 29.3 Å	12 Å
Calcium formate ion pair	1 Ca^{2+}, $HCOO^-$, 107 H_2O	NVT	14.73 Å	6 Å
1 m $CaAc_2$ solution	81 Ca^{2+}, 162 CH_3COO^-, 4581 H_2O	N*p*T	~ 53 Å	12 Å

All force field molecular dynamics simulations were performed with the Gromacs software[26] using the leap frog propagator with a 1 fs time step at constant temperature of 300 K maintained by the Canonical Sampling through Velocity Rescaling (CSVR) thermostat with a time constant of 0.2 ps.[27] Water molecules were described using the SPC/E force field,[28] their geometry being kept rigid by the Settle algorithm.[29] Periodic boundary conditions were employed and long-range electrostatic interactions were treated by the particle mesh Ewald method.[30] When a barostat was employed, the pressure coupling constant was set to 1 ps. Each concentrated solution ($CaCl_2$ or $CaAc_2$) was equilibrated for at least 20 ns before a production run of at least 20 ns long, which was used to evaluate the radial distribution functions required to compute the neutron scattering signal.

In this study, we employed several force fields for calcium and chloride ions. First, we used a common full charge force field for both Ca^{2+} and Cl^-.[31,32] We hereafter denote this force field FULL. We compare this standard force field to the scaled charges calcium force field recently developed in our group, which employs the electronic continuum correction with ionic size refinement (ECCR).[11] Here, the ionic charges are scaled by the factor 0.75 to account for electronic polarization in a mean field way.[17] The ionic Lennard-Jones parameter σ is correspondingly reduced to recover the proper Ca^{2+} - H_2O interaction, as estimated from neutron scattering data. In this work, we propose a refined Ca^{2+} scaled charge force field, denoted as ECCR2, which provides better agreement with our *ab initio* simulation data, leading to a slightly enlarged calcium radius compared to the ECCR model. In combination with both ECCR and ECCR2 scaled charges calcium force fields, we used a scaled charge force field for the chloride

anion, the parameters of which were refined compared to our previous study,[11] which was parametrized in our group using reference neutron scattering data of lithium chloride salt.[33] Details of the force fields used for Ca^{2+} and Cl^- are provided in TABLE III.

TABLE III. Force fields for the calcium and chloride ions.

	Ca^{2+}			Cl^-		
	σ (Å)	ε (kJ/mol)	Charge (*e*)	σ (Å)	ε (kJ/mol)	charge (*e*)
FULL	2.8196	0.5072	+2.00	4.4499	0.4184	-1.00
ECCR	2.5376	0.5072	+1.50	4.1000	0.4928	-0.75
ECCR2	2.66558	0.5072	+1.50	4.1000	0.4928	-0.75

In addition, we designed a force field for the acetate and formate anions based on the Amber ff99 force field,[34] the charges being obtained from a RESP fit of the electrostatic potential on an optimized geometry. These calculations were performed using the Gaussian 09 software[35] with the Hartree-Fock method with the 6-31G* basis set[36] employed both for the geometry optimization and the electrostatic potential calculation. The resulting RESP charges are listed in TABLE S I and TABLE S II in SI. The corresponding scaled charges force field was obtained by scaling the RESP charges by the factor 0.75 without modification of the van der Waals potentials.

2. Ab initio molecular dynamics

Born-Oppenheimer *ab initio* molecular dynamics simulations were performed using the Quickstep module of the CP2K package[37] implementing the hybrid Gaussian functions and plane waves (GPW) method.[38] All simulations were performed with a time step of 0.5 fs in the NVT ensemble, using the CSVR thermostat[27] with a time constant of 50 fs. Periodic boundary conditions were used. The electronic structure of the system was treated at the density functional level of theory with the BLYP functional[39,40] with the Grimme correction scheme (DFT-D2)[41] to account for dispersion interactions. The core electrons were described by norm-conserving GTH pseudopotentials.[42] Kohn-Sham orbitals were expanded in a Gaussian basis set: TZV2P MOLOPT for O, H, C, and DZVP-MOLOPT-SR-GTH-q10 for Ca in the aqueous Ca^{2+} $HCOO^-$ system and in a DZVP-MOLOPT-SR-GTH Gaussian basis set for the aqueous

Ca^{2+} Cl^- system.[38] Previous studies demonstrated that the DZVP basis set yields similar results to TZVP, while being significantly less computationally demanding.[43] A cutoff of 400 Ry was used for the auxiliary plane wave basis set.

Free energy profiles for the ion pairing were obtained for a single CaCl ion pair and a single calcium-formate ion pair at the *ab initio* MD level, and compared to results of force field simulations on the same systems. Converging such *ab initio* free-energy profiles is extremely challenging and computationally expensive, so that only a few attempts in this direction can be found in the literature[21,44] including studies of the Ca-Cl ion pair.[45,46] The free energy profile along the Ca-Cl distance was obtained at the *ab initio* level using two (in principle equivalent) methods: (i) integration of mean force and (ii) umbrella sampling. This allowed us to check the convergence of our computationally very demanding calculations and estimate the associated error. In method (i), the average force between the studied ions is evaluated at different fixed interionic distances and consequently integrated along the Ca-Cl distance to obtain the potential of mean force (PMF). In contrast, the umbrella sampling method uses a set of biasing harmonic potentials along Ca-Cl distance to enhance the sampling of rare ionic configurations. The initial configurations for each window of the *ab initio* simulations were taken from classical MD simulations of the same system. The umbrella sampling windows were then combined and unbiased to obtain the free-energy profile using the WHAM algorithm.[47] The standard volume entropy correction[44] ($+2k_BTln(r)$) was applied to all the obtained free energy profiles.

A total of 38 simulations with fixed Ca-Cl interionic distances ranging from 2.2 Å to 6.0 Å were performed to obtain the average mean force along the Ca-Cl distance, while the umbrella sampling simulations used a total of 10 windows with distances ranging from 2 to 6 Å. Each umbrella sampling window was duplicated starting with two different initial geometries to ensure proper sampling of relevant calcium-water coordination numbers.[46] Each window was equilibrated for 10 ps before a 50 ps production run. The error bars were evaluated as the standard deviations obtained from 10 free-energy profiles generated from 5 ps cuts of the simulation production run.

