

# **Computational Methodology for Solubility Prediction: Application to Sparingly Soluble Organic/Inorganic Materials**

Lunna Li,<sup>1,a)</sup> Tim Totton,<sup>2</sup> and Daan Frenkel<sup>1,b)</sup>

<sup>1</sup>*Department of Chemistry, University of Cambridge, Cambridgeshire, CB2 1EW, United Kingdom*

<sup>2</sup>*BP Exploration Operating Co. Ltd, Sunbury-on-Thames, TW16 7LN, United Kingdom*

The solubility of a crystalline material can be estimated from the absolute free energy of the solid and the excess solvation free energy. In the earlier work we presented a general-purpose molecular-dynamics-based methodology enabling solubility predictions of crystalline compounds, yielding accurate estimates of the aqueous solubilities of naphthalene at various pressures and temperatures. In the present work, we investigate a number of prototypical complex materials, including phenanthrene, calcite and aragonite over a range of temperatures and pressures. Our results provide stronger evidence for the power of the methodology for universal solubility predictions.

## **I. INTRODUCTION**

The knowledge of molecular solubility in solution is fundamentally important in many areas of science and technology<sup>1-7</sup>. Over the last few decades, researchers have been developing numerous theoretical methods to study solubility<sup>8-29</sup>. Two major routes of particular interest are: 1) simulating solid-solution coexistence directly<sup>8-9, 15-17</sup> and 2) the thermodynamic approach: computing the chemical potentials/free energies of the relevant species in the solid and solution phases<sup>14, 16-20, 29</sup> – a species in saturated solution is in equilibrium with its pure crystalline phase when its chemical potentials in the coexisting phases are equal. While direct coexistence does not require knowledge of chemical potentials/free energies, it may require very long simulations (e.g. up to microseconds) to reach equilibrium<sup>16</sup>, and its performance might be influenced by the (initial) local structures

---

<sup>a)</sup> This research was performed while Lunna Li was at Department of Chemistry, University of Cambridge, Cambridgeshire, CB2 1EW, United Kingdom.

<sup>b)</sup> Author to whom correspondence should be addressed. Electronic mail: df246@cam.ac.uk

of solid-solution interfaces; overall direct coexistence might be unattractive for compounds of low solubilities because a large volume of liquid is necessary to observe only a handful of molecules in the solution. For this reason, a generic, free-energy-based solubility prediction can be of considerable practical value. In the context, however, the bottleneck is how to calculate the chemical potentials of both the crystalline and solution phases accurately. It is not so much that there are no techniques to compute these chemical potentials, but that most work is limited to address the problem on a case-to-case basis. This is probably related to the fact that, the scale and complexity of the topic (viz. compounds of drastically different nature in different solvent environments under varying external/internal conditions) means that the reliability of molecular-dynamics (MD)/Monte Carlo (MC)-based studies depends profoundly on the availability of classical force-fields, which are required to work for both phases. As a result, the computational problem is often circumvented by using approximate informatics-based approaches<sup>24-28</sup>. However, overall, informatics-based approaches lack fundamental “molecular” understandings and, consequently, may have little predictive power outside “training sets”. Within the category of thermodynamic approaches, apart from directly computing chemical potentials<sup>14, 16-20</sup>, one can determine the composition of a saturated solution by imposing the solid chemical potential in a grand canonical style simulation of the liquid, such as in the osmotic ensemble method/OEMC/OEMD<sup>10-13, 21</sup>, assuming that the chemical potential of the crystal is known from other sources. This method can be more advantageous for multi-electrolyte solutions.

The present work continues to make use of the first thermodynamic concept and aims to contribute a general-purpose approach to solubility prediction via standard, open-source software such as LAMMPS<sup>30</sup>. As in our earlier work<sup>29</sup>, the numerical techniques utilized are standard alchemical free-energy methods such as free energy perturbation (FEP)<sup>31-32</sup> and thermodynamic integration (TI)<sup>33-36</sup> that are commonly used in MD or MC simulations for

computing free energy differences<sup>37-48</sup>. In devising a generic approach to compute chemical potentials in dense liquids, we aim to resolve the singular behavior of solute addition into the solution during thermodynamic integration steps<sup>49-51</sup>. In reference 29 we have employed a singularity-free cavity method where a softly repulsive cavity is first created in the solvent before inserting the solute in steps. One advantage is that the cavity, once created, can be used for more than one type of solute, and molecules of different sizes and structures. In computing the chemical potentials of solid molecules, we have used a molecular adaptation of the well-tested Einstein crystal method<sup>29, 34, 39-44</sup>. In the present work, we apply our generic methodology to a set of sparingly soluble organic/inorganic compounds, including phenanthrene, calcite, aragonite and caffeine. Our methodology has been explicated in reference 29, here we only sketch the essentials in Appendix A-C and refer the reader to reference 29 for details.

## II. SIMULATION DETAILS

All simulations were performed with the MD software package LAMMPS<sup>30</sup>. We used a selection of classical force-fields depending on their availability for the systems studied, including OPLS-AA (optimized potentials for liquid simulations all atomic)<sup>52</sup>, SPC (simple point charge) water<sup>53</sup>, one carbonate model<sup>54-57</sup> and SPC/Fw (flexible simple point charge) water<sup>58-59</sup>, GAFF (general AMBER force field)<sup>60-61</sup> caffeine in SPC/E (extended simple point charge model) water<sup>62</sup>, all in the forms of 12-6 LJ (Lennard-Jones) potential<sup>63</sup> or the Buckingham potential<sup>64</sup>, and the Lorentz-Berthelot combining rules<sup>65-67</sup> were employed for interatomic mixings when applicable. The integration time step was generally  $dt = 0.5-1.0$  fs. In LAMMPS<sup>30</sup>, the equations of motion are based on the work of Shinoda *et al.*<sup>68</sup>, Martyna *et al.*<sup>69</sup> and Parrinello and Rahman<sup>70</sup>; for time integration schemes, we consistently used the default velocity-Verlet integrator<sup>71</sup>. For isobaric-isothermal  $NpT$  simulations, a Nosé-

Hoover<sup>72-73</sup> thermostat and barostat were used to equilibrate all systems. For canonical *NVT* simulations, both Nosé-Hoover<sup>72-73</sup> and Langevin<sup>74</sup> thermostats were used. For long-range electrostatic interactions, LAMMPS' default Ewald<sup>34,75</sup>/PPPM (Particle-Particle-Particle-Mesh)<sup>34,76</sup> summation with tin-foil boundary conditions<sup>77</sup> was used.

For nonbonding van der Waals interactions, we used standard cut-off values. For phenanthrene we applied  $r_{\text{cut}} = 12.0 \text{ \AA}$  with  $N_{\text{water}} = 2048$  in the solution and  $r_{\text{cut}} = 10.0 \text{ \AA}$  with  $N_{\text{phenanthrene}} = 250$  in the crystal phase, for all pressure ( $p = 0.1, 100$  and  $200 \text{ MPa}$ ) at  $T = 298 \text{ K}$ . For calcite and aragonite, we used  $r_{\text{cut}} = 9.0 \text{ \AA}$  and a tapering function for the last  $3.0 \text{ \AA}$  for all van der Waals potentials, as implemented in references 54-57. We used  $N_{\text{water}} = 2048$  for all carbonate solutions. For carbonate solids, we had  $N_{\text{CaCO}_3} = 540$  for calcite and  $N_{\text{CaCO}_3} = 500$  for aragonite. We studied calcite at  $p = 0.1, 100$  and  $200 \text{ MPa}$  for  $T = 298 \text{ K}$ , and at  $T = 298, 320$  and  $348 \text{ K}$  for  $p = 0.0.1 \text{ MPa}$ . For aragonite, we investigated  $p = 1$  and  $T = 298 \text{ K}$  only. Finally, for GAFF-based caffeine<sup>60-61</sup> in SPC/E<sup>62</sup> water at ambient conditions, we simulated  $N_{\text{water}} = 2048$  with  $r_{\text{cut}} = 12.0 \text{ \AA}$  in the solution, and  $N_{\text{caffeine}} = 200$  with  $r_{\text{cut}} = 10.0 \text{ \AA}$  in the crystal ( $\beta$ -caffeine) phase. For all crystal phases, we started with experimental crystal structures, after which the final configurations were taken as the reference structures for the extended Einstein crystal method. Generally equilibrating simulations were typically above  $1 \text{ ns}$ , after which further runs of more than  $1 \text{ ns}$  were used for data.

Gauss-Legendre quadrature<sup>78</sup> and the trapezoid rule<sup>79</sup> were used for TI in data analysis. In both the extended Einstein crystal method and cavity method, statistical errors were calculated from block averaging<sup>34, 80</sup>, and errors of calculated values are obtained from error propagations of the individual block averaging errors.

### III. RESULTS AND DISCUSSIONS

#### A. Phenanthrene: Excess Solvation Free Energy in Water

In reference 29, we reported the aqueous excess solvation free energies of naphthalene for a range of pressures and temperatures, yielding accurate solubility estimations comparable with experimental data. Phenanthrene represents a more complex aromatic molecule that can be described using the same OPLS-AA<sup>52</sup> force-field parameters.

For  $p = 0.1$ -200 MPa and  $T = 298$  K, our computed excess solvation free energies of phenanthrene are summarized in Table I (as before, for simplicity, one typical statistical error is quoted to represent general solvation errors). For  $T = 298$  K and  $p = 0.1$  MPa, we obtained an average  $\Delta\mu_{\text{excess}} = -12.25(50)$  kJ/mol, in good agreement with  $\Delta\mu_{\text{excess}} = -11.76$  kJ/mol in the simulation reference<sup>38</sup>, and both are close to the experimental<sup>28, 81-83</sup> value of  $-16.53$  kJ/mol. As the two simulations agree well, the difference between the simulations and the experiment is likely to be caused by deficiencies of the force field.

TABLE I Summary of the computed excess solvation free energies of phenanthrene (at infinite dilution) and the average infinitely dilute solution densities for  $p = 0.1$ -200 MPa. Pressures are expressed in MPa, temperature in K, energies in kJ/mol and densities in kg/m<sup>3</sup>. See the supplementary material for details.

p	T	$\Delta\mu_{\text{excess}}$	$\langle\rho_{\text{solution}}\rangle$
0.1	298	-12.25(50)	977.8(2)
100	298	4.85(50)	1023.8(1)
200	298	19.74(50)	1061.5(1)

Figure 1 (a), (b) and (c) show plots of the computed excess solvation free energies at three different pressures and at increasing cavity sizes. We see that for all pressures, result 1 and result 2 converge to each other, and  $\Delta\mu_{\text{excess}}(\lambda)$  becomes independent of  $\lambda$  as  $\lambda$  increases. As emphasized in our earlier work<sup>29</sup>, the systematic trend indicates that as the cavity (to

insert/grow the LJ part of the solute) becomes larger, “end-point” singularity inaccuracy is gradually reduced until removal, and fluctuations of  $\Delta\mu_{\text{excess}}(\lambda)$  become comparable with statistical errors. The threshold size might be around  $\lambda = 1.5$  here, but exact values are not essential.

As in our earlier work, the pressure derivative of  $\Delta\mu_{\text{excess}}$  at fixed  $T$  and infinite dilution is the excess partial molar volume at infinite dilution<sup>84-85</sup>,  $v_{\text{excess}}^{\text{solute}}$ ,

$$\left(\frac{\partial\Delta\mu_{\text{excess}}}{\partial p}\right)_T = v_{\text{excess}}^{\text{solute}} \quad (1)$$

In Fig. 1 (d), the gradient gives a computed partial molar volume of 160.0(3.0) cm<sup>3</sup>/mol compared with the experimental partial molar volume<sup>86</sup> of 159.9 cm<sup>3</sup>/mol.

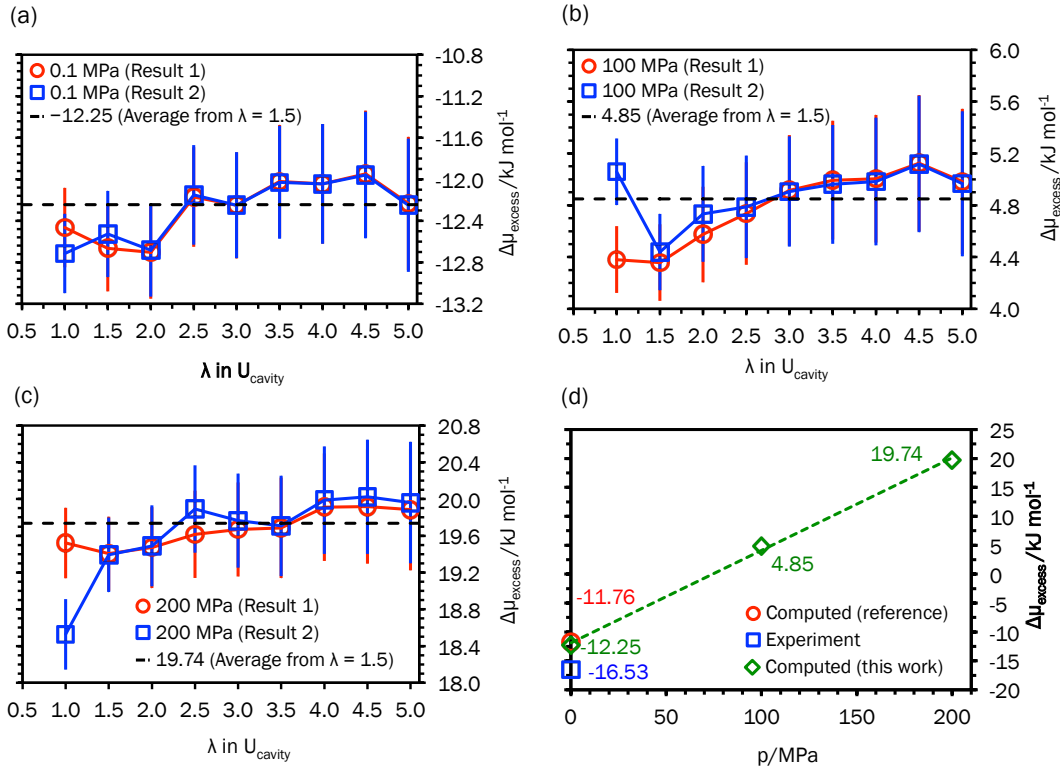


FIG. 1. Computed excess solvation free energies of phenanthrene in water at  $T = 298$  K computed with increasing cavity size  $\lambda$  for  $p$  at (a) 0.1 MPa; (b) 100 MPa; (c) 200 MPa.  $\lambda$  is a parameter in the functional form of the repulsive cavity that can be adjusted to change the cavity size (please refer to Appendix C and reference 29 for details). The black dashed lines in (a), (b) and (c) are the averages from  $\lambda = 1.5$  for the three pressures, and they are also the empty green rhombuses in (d). In (a)-(c), result 1 and result 2 are from the same  $\Delta G_{\text{grow}}$  and  $\Delta G_{\text{shrink}}$  but different  $\Delta G_{\text{insert}}$ , where the van der Waals’ nonbonding contributions are from multi-steps TI and single-step FEP respectively.

## B. Phenanthrene: Absolute Solid Free Energy and Solubility Prediction

Next, we consider computing the absolute free energy of crystalline phenanthrene, via implementing the extended Einstein crystal method as described in reference 29. Phenanthrene has several known polymorphs, but under the conditions studied, the molecule is in its ambient polymorph (II)<sup>87</sup>. Fig. 2 below shows the pressure dependence of our simulated average volume per phenanthrene molecule, which we compare with the experimental data<sup>88</sup>. The average ambient density obtained in the simulations differs by around 1.3 % from the experimental value<sup>88</sup> and the individual lattice parameters  $a$ ,  $b$ ,  $c$  and  $\beta$  differ between 0.6 and 2% different from the experimental values<sup>87</sup>. As exemplified for naphthalene before, we used the same force field to compute the absolute Helmholtz free energies of solid phenanthrene for  $p = 0.1$ -200 MPa and  $T = 298$  K, as summarized in TABLE II.

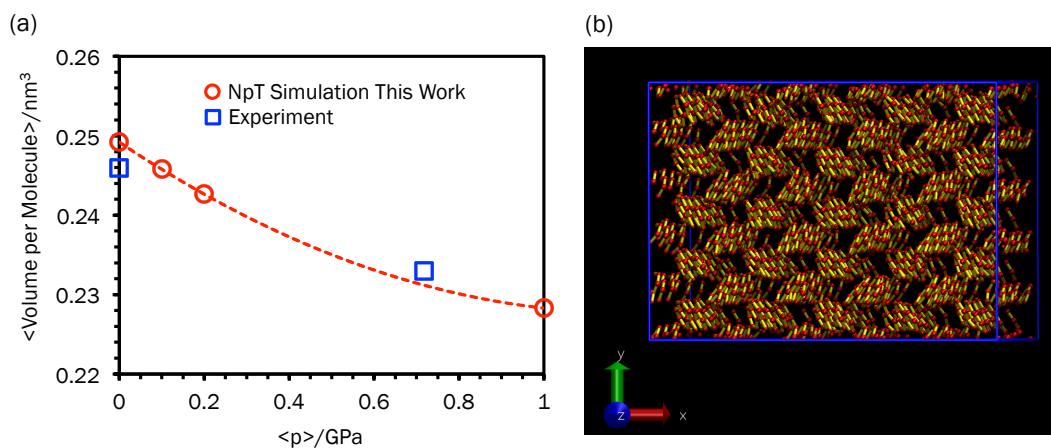


FIG. 2. (a) Simulated and experimental<sup>88</sup> volume per phenanthrene with increasing pressure; (b) snapshot of the phenanthrene crystal during ambient  $NpT$  equilibrations.

TABLE II Summary of the computed absolute solid free energies  $A_{\text{solid}}$  of phenanthrene. Pressures are in MPa, temperatures in K, energies in kJ/mol and densities in kg/m<sup>3</sup>.  $\langle\rho_{\text{solid}}\rangle$  is the average density during  $NpT$  relaxations, and  $\rho_{\text{solid}}$  is the density of the reference structure (final configurations of relaxations). See the supplementary material for details.

p	T	$A_{\text{solid}}/N_{\text{solid}}$	$pV_{\text{solid}}/N_{\text{solid}}$	$G_{\text{solid}}/N_{\text{solid}}$	$\rho_{\text{solid}}$	$\langle\rho_{\text{solid}}\rangle$
0.1	298	-57.86(15)	0.00	-57.86(15)	1185.0	1187.7(5)
100	298	-57.99(15)	14.78	-43.21(15)	1205.7	1204.1(4)
200	298	-58.19(15)	29.16	-29.03(15)	1222.4	1219.8(5)

Along with  $\Delta\mu_{\text{excess}}$  calculated in Section III. A, we can proceed to estimate the aqueous solubility of phenanthrene. We compare our solubilities with experiment evidence<sup>86</sup>, and Fig. 3 is a plot of the pressure dependence of our simulated results and extracted experiment data.

Phenanthrene (like naphthalene) has a low aqueous solubility, so it is valid to take the excess chemical potential at infinite dilution as its excess chemical potential at saturation. In Fig. 3, for  $T = 298$  K and  $p = 0.1$  MPa, the computed mole-fraction solubility is  $3.09(0.81) \times 10^{-7}$ , compared with the experimental<sup>86</sup>  $1.1 \times 10^{-7}$ . The agreement within a factor of three is excellent considering we only use a simple force field. In addition, as in the case of naphthalene, the negative pressure dependence of phenanthrene solubility is well reproduced.

Furthermore, the slopes of the least squares fits in Fig. 3 can give an estimate of the computed volume changes accompanying aqueous dissolution<sup>86</sup>,

$$\Delta v = -RT \left[ \partial \ln x / \partial p \right]_T \quad (2)$$

where  $\Delta v$  is defined as the difference between the partial molar volume at infinite dilution  $v_{\text{excess}}^{\text{solute}}$  (see Eq. (1)) and the partial molar volume in the crystalline phase  $v_{\text{solid}}$ ,

$$\Delta v = v_{\text{excess}}^{\text{solute}} - v_{\text{solid}} \quad (3)$$

Fig. 3 gives  $\Delta v_{\text{computed}} = 16.8(4.6) \text{ cm}^3/\text{mol}$  and  $\Delta v_{\text{experimental}} = 12.4(3) \text{ cm}^3/\text{mol}$ <sup>86</sup>.