Unlike the case of Ca^{2+}-Cl^-, *ab initio* MD simulations are too expensive to allow for satisfactorily converged sampling of all degrees of freedom of the structurally more complex Ca^{2+}-$HCOO^-$ ion pair. It is thus beyond our reach to fully characterize the *ab initio* free energy landscape for the ion-ion interaction. Nevertheless, we were able to obtain the free-energy profile along the Ca-O distance in a monodentate arrangement of the ion pair using umbrella sampling simulations (13 windows) with restraining the COCa angle around 180° using a harmonic force constant of 100 kJ mol^{-1} rad^{-2}. To further investigate the interplay between the mono- and bi-dentate interaction mode, we

generated the free-energy profile along the Ca-C distance (22 windows), restraining Ca^{2+} to the OCO plane using a harmonic restraint on the OCOCa dihedral angle with a force constant of 100 kJ mol^{-1} rad^{-2}. Due to the higher complexity of the phase space and applied restraints, the standard radial volume entropy correction cannot be employed here. The procedure used to correct the free energy profiles for the sampled volume is described in detail in SI.

3. Comparison between ab initio and force field free energy profiles

The various *ab initio* free energy profiles were compared with the corresponding profiles obtained with different force fields, for which full convergence could be reached using longer simulation times and additional umbrella sampling windows. We further characterized with the different force fields the detailed interaction of a calcium cation with the carboxylic group of the formate anion. In particular we computed the full 2D free energy profile in the OCOCa plane along the CaC distance and Ca-C-O_{mid} angle, where O_{mid} is the middle point between the two carboxylate oxygen atoms. The distance coordinate was binned from 2.7 to 6.1 Å and the angular coordinate was binned from 0 to 110°. The sampling was restrained to the formate plane using the same restraints as above.

III. RESULTS AND DISCUSSION

A. Refining the Ca^{2+} force field using *ab initio* simulations

First, we describe the Ca^{2+} - Cl^- ion pairing using AIMD simulations of single ion pair in water. We obtain the *ab initio* free energy profile along the Ca^{2+}-Cl^- distance. FIG. 1 compares two independent free energy methods - umbrella sampling and integration of the mean force. The free energy profiles obtained by the two approaches are consistent with each other within the estimated error bars. Both predict a Ca^{2+} - Cl^- distance of 2.75 ± 0.05 Å at the contact ion pair (CIP), as well as the same position of 3.75 ± 0.05 Å of the barrier between the CIP and the Solvent Shared Ion Pair (SShIP), and an almost symmetrical barrier of 10 ± 1 kJ/mol between the two minima. The estimate of the SShIP position is less accurate because of increasing statistical uncertainties at larger distances, nevertheless it can be estimated to be around 5.0 Å. The free energies of the CIP and SShIP are practically identical within the error bars.

The good agreement between the umbrella sampling and mean force integration methods points to a satisfactory convergence of the obtained Ca^{2+} - Cl^- free energy profile. This is further supported by comparison of the present results to a free energy profile obtained for an analogous system using a similar method recently.[46] With this reassurance, we use below the free energy profile obtained using the umbrella sampling technique (with the slightly smaller statistical error) as the reference for comparison with force field simulations and for further force field refinement.

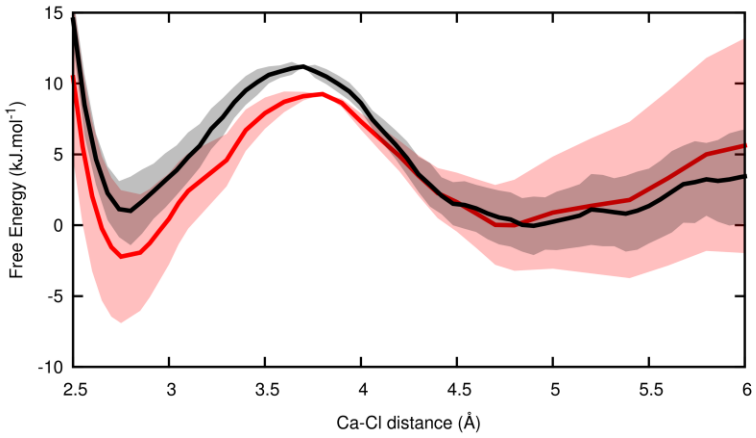

FIG. 1. Free energy profile along the Ca^{2+} - Cl^- distance obtained by mean force integration (red) and umbrella sampling (black).

We compared an AI results with three empirical force field simulations. The first one is a standard full charges force field (denoted here as "FULL") with full integer charges on both ions. This full charges force field provides a free energy profile significantly different from the AIMD reference (FIG. 1). In particular, this force field significantly underestimates the stability of the CIP. It is found higher in free energy by about 8 kJ/mol compared to the SShIP, with the barrier between SShIP and CIP being 19 kJ/mol, i.e., 8 kJ/mol higher than using AIMD. The Ca - Cl distance at the CIP is only slightly shifted compared to AIMD and the transition barrier is found at 3.6 Å, which is close to the AIMD value of 3.7 Å.

The second force field, ECCR, previously[11] designed by us, employs the electronic continuum correction to account for water polarization, with the size of the calcium ion (i.e., the Lennard-Jones σ-parameter) fitted to experimental neutron scattering data. While this force field reproduces quantitatively the neutron scattering data available in the literature[20] (FIG. 3), it does not compare as well with the present *ab initio* free energy profile (FIG. *2*). Namely, the Ca - Cl distance at the CIP is slightly shorter than that found in the *ab initio* simulations, 2.72 vs 2.80 Å. In addition, the barrier height is found significantly larger than for the *ab initio* reference, 20 vs 11 kJ/mol.

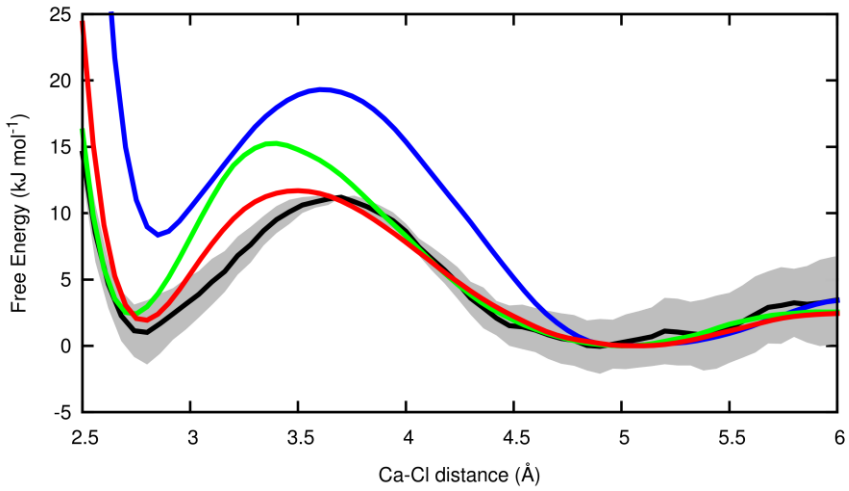

FIG. 2. Free energy profile along the Ca^{2+} - Cl^- distance obtained by umbrella sampling (black) compared with force field simulations performed using the full charges force field (blue), the ECCR (green) and the ECCR2 force field (red).