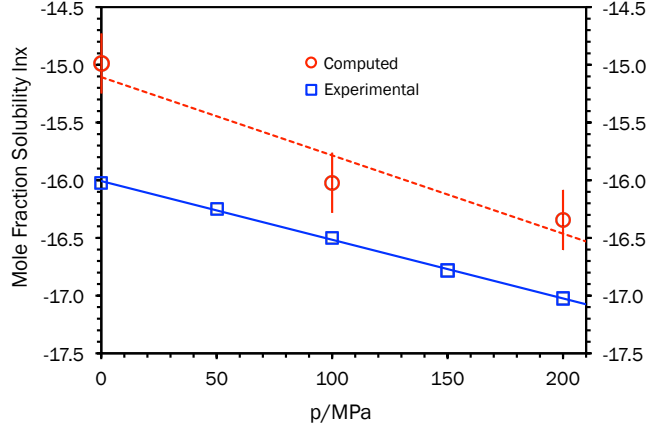


FIG. 3. Computed and experimental<sup>86</sup> mole fraction aqueous solubility of phenanthrene with increasing pressure. See the supplementary materials for details of error propagations.

### C. Calcite: Single-ion Excess Solvation Free energies

Thus far, we have only worked with neutral organic molecules, but the solubility of inorganic compounds is also important. Here we address the solubility problem of  $\text{CaCO}_3$  calcite. Calcite is also sparingly soluble in water so we can assume an infinite dilution (viz. mean ionic activity coefficient = 1) when computing solvation free energies. In practice, this can be achieved by solvating a single ion in a large simulation box. In short, we need to compute the single-ion excess solvation free energies  $\Delta\mu_{\text{excess}}^{\text{Ca}^{2+}}$  and  $\Delta\mu_{\text{excess}}^{\text{CO}_3^{2-}}$ , and the calcite solid free energies  $G_{\text{CaCO}_3}$ . In contrast to neutral solutes, computing single ionic solvation free energies involves two steps<sup>89</sup>: 1) first we calculate the nonbonding van der Waals contribution (via the cavity method); 2) next, we calculate the ionic charging free energy; the total single-ion excess solvation free energy is the sum of the two.

The total single-ion excess solvation free energy is,

$$\Delta\mu_{\text{excess}} = \Delta G_{\text{excess}}^{\text{vdwl}} + \Delta G_{\text{excess}}^{\text{charging}} \quad (4)$$

In Eq. (4),  $\Delta\mu_{\text{excess}}$  refers to  $\Delta\mu_{\text{excess}}^{\text{Ca}^{2+}}$  or  $\Delta\mu_{\text{excess}}^{\text{CO}_3^{2-}}$ .

We compute the ionic charging free energy  $\Delta G_{\text{excess}}^{\text{charging}}$  in  $NpT$  ensemble via TI (see the supplementary material for details),

$$\Delta G_{\text{excess}}^{\text{charging}}(q) = \Delta G_{\text{excess}}^{\text{solute-solvent}}(q) + 1/2 q^2 \langle \xi_{\text{EW}} \rangle \quad (5)$$

$$\Delta G_{\text{excess}}^{\text{solute-solvent}}(q) = \int_0^1 \langle U_{\text{solute-solvent}}(\alpha) / \alpha \rangle_{\alpha, NpT} d\alpha \quad (6)$$

In Eq. (5),  $q$  is the full ionic charge (viz.  $q = +2/-2$  for  $\text{Ca}^{2+}/\text{CO}_3^{2-}$ ), and  $\Delta G_{\text{excess}}^{\text{solute-solvent}}(q)$  is the ion-solvent contribution that can be obtained from the Kirkwood linear coupling method<sup>36</sup> as  $q$  is increased from zero to its full value. In contrast to neutral molecules, the second term,  $1/2 q^2 \langle \xi_{\text{EW}} \rangle$  is the essential “self-interaction term” or “finite-size correction”, and it is a significant contribution in computing ionic solvation energies when using lattice summation techniques such as Ewald/PPPM. Including this term in the final  $\Delta G_{\text{excess}}^{\text{charging}}(q)$  removes most electrostatic finite-size effects in periodic boundary conditions<sup>90-91</sup>. Here, we adopt the simplest finite-size correction  $1/2 q^2 \langle \xi_{\text{EW}} \rangle$  without considering the solvent dielectric constant  $\epsilon_r$  or the finite ionic size  $R$ . It has been reported<sup>92</sup> that for aqueous solutions with more than 100 water molecules, the difference incurred to  $\Delta G_{\text{excess}}^{\text{solute-solvent}}(q)$  between using  $1/2 q^2 \xi_{\text{EW}}$  and  $1/2 q^2 (\xi_{\text{EW}} + 2\pi R^2 / L^3)(1 - 1/\epsilon_r)$  is  $\leq 1-2\%$ . When conducting simulations in constant-pressure  $NpT$  ensemble, the average  $1/2 q^2 \langle \xi_{\text{EW}} \rangle$  should be used<sup>93</sup>.

Table III summarizes our computed  $\Delta\mu_{\text{excess}}^{\text{Ca}^{2+}}$  and  $\Delta\mu_{\text{excess}}^{\text{CO}_3^{2-}}$  for  $p = 0.1$  MPa and  $T = 298-348$  K. The total statistical error is around 0.8 kJ/mol, of which the charging error contributes around 0.3-0.5 kJ/mol. In summary, for ambient conditions, we have obtained  $\Delta\mu_{\text{excess}}^{\text{Ca}^{2+}} = -1446$  kJ/mol and  $\Delta\mu_{\text{excess}}^{\text{CO}_3^{2-}} = -1310$  kJ/mol in comparison with the simulation reference<sup>57</sup> with  $\Delta\mu_{\text{excess}}^{\text{Ca}^{2+}} = -1444$  kJ/mol and  $\Delta\mu_{\text{excess}}^{\text{CO}_3^{2-}} = -1312$  kJ/mol, which has been parameterized against the experimental  $\Delta\mu_{\text{excess}}^{\text{Ca}^{2+}} = -1444$  kJ/mol<sup>94</sup> and  $\Delta\mu_{\text{excess}}^{\text{CO}_3^{2-}} = -1315$  kJ/mol<sup>95</sup>

TABLE III Computed single-ion excess solvation free energies of  $\text{Ca}^{2+}$  and  $\text{CO}_3^{2-}$  at  $p = 0.1\text{-}200$  MPa and  $T = 298\text{-}348$  K. Pressure are in MPa, temperatures in K, energies are in kJ/mol, and densities in  $\text{kg/m}^3$ . The densities are the average solution densities.

$p$	$T$	$\Delta\mu_{\text{excess}}^{\text{Ca}^{2+}}$	$\langle\rho_{\text{solution}}^{\text{Ca}^{2+}}\rangle$	$\Delta\mu_{\text{excess}}^{\text{CO}_3^{2-}}$	$\langle\rho_{\text{excess}}^{\text{CO}_3^{2-}}\rangle$	$\langle\rho_{\text{SPC/Fw}}\rangle$
0.1	298	-1446.0(8)	1010.3(1)	-1310.0(8)	1011.3(1)	1010.7(1)
100	298	-1448.6(8)	1050.6(1)	-1311.8(8)	1051.0(1)	1049.9(1)
200	298	-1449.5(8)	1084.5(1)	-1312.5(8)	1085.1(1)	1084.4(1)
0.1	323	-1441.1(8)	998.2(3)	-1304.4(8)	997.8(1)	997.0(1)
0.1	348	-1433.5(8)	979.7(1)	-1297.7(8)	980.8(1)	979.8(1)

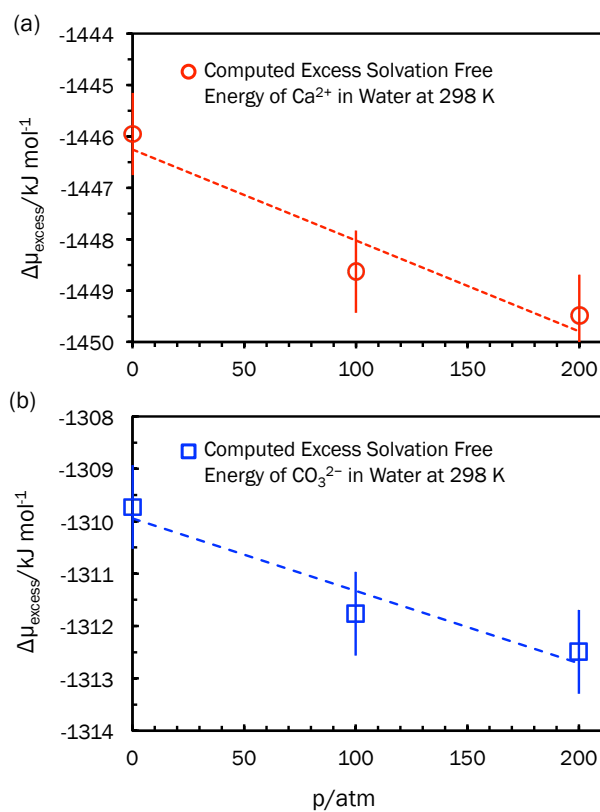


FIG. 4. Computed single-ion excess solvation free energies with increasing pressure at  $T = 298$  K for (a)  $\text{Ca}^{2+}$ ; (b)  $\text{CO}_3^{2-}$ .

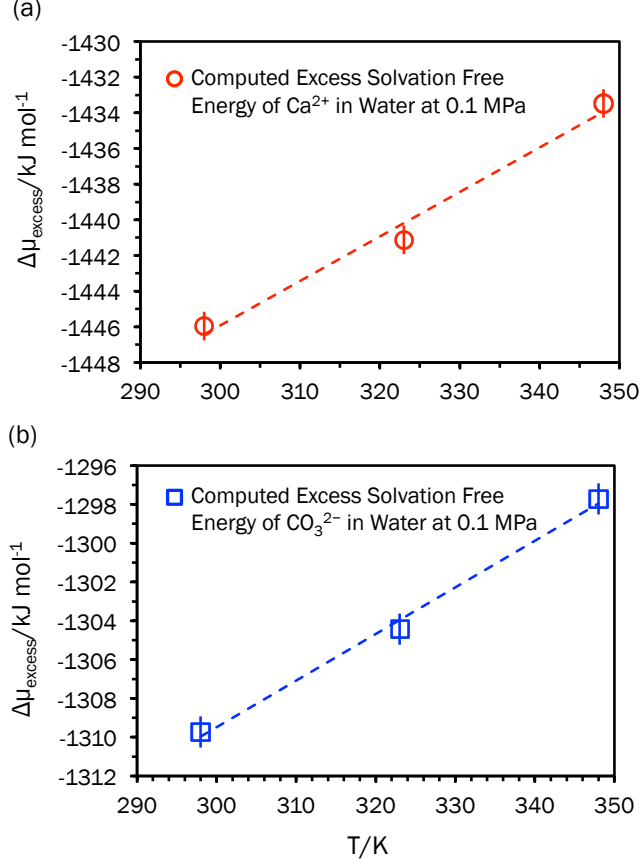


FIG. 5. Computed single-ion excess solvation free energies with increasing temperature at  $p = 0.1$  MPa for (a)  $\text{Ca}^{2+}$ ; (b)  $\text{CO}_3^{2-}$ .

In Fig.4, similarly, the  $p$  derivative of  $\Delta\mu_{\text{excess}}$  at constant  $T$  at infinite dilution is the excess ionic partial molar volume at infinite dilution  $(\partial\Delta\mu_{\text{excess}}/\partial p)_T = v_{\text{excess}}^{\text{ion}}$ . The slopes of Fig. 4 (a) and (b) thus gives  $v_{\text{excess}}^{\text{Ca}^{2+}} = -17.69(5.66) \text{ cm}^3/\text{mol}$  and  $v_{\text{excess}}^{\text{CO}_3^{2-}} = -12.45(5.66) \text{ cm}^3/\text{mol}$  respectively. It is known that the ionic partial molar volume is made up of two terms, one describing the (positive) intrinsic ionic volume and the other describing the (negative) volume change due to the electrostatic interaction between the solvent and the ion<sup>96</sup>, as first suggested by Drude and Nernst<sup>97</sup>. Here, we also find that the van der Waals contribution increases with increasing pressure, namely  $(\partial\Delta\mu_{\text{excess}}^{\text{vdwl}}/\partial p)_T$  is positive, and the charging contribution decreases with increasing pressure, namely  $(\partial\Delta\mu_{\text{excess}}^{\text{charging}}/\partial p)_T$  is negative and

dominating (see the supplementary material) – this is in line with the negative pressure dependence of overall  $\Delta\mu_{\text{excess}}$ . Experimentally, Marcus<sup>95-96</sup> quotes  $-28.9/-28.67 \text{ cm}^3/\text{mol}$  for  $\text{Ca}^{2+}$  and  $+6.7 \text{ cm}^3/\text{mol}$  for  $\text{CO}_3^{2-}$ , while Han<sup>98-99</sup> quotes  $-25.5 \text{ cm}^3/\text{mol}$  for  $\text{Ca}^{2+}$  and  $+5.5 \text{ cm}^3/\text{mol}$  for  $\text{CO}_3^{2-}$ , resulting in the experimental sum of the pair between  $-20$  and  $-22.2 \text{ cm}^3/\text{mol}$  versus our value of around  $-30 \text{ cm}^3/\text{mol}$ . Hereby the good news is that the negative sum of the partial molar volumes is reproduced by our simulations, and hence the direction of increasing solubility of calcite with increasing pressure; on the other hand, our simulations also suggest that  $\text{Ca}^{2+}$  and  $\text{CO}_3^{2-}$  should both exhibit negative partial molar volumes in aqueous solution – and this is not true. The failure to capture precisely the positive nature of the  $\text{CO}_3^{2-}$  partial molar volume might be related to the calcite-water model itself, but the magnitude of variation with increasing pressure is small and comparable within statistical errors, and thus one cannot give sound speculations without thorough future investigations. Also, this discrepancy could be related to many other factors that are not primarily studied in the present work, such as dielectric constants, ionic sizes, and second-order accuracy in ionic charging free energy. For example,  $\Delta\mu_{\text{excess}}$  would decrease with increasing pressure at a slightly faster rate than depicted in Fig. 4 due to  $\epsilon_r$  of SPC/Fw water increasing with higher pressure<sup>100</sup>.

Finally, we have also computed the excess solvation free energies of  $\text{Ca}^{2+}$  and  $\text{CO}_3^{2-}$  in water at  $T = 323 \text{ K}$  and  $T = 348 \text{ K}$  for  $p = 0.1 \text{ MPa}$ , all summarized in Table III. For both  $\text{Ca}^{2+}$  and  $\text{CO}_3^{2-}$ , we see that  $\Delta\mu_{\text{excess}}$  increases with increasing temperature. It is known that the ionic partial molar entropy of ionic solvation can be estimated from the thermodynamic relationship  $\left(\partial\Delta\mu_{\text{excess}}/\partial T\right)_p = -\Delta S_{\text{excess}}^{\text{solute}}$ . In Fig. 5, we plot our computed  $\Delta\mu_{\text{excess}}$  against the three temperatures, and our least-square fit slopes give  $\Delta S_{\text{excess}}^{\text{Ca}^{2+}} = -250 \text{ Jmol}^{-1}/\text{K}$  comparing with the experimental reference  $-271 \text{ Jmol}^{-1}/\text{K}$ <sup>95</sup> and the numerical reference  $-295 \text{ Jmol}^{-1}/\text{K}$ <sup>57</sup>,

and  $\Delta s_{\text{excess}}^{\text{CO}_3^{2-}} = -246 \text{ Jmol}^{-1}/\text{K}$  comparing with the experimental reference  $-264 \text{ Jmol}^{-1}/\text{K}$ <sup>95</sup> and the numerical reference  $-260 \text{ Jmol}^{-1}/\text{K}$ <sup>57</sup>. The difference could be attributed to many factors that are not primarily investigated here, such as the simulation lengths, number of temperatures and elements used in charging free energies such as finite-size corrections, dielectric constants and second-order accuracy. Despite this, the negative nature of  $\Delta s$  for both ions is well reproduced.

#### D. Calcite: Absolute Solid Free Energy and Solubility Prediction

As before, in order to estimate the aqueous solubility of calcite, we computed the absolute solid free energies of calcite crystals at  $p = 0.1\text{-}200 \text{ MPa}$  and  $T = 298\text{-}423 \text{ K}$ . Using the designated calcite force field<sup>54-57</sup>, lattice parameters of the ambient calcite crystal from our  $NpT$  simulations agree with both the simulation reference<sup>57</sup> as well as experiments<sup>101</sup> (see the supplementary material). In addition, we also checked the bulk modulus at  $T = 298 \text{ K}$ . Using 0.1, 100 and 200 MPa, we obtained  $K_m = 81 \text{ GPa}$ , compared with  $K_m = 85 \text{ GPa}$  from the simulation reference<sup>57</sup> and  $K_m = 67 \text{ GP}$  from the experiment<sup>102</sup>. We also calculated the thermal expansion coefficients  $\bar{\alpha}_a = -2.2 \times 10^{-6} \text{ K}^{-1}$ ,  $\bar{\alpha}_b = -2.1 \times 10^{-6} \text{ K}^{-1}$  and  $\bar{\alpha}_c = 27.0 \times 10^{-6} \text{ K}^{-1}$ , comparing with  $\bar{\alpha}_a = \bar{\alpha}_b = -2.1 \times 10^{-6} \text{ K}^{-1}$  and  $\bar{\alpha}_c = 26.2 \times 10^{-6} \text{ K}^{-1}$  from the simulation reference<sup>57</sup> and  $\bar{\alpha}_a = \bar{\alpha}_b = -2.8 \times 10^{-6} \text{ K}^{-1}$  and  $\bar{\alpha}_c = 32.3 \times 10^{-6} \text{ K}^{-1}$  from the experiment<sup>103</sup>.

For inorganic salts with low aqueous solubilities, we consider the solubility product  $K_{\text{sp}}$  for its solubility information. The congruent dissolution reaction of  $\text{CaCO}_3$  calcite in water is,



In the limit of infinite dilution, we can write the solubility product  $K_{\text{sp}}$  as the product of the (equal) molar concentration of the two constituting ions,

$$K_{\text{sp}} = [\text{Ca}^{2+}(\text{aq})][\text{CO}_3^{2-}(\text{aq})] \quad (8)$$

It is well known that in aqueous solutions  $\text{CO}_3^{2-}(\text{aq})$  occurs in equilibrium with  $\text{HCO}_3^-$  (aq) and  $\text{CO}_2(\text{aq})$ , thus the actual solubility of calcite depends on both the pH of the solution and the  $\text{CO}_2$  partial pressure<sup>104-105</sup>. As a consequence, actual measured solubilities can often be different from  $K_{\text{sp}}$  solubilities calculated via Eq. (8). Here, we only study the effects of hydrostatic pressure and temperature in pure water, and do not investigate the effect of pH hence we only compute and compare  $K_{\text{sp}}$  solubilities.

Using the extended Einstein crystal method, we calculated the absolute solid free energies of calcite  $G_{\text{CaCO}_3}$  for  $p = 0.1\text{-}200$  MPa and  $T = 298\text{-}348$  K: the results are summarized in Table IV below. Because calcite crystal contains atomic cations and molecular anions, the actual implementation is slightly different from naphthalene and phenanthrene: in summary, we treat one cation-anion pair as “one molecule”, and we only need to consider the molecular orientation contribution of the carbonate.

TABLE IV Summary of the computed absolute solid Gibbs free energies and number density solubilities at  $T = 298\text{-}423$  K and  $p = 0.1\text{-}200$  MPa for calcite. Pressures are in MPa, temperatures in K and energies in kJ/mol.

p	T	$G_{\text{CaCO}_3}/N_{\text{CaCO}_3}$
0.1	298	-2838.88(8)
100	298	-2835.17(8)
200	298	-2831.46(8)
0.1	323	-2835.07(8)
0.1	348	-2831.23(8)

In order to calculate solubility, we write Eq. (15) in the context of calcite,

$$\Delta\mu_{\text{excess}}^{\text{Ca}^{2+}} + \Delta\mu_{\text{excess}}^{\text{CO}_3^{2-}} + 2k_{\text{B}}T \ln(\rho_{\text{CaCO}_3}) = (A_{\text{CaCO}_3} + pV_{\text{CaCO}_3})/N_{\text{CaCO}_3} \quad (9)$$

As in the cases of naphthalene and phenanthrene, the computed solubility  $\rho_{\text{CaCO}_3}$  should correspond immediately to the number of solute molecules (ion pairs) per unit volume (viz.  $\text{\AA}^3$ ), we need to convert it into molar solubility and then  $K_{\text{sp}}$  for direct comparison with experimental data, as in Table V and VI below.

TABLE V Summary of the computed calcite solubility  $\rho_{\text{CaCO}_3}$  and  $\ln(K_{\text{sp}})$  for calcite in water.

p	T	$\rho_{\text{CaCO}_3}/\text{M}$	$\ln(K_{\text{sp}})$
0.1	298	$8.97 \times 10^{-5}$	-18.64(81)
100	298	$4.65 \times 10^{-4}$	-15.35(81)
200	298	$1.35 \times 10^{-3}$	-13.21(81)
0.1	323	$9.60 \times 10^{-5}$	-18.50(74)
0.1	348	$5.17 \times 10^{-5}$	-19.74(69)

TABLE VI Summary of the experimental  $K_{\text{sp}}$  and  $\ln(K_{\text{sp}})$  for calcite in pure water. Pressures are in MPa, temperatures in K and energies in kJ/mol.

p	T	$K_{\text{sp}}$	$\ln(K_{\text{sp}})$
0.1	298	$3.31 \times 10^{-9}$	-19.53
20	298	$5.23 \times 10^{-9}$	-19.07
40	298	$8.08 \times 10^{-9}$	-18.63
60	298	$1.22 \times 10^{-8}$	-18.22
80	298	$1.83 \times 10^{-8}$	-17.82
100	298	$2.68 \times 10^{-8}$	-17.43
0.1	324	$2.12 \times 10^{-9}$	-19.97
0.1	346	$1.19 \times 10^{-9}$	-20.55
0.1	363	$7.62 \times 10^{-10}$	-20.99

In Table V, for the ambient conditions, experiments<sup>106-107</sup> yield  $K_{\text{sp}} = 10^{-8.48} = 3.31 \times 10^{-9}$ , giving  $[\text{Ca}^{2+}(\text{aq})] = [\text{CO}_3^{2-}(\text{aq})] = \sqrt{K_{\text{sp}}} = \rho_{\text{CaCO}_3} = \sqrt{3.31 \times 10^{-9}} = 5.75 \times 10^{-5}$  M. At the same time, our simulations give  $[\text{Ca}^{2+}(\text{aq})] = [\text{CO}_3^{2-}(\text{aq})] = \rho_{\text{CaCO}_3} = 8.97(3.62) \times 10^{-5}$  M, which is around 1.6 times larger than the experimental  $\rho_{\text{CaCO}_3} = \sqrt{K_{\text{sp}}}$ . From  $\rho_{\text{CaCO}_3} = 8.97 \times 10^{-5}$  M, we calculated the solubility product  $K_{\text{sp}} = \rho_{\text{CaCO}_3}^2 = 8.05 \times 10^{-9}$ .



In addition, we estimate the dissolution free energy of calcite in water  $\Delta G_{\text{dissolution}} = -RT \ln(K_{\text{sp}}) = -46.41 (2.00) \text{ kJ/mol}$  comparing with the experimental  $-48.38 \text{ kJ/mol}$  (calculated from  $K_{\text{sp}} = 3.31 \times 10^{-9}$ ). Fig. 6 below is a plot of the computed and experimental  $\ln(K_{\text{sp}})$  with increasing temperature. Again, our simulations reproduce the experimental negative dependence of solubility product upon increasing temperature<sup>106</sup>.

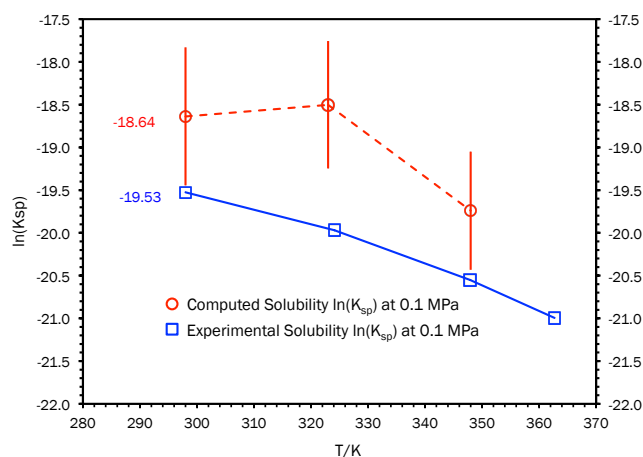


FIG. 6. Computed and experimental<sup>106-107</sup>  $\ln(K_{\text{sp}})$  with increasing temperature at 0.1 MPa. Experimental error bars are of similar magnitude to the data marker sizes.

Next, we turn to the computation of the variation of calcite solubility with pressure. The results are plotted in Fig. 7. In Fig. 7, our simulations reproduce the experimental observation that the calcite solubility product increases with pressure<sup>108-109</sup>.

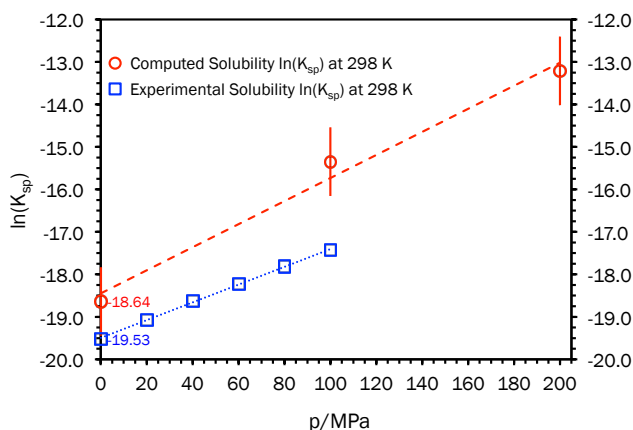


FIG 7 Computed  $\ln(K_{sp})$  with increasing pressure in comparison with the experimental<sup>108-109</sup>  $\ln(K_{sp})$ .

As with our earlier work<sup>29</sup>, it is crucial to scrutinize our results for self-consistency. To test the internal consistency, we check  $(A_{CaCO_3}/N_{CaCO_3})/T$  versus  $1/T$  to verify if we satisfy the thermodynamic relationship  $\partial[(A_{solid}/N_{solid})/T]/[\partial(1/T)] = U_{solid}/N_{solid}$  where  $U_{solid}/N_{solid}$  is the internal energy. The slope of the least-squares fit give  $-2884.5(8)$  kJ/mol, and our reference crystal (viz. used as the initial structure for the extended Einstein crystal) at  $T = 298$  K and  $p = 0.1$  MPa has a static internal energy of  $-2880.3$  kJ/mol per ion pair. Similarly, by checking  $G_{CaCO_3}/N_{CaCO_3}$  versus  $p$ , we estimate the slope  $\partial(G_{CaCO_3}/N_{CaCO_3})/\partial p = v_{CaCO_3} = 37.14(59)$  cm<sup>3</sup>/mol, and our reference crystal has a fixed molar volume of  $v_{CaCO_3} = 36.50$  cm<sup>3</sup>/mol. From the slope of Fig. 7, we can estimate of the difference between the partial molar volume of calcite in the solution and its molar volume in the solid<sup>108-109</sup>,

$$\Delta v = -RT \left( \partial \ln K_{sp} / \partial p \right)_T \quad (10)$$

We obtain  $\Delta v = -67.28(14.15)$  cm<sup>3</sup>, and the experiment line yields  $\Delta v = -51.8$  cm<sup>3</sup>/mol.

We also known that,

$$\Delta v = v_{excess}^{solute} - v_{solid} \quad (11)$$

From previous analysis, we already know that our simulations give  $v_{excess}^{Ca^{2+}} = -17.69(5.66)$  cm<sup>3</sup>/mol,  $v_{excess}^{CO_3^{2-}} = -12.45(5.66)$  cm<sup>3</sup>/mol and  $v_{CaCO_3} = 37.14$  cm<sup>3</sup>/mol, so here we calculate

$$\Delta v = v_{excess}^{Ca^{2+}} + v_{excess}^{CO_3^{2-}} - v_{CaCO_3} = -67.28 \text{ cm}^3/\text{mol} - \text{matching the } \Delta v \text{ estimated just above.}$$

We note that the solubility that we have calculated is strictly speaking the solubility product  $K_{sp}$ , which can be viewed as the equilibrium constant for the calcite dissociation reaction into  $Ca^{2+}$  (aq) and  $CO_3^{2-}$  (aq). As  $K_{sp}$  is determined by the Gibbs free energy change

of the reaction, it should only be pressure/temperature-dependent in the context. Of course,  $\text{CO}_3^{2-}$  is in equilibrium with  $\text{HCO}_3^-$  and  $\text{CO}_2$  (aq), and this affects the actual solubilities of calcite and aragonite – a lower pH increases the measured  $\text{Ca}^{2+}$  solubility from its value as obtained from  $K_{\text{sp}}$  because  $\text{CO}_3^{2-}$  can combine with  $\text{H}^+$  to form  $\text{HCO}_3^-$ . Nevertheless, for example, whilst a decrease of 0.3 in the pH yields a 60% increase in the total concentration of dissolved  $\text{Ca}^{2+}$  from calcite<sup>106</sup>, the value for  $K_{\text{sp}}$  is the same to within experimental error. Additional calculations would be needed to compute the total  $\text{Ca}^{2+}$  concentration in solution.

### **E. Other Carbonate Crystals: Solubility Predictions and Relative Stability**

In this short section, we will briefly discuss the aqueous solubility of aragonite under ambient conditions. To this end we make use of the same single-ion excess solvation free energies  $\Delta\mu_{\text{excess}}^{\text{Ca}^{2+}}$  and  $\Delta\mu_{\text{excess}}^{\text{CO}_3^{2-}}$  as before, but we further calculate the absolute solid free energy of aragonite. The aqueous solubility limit for aragonite is thus estimated to be  $1.13 (4.56 \times 10^{-5}) \times 10^{-4}$  M, compared with the slightly lower experimental value  $\rho_{\text{aragonite}} = 6.79 \times 10^{-5}$  M (calculated from the experimental  $K_{\text{sp}}^{\text{aragonite}} = 10^{-8.336}$ <sup>106</sup>). From the last section we have  $\rho_{\text{calcite}} = 8.97 \times 10^{-5}$  M with the experimental  $\rho_{\text{calcite}} = 5.75 \times 10^{-5}$  M. Although numerically the aqueous solubilities of calcite and aragonite are both slightly overestimated, our results reproduce the marginally higher aqueous solubility of aragonite than calcite under ambient conditions. In fact, our results show that calcite is 1.09(8) kJ/mol more stable than aragonite, very close to the experimental free energy difference of 0.84 kJ/mol<sup>54, 110</sup>. In summary, this calculation reveals another merit of our free-energy route to solubility (e. g. over the direct coexistence course), we only have to compute the aqueous properties once; moreover we gain extra knowledge of the relative thermodynamic stability of the polymorphs under investigation.

## F. Caffeine: Solubility Prediction

We have already analyzed a few examples of organic and inorganic molecules, whose numerical solubilities agree fairly well with experiments. However, the largest organic molecules considered thus far, naphthalene and phenanthrene, are both rather apolar. In order to get a deeper glimpse into the generic nature of the method, we now consider a more complex “polar” organic molecule, caffeine. Below, we present solubility calculations of a GAFF-based<sup>60-61</sup>  $\beta$ -caffeine (the most stable ambient polymorph<sup>111-112</sup>) at  $p = 0.1$  MPa and  $T = 298$  K in SPC/E water<sup>62</sup>.

First, we computed the excess solvation free energy of caffeine. We have adopted an approach similar to the one we used for ionic substance: 1) in the first stage, we calculated the nonbonding van der Waals contribution via the cavity method; 2) next, we estimated the electrostatic charging free energy via a Kirkwood TI procedure by charging all the caffeine atoms by a linear charge ratio  $\alpha$ ; the total excess solvation free energy is hence the sum. In such manner, we have obtained  $\Delta\mu_{\text{excess}} = -61.8(8)$  kJ/mol (see the supplementary material), comparing with the numerical reference of  $-64.7(1)$  kJ/mol<sup>61</sup> and the experimental reference of  $-53.7(6)$  kJ/mol<sup>113-114</sup>. We also notice that computed solvation free energies of caffeine reported in the literature can vary by as much as 42 kJ/mol within themselves, and may differ by up to 30 kJ/mol from experimental data. The computed values depend strongly on the atomic partial-charge distribution, the solvent model and the computational methods used<sup>114</sup>.

To compute the absolute solid free energies, again we used the final configuration from a  $5n NpT$  simulation as our reference structure. Our reference structure has a density of  $1.468$  gcm<sup>-3</sup>, which is close to the experimental density<sup>115</sup> of  $1.448$  gcm<sup>-3</sup>; our average simulated lattice parameters  $a$ ,  $b$ ,  $c$  and  $\beta$  deviate from the experimental values by 0.5-2.5 %. We carried out Einstein crystal method calculations using five different “Einstein arrangements”, and obtained  $A_{\text{caffeine}}/N_{\text{caffeine}}$  ranging from  $-100.6(1)$  kJ/mol to  $-104.1(1)$  kJ/mol (see the

supplementary material). The largest variation in  $A_{\text{caffeine}}/N_{\text{caffeine}}$  is no more than 3.6 kJ/mol, which is merely a factor of 4 in solubility, but it is apparently larger than the typical statistical error of 0.1 kJ/mol. In our earlier work<sup>29</sup>, we discuss that for naphthalene, the largest discrepancy is 0.8 kJ/mol. We have hypothesized that the problem could be related to using flexible molecular models in the Einstein crystal method scheme, where harmonic bonds are simultaneously used to constrain atomic positions/orientations. Here the discrepancies in free energies resulting from utilizing different Einstein arrangement choices for seem to be more pronounced for larger molecules with a higher complexity (viz. with polarity and heteroatoms). It should be possible to remedy this problem by including the intramolecular parts in the Einstein crystal Hamiltonian, so as to compute explicitly the intramolecular contributions to the “Einstein free energy”, or by carrying out the simulations for a rigid caffeine model. Within the current, LAMMPS-based approach, both approaches would not be simple – although with a custom-made code they should be attainable. The problem we highlight here is thus an illustration of the limitation of using standard open-source software for the calculation of the solubility of complex molecules.

Experimentally, caffeine is reported to have an aqueous solubility limit of 0.108 M<sup>116</sup>. The results that we have obtained are only useful to give an estimate of the coexistence concentration of “caffeine monomers”, in equilibrium with dimers, trimers etc. We note that the assumption of infinite dilution is most likely no longer valid at the saturation concentration. To compare with experiments, we need to know the experimental fraction of caffeine monomers present at the solubility limit. For this purpose, we used the caffeine solubility information in D<sub>2</sub>O from FT-IR and ultraviolet studies<sup>117</sup>. With the excess solvation free energy of caffeine monomer  $-61.8(8)$  kJ/mol and the average solid free energy  $A_{\text{caffeine}}/N_{\text{caffeine}} = -103.0(1)$  kJ/mol, the solubility limit of caffeine monomers (in equilibrium with other polymers) is estimated to be  $\rho_{\text{caffeine-monomer}} = 1.0(4) \times 10^{-4}$  M. In order to have a

rough idea of how this number might compare with experiments, we fictitiously take  $\rho_{\text{caffeine-total}} = 0.100 \text{ M}$  as the solubility limit<sup>117</sup>, and based upon the reported weight fraction of monomers in D<sub>2</sub>O, we estimate  $\rho_{\text{caffeine-monomer}}^{\text{exp}}$  to be around 0.01 M – around 100 times more soluble than our numerical prediction which is equivalent to a free energy difference of 11 kJ/mol. As we can see, the discrepancy is larger than that can be caused by the variation of  $A_{\text{caffeine}}/N_{\text{caffeine}}$  incurred by molecular flexibility. In conclusion, the solubility prediction for caffeine is poor, and at this stage, it is not easy to pinpoint a single reason behind this inaccuracy, as there are likely to be causes from both the solution and solid parts, e.g. force-fields parameterizations, molecular flexibility and electrostatic accuracy. The other primary factor is how different polymorphs might influence each other in the aspects of their individual dissolution limits chemical-potential-wise, which requires a more thorough future investigation.

#### IV. CONCLUSION

We have applied our generic solubility calculation approach to a range of rather different molecular crystals including phenanthrene, calcite and aragonite and caffeine. Most predictions are in fair agreement with experiments. Moreover, our approach allows us to quantitatively compare the relative stability of the polymorphs of the same compound.

Nevertheless, our methodology shows several “external” or “internal” limitations.

One “external” obstacle is the scarcity of good force-fields to describe both the crystalline and aqueous (solution) compounds, in particular for complex molecules. Clearly, the problem associated with force field accuracy is not limited to our work but is generic to all crystal simulations: for example, direct (solid-liquid) coexistence simulations are similarly susceptible to the quality of force-fields, as well as other thermodynamic approaches. In order to compute solubilities, the force-fields must be good enough to ensure that the target crystal

structure is mechanically stable. To test whether it is also thermodynamically more stable than other possible polymorphs, we would have to compute the free energy of these other (mechanically stable) structures; with the current approach, the computation procedure is straightforward). Popular general force-fields, such as GAFF<sup>60-61</sup> and OPLS-AA<sup>52</sup>, are mostly targeted at organic liquid simulations, and their overall performances in solubility predictions are difficult to predict. There is a similar circumstance for inorganic materials. When computing inorganic crystals free energies, we found that generally there was a lack of published accurate force-fields, especially for those with multivalent ions. Due to the simple crystalline structures and monovalent nature, it is somewhat easier to design high-quality force-fields for materials like NaCl and KF<sup>8-21</sup>, and these calculations are typically in good agreement with experiments. In summary, to compute the solubility limit of experimentally observed crystal structures, our models should at the very least be able to reproduce the mechanical stability of that structure. In general, this is a less stringent requirement than requiring that the experimental crystal structure is also the model structure that is thermodynamically most stable. Reproducing the thermodynamic stability of the experimental crystal structure is more demanding: it is most likely for force-fields that have been designed to reproduce properties of both the liquid and the crystal phase. In addition, free energy calculations for all candidate crystal phases would be needed to settle this issue. However, in this work we only focus on the prediction of the solubility of a known crystal phase, hence we need not solve the full “polymorph problem”. In conclusion, overall accurate free energy (and hence solubility) estimates of complex molecular crystals are still relatively rare – despite the fact that numerical techniques to compute crystal free energies have been around since the 1960s.

For inorganic crystals and their aqueous ions, experimental single-ion solvation free energies are mostly large and negative. What adds to the challenge is that sometimes there is

even a lack of consensus for certain ions – for example, even for the most studied NaCl, Noyes<sup>118</sup> gives  $-412$  kJ/mol for  $\text{Na}^+$  and  $-318$  kJ/mol for  $\text{Cl}^-$ ; Tissandier<sup>119</sup> gives  $-424$  kcal/mol for  $\text{Na}^+$  and  $-304$  kcal/mol for  $\text{Cl}^-$ ; Marcus<sup>95</sup> gives  $-365$  kJ/mol for  $\text{Na}^+$  and  $-340$  kJ/mol for  $\text{Cl}^-$ . If force-fields are parameterized to fit experimental data, the discrepancy between the various experimental data is a problem that cannot simply be ignored. In some cases, even when force-fields are specifically parameterized, the “usability” is not guaranteed. For instance, the MSXX force field<sup>109</sup> and its subsequent modifications<sup>120-122</sup> have been parameterized to describe barite-barite and barite-water interactions. However, to our best knowledge, overall they do not give satisfying single-ion free energies to achieve accurate solubility predictions, despite their significance in studying barite-water interfaces.

An “internal” limitation is that the current extended Einstein-crystal method ignores molecular flexibility. The problem caused by this approximation shows up in the example of caffeine, while it is less evident for naphthalene and phenanthrene. We suggest that, for future work on large organic/polar molecules one will either need to use rigid models, or explicitly consider the intramolecular contributions in the extended Einstein crystal approach, or using different strategies for aligning flexible molecules in an Einstein crystal. Within LAMMPS the latter two options might not be easy to implement. For polar organic or inorganic compounds with (relatively simple) molecular structures, the present scheme should be adequate for achieving good accuracy.