Analysis of the simulations shows that the relatively small discrepancy between the *ab initio* and ECCR force field profiles stems mainly from different Ca - O(water) distances. Namely, the average Ca - O distance obtained from the AIMD simulation is 2.43 Å, in contrast with 2.35 Å with the ECCR force field, which was fitted to reproduce the first peak of the neutron scattering pattern (2.38 Å) corresponding to the Ca - O and Ca - Cl correlations. To ensure that our DFT setup provides reliable Ca - O optimal distances, we optimized the geometry of a perfectly symmetric $Ca^{2+}(H_2O)_6$ cluster at different levels of theory, employing DFT methods (BLYP and B3LYP) as well as HF, full MP2, and CCSD with basis sets of various sizes (6-31G*, 6-31+G*, 6-311++g**, cc-pVDZ, and cc-pVTZ). The Ca - O distances at the optimized geometries are provided in SI, showing that the Ca-O distance is practically insensitive to the level of theory, which means that the BLYP-D2 level of description used in our AIMD simulations does not suffer

from any significant systematic deviation in comparison to higher levels of theory or larger basis sets. Moreover, a very similar AIMD setup was already shown to correctly reproduce the structure of pure water[48] and a very recent study[46] showed, that the EXAFS spectrum of $CaCl_2$ solutions was extremely well reproduced using the same AIMD setup as in the present study.

Hence, we suggest here a refined parametrization for Ca^{2+} based on the previous ECCR one (hence we call it ECCR2), with a calcium cation σ-parameter enlarged by 5 % to reach a better agreement with the AIMD free energy profile.[46] As seen in FIG. 2, the position of CIP at 2.80 Å, is now fully consistent with the AIMD result. The free energy of the CIP and SShIP are found very similar to each other, with the CIP being only 1.8 kJ/mol higher in energy (compared to 0.9 kJ/mol with AIMD, well within the AIMD error bars). The SShIP and CIP are separated by a 11.4 kJ/mol high transition barrier, which is in much better agreement with the AIMD result of 11 kJ/mol than the original ECCR force field. Finally, we show (SI, FIG S4) that the free energy profile obtained with the ECCR2 scaled charges force field is in quantitative agreement with that obtained using the fully polarizable force field AMOEBA itself within the error bars of the AIMD calculation. The price is that the ECCR2 description does not compare as well as the original ECCR with the neutron scattering data. Nevertheless, the comparison is still good and presents a significantly improvement compared to the full charges force field (FIG. 3). The new ECCR2 calcium parametrization thus appears to be the best compromise between agreement with the available neutron scattering data on the one hand and with EXAFS experiments and advanced AIMD simulations on the other hand.

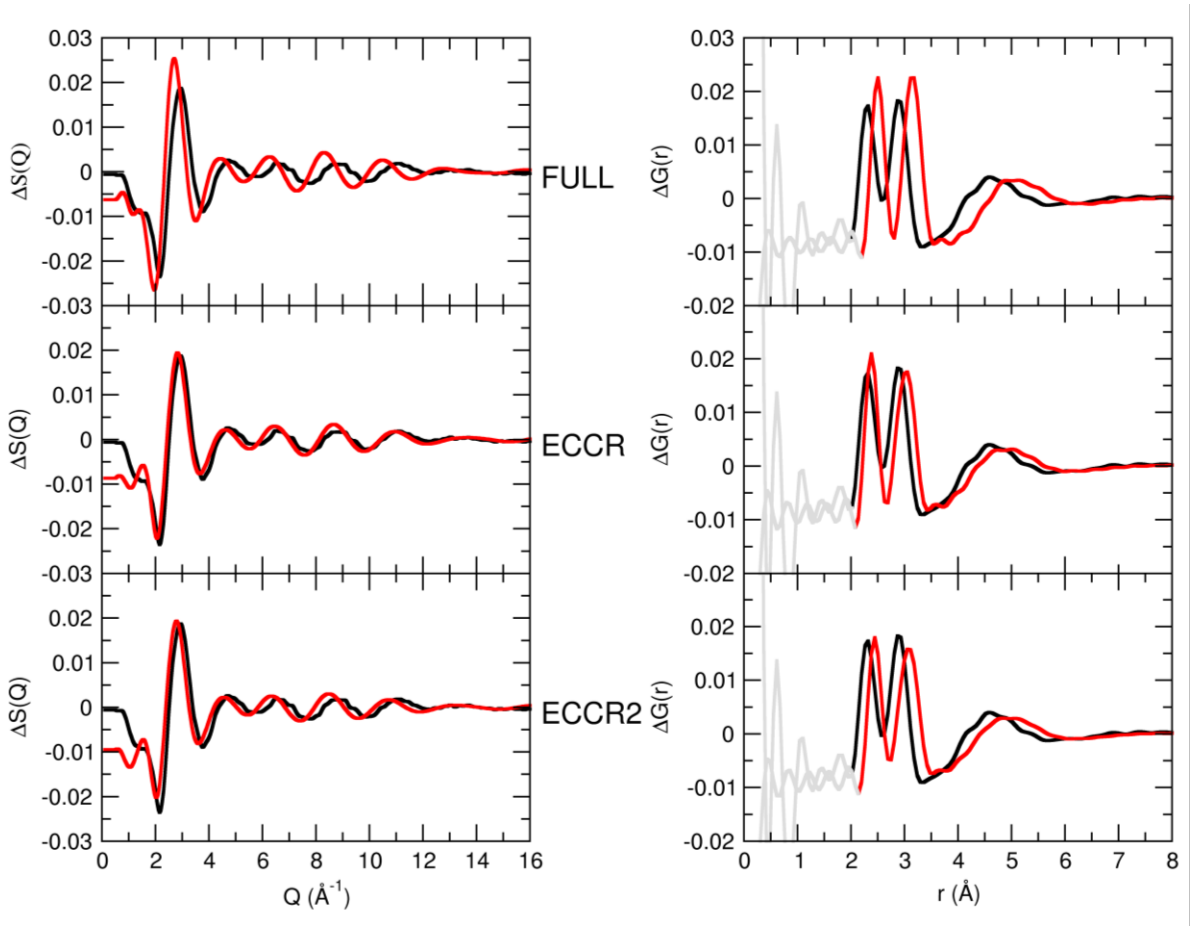

FIG. 3. First order difference (on calcium) in Q-space (left hand side) and in r space (right hand side) for a concentrated 4 m $CaCl_2$ solution, as obtained from the neutron scattering experiment (black) and from simulations with different force field, full charges (top), ECCR (middle) and ECCR2 (bottom). Since the experimental signal was obtained by subtracting the neutron scattering patterns obtained with different calcium isotopes, the signal only reports on correlations involving the Ca atom. See more details in ref[11,20].