In summary, free-energy-based solubility calculations, based on well-known numerical techniques and theoretical approaches, have proffered the prospect of achieving universal solubility predictions; however one crucial reason why they have not become widely used as standard procedures as yet is that their compound-specific machinery is cumbersome to program, and often take longer than the actual solubility calculations. Based on precedent work in the field, we have addressed this problem by presenting a “black-box” approach,



with an integrated cycle of extended Einstein crystal steps where free energy changes are explicitly calculated for each intermediate crystalline stage<sup>29</sup>. Our tailor-made protocols make use of commercially available software where simple setups such as initial configurations and adequate force-fields are sufficient to start calculations. In the present work, we advance to explore the methodology by applying it to a wider range of prototypical complex materials, including phenanthrene, calcite and aragonite over many conditions. Of course, extra care may be required to check for molecular flexibility (when it is indispensable), unexpected flipping of equivalent atoms (occasionally in non-rotator crystalline phases at weak Einstein crystal springs when energy barriers are low), or high solubility limits (the excess chemical potential would correspond to a certain concentration rather than the infinite dilution); and as discussed, much accuracy is dependent upon the availability of good-quality force-fields – these are the challenges of the problem. However, the positive message that follows from the present paper and its predecessors is that despite these difficulties, developments of computational efficiency and high-quality force-fields has made it more and more conceivable towards a unified methodology of universal solubility predictions.

## **SUPPLEMENTARY MATERIAL**

The supplementary material contains the LAMMPS protocol to calculate the aqueous solubility of calcite at  $T = 298$  K and  $p = 0.1$  MPa.

## **ACKNOWLEDGMENTS**

The authors would like to acknowledge the funding and technical support from BP through the BP International Centre for Advanced Materials (BP-ICAM) which made this research possible.

## Appendix A

### A. Theoretical Background

We start from the standard thermodynamic expressions<sup>46</sup>,  $G$ ,  $\mu$ ,  $A$ ,  $V$  are the Gibbs free energy, chemical potential, Helmholtz free energy and total volume. At constant pressure  $p$  and temperature  $T$ , the equilibrium between the crystal phase of the solute and its solution requires:

$$\mu_{\text{solution}}^{\text{solute}}(p, T) = \mu_{\text{solid}}^{\text{solute}}(p, T) \quad (12)$$

We compute  $\mu_{\text{solution}}^{\text{solute}}(p, T)$  and  $\mu_{\text{solid}}^{\text{solute}}(p, T)$ .

The chemical potential of the solute in the solution is

$$\begin{aligned} \mu_{\text{solution}} = & k_B T \ln(\rho_{\text{solute}}^0 \Lambda_{\text{solute}}^3 q_{\text{solute}}^{-1}) + k_B T \ln(\rho_{\text{solute}} / \rho_{\text{solute}}^0) \\ & - k_B T \ln \left\langle \exp \left[ -\beta U_{\text{solute-solution}}^{\rho_{\text{solute}}}(\mathbf{R}_{\text{solute}}) \right] \right\rangle_0 \end{aligned} \quad (13)$$

In Eq. (13), the only part that requires computation is the last term on the right, i.e. the excess chemical potential  $\Delta\mu_{\text{excess}}$ .  $\rho_{\text{solute}} = N_{\text{solute}}/V$  denotes the number density of the solute, with the common ideal gas reference state  $\rho_{\text{solute}}^0 = 1$  (i.e. 1 molecule in  $1 \text{ \AA}^3 = 0.001 \text{ nm}^3$ ), where  $N_{\text{solute}}$  is the number of solute molecules.  $\Lambda_{\text{solute}}$  is the solute thermal de Broglie wavelength and  $q_{\text{solute}}$  is its intramolecular (rotational, vibrational and electronic) partition functions;  $\mathbf{R}_{\text{solute}}$  is the fixed position of the extra single solute and  $U_{\text{solute-solution}}(\mathbf{R}_{\text{solute}})$  is thus the interaction energy of the extra solute with the rest of the solution. As explained in the reference<sup>29</sup>, the constant term  $\ln(\rho_{\text{solute}}^0 \Lambda_{\text{solute}}^3 q_{\text{solute}}^{-1})$  can be ignored because it is the same in both coexisting phases and further information might require quantum knowledge.

The chemical potential of the solute in the solid is,

$$\mu_{\text{solid}}^{\text{solute}} = G_{\text{solid}} / N_{\text{solid}} = (A_{\text{solid}} + pV_{\text{solid}}) / N_{\text{solid}} \quad (14)$$

where  $N_{\text{solid}}$  is the number of solute molecules in the crystal.

Eq. (12) thus becomes,

$$k_B T \ln(\rho_{\text{solute}}) - k_B T \ln \left\langle \exp \left[ -\beta U_{\text{solute-solution}}(\mathbf{R}_{\text{solute}}) \right] \right\rangle_0 = (A_{\text{solid}} + pV_{\text{solid}}) / N_{\text{solid}} \quad (15)$$

In the case of a sparingly soluble solute, we may set the second term on the left to the solvation free energy of a single solute molecule in the solution (assuming infinite dilution), with  $\Delta\mu_{\text{excess}} = -k_B T \ln \left\langle \exp \left[ -\beta U_{\text{solute-solution}}(\mathbf{R}_{\text{solute}}) \right] \right\rangle_0$  accordingly, after which we can solve  $\rho_{\text{solute}}$  analytically. In the case of higher solubilities, the solvation free energy becomes a function of composition, but otherwise the approach remains the same. In the current scheme, we neglect intramolecular contributions to the solubility, and Eq. (15) does not take into the intramolecular energies of the solute in its solution/crystal phase into account, as justified in our previous work<sup>29</sup>.

## B. Extended Einstein crystal method

We compute absolute crystal Helmholtz free energy  $A_{\text{solid}}$  using a straightforward extension of the Einstein crystal method. The method is based upon the original Einstein crystal method<sup>34, 39-40</sup> and its recent adaptations<sup>41-45</sup>. The absolute Helmholtz free energy of a molecular crystal is computed by TI from the fully interaction crystal to an atomic Einstein crystal for which the free energy can be calculated analytically. In the extended Einstein crystal method, we transform the molecular crystal into an Einstein crystal of independent molecules that are tethered by harmonic springs to initial reference locations. In our work, the method has been adapted to fit to MD simulations in LAMMPS<sup>30</sup> without extra codes. We have described the method for the case of naphthalene in reference 29. Fig. 8 below shows a

pictorial representation of the extended Einstein crystal method for calcite. In summary, we have,

$$A_{\text{solid}} = A_0 + \Delta A_0 + \Delta A_1 + \Delta A_2 + \Delta A_{\text{symmetry}} \quad (16)$$

$\Delta A_{\text{symmetry}}$  is the explicit consideration of the molecular point group because a molecule in a non-rotator phase in MD simulations cannot occupy all the symmetrically equivalent orientations due to kinetic reasons, in contrast to the same solute in the solution. As in reference 29, here we have also included the relevant finite-size corrections<sup>47-48</sup>.

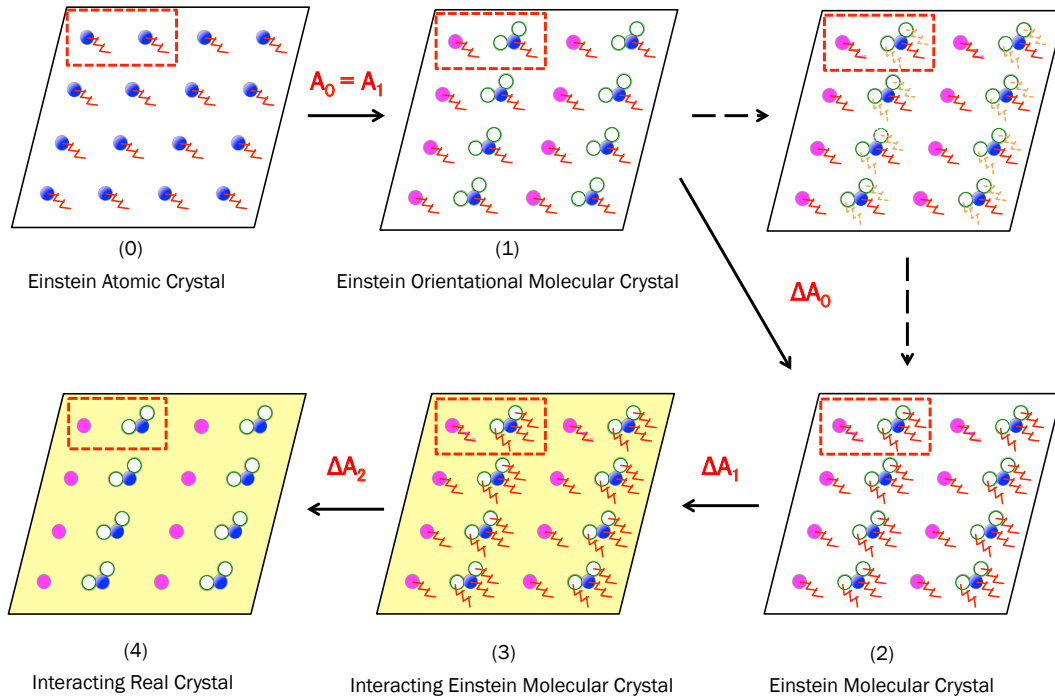


Figure 8 Extended Einstein crystal method for calcite, a molecular crystal comprised of ionic pairs. As in our previous work, (0) is the Einstein atomic crystal for which we know the analytic Helmholtz free energy  $A_0$  and (4) is the real molecular crystal whose  $A_{\text{solid}}$  we wish to compute. Steps (0)-(4) include the various thermodynamically reversible paths and their accompanying free energy changes  $\Delta A$ . The solid pink circles represent  $\text{Ca}^{2+}$  cations, and the solid blue circles and empty green circles together represent the molecular  $\text{CO}_3^{2-}$  anions; the dashed red rectangle labels an ionic pair as a “calcite molecule”. In the extended Einstein crystal scheme, we have introduced the idea of “Central Atom” and “Orientational Atoms”, controlling the location and orientation of a molecule respectively. In calcite, for  $\text{CO}_3^{2-}$  only, the solid blue carbon atoms are the central carbon atoms, and the empty green oxygen atoms are the orientational atoms (only two instead of three oxygen atoms are drawn to avoid crowding, and only two O atoms are required for fixing orientation); for the pink solid  $\text{Ca}^{2+}$ , there is no need for considering orientation. From (0)-(1), (1) is the Einstein orientational atomic crystal, where each molecule is freely rotatable, Here we have  $A_0 = A_1$  as long as the solute is also considered to be freely rotatable in the solution phase. Throughout (1)-(3), all the “Central Atoms” are attached to their initial positions with the strongest Einstein springs  $K_{\text{max}}$  represented by the red solid

springs; during (1)-(2), the “Orientational Atoms” are attached with varying  $K$  from  $K_{\min} = 0$  to  $K_{\max}$ , represented by the orange dashed springs – this is for calculating the orientational contribution  $\Delta A_0$ . During (2)-(3), both the “Central Atoms” and “Orientational Atoms” are attached with the strongest Einstein springs  $K_{\max}$ , which are both represented by the red solid springs, in order to sample the intermolecular energy contribution  $\Delta A_1$ . Finally, during (3)-(4), all the Einstein springs are reduced from  $K_{\max}$  to  $K_{\min} = 0$  to retrieve the real crystal, with the yellow background in (3)-(4) indicating that the intermolecular interaction is active in (3)-(4) but absent in (0)-(2). Such a central-orientational spring arrangement is termed the “Einstein Arrangement” or “EA”<sup>29</sup>.

### C. Cavity method

We use the cavity method to compute the excess solvation free energy  $\Delta\mu_{\text{excess}}$  of general solutes in isothermal-isobaric  $NpT$  ensembles. As a typical mean in computing solvation free energies, directly growing a solute in solution can suffer from numerical inaccuracies, mainly due to the well-known, “end-point” singularity associated with LJ solute-solvent interactions<sup>49-50</sup>. The cavity method eliminates the “end-point” singularity by creating a repulsive isotropic cavity in the solution, followed by the insertion of the solute molecule, after which a cavity-free solution is retrieved,

$$\Delta\mu_{\text{excess}} = \Delta G_{\text{grow}} + \Delta G_{\text{insert}} + \Delta G_{\text{shrink}} \quad (17)$$

$\Delta G_{\text{grow}}$ ,  $\Delta G_{\text{insert}}$  and  $\Delta G_{\text{shrink}}$  are the Gibbs free energy changes of creating the cavity, inserting/growing the solute and shrinking the (same) cavity.

There is no fixed form for the functional form for the cavity as long as it does not suffer from the same divergence problem as the LJ potential. For example, Postma *et al.*<sup>51</sup> have studied the thermodynamics of cavity in the form of  $U_{\text{cavity}}(\lambda) = \lambda(B/r)^{12}$ , where  $\lambda$  is a linear scaling parameter, but we cannot use this potential form because it retains the end-point singularity property. As in our earlier work<sup>29</sup>, we consistently choose  $U_{\text{cavity}}(\lambda) = \sum_i^{N_{\text{solvent}}} A \exp(-r_{\text{solute-}i}/B + \lambda)$ , where  $A$  and  $B$  are pre-determined constants (in our previous work, we used  $A = 1255$  kJ/mol and  $B = 0.1$  nm for the typical case of a single solute molecule in  $N_{\text{water}} = 864$  water molecules; we used the same  $A$  and  $B$  in the present

work),  $r_{\text{solute-}i}$  is the distance between (one particular atom of the) solute molecule and (one particular atom of the) solvent molecule, and  $\lambda$  is our variable to adjust the cavity size.

If we use more than one cavity size (i.e. varying  $\lambda$ ) to compute  $\Delta\mu_{\text{excess}}$ , we should have,

$$\Delta\mu_{\text{excess}}(\lambda) = \Delta G_{\text{grow}}(\lambda) + \Delta G_{\text{insert}}(\lambda) + \Delta G_{\text{shrink}}(\lambda) = \Delta\mu_{\text{excess}}(\lambda') \quad (18)$$

where  $\lambda \neq \lambda'$ . Since we use TI/FEP,  $\Delta G_{\text{grow}}(\lambda')$  and  $\Delta G_{\text{grow}}(\lambda)$  can be built from  $\Delta G_{\text{grow}}(\lambda)$  and  $\Delta G_{\text{grow}}(\lambda)$  for  $\lambda' > \lambda$  to save computational time and expenses.

The cavity method is discussed in detail in reference 29.

## REFERENCES

<sup>1</sup>G. L. Amidon, H. Lennernäs, V. P. Shah, and J. R. Crison, Pharm. Res. An Off. J. Am. Assoc. Pharm. Sci. **12**, 413 (1995).

*A Theoretical Basis for a Biopharmaceutic Drug Classification: The Correlation of in Vitro Drug Product Dissolution and in Vivo Bioavailability*

<sup>2</sup>W. L. Jorgensen and E. M. Duffy, Adv Drug Deliv Rev. **54**, 355-66 (2002).

*Prediction of Drug Solubility from Structure*

<sup>3</sup>W. L. Jorgensen and E. M. Duffy, Bioorg. Med. Chem. Lett. **10** (2000).

*Prediction of Drug Solubility from Monte Carlo Simulations*

<sup>4</sup>S. J. Park and G. A. Mansoori, Int. J. Energy Sources **10**, 109 (1988).

*Aggregation and Deposition of Heavy Organics in Petroleum Crudes*

<sup>5</sup>K. Raju and G. Atkinson, J. Chem. Eng. Data **33**, 490 (1988).

*Thermodynamics of "Scale" Mineral Solubilities. 1. BaSO<sub>4</sub>(s) in H<sub>2</sub>O and Aqueous NaCl*

<sup>6</sup>K. U. G. Raju and G. Atkinson, J. Chem. Eng. Data **35**, 361 (1990).

*Thermodynamics of "Scale" Mineral Solubilities. 3. Calcium Sulfate in Aqueous NaCl*

<sup>7</sup>T. Chen, A. Neville and M. Yuan, J. Petrol. Sci. Eng. **46**, 185-194 (2005).

*Calcium Carbonate Scale Formation—Assessing the Initial Stages of Precipitation and Deposition*

<sup>8</sup>J. R. Espinosa, J. M. Young, H. Jiang, D. Gupta, C. Vega, E. Sanz, P. G. Debenedetti and A. Z. Panagiotopoulos, *J. Chem. Phys.* **145**, 154111 (2016).

*On the Calculation of Solubilities via Direct Coexistence Simulations: Investigation of NaCl Aqueous Solutions and Lennard-Jones Binary Mixtures.*

<sup>9</sup>J. Kolafa, *J. Chem. Phys.* **145**, 204509 (2016).

*Solubility of NaCl in Water and Its Melting Point by Molecular Dynamics in the Slab Geometry and a New BK3-compatible Force Field.*

<sup>10</sup>M. Lísal, W. R. Smith and J. Kolafa, *J. Phys. Chem. B* **109**, 12956 (2005).

*Molecular Simulations of Aqueous Electrolyte Solubility: 1. The Expanded-ensemble Osmotic Molecular Dynamics Method for the Solution Phase.*

<sup>11</sup>F. Moučka, M. Lísal and W. R. Smith, *J. Phys. Chem. B* **116**, 5468 (2012).

*Molecular Simulation of Aqueous Electrolyte Solubility. 3. Alkali-halide Salts and Their Mixtures in Water and in Hydrochloric Acid.*

<sup>12</sup>F. Moučka, I. Nezbeda and W. R. Smith, *J. Chem. Theory Comput.* **11**, 1756(2015).

*Chemical Potentials, Activity Coefficients, and Solubility in Aqueous NaCl Solutions: Prediction by Polarizable Force Fields.*

<sup>13</sup>I. Nezbeda, F. Moučka and W. R. Smith, *Molec. Phys.* **114**, 1665(2016).

*Recent Progress in Molecular Simulation of Aqueous Electrolytes: Force Fields, Chemical Potentials and Solubility.*

<sup>14</sup>M. Ferrario, G. Ciccotti, E. Spohr, T. Cartailier, and P. Turq, *J. Chem. Phys.* **117**, 4947 (2002).

*Solubility of KF in water by Molecular Dynamics Using the Kirkwood Integration Method*

<sup>15</sup>H. M. Manzanilla-Granados, H. Saint-Martín, R. Fuentes-Azcatl, and J. Alejandre, J. Phys. Chem. B **119**, 8389 (2015).

*Direct Coexistence Methods to Determine the Solubility of Salts in Water from Numerical Simulations. Test Case NaCl*

<sup>16</sup>J. L. Aragoñes, E. Sanz, and C. Vega, J. Chem. Phys. **136**, 244508 (2012).

*Solubility of NaCl in Water by Molecular Simulation Revisited*

<sup>17</sup>A. L. Benavides, J. L. Aragoñes and C. Vega, J. Chem. Phys. **144**, 124504 (2016)

*Consensus on the Solubility of NaCl in Water from Computer Simulations Using the Chemical Potential Route*

<sup>18</sup>A. S. Paluch, S. Jayaraman, J.K. Shah, and E.J. Maginn, J. Chem. Phys. **133**, 124504 (2010).

*A Method for Computing the Solubility Limit of Solids: Application to Sodium Chloride in Water and Alcohols*

<sup>19</sup>Z. Mester and A. Z. Panagiotopoulos. J. Chem. Phys., **142**, 044507 (2015).

*Mean Ionic Activity Coefficients in Aqueous NaCl Solutions from Molecular Dynamics Simulations.*

<sup>20</sup>Z. Mester and A. Z. Panagiotopoulos, J. Chem. Phys., **143**, 044505 (2015)

*Temperature-Dependent Solubilities and Mean Ionic Activity Coefficients of Alkali Halides in Water from Molecular Dynamics Simulations.*

<sup>21</sup>F. Moucka, M. Lisal, J. Skvor, J. Jirsak, I. Nezbeda, and W. R. Smith, J. Phys. Chem. B **115**, 7849 (2011).

*Molecular Simulation of Aqueous Electrolyte Solubility. 2. Osmotic Ensemble Monte Carlo Methodology for Free Energy and Solubility Calculations and Application to NaCl*

<sup>22</sup>M.J. Schnieders, J. Baltrusaitis, Y. Shi, G. Chattree, L. Zheng, W. Yang, and P. Ren, J. Chem. Theory Comput. **8**, 1721 (2012).



*The Structure, Thermodynamics, and Solubility of Organic Crystals from Simulation with a Polarizable Force Field*

<sup>23</sup>D. S. Palmer, J.L. McDonagh, J.B.O. Mitchell, T. Van Mourik, and M. V. Fedorov, J. Chem. Theory Comput. **8**, 3322 (2012).

*First-Principles Calculation of the Intrinsic Aqueous Solubility of Crystalline Druglike Molecules*

<sup>24</sup>J. C. Dearden, Expert Opin. Drug. Discov. **1**, 31 (2006).

*In Silico Prediction of Aqueous Solubility*

<sup>25</sup>A. S. Paluch and E. J. Maginn, AIChE J. **59**, 2647(2013).

*Predicting the Solubility of Solid Phenanthrene: A Combined Molecular Simulation and Group Contribution Approach*

<sup>26</sup>Ind. Eng. Chem. Res. **44**, 4075 (2005).

*Revision of MOSCED Parameters and Extension to Solid Solubility Calculations*

<sup>27</sup>R. T. Ley, G. B. Fuerst, B. N. Redeker, and A. S. Paluch, Ind. Eng. Chem. Res. **55**, 5415 (2016).

*Developing a Predictive Form of MOSCED for Nonelectrolyte Solids Using Molecular Simulation: Application to Acetanilide, Acetaminophen, and Phenacetin*

<sup>28</sup>S. Cabani, G. Mollica, and V. Lepori, J. Sol. Chem. **10**, 563 (1981).

*Group Contributions to the Thermodynamic Properties of Non-Ionic Organic Solutes in Dilute Aqueous Solution*

<sup>29</sup>L. Li, T. Totton and D. Frenkel, J. Chem. Phys. **146**, 214110 (2017).

*Computational Methodology for Solubility Prediction: Application to Simple Aromatic Crystals*

<sup>30</sup>S. Plimpton, J Comp Phys. **117**, 1 (1995).

*Fast Parallel Algorithms for Short-Range Molecular Dynamics*

<sup>31</sup>R. W. Zwanzig, J. Chem. Phys. **22**, 1420 (1954).

*High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases*

<sup>32</sup>R. W. Zwanzig, J. Chem. Phys. **23**, 1915 (1955).

*High-Temperature Equation of State by a Perturbation Method. II. Polar Gases*

<sup>33</sup>D. A. Kofke and P. T. Cummings, Mol. Phys. **92**, 973 (1997).

*Quantitative Comparison and Optimization of Methods for Evaluating the Chemical Potential by Molecular Simulation*

<sup>34</sup>D. Frenkel and B. Smith, *Understanding Molecular Simulations*, 2nd ed. (Academic, London, 2002).

<sup>35</sup>P. Kollman, Chem. Rev. **93**, 2395 (1993).

*Free Energy Calculations: Applications to Chemical and Biochemical Phenomena*

<sup>36</sup>J. G. Kirkwood, J. Chem. Phys. **3**, 300 (1935).

*Statistical Mechanics of Fluid Mixtures.*

<sup>37</sup>W. L. Jorgensen, J. F. Blake and J. K. Buckner, Chem. Phys. **129**, 193 (1989).

*Free Energy of TIP4P Water and the Free Energies of Hydration of CH<sub>4</sub> and Cl<sup>-</sup> from Statistical Perturbation Theory*

<sup>38</sup>D. Shivakumar, J. Williams, Y. J. Wu, W. Damm, J. Shelley, and W. Sherman, J. Chem. Theory Comput. **6**, 1509 (2010).

*Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field*

<sup>39</sup>D. Frenkel and A. J. Ladd, J. Chem. Phys. **81**, 3188 (1984).

*New Monte Carlo Method to Compute the Free Energy of Arbitrary Solids. Application to the Fcc and Hcp Phases of Hard Spheres*

<sup>40</sup>E. J. Meijer, D. Frenkel, R. A. LeSar and A. J. C. Ladd, J. Chem. Phys. **92**, 7570 (1990)

*Location of Melting Point at 300 K of Nitrogen by Monte Carlo Simulation*

<sup>41</sup>C. Vega, E. Sanz, J. L. F. Abascal, and E. G. Noya, *J. Phys.: Condens. Matter* **20**, 153101 (2008).

*Determination of Phase Diagrams via Computer Simulation: Methodology and Applications to Water, Electrolytes and Proteins*

<sup>42</sup>J. L. Aragones, C. Valeriani, and C. Vega, *J. Chem. Phys.* **137**, 146101 (2012).

*Note: Free Energy Calculations for Atomic Solids through the Einstein Crystal/molecule Methodology Using GROMACS and LAMMPS*

<sup>43</sup>J. L. Aragones, E. G. Noya, C. Valeriani, and C. Vega, *J. Chem. Phys.* **139**, 034104 (2013).

*Free Energy Calculations for Molecular Solids Using GROMACS*

<sup>44</sup>K. Ioannidou, M. Kanduč, L. Li, D. Frenkel, J. Dobnikar and E. D. Gado, *Nat. Commun.* **7**, 12106 (2016).

*The Crucial Effect of Early-stage Gelation on the Mechanical Properties of Cement Hydrates*

<sup>45</sup>M. S. Sellers, M. Lísal, and J. K. Brennan, *Phys. Chem. Chem. Phys.*, **18**, 7841 (2016).

*Free-energy Calculations Using Classical Molecular Simulation: Application to the Determination of the Melting Point and Chemical Potential of a Flexible RDX Model.*

<sup>46</sup>D A McQuarrie, *Statistical Mechanics*, (New York: Harper and Row 1976).

<sup>47</sup>W. G. Hoover, *J. Chem. Phys.* **49**, 1981 (1968).

*Entropy for Small Classical Crystals*

<sup>48</sup>J. Polson, E. Trizac, S. Pronk, and D. Frenkel, *J. Chem. Phys.* **112**, 5339 (2000).

*Finite-size Corrections to the Free Energies of Crystalline Solids*

<sup>49</sup>T. Steinbrecher, D. L. Mobley, and D. A. Case, *J. Chem. Phys.* **127**, 214108 (2007).

*Nonlinear Scaling Schemes for Lennard-Jones Interactions in Free Energy Calculation*

<sup>50</sup>T. Simonson, *Mol. Phys.* **80**, 441 (1993).

*Free Energy of Particle Insertion*

<sup>51</sup>J. P. M. Postma, H. J. C. Berendsen and J. R. Haak, *Faraday Symp. Chem. Soc.* **17**, 55(1982).

*Thermodynamics of Cavity Formation in Water a Molecular Dynamics Study.*

<sup>52</sup>W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, *J. Am. Chem. Soc.* **118**, 11225 (1996).

*Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids*

<sup>53</sup>H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, *In Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht, 1981), p. 331.

<sup>54</sup>P. Raiteri, J. D. Gale, D. Quigley and P. M. Rodger, *J. Phys. Chem. C* **114**, 5997 (2010).  
*Derivation of an Accurate Force-Field for Simulating the Growth of Calcium Carbonate from Aqueous Solution: A New Model for the Calcite-Water Interface*

<sup>55</sup>P. Raiteri, J. D. Gale, *J. Am. Chem. Soc.*, **132**, 17623 (2010).

*Water Is the Key to Nonclassical Nucleation of Amorphous Calcium Carbonate*

<sup>56</sup>P. Raiteri, J. D. Gale, D. Quigley and D. Gebauer, *Nat. Commun.* **2**, 590 (2011).

*Stable Prenucleation Mineral Clusters are Liquid-like Ionic Polymers*

<sup>57</sup>P. Raiteri, R. Demichelis and J. D. Gale, *J. Phys. Chem. C* **119**, 24447 (2015).

*A Thermodynamically Consistent Force Field for Molecular Dynamics Simulations of Alkaline-Earth Carbonates and Their Aqueous Speciation*

<sup>58</sup>Y. Wu, H. L. Tepper and G. A. Voth, *J. Chem. Phys.* **124**, 024503 (2006).

*Flexible Simple Point-charge Water Model with Improved Liquid State Properties.*

<sup>59</sup>Y. Wu, H. Chen, F. Wang, F. Paesani, G. A. Voth, *J. Phys. Chem. B* **112**, 467 (2008).

*An Improved Multistate Empirical Valence Bond Model for Aqueous Proton Solvation and Transport*

<sup>60</sup>J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.* **25**, 1157 (2004).

*Development and Testing of a General Amber Force Field*

<sup>61</sup>A. S. Paluch, S. Parameswaran, S. Liu, A. Kolavennu and D. L. Mobley, *J. Chem. Phys.* **142**, 044508 (2015).

*Predicting the Excess Solubility of Acetanilide, Acetaminophen, Phenacetin, Benzocaine, and Caffeine in Binary Water/ethanol Mixtures via Molecular Simulation*

<sup>62</sup>H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma, *J. Phys. Chem.* **91**, 6269 (1987).

*The Missing Term in Effective Pair Potentials*

<sup>63</sup>J. E. Lennard-Jones, *Proc. R. Soc. Lond. A* **106**, 463 (1924).

*On the Determination of Molecular Fields*

<sup>64</sup>R. A. Buckingham, *P. Roy. Soc. Lon. A Mat.* 168, **264** (1938).

*The Classical Equation of State of Gaseous Helium, Neon and Argon*

<sup>65</sup>D. Berthelot, *Compt. Rend. Acad. Sci. Paris* **126** 1703 (1889).

<sup>66</sup>H. A. Lorentz, *Ann. Phys.* **12** 127 (1881).

<sup>67</sup>M. P. Allen and D. J. Tildesley, *Computer simulation of liquids* (Oxford University Press, 1987).

<sup>68</sup>W. Shinoda, M. Shiga, and M. Mikami, *Phys. Rev. B* **69**, 134103 (2004).

*Rapid Estimation of Elastic Constants by Molecular Dynamics Simulation under Constant Stress*

<sup>69</sup>G. J. Martyna, D. J. Tobias and M. L. Klein, *J. Chem. Phys.* **101**, 4177 (1994).

*Constant Pressure Molecular Dynamics Algorithms*

<sup>70</sup>M. Parrinello and A. Rahman, *J. Appl. Phys.* **52**, 7182 (1981).

*Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method*

<sup>71</sup>L. Verlet, *Phys. Rev.* **159**, 98 (1967).

*Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules*

<sup>72</sup>W. G. Hoover, Phys. Rev. A. **31**, 1695 (1985).

*Canonical Dynamics: Equilibrium Phase-space Distributions*

<sup>73</sup>W. G. Hoover, Phys. Rev. A **34**, 2499 (1986).

*Constant-pressure Equations of Motion*

<sup>74</sup>T. Schneider and E. Stoll, Phys Rev B, **17**, 1302 (1978).

*Molecular-dynamics Study of a Three-dimensional One-component Model for Distortive Phase Transitions*

<sup>75</sup>P. P. Ewald, Ann. Phys. (Leipzig) **64**, 253 (1921).

<sup>76</sup>R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles* (McGraw-Hill, New York, 1989).

<sup>77</sup>S. W. de Leeuw, J. W. Perram and E. R. Smith, Proc. R. Soc. London A **373**, 27 (1980).

*Simulation of Electrostatic Systems in Periodic Boundary Conditions. I. Lattice Sums and Dielectric Constants*

<sup>78</sup>M. Abramowitz and A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).

<sup>79</sup>K. E. Atkinson, *An Introduction to Numerical Analysis*, 2nd ed. (New York, 1989).

<sup>80</sup>H. Flyvbjerg and H. G. Petersen, J. Chem. Phys. **91**, 461 (1989).

*Error Estimates on Averages of Correlated Data*

<sup>81</sup>J. Hine and P. K. Mookerjee, J. Org. Chem. **40**, 292 (1975).

*Structural Effects on Rates and Equilibriums. XIX. Intrinsic Hydrophilic Character of Organic Compounds. Correlations in Terms of Structural Contributions*

<sup>82</sup>R. D. Wauchope and R. Haque, Can. J. Chem. **50**, 133 (1972)

*Aqueous Solubility of Polynuclear Aromatic Hydrocarbons*

<sup>83</sup>W. Shiu and K. Ma, *J. Phys. Chem. Ref. Data* **29**, 41 (2000).

*Temperature Dependence of Physical–Chemical Properties of Selected Chemicals of Environmental Interest. I. Mononuclear and Polynuclear Aromatic Hydrocarbons*

<sup>84</sup>A. Ben-Naim and Y. Marcus, *J. Chem. Phys.* **81**, 2016 (1984).

*Solvation Thermodynamics of Nonionic Solutes*

<sup>85</sup>J. Z. Vilseck, J. Tirado-Rives, and W. L. Jorgensen, *Phys. Chem. Chem. Phys.* **17**, 8407 (2015).

*Determination of Partial Molar Volumes from Free Energy Perturbation Theory*

<sup>86</sup>S. Sawamura, *J. Solution Chem.* **29**, 369 (2000).

*Pressure Dependence of the Solubilities of Anthracene and Phenanthrene in Water at 25°C*

<sup>87</sup>F. P. A. Fabbiani, D. R. Allan, S. Parsons, and C. R. Pulham, *Acta Crystallogr., Sect. B: Struct. Sci.* **62**, 826 (2006).

*Exploration of the High-pressure Behaviour of Polycyclic Aromatic Hydrocarbons: Naphthalene, Phenanthrene and Pyrene*

<sup>88</sup>Q-W. Huang, J. Zhang, A. Berlie, Z-X. Qin, X-M. Zhao, J-B. Zhang, L-Y. Tang, J. Liu4, Chao. Zhang, G-H. Zhong, H-Q. Lin, and X-J. Chen. *Phys.* **139**, 104302 (2013).

*Structural and Vibrational Properties of Phenanthrene under Pressure*

<sup>89</sup>I. S. Joung and T. E. Cheatham, III, *J. Phys. Chem. B* **112**, 9020 (2008). *Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations*

<sup>90</sup>G. Hummer, L. R. Pratt and A. E. García, *J. Phys. Chem.* **100**, 1206 (1996).

*Free Energy of Ionic Hydration*

<sup>91</sup>F. Figueirido, G. S. Del Buono, and R. M. Levy, *J. Phys. Chem. B* **101**, 5622 (1997).

*On Finite-Size Corrections to the Free Energy of Ionic Hydration*

<sup>92</sup>G. Hummer, L. R. Pratt and A. E. García, *J. Chem. Phys.* **107**, 9275 (1997).

*Ion sizes and Finite-size Corrections for Ionic-solvation Free Energies*

<sup>93</sup>G. Hummer, L. R. Pratt and A. E. Garcia, *J. Phys. Chem. A* **102**, 7885 (1998). *Molecular Theories and Simulation of Ions and Polar Molecules in Water*

<sup>94</sup>F. David, V. Vokhmin, G. J. Ionova, *Mol. Liq.* **90**, 45 (2001).

*Water Characteristics Depend on the Ionic Environment. Thermodynamics and Modelisation of the Aquo Ions*

<sup>95</sup>Y. Marcus, *Biophys. Chem.* **51**, 111 (1994).

*A Simple Empirical Model Describing the Thermodynamics of Hydration of Ions of Widely Varying Charges, Sizes, and Shapes*

<sup>96</sup>Y. Marcus, *J. Phys. Chem. B*, 113, 10285 (2009).

*The Standard Partial Molar Volumes of Ions in Solution. Part 4. Ionic Volumes in Water at 0-100 °C*

<sup>97</sup>P. Drude, W. Z. Nernst, *Phys. Chem.* 15, 79 (1894).

<sup>98</sup>K. N. Han, *Fundamentals of Aqueous Metallurgy* (SME, 2002). p. 81.

<sup>99</sup>K. N. Han, *Metals Mater. Int.* **4**, 1097 (1998).

*Effect of Temperature and Pressure on Equilibrium Constant Involving Metal Ions*

<sup>100</sup>G. Raabe and R. J. Sadus, *J. Chem. Phys.* **134**, 234501 (2011).

*Molecular Dynamics Simulation of the Dielectric Constant of Water: The Effect of Bond Flexibility*

<sup>101</sup>H. Sitepu, B. H. O'Connor and D. Li, *J. Appl. Cryst.* **38**, 158 (2005).

*Comparative Evaluation of the March and Generalized Spherical Harmonic Preferred Orientation Models Using X-ray Diffraction Data for Molybdite and Calcite Powders*

<sup>102</sup>J. Zhang, R. Reeder, *J. Am. Mineral.* **84**, 861 (1999).

*Comparative Compressibilities of Calcite-structure Carbonates: Deviations from Empirical Relations*



<sup>103</sup>S. A. Markgraf and R. J. Reeder, *Am. Mineral.* **70**, 590 (1985).

*High-Temperature Structure Refinements of Calcite and Magnesite*

<sup>104</sup>J. Kendall, *Philos. Mag.* **23**, 958 (1912).

XCVI. The Solubility of Calcium Carbonate in Water

<sup>105</sup>J. Phys. Chem. Ref. Data **41**, 023105 (2012).

*IUPAC-NIST Solubility Data Series. 95. Alkaline Earth Carbonates in Aqueous Systems. Part*

*2. Ca*

<sup>106</sup>L. N. Plummer and E. Busenberg, *Geochim. Cosmochim. Acta*, **46**, 1011 (1982). *The Solubilities of Calcite, Aragonite, and Vaterite in CO<sub>2</sub>-H<sub>2</sub>O Solutions between 0 and 90°C, and an Evaluation of the Aqueous Model for the System CaCO<sub>3</sub>-CO<sub>2</sub>- H<sub>2</sub>O*

<sup>107</sup>A. Mucci, *Am. J. Sci.* **283**, 780 (1983).

*The Solubility of Calcite and Aragonite in Seawater at Various Salinities, Temperatures, and One Atmosphere Total Pressure*

<sup>108</sup>Owen and S. R. Brinkley, *Chem. Rev.* **29**, 461 (1941).

*Calculation of the Effect of Pressure upon Ionic Equilibria in Pure Water and in Salt Solutions*

<sup>109</sup>F. J. Millero, *Geochim. Cosmochim. Acta* **46**, 11 (1982).

*The Effect of Pressure on the Solubility of Minerals in Water and Seawater*

<sup>110</sup>G. Wolf, J. Lerchner, H. Schmidt, H. Gamsjäger, E. Königsberger and P. Schmidt, *J. Therm. Anal.* **46**, 353 (1996).

*Thermodynamics of CaCO<sub>3</sub> phase transitions*

<sup>111</sup>H. Bothe and H. K. Cammenga, *J. Therm. Anal.* **16**, 267 (1979).

*Phase Transitions and Thermodynamic Properties of Anhydrous Caffeine*

<sup>112</sup>C. W. Lehmann and F. Stowasser, *Chem.–Eur. J.*, **13**, 2908 (2007).

*The Crystal Structure of Anhydrous  $\beta$ -Caffeine as Determined from X-ray Powder-Diffraction Data*

<sup>113</sup>M. T. Geballe, A. G. Skillman, A. Nicholls, J. P. Guthrie and P. J. Taylor, *J. Comput. Aided Mol. Des.* **24**, 259 (2010).

*The SAMPL2 Blind Prediction Challenge: Introduction and Overview*

<sup>114</sup>J. P. M. Jämbeck, F. Mocci, A. P. Lyubartsev and A. Laaksonen, *J. Comput. Chem.* **34**, 187 (2013).

*Partial Atomic Charges and Their Impact on the Free Energy of Solvation*

<sup>115</sup>C. W. Lehmann and F. Stowasser, *Chem.–Eur. J.*, **13**, 2908 (2007).

*The Crystal Structure of Anhydrous  $\beta$ -Caffeine as Determined from X-ray Powder-Diffraction Data*

<sup>116</sup>T. H. Lilley, H. Linsdell and A. Maestre, *J. Chem. Soc., Faraday Trans.*, **88**, 2865 (1992).

*Association of Caffeine in Water and in Aqueous Solutions of Sucrose*

<sup>117</sup>M. Falk, M. Gil and N. Iza, *Can. J. Chem.* **68**, 1293 (1990).

*Self-association of Caffeine in Aqueous Solution: an FT-IR Study*

<sup>118</sup>R. M. Noyes, *J. Am. Chem. Soc.* **84**, 513 (1962).

*Thermodynamics of Ion Hydration as a Measure of Effective Dielectric Properties of Water*

<sup>119</sup>M. D. Tissandier, K. A. Cowen, W. Y. Feng, E. Gundlach, M. H. Cohen, A. D. Earhart and J. V. Coe, *J. Phys. Chem. A* **102**, 7787 (1998).

*The Proton's Absolute Aqueous Enthalpy and Gibbs Free Energy of Solvation from Cluster-Ion Solvation Data*

<sup>120</sup>Y. H. Jang, X. Y. Chang, M. Blanco, S. Hwang, Y. Tang, P. Shuler and W. A. Goddard III, *J. Phys. Chem. B* **106**, 9951 (2002). *The MSXX Force Field for the Barium Sulfate-Water Interface*

<sup>121</sup>A. G. Stack and J. R. Rustad, *J. Phys. Chem. C* 111, 16387 (2007). *Structure and Dynamics of Water on Aqueous Barium Ion and the {001} Barite Surface*

<sup>122</sup>A. G. Stack, *J. Phys. Chem. C* 113, 2104 (2009). *Molecular Dynamics Simulations of Solvation and Kink Site Formation at the {001} Barite-Water Interface*

## SUPPLEMENTARY MATERIAL

### I. PHENANTHRENE (SOLUTION)

TABLE S1 Computed excess solvation free energy of phenanthrene in water at  $p = 0.1$  MPa and  $T = 298$  K (Result 1 in Fig. 1 (a)). Energies are in kJ/mol.