B. Ca^{2+} interaction with the carboxylate group

Having in hand a refined calcium force field that provides a satisfactory agreement with both AIMD and neutron data, we now aim at testing the transferability toward a description of the interaction of calcium with negatively charged protein residues. While the calcium ion is also expected to interact with the polar protein backbone,[49] the interaction with the negatively charged carboxylates of aspartate and glutamate is expected to be the strongest. We thus focus here on the carboxylate group using the acetate anion as a simple proxy.

The total neutron scattering patterns *S(Q)* (FIG. 4a) were obtained for four 1 m calcium acetate solutions with different isotopic compositions (see Methods, TABLE I). These patterns exhibit a large background slope, due to the Placzek effect.[50] This effect is more pronounced for H_2O solutions than for D_2O, as expected from the higher incoherent scattering cross section of ^1H versus ^2H. The total *S(Q)* is largely dominated by contributions from water-water interactions and the calcium acetate interaction is thus mostly hidden. The use of H/D isotopic substitutions, both on water and on the non-exchange hydrogen of the acetate, thus allows us to focus on the acetate hydration properties and indirectly on the ion pairing with calcium ion, since ion pairing displaces water molecules from the hydration shell.

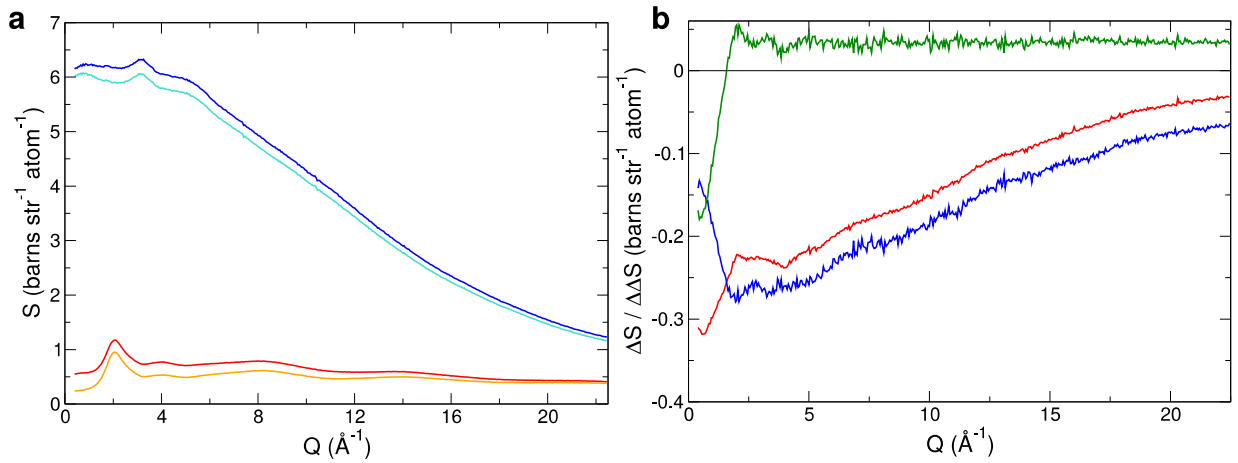

FIG. 4. a) Total neutron scattering patterns for 1 m solutions of calcium h3-acetate (dark blue) in H_2O, calcium d3-acetate (light blue) in H_2O, calcium h3-acetate (red) in D_2O, calcium d3-acetate (orange) in D_2O. b) First order differences in H_2O, $\Delta S^{H_2O}_{\mathrm{HsubX}}(Q)$ (blue), and in D_2O, $\Delta S^{D_2O}_{\mathrm{HsubX}}(Q)$ (red), together with the second order difference, $\Delta\Delta S(Q)$ (green).

The first order differences in H_2O and D_2O $\Delta S^{H_2O}_{\mathrm{HsubX}}(Q)$ and $\Delta S^{D_2O}_{\mathrm{HsubX}}(Q)$, are obtained by direct difference between pairs of solutions with different isotopes on the acetate.They only contain the correlation between the substituted acetate hydrogens *Hsub* and any other atom in the system, *X* (Eq. 1-2). This subtraction thus cancels most of the background slope and removes the large water-water correlations from the signal. However, the first order differences still contain strong intramolecular correlations (*Hsub*-C and *Hsub*-O from the same acetate molecule), which dominate the signal but are of little interest in this work. This problem is solved by taking the difference between the two first order differences (Eq. 3) yielding the second order difference *ΔΔS(Q)* (FIG. 4). This has no residual slope and reports exclusively on the part of the system we are directly interested in, i.e., the acetate hydration shell, since it

contains only correlations between the substituted hydrogens *Hsub* on acetate and the exchangeable water hydrogens *Hex* (see Eq. 3). In all that follows, the constant offset is subtracted from the double difference so that it oscillates around zero, consistently with the way we defined $\Delta\Delta S(Q)$ in Eq. 3.

Comparison of the experimental neutron second order difference with that computed with different force fields constitutes a stringent test of the quality of a force field and of its ability to capture the hydration and ion pairing properties of acetate in concentrated calcium acetate solutions. FIG. 5a shows that while a standard full charges force field correctly captures the structure at large values of Q, it exhibits a spurious sharp peak at low Q (below 1 $Å^{-1}$), which is not present in the experimental signal. Such low Q features typically point to clustering in the solution, which is confirmed by visual examination of the MD simulations (FIG. 6a). Indeed, when a full charges description is used, all the ions form contact ion pairs with no free ion left in solution. This behavior results in an unrealistic depletion at short distance of the simulated r-space signal (FIG. 5), since the clustering of ion pairs effectively reduces the number of water molecules (and thus *Hex*) in the vicinity of each acetate molecule.