$\lambda$	$G_{\text{grow}}$	$G_{\text{insert}}$	$G_{\text{shrink}}$	$\mu_{\text{excess}}^{\text{vdwl}}$
1	258.15	-29.74	-240.88	-12.47(38)
1.5	297.34	-21.71	-288.29	-12.66(42)
2	338.64	-15.85	-335.50	-12.70(45)
2.5	382.65	-12.05	-382.78	-12.17(48)
3	428.07	-9.45	-430.87	-12.25(51)
3.5	475.15	-7.67	-479.49	-12.02(54)
4	523.19	-6.43	-528.81	-12.04(58)
4.5	572.24	-5.56	-578.63	-11.95(61)
5	620.67	-4.94	-627.96	-12.23(64)

TABLE S2 Computed excess solvation free energy of phenanthrene in water at  $p = 100$  MPa and  $T = 298$  K (Result 1 in Fig. 1 (b)). Energies are in kJ/mol.

$\lambda$	$G_{\text{grow}}$	$G_{\text{insert}}$	$G_{\text{shrink}}$	$\mu_{\text{excess}}^{\text{vdwl}}$
1	345.33	-39.45	-301.50	4.38(26)
1.5	407.29	-29.99	-372.94	4.36(30)
2	474.75	-22.16	-448.02	4.57(37)
2.5	547.89	-16.61	-526.54	4.74(40)
3	626.39	-12.81	-608.66	4.92(43)
3.5	710.05	-10.19	-694.86	4.99(46)
4	798.38	-8.32	-785.05	5.00(49)
4.5	890.73	-7.02	-878.59	5.13(53)
5	985.55	-6.10	-974.47	4.98(56)

TABLE S3 Computed excess solvation free energy of phenanthrene in water at  $p = 200$  MPa and  $T = 298$  K (Result 1 in Fig. 1 (c)). Energies are in kJ/mol.

$\lambda$	$G_{\text{grow}}$	$G_{\text{insert}}$	$G_{\text{shrink}}$	$\mu_{\text{excess}}^{\text{vdwl}}$
1	415.53	-46.32	-349.69	19.52(38)
1.5	496.11	-36.35	-440.35	19.40(40)
2	585.41	-27.34	-538.60	19.47(44)
2.5	683.55	-20.44	-643.49	19.62(48)
3	789.98	-15.66	-754.66	19.67(51)
3.5	905.12	-12.32	-873.12	19.69(55)
4	1028.55	-9.96	-998.68	19.91(59)
4.5	1159.15	-8.21	-1131.02	19.92(62)
5	1295.67	-6.97	-1268.81	19.89(66)

TABLE S4 Computed  $\Delta G_{\text{insert}}$  for phenanthrene in water at  $p = 0.1\text{-}200$  MPa and  $T = 298$  K at three different cavity sizes  $\lambda = 1, 1.5$  and  $2$ . Energies are in kJ/mol. “Multi-steps TI” and “Single-step FEP” corresponds to Result 1 and 2 in Fig. 1 (a), (b) and (c) respectively.

p = 0.1 MPa		
$\lambda$	Single-step FEP	Multi-steps TI
1	-29.988	-29.739
1.5	-21.575	-21.713
2	-15.831	-15.851
p = 100 MPa		
$\lambda$	Single-step FEP	Multi-steps TI
1	-38.767	-39.445
1.5	-29.910	-29.992
2	-21.997	-22.156
p = 200 MPa		
$\lambda$	Single-step FEP	Multi-steps TI
1	-47.311	-46.317
1.5	-36.364	-36.352
2	-27.319	-27.338

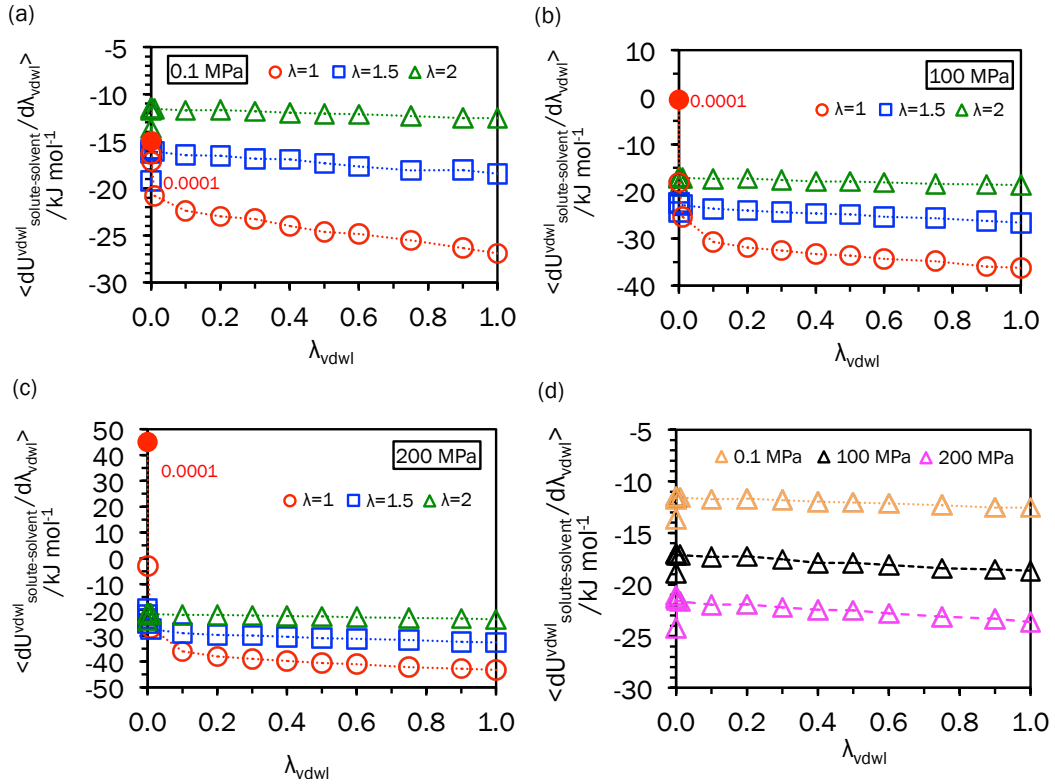


FIG S1 Thermodynamic integration contributions,  $\langle dU_{\text{solute-solvent}}^{\text{vdwl}}/d\lambda_{\text{vdwl}} \rangle$  of  $\Delta G_{\text{insert}}^{\text{vdwl}}$  (excluding tail corrections) for phenanthrene in water at  $p = 0.1\text{-}200$  MPa and  $T = 298$  K at three different cavity sizes  $\lambda = 1, 1.5$  and  $2$ , at (a)  $p = 0.1$  MPa; (b)  $p = 100$  MPa; (c)  $p = 200$  MPa; (d)  $\lambda = 2$  for  $p = 0.1, 100$  and  $200$  MPa. Larger cavities ( $\lambda$ ) inhibit end-point singularity.

## II. PHENANTHRENE (CRYSTAL)

In the extended Einstein crystal scheme, we have introduced the idea of “Central Atom” and “Orientational Atoms”, controlling the location and orientation of a molecule respectively<sup>1</sup>. Such a central-orientational spring arrangement is termed the “Einstein Arrangement” or “EA” in our work. We fix a set of atoms to their original coordinates, via harmonic springs of varying strength ( $K_{\max}$  for central atoms, and  $0 < K < K_{\max}$  depending on thermodynamics states). Fig. S2 below shows the two Einstein arrangements used for phenanthrene in the present work, and Table S5 below summarize our solid free energies.

TABLE S5 Computed absolute solid free energies of phenanthrene and their averages at  $p = 0.1-200$  MPa and  $T = 298$  K. Energies are in kJ/mol and  $K_{\max}$  is in kJ/mol/Å<sup>2</sup>.

p/MPa	EA	Thermostat <sup>a</sup>	$K_{\max}$	$A_{\text{solid}}/N_{\text{solid}}$	$pV_{\text{solid}}/N_{\text{solid}}$	$G_{\text{solid}}/N_{\text{solid}}$
0.1	1	Nosé–Hoover 1	10460	-57.76	0.00	-57.76
0.1	1	Nosé–Hoover 2	10460	-57.91	0.00	-57.91
0.1	1	Langevin 1	10460	-57.54	0.00	-57.54
0.1	1	Langevin 2	10460	-57.47	0.00	-57.47
0.1	2	Nosé–Hoover	10460	-58.23	0.00	-58.23
0.1	2	Langevin	10460	-58.24	0.00	-58.24
Average						-57.86
100	2	Nosé–Hoover	10460	-58.04	14.78	-43.26
100	2	Langevin	10460	-57.95	14.78	-43.17
Average						-43.21
200	2	Nosé–Hoover	10460	-58.17	29.16	-29.01
200	2	Langevin	10460	-58.21	29.16	-29.05
Average						-29.03

<sup>a</sup>The number labeling after the thermostat type indicates different constants  $c$  in computing  $\Delta A_1$  and  $\Delta A_2$ . See reference 1 for usage of  $c$  for (different)  $\Delta A_2$  and  $\Delta A_0$  in the Einstein crystal method;  $c$  is generally between 0-6 in this work.

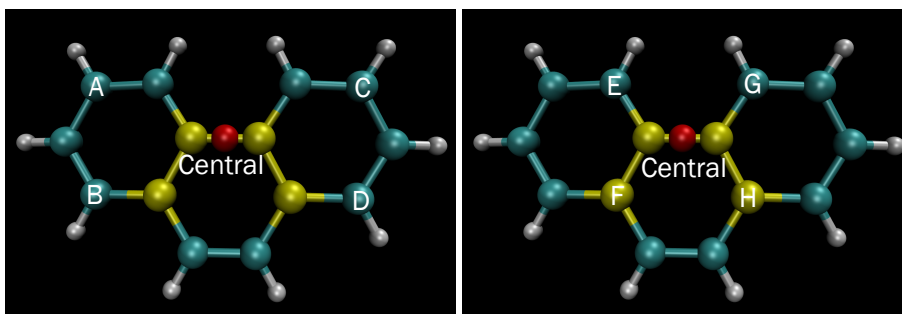


FIG. S2. (a) Einstein arrangement 2; (b) Einstein arrangement 1. Different atom types are colored differently: the white atoms are hydrogen atoms; the cyan atoms are non-aromatic carbon atoms; the yellow atoms are the aromatic carbon atoms; the red atom is a virtual non-interacting central atom in the Einstein crystal.

### III. ERROR CALCULATION FOR FIG.3 – PHENANTHRENE SOLUBILITY

Here we show how we calculated the statistical error of the aqueous solubility of phenanthrene in Fig. 3.

The procedure is the same as in reference 1.

To calculate solubilities in mole fractions, the following relationship was used:

$$x = \rho_{\text{solute}} / \rho_{\text{solution}}$$

where  $\rho_{\text{solution}}$  and  $\rho_{\text{solute}}$  are number densities of the solution and solute in the solution respectively, and  $x$  is the mole fraction. At infinitely dilution limit,  $\rho_{\text{solution}} = \langle \rho_{\text{solvent}} \rangle = 977.8(2) \text{ kg/m}^3 = 0.032684(1) \text{ water molecules per } \text{\AA}^3$ .

The number density solubility of phenanthrene was calculated from Eq. (A4) in the main text,

$$k_B T \ln(\rho_{\text{solute}}) + \Delta\mu_{\text{excess}} = (A_{\text{solid}} + pV_{\text{solid}}) / N_{\text{solid}}$$

For  $p = 0.1 \text{ MPa}$  and  $T = 298 \text{ K}$ ,  $G_{\text{solid}} = A_{\text{solid}} = -57.858 \text{ kJ/mol}$ , and  $\Delta\mu_{\text{excess}} = -12.246 \text{ kJ/mol}$ . The statistical error in  $G_{\text{solid}}$  and  $\Delta\mu_{\text{excess}}$  can be taken to be around  $0.15 \text{ kJ/mol}$  and  $0.50 \text{ kJ/mol}$ , so we can take the total error to be  $0.65 \text{ kJ/mol}$ , hence,

$$\ln(\rho_{\text{solute}}) = -45.612 (650) \text{ kJ/mol} = -18.409(262) k_B T \text{ (per molecule)}$$

If  $y = e^x$ , then the error in  $y$ ,  $\Delta y = \Delta x \times e^x$ ; we have  $\rho_{\text{solute}} = \exp(-18.409) = 1.01 \times 10^{-8}$  (number of solute molecules in  $1 \text{ \AA}^3 = 0.001 \text{ nm}^3$ ), so  $\Delta\rho_{\text{solute}} = 0.262 \times 1.01 \times 10^{-8} = 2.65 \times 10^{-9}$  (number of solute molecules in  $1 \text{ \AA}^3 = 0.001$ ).

Neglecting the error of ambient water density and taking  $\rho_{\text{solution}} = \langle \rho_{\text{solvent}} \rangle$  gives the mole fraction solubility as,

$$x = \rho_{\text{solute}} / \rho_{\text{solution}} = (1.01 \times 10^{-8} / 0.032684) = 3.09 \times 10^{-7}$$

$$\Delta x = \Delta\rho_{\text{solute}} / \rho_{\text{solution}} = (2.65 \times 10^{-9} / 0.032684) = 8.10 \times 10^{-8}$$



If  $y = \ln x$ , the error in  $y$ ,  $\Delta y = \Delta x/x$ .

In Fig. 3 (a), the error of the natural logarithm of mole fraction solubility  $\ln x$  at ambient conditions is,  $\Delta \ln x = 8.10 \times 10^{-8}/3.09 \times 10^{-7} = 0.262$ .

The same procedure of error propagation was applied to data points at other pressure/temperature and other compounds.

#### IV. CALCITE – IONIC CHARGING FREE ENERGY

This session explains how we obtained  $\Delta G_{\text{excess}}^{\text{solute-solvent}}$  in Eq. (5) in the main text.

For a single ion with charge  $q_i$  at position  $\mathbf{r}$  in solution, its electrostatic potential energy is<sup>2-3</sup>,

$$U(\mathbf{r}, q_i) = q_i \varphi(\mathbf{r}) + 1/2 q_i^2 \xi_{\text{EW}} \quad (\text{S1})$$

where  $\varphi(\mathbf{r})$  is the total electrostatic potential at  $\mathbf{r}$ .

We use  $0 \leq \alpha \leq 1$  as our linear atomic (partial) charge ratio, with  $q_i = \alpha q$ ,

$$U(\mathbf{r}, \alpha) = U(\alpha) = \alpha q \varphi(\mathbf{r}) + 1/2 (\alpha q)^2 \xi_{\text{EW}} \quad (\text{S2})$$

Separating the ion-solvent interaction and ion-ion self-interaction in Eq. (S2),

$$U_{\text{solute-solvent}}(\alpha) = \alpha q \varphi(\mathbf{r}) \quad (\text{S3})$$

$$U_{\text{solute-solute}}(\alpha) = 1/2 (\alpha q)^2 \xi_{\text{EW}} \quad (\text{S4})$$

For the total charging free energy  $\Delta G_{\text{total}}$ ,

$$\Delta G_{\text{total}} = \Delta G_{\text{solute-solvent}} + \Delta G_{\text{solute-solute}} \quad (\text{S5})$$

The solute-solvent  $\Delta G_{\text{solute-solvent}}$  is,

$$\Delta G_{\text{solute-solvent}} = \int_0^1 \langle \partial U_{\text{solute-solvent}}(\alpha) / \partial \alpha \rangle d\alpha \quad (\text{S6})$$

The integrand in Eq. (S6) for  $\Delta G_{\text{solute-solvent}}$  is,

$$\langle \partial U_{\text{solute-solvent}}(\alpha) / \partial \alpha \rangle = \langle q \varphi(\mathbf{r}) \rangle = \langle U_{\text{solute-solvent}}(\alpha) / \alpha \rangle \quad (\text{S7})$$

Thus we sample  $\langle U_{\text{solute-solvent}}(\alpha) / \alpha \rangle$  in our simulations.

For the finite-size correction  $\Delta G_{\text{solute-solute}}^2$ , we use,

$$\Delta G_{\text{solute-solute}} = 1/2 (\alpha q)^2 \xi_{\text{EW}} \Big|_{\alpha=1} - 1/2 (\alpha q)^2 \xi_{\text{EW}} \Big|_{\alpha=0} = 1/2 q^2 \xi_{\text{EW}} \quad (\text{S8})$$

At constant pressure simulation where the box volume changes,  $\xi_{EW}$  should also be averaged<sup>4</sup>. So we sample  $1/2q^2\langle\xi_{EW}\rangle$  in our simulations.

In reference 2, the excess single-ion solvation free energies were computed via the second-order perturbation, where the second-order term in the Taylor expansion of the ionic charge was included in the free energy. However, for accurate extractions, this requires longer runs – 100000 Monte Carlo data with were sampled for each  $\alpha$  in reference 2 compared to 2000 data our work (every two data points are 500 fs apart), while the improvement is mild for the purpose of solubility estimation. Table S6 below shows our results (1000 data and 8 values of  $\alpha$ ) of repeating Hummer’s work in computing the free energy of charging Straatsm-Berendsen<sup>4</sup> Na<sup>+</sup> in SPC (simple point charge) water and  $NVT$  ensemble, as well as Hummer’s results (100000 data and 8 values of  $\alpha$ ). In this case the Helmholtz free energy  $A$  is used instead of  $G$  because reference 2 has employed a fixed volume. The difference between the two works is a few kJ/mol and of similar magnitude to the variation within Hummer’s own results.

Table S6 Charging free energies of Na<sup>+</sup> compared with Hummer’s work<sup>2</sup>. Energies are in kJ/mol.

	Source	N <sub>SPC</sub>	$\Delta A_{\text{solute-solute}}$	$\Delta A_{\text{solute-solvent}}$	$\Delta A_{\text{total}}$
TI (charging)	This Work	256	-100	-301	-402
TI (charging)	Hummer	256	-100	-304	-404
TI (decharging)	Hummer	256	-100	-306	-406
TI (charging)	This Work	32	-198	-209	-407
TI (charging)	Hummer	32	-198	-205	-403
TI (decharging)	Hummer	32	-198	-209	-407

## V. CALCITE – VAN DER WAALS SOLVATION FREE ENERGY

### A. $p = 0.1$ MPa and $T = 298$ K

Fig. S3 (a) and (b) below are plots of the computed excess van der Waals solvation free energies of the charges-free  $\text{Ca}^0$  and  $\text{CO}_3^0$ , i. e.  $\Delta G_{\text{excess}}^{\text{vdwl}}$ , with varying cavity size  $\lambda$ , at  $p = 0.1$  MPa and  $T = 298$  K. Data are also summarized in Table S7 and S8 below.

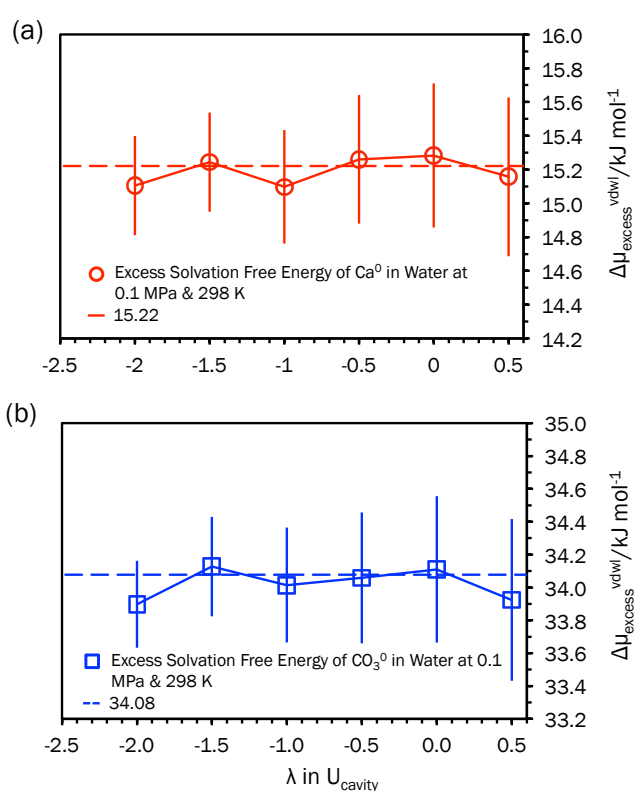


FIG S3 (a) Computed van der Waals excess solvation free energies of  $\text{Ca}^0$  in SPC/Fw water at  $p = 0.1$  MPa and  $T = 298$  K via the cavity method. The dashed red line is the average of the four data points in the range of  $\lambda = -1.5$  and  $\lambda = 0.0$ ; (b) computed van der Waals excess solvation free energies of  $\text{CO}_3^0$  in SPC/Fw water at  $p = 1$  atm and  $T = 298$  K via the cavity method. The dashed blue line is the average of the four data points in the range of  $\lambda = -1.5$  and  $\lambda = 0.0$ .

TABLE S7 Computed van der Waals excess solvation free energies of  $\text{Ca}^0$  in SPC/Fw water at  $p = 0.1$  MPa and  $T = 298$  K. Energies are in kJ/mol.

$\lambda$	$G_{\text{grow}}$	$G_{\text{insert}}$	$G_{\text{shrink}}$	$\Delta\mu_{\text{excess}}^{\text{vdwl}}$
-2	81.4	-0.9	-65.4	15.1(3)
-1.5	107.1	-0.8	-91.1	15.2(3)
-1	136.1	-0.5	-120.5	15.1(4)
-0.5	168.4	-0.3	-152.9	15.3(4)
0	203.8	-0.2	-188.3	15.3(5)
0.5	241.7	-0.1	-226.5	15.2(5)

TABLE S8 Computed van der Waals excess solvation free energies of  $\text{CO}_3^0$  in SPC/Fw water at  $p = 0.1$  MPa and  $T = 298$  K. Energies are in kJ/mol.

$\lambda$	$G_{\text{grow}}$	$G_{\text{insert}}$	$G_{\text{shrink}}$	$\Delta\mu_{\text{excess}}^{\text{vdwl}}$
-2	81.4	3.3	-50.8	33.9(3)
-1.5	107.1	1.1	-74.1	34.1(3)
-1	136.1	0.2	-102.4	34.0(3)
-0.5	168.4	0.0	-134.4	34.1(4)
0	203.8	0.0	-169.7	34.1(4)
0.5	241.7	0.0	-207.8	33.9(5)

## B. $p = 0.1$ MPa and $T = 323$ K

Fig. S3 (a) and (b) below are plots of the computed excess van der Waals solvation free energies of the charges-free  $\text{Ca}^0$  and  $\text{CO}_3^0$ , i. e.  $\Delta\mu_{\text{excess}}^{\text{vdw}}$ , with varying cavity size  $\lambda$ , at  $p = 0.1$  MPa and  $T = 323$  K. The results are also summarized in Table S9 and S10 below.

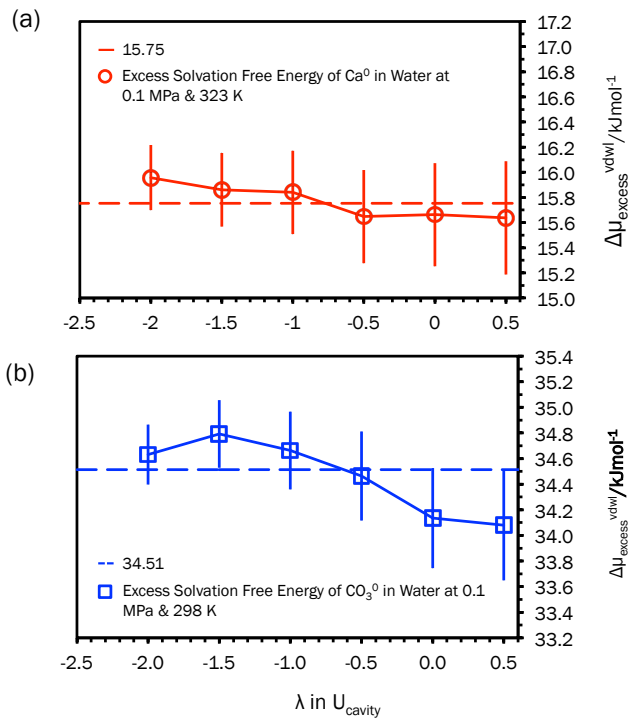


FIG S4 (a) Computed van der Waals excess solvation free energies of  $\text{Ca}^0$  in SPC/Fw water at  $p = 0.1$  MPa and  $T = 323$  K via the cavity method. The dashed red line is the average of the four data points in the range of  $\lambda = -1.5$  and  $\lambda = 0.0$ ; (b) computed van der Waals excess solvation free energies of  $\text{CO}_3^0$  in SPC/Fw water at  $p = 1$  atm and  $T = 298$  K via the cavity method. The dashed blue line is the average of the four data points in the range of  $\lambda = -1.5$  and  $\lambda = 0.0$ .

TABLE S9 Computed van der Waals excess solvation free energies of  $\text{Ca}^0$  in SPC/Fw water at  $p = 0.1$  MPa and  $T = 323$  K. Energies are in kJ/mol.

$\lambda$	$G_{\text{grow}}$	$G_{\text{insert}}$	$G_{\text{shrink}}$	$\Delta\mu_{\text{excess}}^{\text{vdwl}}$
-2	80.1	-0.8	-63.4	16.0(3)
-1.5	105.1	-0.8	-88.5	15.9(3)
-1	133.2	-0.5	-116.9	15.8(3)
-0.5	164.2	-0.3	-148.3	15.6(4)
0	198.1	-0.2	-182.2	15.7(4)
0.5	234.6	-0.1	-218.8	15.6(5)

TABLE S10 Computed van der Waals excess solvation free energies of  $\text{CO}_3^0$  in SPC/Fw water at  $p = 0.1$  MPa and  $T = 323$  K. Energies are in kJ/mol.

$\lambda$	$G_{\text{grow}}$	$G_{\text{insert}}$	$G_{\text{shrink}}$	$\Delta\mu_{\text{excess}}^{\text{vdwl}}$
-2	80.1	3.7	-49.2	34.6(2)
-1.5	105.1	1.2	-71.6	34.8(3)
-1	133.2	0.3	-98.8	34.7(3)
-0.5	164.2	0.0	-129.8	34.5(3)
0	198.1	0.0	-163.9	34.1(4)
0.5	234.6	0.0	-200.5	34.1(4)

### C. $p = 0.1$ MPa and $T = 348$ K

Fig. S3 (a) and (b) below are plots of the computed excess van der Waals solvation free energies of the charges-free  $\text{Ca}^0$  and  $\text{CO}_3^0$ , i. e.  $\Delta G_{\text{excess}}^{\text{vdwl}}$ , with varying cavity size  $\lambda$ , at  $p = 0.1$  MPa and  $T = 348$  K. The results are also summarized in Table S9 and S10 below.

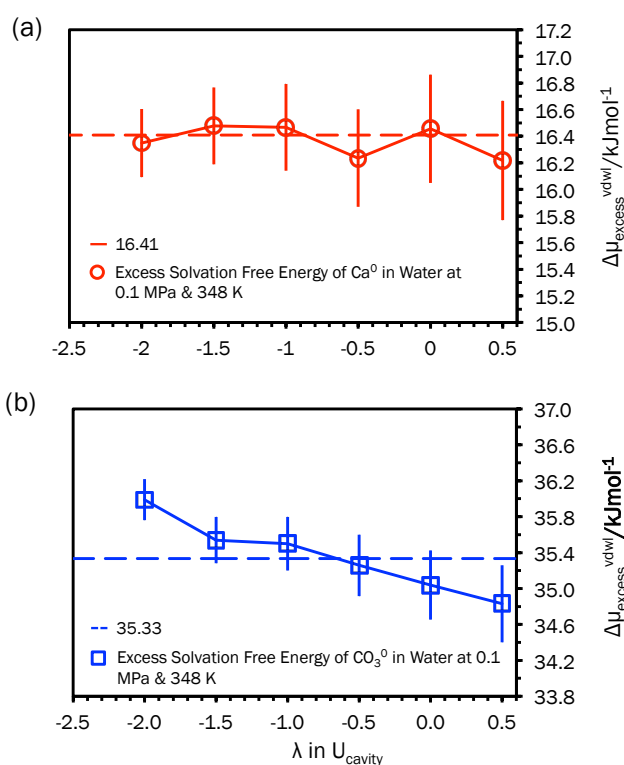


FIG S5 (a) Computed van der Waals excess solvation free energies of  $\text{Ca}^0$  in SPC/Fw water at  $p = 0.1$  MPa and  $T = 348$  K via the cavity method. The dashed red line is the average of the four data points in the range of  $\lambda = -1.5$  and  $\lambda = 0.0$ ; (b) computed van der Waals excess solvation free energies of  $\text{CO}_3^0$  in SPC/Fw water at  $p = 1$  atm and  $T = 348$  K via the cavity method. The dashed blue line is the average of the four data points in the range of  $\lambda = -1.5$  and  $\lambda = 0.0$ .



TABLE S11 Computed van der Waals excess solvation free energies of  $\text{Ca}^0$  in SPC/Fw water at  $p = 0.1$  MPa and  $T = 348$  K. Energies are in kJ/mol.

$\lambda$	$G_{\text{grow}}$	$G_{\text{insert}}$	$G_{\text{shrink}}$	$\Delta\mu_{\text{excess}}^{\text{vdwl}}$
-2	78.6	-0.7	-61.6	16.3(3)
-1.5	102.8	-0.7	-85.6	16.5(3)
-1	129.8	-0.5	-112.9	16.5(3)
-0.5	159.7	-0.5	-143.0	16.2(4)
0	192.2	-0.2	-175.6	16.5(4)
0.5	227.1	-0.2	-210.7	16.2(4)

TABLE S12 Computed van der Waals excess solvation free energies of  $\text{CO}_3^0$  in SPC/Fw water at  $p = 0.1$  MPa and  $T = 348$  K. Energies are in kJ/mol.

$\lambda$	$G_{\text{grow}}$	$G_{\text{insert}}$	$G_{\text{shrink}}$	$\Delta\mu_{\text{excess}}^{\text{vdwl}}$
-2	78.6	4.5	-47.1	36.0(2)
-1.5	102.8	1.2	-68.5	35.5(3)
-1	129.8	0.3	-94.6	35.5(3)
-0.5	159.7	0.0	-124.5	35.3(3)
0	192.2	0.0	-157.2	35.0(4)
0.5	227.1	0.0	-192.3	34.8(4)

#### D. $p = 100$ MPa and $T = 298$ K

Fig. S6 (a) and (b) below are plots of the computed excess van der Waals solvation free energies of the charges-free  $\text{Ca}^0$  and  $\text{CO}_3^0$ , i. e.  $\Delta G_{\text{excess}}^{\text{vdw}}$ , with varying cavity size  $\lambda$ , at  $p = 100$  MPa and  $T = 298$  K. The results are also summarized in Table S13 and S14 below.

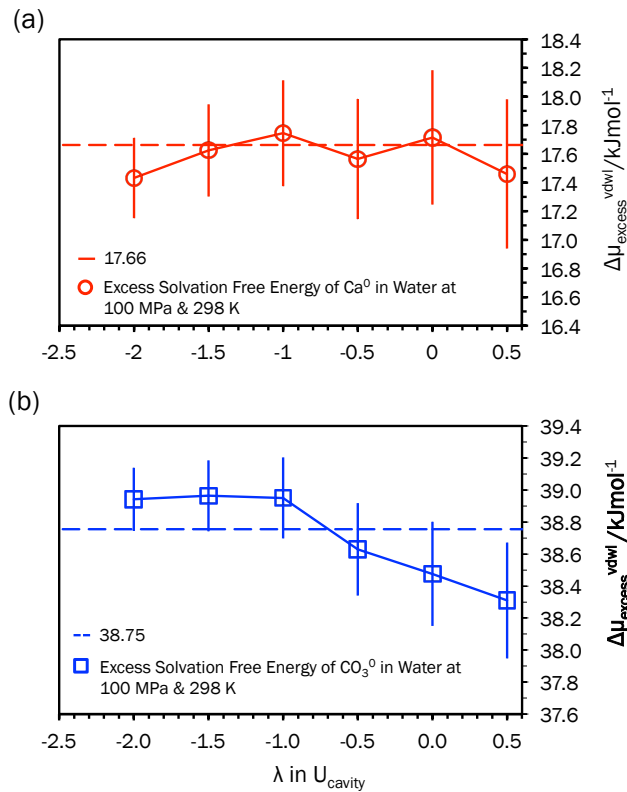


FIG S6 (a) Computed van der Waals excess solvation free energies of  $\text{Ca}^0$  in SPC/Fw water at  $p = 100$  MPa and  $T = 298$  K via the cavity method. The dashed red line is the average of the four data points in the range of  $\lambda = -1.5$  and  $\lambda = 0.0$ ; (b) computed van der Waals excess solvation free energies of  $\text{CO}_3^0$  in SPC/Fw water at  $p = 1$  atm and  $T = 348$  K via the cavity method. The dashed blue line is the average of the four data points in the range of  $\lambda = -1.5$  and  $\lambda = 0.0$ .

TABLE S13 Computed van der Waals excess solvation free energies of  $\text{Ca}^0$  in SPC/Fw water at  $p = 100$  MPa and  $T = 298$  K. Energies are in kJ/mol.

$\lambda$	$G_{\text{grow}}$	$G_{\text{insert}}$	$G_{\text{shrink}}$	$\Delta\mu_{\text{excess}}^{\text{vdwl}}$
-2	90.8	-0.8	-72.6	17.4(3)
-1.5	122.2	-1.0	-103.6	17.6(3)
-1	158.8	-0.7	-140.4	17.7(4)
-0.5	200.7	-0.5	-182.7	17.6(4)
0	248.0	-0.3	-230.0	17.7(5)
0.5	300.4	-0.2	-282.8	17.5(5)

TABLE S14 Computed van der Waals excess solvation free energies of  $\text{CO}_3^0$  in SPC/Fw water at  $p = 100$  MPa and  $T = 298$  K. Energies are in kJ/mol.

$\lambda$	$G_{\text{grow}}$	$G_{\text{insert}}$	$G_{\text{shrink}}$	$\Delta\mu_{\text{excess}}^{\text{vdwl}}$
-2	90.8	5.2	-57.1	38.9(2)
-1.5	122.2	1.9	-85.1	39.0(2)
-1	158.8	0.4	-120.3	39.0(3)
-0.5	200.7	0.1	-162.2	38.6(3)
0	248.0	0.0	-209.5	38.5(3)
0.5	300.4	0.0	-262.1	38.3(4)

### E. $p = 200$ MPa and $T = 298$ K

Fig. S6 (a) and (b) below are plots of the computed excess van der Waals solvation free energies of the charges-free  $\text{Ca}^0$  and  $\text{CO}_3^0$ , i. e.  $\Delta G_{\text{excess}}^{\text{vdwl}}$ , with varying cavity size  $\lambda$ , at  $p = 200$  MPa and  $T = 298$  K. The results are also summarized in Table S15 and S16 below.

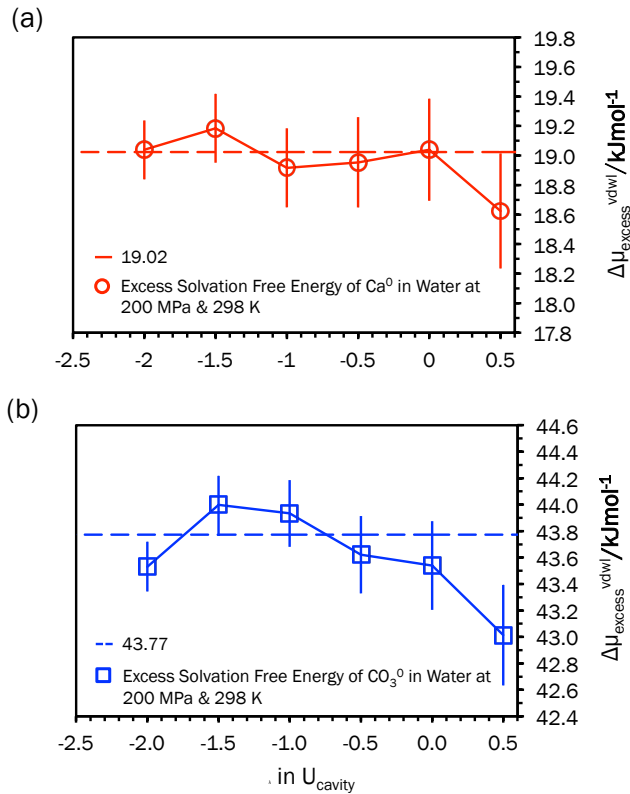


FIG S7 (a) Computed van der Waals excess solvation free energies of  $\text{Ca}^0$  in SPC/Fw water at  $p = 200$  MPa and  $T = 298$  K via the cavity method. The dashed red line is the average of the four data points in the range of  $\lambda = -1.5$  and  $\lambda = 0.0$ ; (b) computed van der Waals excess solvation free energies of  $\text{CO}_3^0$  in SPC/Fw water at  $p = 1$  atm and  $T = 348$  K via the cavity method. The dashed blue line is the average of the four data points in the range of  $\lambda = -1.5$  and  $\lambda = 0.0$ .

TABLE S15 Computed van der Waals excess solvation free energies of  $\text{Ca}^0$  in SPC/Fw water at  $p = 200$  MPa and  $T = 298$  K. Energies are in kJ/mol.

$\lambda$	$G_{\text{grow}}$	$G_{\text{insert}}$	$G_{\text{shrink}}$	$\Delta\mu_{\text{excess}}^{\text{vdwl}}$
-2	98.6	-0.7	-78.8	19.0(2)
-1.5	134.6	-1.0	-114.5	19.2(2)
-1	177.6	-0.8	-157.9	18.9(3)
-0.5	227.8	-0.6	-208.3	19.0(3)
0	285.6	-0.4	-266.2	19.0(4)
0.5	350.3	-0.2	-331.5	18.6(4)

TABLE S16 Computed van der Waals excess solvation free energies of  $\text{CO}_3^0$  in SPC/Fw water at  $p = 100$  MPa and  $T = 298$  K. Energies are in kJ/mol.

$\lambda$	$G_{\text{grow}}$	$G_{\text{insert}}$	$G_{\text{shrink}}$	$\Delta\mu_{\text{excess}}^{\text{vdwl}}$
-2	98.6	7.2	-62.2	43.5(2)
-1.5	134.6	2.9	-93.6	44.0(2)
-1	177.6	0.8	-134.5	43.9(3)
-0.5	227.8	0.1	-184.3	43.6(3)
0	285.6	0.0	-242.0	43.5(3)
0.5	350.3	0.0	-307.3	43.0(4)

## VI. CALCITE – CHARGING FREE ENERGY

Fig. S8 blow is a plot of the linear integrands  $\left\langle U_{\text{solute-solvent}}(\alpha) / \alpha \right\rangle_{\alpha, NpT}$  at different linear charge ratios for both  $\text{Ca}^{2+}$  and  $\text{CO}_3^{2-}$ . Ions at other temperatures and pressures show similar trends.

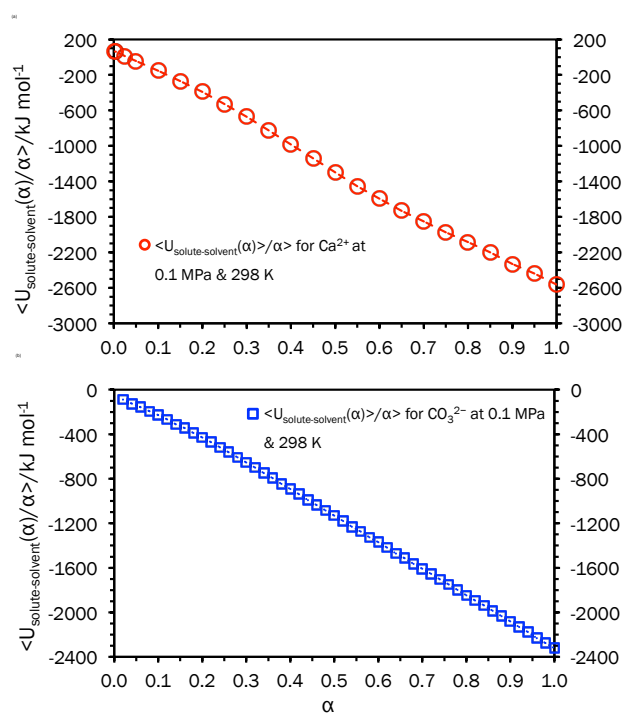


FIG S8  $\left\langle U_{\text{solute-solvent}}(\alpha) / \alpha \right\rangle_{\alpha, NpT}$  during ionic charging at  $p = 1$  atm and  $T = 298$  K for (a)  $\text{Ca}^{2+}$ ; (b)  $\text{CO}_3^{2-}$ . Error bars are smaller than the data marker sizes; plots of other temperatures and pressures are similar.