Scaling only the charge of the calcium ion reduces marginally this spurious peak at low Q and the structure of the solution remains very much the same as for the full charge model (see SI FIG S5 for comparison of the ECCR and ECCR2 parameterizations yielding essentially the same result). A good agreement with experiment is obtained only upon scaling both calcium and acetate ions. FIG. 5c shows that in this case the experimental Q-space signal is almost perfectly reproduced with the proper low-Q limit. The r-space signal is no longer depleted at short distances, and the agreement with experiment becomes very good. A snapshot of the simulation (FIG. 6c) shows that ion pairing is much weaker than in the previous two simulations, now with many free ions in solution. As a result, the solution is much more homogeneous with no sign of strong ion clustering. This study thus illustrates an important point for biological simulations - when using the scaled charge description, one should scale not only the charges of the ions in the solution, but also those of *charged* protein residues, as already suggested in the literature.[17]

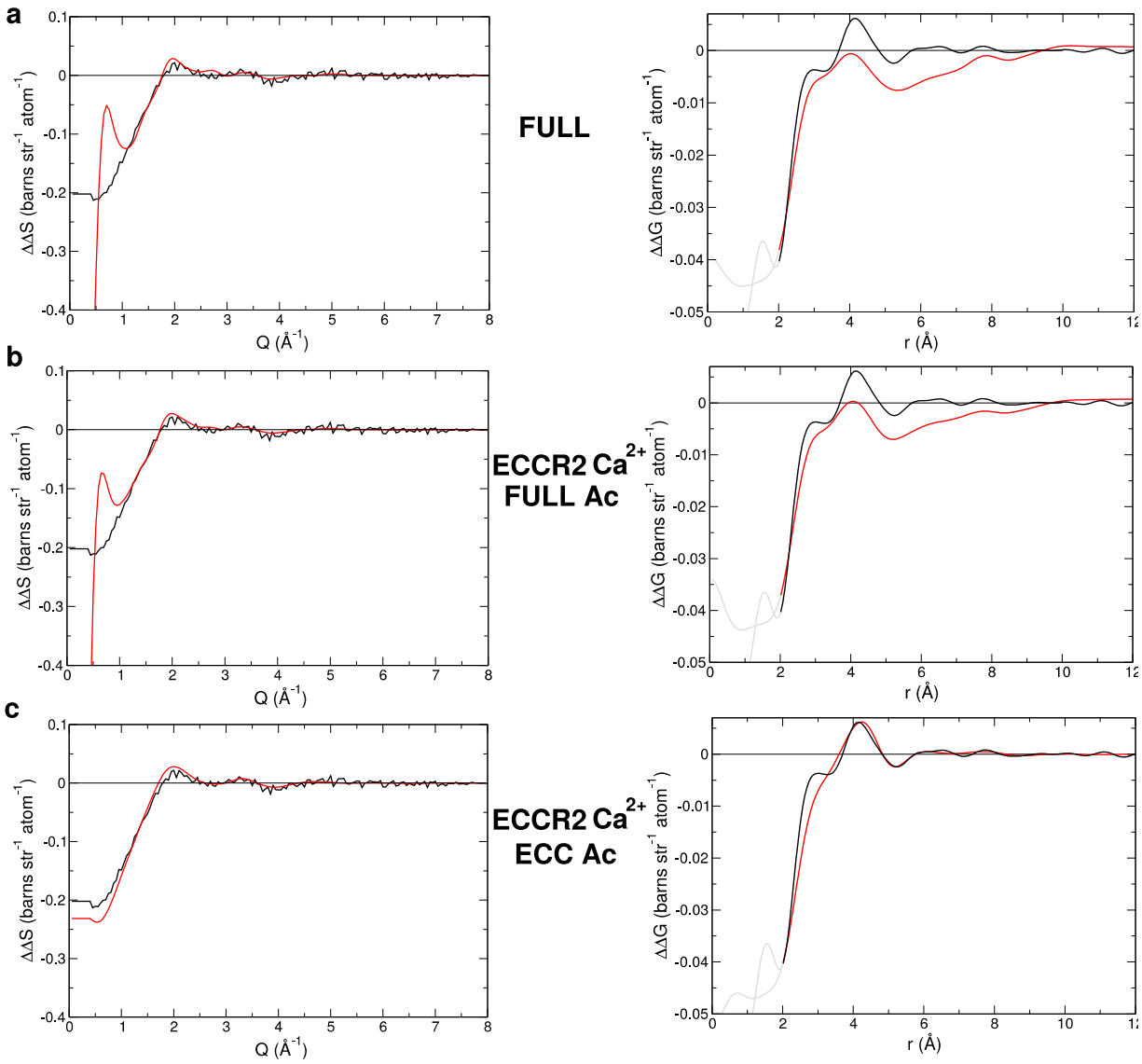

FIG. 5. Comparison between the experimental neutron double difference (black) in Q space, $\Delta\Delta S(Q)$ (left), and in r space, $\Delta\Delta G(r)$ (right), and the signal calculated from molecular dynamics simulations (red) using a) a full charges force field, b) the ECCR2 force field for Ca^{2+} and full charges for the acetate and c) the ECCR2 force field for Ca^{2+} and a ECC description for the acetate anion.

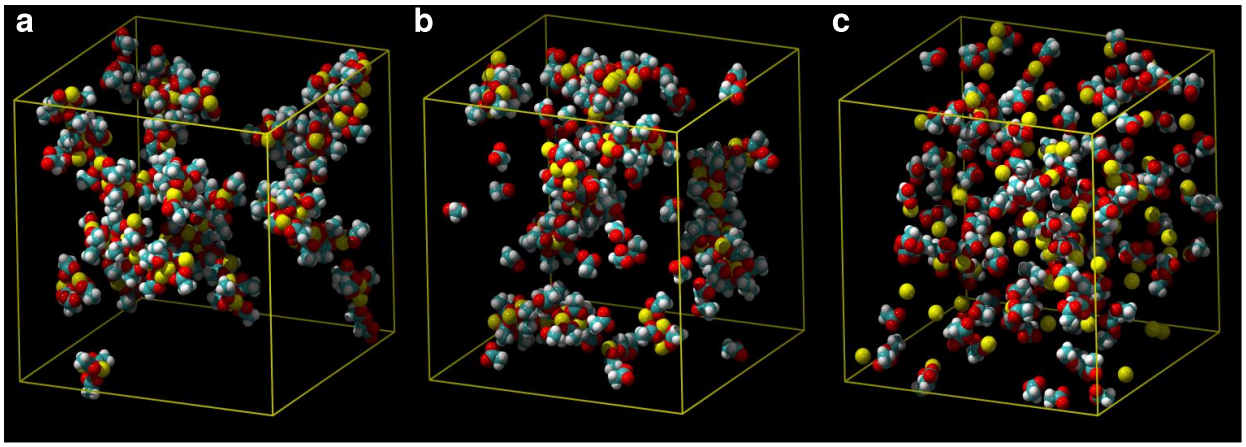

FIG. 6. Snapshots from 1 m calcium acetate MD simulations performed using a) a full charges force field, b) the ECCR2 force field for Ca^{2+} and full charges for the acetate and c) the ECCR2 force field for Ca^{2+} and a ECC description for the acetate anion.