## VII. CALCITE- VAN DER WAALS/CHARGING CONTRIBUTIONS

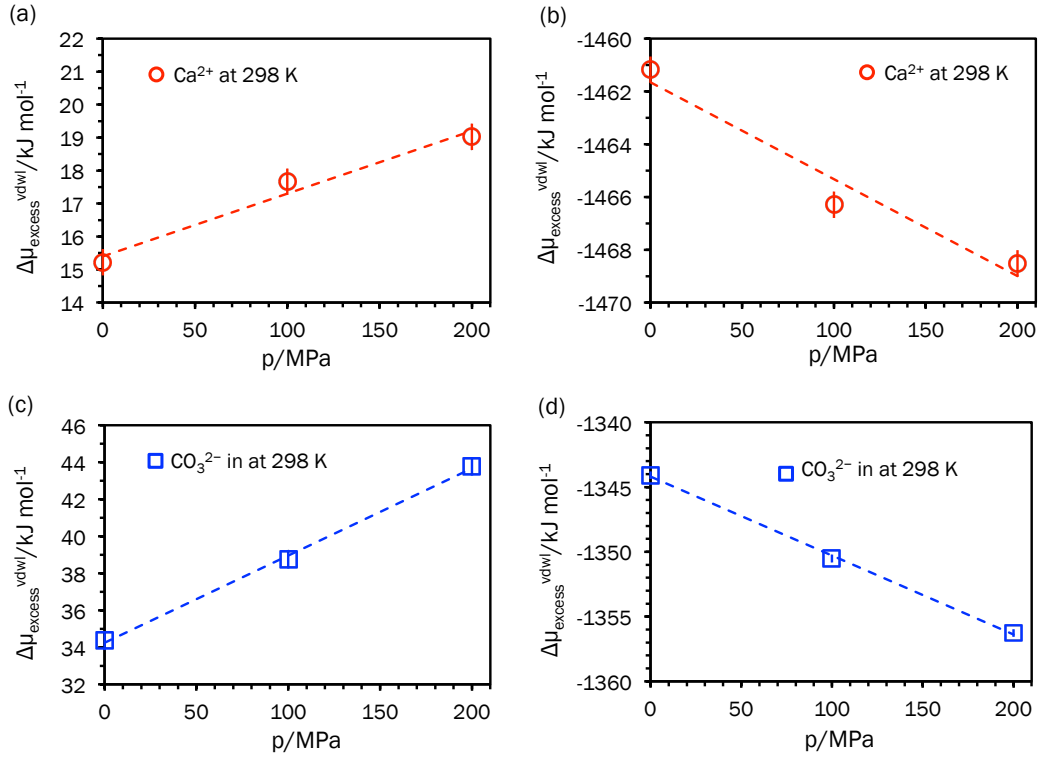


FIG S9 The pressure dependence of the van der Waals contribution  $\left(\partial\Delta\mu_{\text{excess}}^{\text{vdwl}}/\partial p\right)_T$  and the charging contribution  $\left(\partial\Delta\mu_{\text{excess}}^{\text{charging}}/\partial p\right)_T$  for  $\Delta\mu_{\text{excess}} = \Delta\mu_{\text{excess}}^{\text{vdwl}} + \Delta\mu_{\text{excess}}^{\text{charging}}$  at  $T = 298$  K for  $\text{Ca}^{2+}$  and  $\text{CO}_3^{2-}$ . Error bars for the van der Waals contributions are typically around 0.3-0.4 kJ/mol and error bars for the charging contributions are typically around 0.3-0.5 kJ/mol.

## VIII. CALCITE- RADIAL DISTRIBUTION FUNCTIONS

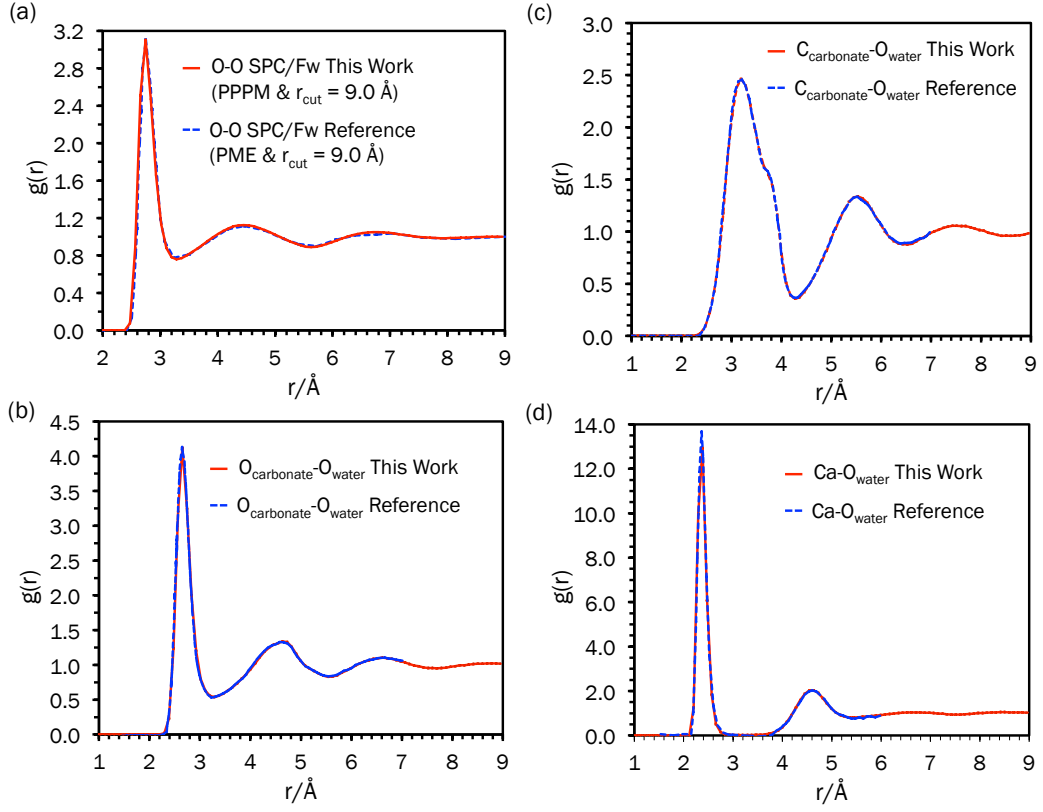


FIG S10 (a) Radial distribution functions  $g_{\text{O-O}}(r)$  for SPC/Fw oxygen in the SPC/Fw reference and in our work at ambient conditions<sup>4-5</sup> computed in PME<sup>6</sup>/PPPM<sup>7</sup>; (b) radial distribution functions  $g_{\text{O}_{\text{carbonate}}-\text{O}_{\text{water}}}(r)$  between the carbonate oxygen and SPC/Fw oxygen<sup>8</sup>; (c) radial distribution functions  $g_{\text{C}_{\text{carbonate}}-\text{O}_{\text{water}}}(r)$  between the carbonate carbon and SPC/Fw oxygen<sup>8</sup>; (d) radial distribution functions  $g_{\text{Ca}-\text{O}_{\text{water}}}(r)$  between the calcium ion and SPC/Fw oxygen<sup>8</sup>. We have obtained the SPC/Fw water density of 1010.7(1) kg/m<sup>3</sup>, compared with 1012(16) in the literature<sup>4</sup> and 1009.9(2) kg/m<sup>3</sup><sup>9</sup>.



## X. CALCITE – CRYSTALS

Table S17 below summarizes our average lattice parameters of the calcite lattice parameters at ambient conditions, in comparison with the simulations of reference 8 and with experimental data<sup>10</sup>. The agreement with the simulation reference is good, although comparison with experiment shows that the calcite model slightly overestimates the crystal density, and our average equilibrated lattice parameters  $a$ ,  $b$  and  $c$  are around 0.2-0.5 % larger than the experiments values. In our 3.5 ns  $NpT$  equilibrations, we allowed all six parameters to change during the relaxation while the simulation reference has kept the cell angles fixed during 5 ns  $NpT$  simulations.

TABLE S17 Average simulated lattice parameters at  $p = 0.1$  MPa and  $T = 298$  K in comparison with the simulation reference<sup>7</sup> and experiments<sup>8</sup>.

Lattice Parameters	This work (average)	Simulation Reference	Experiment
$a/\text{\AA}$	4.9362(2)	4.935	4.991
$b/\text{\AA}$	4.9359(2)	4.935	4.991
$c/\text{\AA}$	17.2291(7)	17.221	17.068
$\gamma/^\circ$	119.999(3)	120.000	120.000

## XI. CALCITE – ABSOLUTE SOLID FREE ENERGY

As a test of the robustness of our approach, we have also computed the ambient  $\Delta A_0$  using two different Einstein arrangements (see Table S18 below). We notice that for the carbonate ion, the two values agree within a few statistical errors.

TABLE S18  $\Delta A_0$  for ambient calcite with two different Einstein arrangements on the carbonate ion. Energies are in kJ/mol.

Einstein Arrangement of $\text{CO}_3^{2-}$	$\Delta A_0/N_{\text{CaCO}_3}$
C atom as the "Central Atom"	-67.81(4)
One of the O atoms as the "Central Atom"	-67.94(4)

In addition, for  $\Delta A_1$  at ambient conditions, we have tried changing the order in which the thermodynamic integration (switching on charges/van der Waals interactions first) were carried out, and the results are shown in Table S19 below.

TABLE S19  $\Delta A_1$  for ambient calcite with two different orders. Energies are in kJ/mol.

	$\Delta A_1/N_{\text{CaCO}_3}$
Growing Charges First	-2879.688(2)
Growing van der Waals Interactions First	-2879.682(2)

TABLE S20 The absolute solid free energy of calcite and its components at ambient conditions (C atom as the Central atom & growing van der Waals first). Energies are in kJ/mol.

$A_{\text{CaCO}_3} = A_0 + \Delta A_0 + \Delta A_1 + \Delta A_2 + \Delta A_{\text{symmetry}}$	kJ/mol
$A_0/N_{\text{CaCO}_3}$	53.543
$\Delta A_0/N_{\text{CaCO}_3}$	67.85
$\Delta A_1/N_{\text{CaCO}_3}$	-2879.688
$\Delta A_2/N_{\text{CaCO}_3}$	-76.145
$\Delta A_{\text{symmetry}}/N_{\text{CaCO}_3}$	-4.44
$A_{\text{CaCO}_3}$	-2838.88

In the main text, we have briefly compared the solid free energies and thermodynamic stability of calcite and aragonite in agreement with experiments. In Table S21 below, we have included the solid free energies of calcite and aragonite.

TABLE S21 Computed absolute solid free energies of calcite and aragonite at  $p = 0.1$  MPa and  $T = 298$  K. Energies are in kJ/mol per formula unit. The formula units are included in the brackets.

	$A_{\text{solid}}/N_{\text{solid}}$
Calcite( $\text{CaCO}_3$ )	-2838.84
Aragonite( $\text{CaCO}_3$ )	-2837.74

## CAFFEINE

Fig. S4 shows a plot of our computed van der Waals contributions for an atomic-partial-charges-free caffeine molecule in water at various cavity size  $\lambda$ . The red dashed line give an average value of  $-4.29(40)$  kJ/mol (between  $\lambda = 0.5$  to  $\lambda = 3.5$ ).

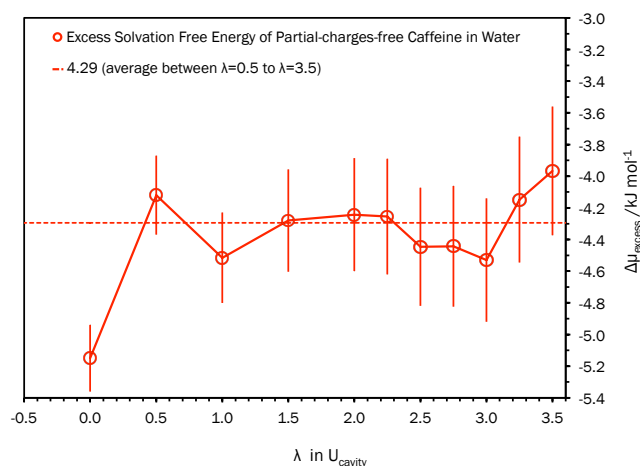


FIG S11 Computed excess solvation free energies of a partial-charges-free caffeine in water.

TABLE S21 Computed van der Waals excess solvation free energies of an atomic-partial-charges-free caffeine in water at  $p = 0.1$  MPa and  $T = 298$  K. Energies are in kJ/mol.

	$G_{\text{grow}}$	$G_{\text{insert}}$	$G_{\text{shrink}}$	
3.5	519.0	-8.0	-515.0	-4.0(4)
3.25	492.6	-8.8	-487.9	-4.1(4)
3	466.6	-9.9	-461.3	-4.5(4)
2.75	441.2	-11.0	-434.6	-4.4(4)
2.5	416.3	-12.4	-408.3	-4.4(4)
2.25	391.8	-14.3	-381.8	-4.3(4)
2	367.9	-16.4	-355.8	-4.2(4)
1.5	322.1	-22.2	-304.2	-4.3(3)
1	278.6	-30.0	-253.1	-4.5(3)
0.5	238.1	-38.6	-203.6	-4.1(2)
0	200.6	-48.4	-157.3	-5.1(2)

Fig. S5 below shows a plot of  $\langle U_{\text{solute-solvent}}(\alpha)/\alpha \rangle_{\alpha, NpT}$  for three series of charge ratio  $\alpha$ . The average sum via Trapezoid rule gives a charging free energy of  $-57.5(4)$  kJ/mol (trapezoid rule).

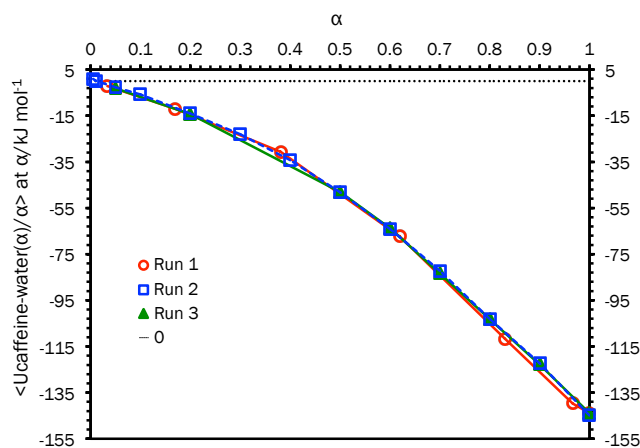


FIG S12 Charging free energy of caffeine in water:  $\langle U_{\text{solute-solvent}}(\alpha)/\alpha \rangle_{\alpha}$  for three sets of  $\alpha$ .

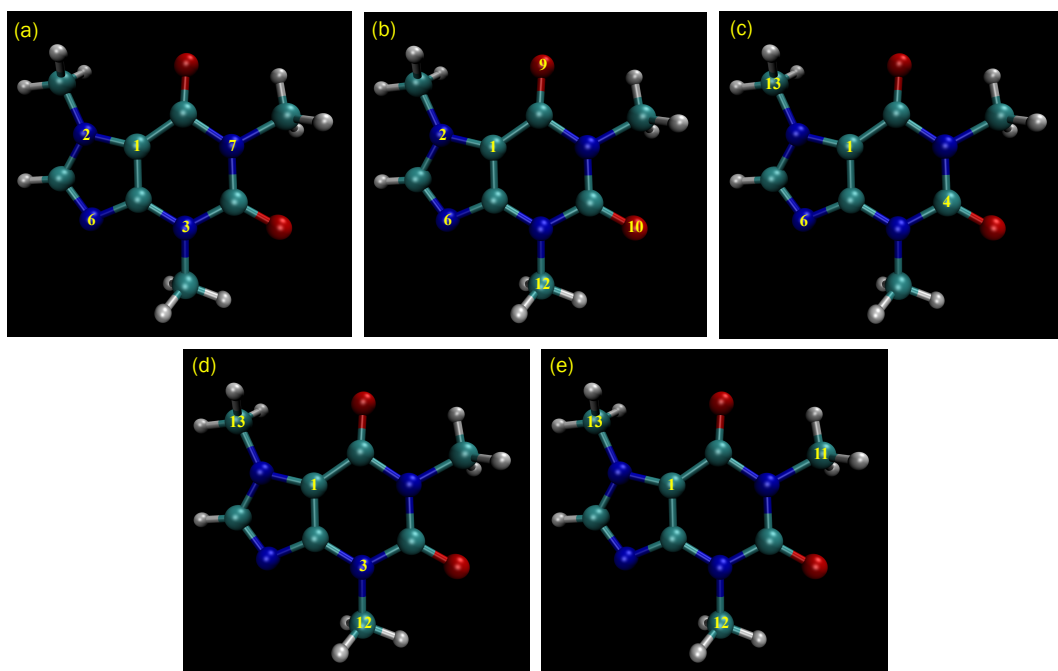


FIG S13 Einstein arrangements for computing the crystal free energies  $A_{\text{solid}}$  of  $\beta$ -caffeine: (a) A; (b) B; (c) C; (d) D; (e) E. The cyan atoms are carbon, the white atoms are hydrogen, the blue atoms are nitrogen and the red atoms are oxygen. Carbon atom 1 has been consistently used as the “Central Atom”, while the other labeled atoms are “Orientational Atoms” for the corresponding Einstein arrangement respectively. Methyl groups are freely rotating in all solid simulations.

TABLE S22 Computed  $A_{\text{solid}}$  of  $\beta$ -caffeine with different Einstein arrangements (EA).  $A_{\text{solid}}$  is in kJ/mol and  $K_{\text{max}}$  is in kJ/mol/Å<sup>2</sup>.

EA	$K_{\text{max}}$	$A_{\text{caffeine}}/N_{\text{caffeine}}$
A	10460	-103.32(8)
A	20920	-103.16(8)
B	10460	-103.73(8)
B	20920	-103.49(8)
C	10460	-104.14(8)
D	10460	-102.41(8)
E	10460	-100.57(8)

## REFERENCES

- <sup>1</sup>L. Li, T. Totton and D. Frenkel, submitted to J. Chem. Phys.  
*Computational Methodology for Solubility Prediction: Application to Simple Aromatic Crystals*
- <sup>2</sup>G. Hummer, L. R. Pratt and A. E. García, J. Phys. Chem. **100**, 1206 (1996).  
*Free Energy of Ionic Hydration*
- <sup>3</sup>F. Figueirido, G. S. Del Buono, and R. M. Levy, J. Phys. Chem. B **101**, 5622 (1997).  
*On Finite-Size Corrections to the Free Energy of Ionic Hydration*
- <sup>4</sup>Y. Wu, H. L. Tepper and G. A. Voth, J. Chem. Phys. **124**, 024503 (2006).  
*Flexible Simple Point-charge Water Model with Improved Liquid State Properties.*
- <sup>5</sup>Y. Wu, H. Chen, F. Wang, F. Paesani, G. A. Voth, J. Phys. Chem. B **112**, 467 (2008).  
*An Improved Multistate Empirical Valence Bond Model for Aqueous Proton Solvation and Transport*
- <sup>6</sup>T. Darden, D. York and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).  
*Particle mesh Ewald: An  $N \cdot \log(N)$  Method for Ewald Sums in Large Systems*

<sup>7</sup>R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles* (McGraw-Hill, New York, 1989).

<sup>8</sup>P. Raiteri, R. Demichelis and J. D. Gale, *J. Phys. Chem. C* **119**, 24447 (2015).

*A Thermodynamically Consistent Force Field for Molecular Dynamics Simulations of Alkaline-Earth Carbonates and Their Aqueous Speciation*

<sup>9</sup>G. Raabe and R. J. Sadus, *J. Chem. Phys.* **134**, 234501 (2011).

*Molecular Dynamics Simulation of the Dielectric Constant of Water: The Effect of Bond Flexibility*

<sup>10</sup>H. Sitepu, B. H. O'Connor and D. Li, *J. Appl. Cryst.* **38**, 158 (2005).

*Comparative Evaluation of the March and Generalized Spherical Harmonic Preferred Orientation Models Using X-ray Diffraction Data for Molybdite and Calcite Powders*