While the neutron scattering data allow us to assess the quality of the employed force field, they represent only an indirect probe of ion pairing. Namely, the amount of ion pairing can be assessed via depletion of the number of hydrating water molecules and through the presence or absence of low Q signal due to strong ion pairing and clustering. We thus complement the neutron scattering experiment with a molecular level study of the calcium-carboxylate interaction using *ab initio* MD simulations. Since such simulations are extremely computationally expensive, we opted for the smallest carboxylate group containing species, the formate anion, which allows us to use a somewhat smaller simulation box.

An interesting property of the carboxylate group is that it can pair with Ca^{2+} either in a monodentate fashion (i.e., interacting with only one of the two carboxylate oxygens) or in a bidentate fashion (i.e., interacting simultaneously with both oxygens). We first focus on the monodentate interaction mode and obtain the free energy profile along the Ca - O(formate) distance using *ab initio* MD umbrella sampling simulations (FIG. 7a). We then compare the AIMD free energy profile with those obtained with a full charge force field for both calcium and formate moieties vs. a scaled charges force field (i.e., ECCR2 for Ca^{2+} and simple charge scaling for formate (see Methods)). At the *ab initio* level, the contact monodentate ion pair at a Ca - O distance of 2.3 Å is found to be 6 kJ/mol more stable than the SShIP. The height of the transition barrier from SShIP to CIP is around 11.5 kJ/mol. The stability of the contact ion pair is significantly overestimated, by 10 kJ/mol, with the full charge force field. This is in accord with the above simulations of concentrated calcium acetate, which yield a strong pairing in solutions described with a full

charge force field. In contrast, the use of a scaled charge force field both for calcium and acetate provides a quantitative agreement with AIMD, both for the relative stability of the contact ion pair and the height of the barrier (FIG. 7a).

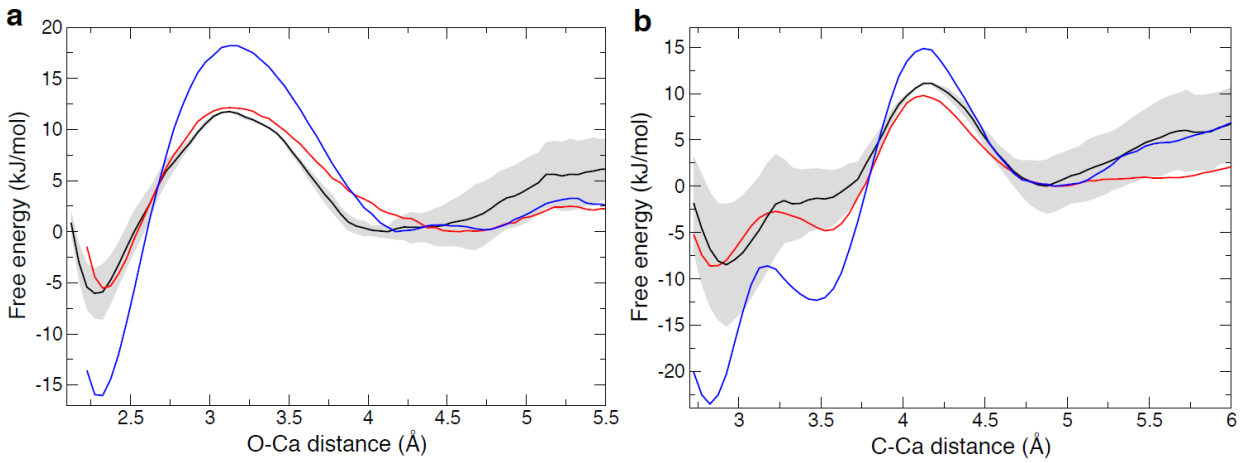

FIG. 7. a) Free energy profile along the Ca-O(formate) distance in a linear monodentate fashion and b) free energy profile along the C-Ca distance in the formate plane using AIMD (black), the full charges force field (blue) or a scaled charges force field both for calcium and formate (red).

Since the neutron scattering measurements cannot provide full information about the mode of calcium – carboxylate interaction (monodentate vs bidentate), we explore it in more details computing the free energy profile along the Ca - C(formate) distance in the COO plane with *ab initio* MD (FIG. 7b). We see that upon prolonging this distance the system is progressively driven from a bidentate geometry (Ca - C distances of 2.9 Å) to a monodentate interaction (Ca - C distance of 3.5 Å). We find the bidentate geometry more stable than the monodentate one by about 7 kJ/mol. The full charge force field again overestimates the stability of the bidentate arrangement — the bidentate well being found 11 kJ/mol lower in free energy than the monodentate arrangement as well as the height of the barrier toward the SShIP. In contrast, the scaled charge force field is in quantitative agreement with the AIMD profile, both for the relative position of the two wells and for the barrier height.

Finally, we used this scaled charge force field, to fully characterize the calcium ion pairing in the formate plane by generating the 2D free energy landscape (FIG. 8). Note that such a 2D free energy landscape requires too much sampling to be currently obtained at the AIMD level. FIG. 8 demonstrates that the global minimum is at a Ca - C distance just below 3 Å with a nearly collinear Ca – C - O_{mid} arrangement, thus corresponding to ion pairing in a bidentate fashion. Following the low energy areas (blue), we see that the dissociation from the bidentate ion pair

proceeds via a rearrangement into a monodentate conformation (broad minimum with Ca – C - O_{mid} angle around 50° and a Ca - C distance of about 3.5 Å), with a very small barrier of 6 kJ/mol. From this point on, solvent shared configurations are accessible with a 14 kJ/mol barrier. The SShIP local minimum is found at around 4.8 - 5.4 Å CaC with a weak preference for a collinear Ca – C - O_{mid} arrangement.

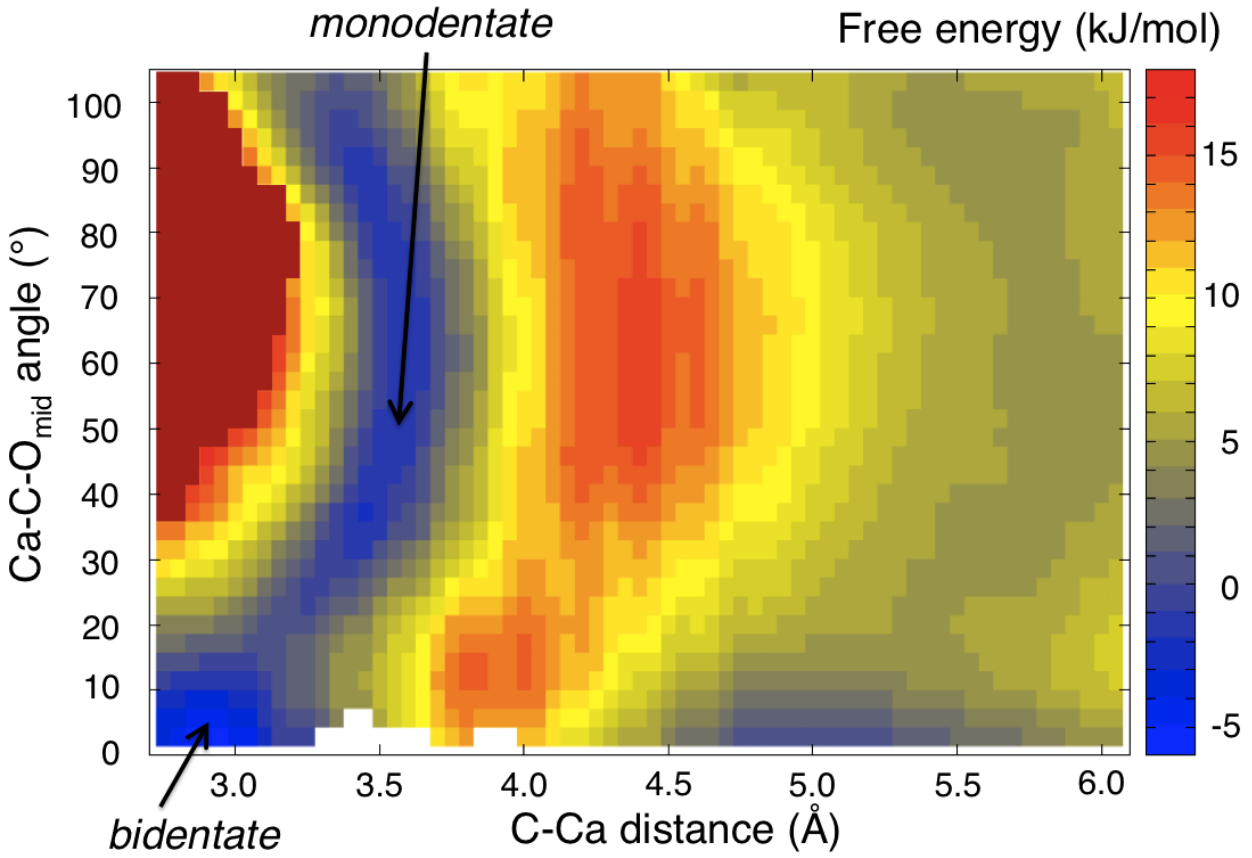

FIG. 8. 2D Free energy landscape along the C(formate)-Ca distance and the Ca-C-O_{mid} angle in the formate plane.

IV. CONCLUSIONS

We have employed *ab initio* molecular dynamics simulations together with neutron scattering experiments to quantify ion hydration and pairing in aqueous calcium chloride and formate/acetate solutions and to refine a calcium force field applicable for simulations involving this key biological ion. By inclusion of electronic polarization effects in a mean-field way via charge rescaling of both calcium and the counter-ion (be it a free anion in the solution or a

negatively charged side chain group on a protein), we are able to accurately describe calcium ions in aqueous solutions. On the one hand, we quantitatively reproduced structural features of concentrated aqueous solutions, as revealed by neutron scattering with isotopic substitution. On the other hand, the refined force field recovered within the statistical error ion-pairing free energy profiles from *ab initio* molecular dynamics. We are thus making available to the scientific community a physically well justified transferable calcium parameterization, which, thanks to its standard format, can be directly used within common molecular dynamics programs and at the same time provides an accurate description including electronic polarization effects of this ion in aqueous biological environments.

SUPPLEMENTARY MATERIAL

See supplementary material for the acetate and formate force fields, the procedure used to perform volume correction on the free energy profiles, QM calculations on a $Ca(H_2O)_6^{2+}$ cluster, and comparison of our results with additional free energy profiles, including that obtained with the AMOEBA force field.

ACKNOWLEDGMENTS

P.J. acknowledges support from the Czech Science Foundation (grant no. P208/12/G016). EDD acknowledges support from the EMBO and Marie Curie Actions (fellowship ALTF 952-2015). E.P. thanks the Czech Science Foundation (Grant No. 17-01982Y).

We thank the Institute Laue Langevin for allocating us time on the neutron scattering facility (proposal number 8-03-807). We particularly thank the staff of D4C and of the ILL chemistry laboratory for their help with the neutron scattering experiments. Calculations of the free energy profiles were made possible through a generous allocation of computer time from the North-German Supercomputing Alliance (HLRN).

REFERENCES

[1] A.G. Szent-Györgyi, Biophys. J. **15**, 707 (1975).

[2] T.M. Perney, L.D. Hirning, S.E. Leeman, and R.J. Miller, Proc. Natl. Acad. Sci. U. S. A. **83**, 6656 (1986).

[3] D.E. Clapham, Cell **131**, 1047 (2007).

[4] H. Tidow and P. Nissen, FEBS J. **280**, 5384 (2013).

[5] P.M. Kekenes-Huskey, V.T. Metzger, B.J. Grant, and J. Andrew McCammon, Protein Sci. **21**, 1429 (2012).

[6] A.O. Aykut, A.R. Atilgan, C. Atilgan, M. Ikura, and J. Maune, PLoS Comput. Biol. **9**, e1003366 (2013).

[7] A. Amcheslavsky, M.L. Wood, A.V. Yeromin, I. Parker, J.A. Freites, D.J. Tobias, and M.D. Cahalan, Biophys. J. **108**, 237 (2015).

[8] I. Frischauf, V. Zayats, M. Deix, A. Hochreiter, I.J. Polo, M. Muik, B. Lackner, B. Svobodová, T. Pammer, M. Litviňuková, A.A. Sridhar, I. Derler, I. Bogeski, C. Romanin, R.H. Ettrich, and R. Schindl, Sci. Signal. **8**, ra131 (2015).

[9] Y.I. Neela, a S. Mahadevi, and G.N. Sastry, J. Phys. Chem. B **114**, 17162 (2010).

[10] J. Adiban, Y. Jamali, and H. Rafii-Tabar, Mol. BioSyst. **13**, 208 (2017).

[11] M. Kohagen, P.E. Mason, and P. Jungwirth, J. Phys. Chem. B **118**, 7902 (2014).

[12] S. Mamatkulov, M. Fyta, and R.R. Netz, J. Chem. Phys. **138**, 24505 (2013).

[13] M. Kohagen, M. Lepšík, and P. Jungwirth, J. Phys. Chem. Lett. **5**, 3964 (2014).

[14] A. Melcrová, S. Pokorna, S. Pullanchery, M. Kohagen, P. Jurkiewicz, M. Hof, P. Jungwirth, P.S. Cremer, and L. Cwiklik, Sci. Rep. **6**, 38035 (2016).

[15] A. Magarkar, P. Jurkiewicz, C. Allolio, M. Hof, and P. Jungwirth, J. Phys. Chem. Lett. **8**, 518 (2017).

[16] J. Kahlen, L. Salimi, M. Sulpizi, C. Peter, and D. Donadio, J. Phys. Chem. B **118**, 3960 (2014).

[17] I. Leontyev and A. Stuchebrukhov, Phys. Chem. Chem. Phys. **13**, 2613 (2011).

[18] I. V. Leontyev and A.A. Stuchebrukhov, J. Chem. Theory Comput. **6**, 1498 (2010).

[19] I. V. Leontyev and A.A. Stuchebrukhov, J. Chem. Theory Comput. **8**, 3207 (2012).

[20] Y.S. Badyal, A.C. Barnes, G.J. Cuello, and J.M. Simonson, J. Phys. Chem. A **108**, 11819 (2004).

[21] E. Pluhařová, O. Marsalek, B. Schmidt, and P. Jungwirth, J. Phys. Chem. Lett. **4**, 4177 (2013).

[22] H.E. Fischer, G.J. Cuello, P. Palleau, D. Feltin, A.C. Barnes, Y.S. Badyal, and J.M. Simonson, Appl. Phys. A Mater. Sci. Process. **74**, s160 (2002).

[23] P. Mason, C. Dempsey, H.E. Fischer, J. Hladilkova, P. Jungwirth, and S. Timr, *Towards a Fuller Understanding of Salt-Bridges in Proteins. Institut Laue-Langevin (ILL), DOI: 10.5291/ILL-DATA.8-03-807* (2014).

[24] A.C. Barnes, S.B. Lague, P.S. Salmon, and H.E. Fischer, J. Phys. Condens. Matter **9**, 6159 (1997).

[25] G.W. Neilson, P.E. Mason, S. Ramos, and D. Sullivan, Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. **359**, 1575 (2001).

[26] B. Hess, C. Kutzner, D. Van Der Spoel, and E. Lindahl, J. Chem. Theory Comput. **4**, 435 (2008).

[27] G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126**, 14101 (2007).

[28] H.J.C. Berendsen, J.R. Grigera, and T.P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).

[29] S. Miyamoto and P.A. Kollman, J. Comput. Chem. **13**, 952 (1992).

[30] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, and L.G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).

[31] L.X. Dang and D.E. Smith, J. Chem. Phys. **102**, 3483 (1995).

[32] C. Oostenbrink, A. Villa, A.E. Mark, and W.F. Van Gunsteren, J. Comput. Chem. **25**, 1656 (2004).

[33] E. Pluhařová, H.E. Fischer, P.E. Mason, and P. Jungwirth, Mol. Phys. **112**, 1230 (2014).

[34] J. Wang, P. Cieplak, and P.A. Kollman, J. Comput. Chem. **21**, 1049 (2000).

[35] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, G.A. Petersson, H. Nakatsuji, M.C. X. Li, A. Marenich, J. Bloino, B.G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, J. V. Ortiz, A.F. Izmaylov, J.L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V.G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. J. A. Montgomery, J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, J.M. Millam, M. Klene, C. Adamo, R. Cammi, J.W. Ochterski, R.L. Martin, K. Morokuma, O. Farkas, J.B. Foresman, and D.J. Fox, *Gaussian 09* (Gaussian, Inc., Wallingford CT, 2009).

[36] G.A. Petersson and M.A. Al-Laham, J. Chem. Phys. **94**, 6081 (1991).

[37] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, Comput. Phys. Commun. **167**, 103 (2005).

[38] J. VandeVondele and J. Hutter, J. Chem. Phys. **127**, 114105 (2007).

[39] C. Lee, W. Yang, and R.G. Parr, Phys. Rev. B **37**, 785 (1988).

[40] A.D. Becke, Phys. Rev. A **38**, 3098 (1988).

[41] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).

[42] S. Goedecker, M. Teter, and J. Hutter, Phys. Rev. B **54**, 1703 (1996).

[43] M. Galib, T.T. Duignan, Y. Misteli, M.D. Baer, G.K. Schenter, J. Hutter, and C.J. Mundy, J. Chem. Phys. **146**, 244501 (2017).

[44] J. Timko, D. Bucher, and S. Kuyucak, J. Chem. Phys. **132**, 114510 (2010).

[45] J. Timko, A. De Castro, and S. Kuyucak, J. Chem. Phys. **134**, 204510 (2011).

[46] M.D. Baer and C.J. Mundy, J. Phys. Chem. B **120**, 1885 (2016).

[47] S. Kumar, J.M. Rosenberg, D. Bouzida, R.H. Swendsen, and P.A. Kollman, J. Comput. Chem. **13**, 1011 (1992).

[48] J. Schmidt, J. VandeVondele, I.-F.W. Kuo, D. Sebastiani, J.I. Siepmann, J. Hutter, and C.J. Mundy, J. Phys. Chem. B **113**, 11959 (2009).

[49] E. Pluhařová, M.D. Baer, C.J. Mundy, B. Schmidt, and P. Jungwirth, J. Phys. Chem. Lett. **5**, 2235 (2014).

[50] G. Placzek, Phys. Rev. **86**, 377 (1952).