DOI: 10.1002/ange.201403541

Solubility Determination

Predicting the Relative Solubilities of Racemic and Enantiopure Crystals by Density-Functional Theory**

*Alberto Otero-de-la-Roza, Blessing Huynh Cao, Ivy K. Price, Jason E. Hein, and Erin R. Johnson**

Abstract: *Isolation of chiral molecules as pure enantiomers remains a fundamental challenge in chemical research. Enantioselective enrichment through preferential crystallization is an efficient method to achieve enantiopure compounds, but its applicability depends on the relative stability of the enantiopure and racemic crystal forms. Using a simple thermodynamic model and first-principles density-functional calculations, it is possible to predict the difference in solubility between the enantiopure and racemic solid phases. This approach uses dispersion-corrected density functionals and is capable of accurately predicting the solution-phase entantiomeric excess to within about 10% of experimental measurements on average. The accuracy of the exchange-hole dipole moment (XDM) model of dispersion enables the viability of the proposed method.*

Chirality has fascinated chemists ever since Pasteur explained the phenomenon at a molecular level by completing the first racemic resolution in 1848.[1] The efficient resolution of chiral mixtures into pure enantiomers has great relevance both in fundamental chemistry (for example, explaining the origin of homochirality in biological molecules[2–4]) as well as in the pharmaceutical and agrochemical industries.[5,6] Ideally, employing asymmetric catalysis can deliver the products in enantiopure form; however, scalemic mixtures are more commonly obtained.[36] One of the most efficient methods of accomplishing enantiopurification of a scalemic mixture is preferential crystallization.[7] However, only a small fraction of chiral organic molecules can be separated using this technique[8] because a conglomerate of enantiopure crystals must form and be more stable than the corresponding racemic solid.

Conglomerate-forming compounds are relatively rare, as most molecules (ca. 90% of chiral crystals[8]) form a racemic crystalline solid (the racemate phase) made up of both enantiomers. This acute but subtle difference between compounds forming enantiopure and racemate phases is best illustrated by examining the ternary phase diagrams for the equilibria between the crystalline and solution-phase enantiomers, shown in Figure 1. Conglomerates (Figure 1a) have ternary phase diagrams involving solely enantiopure crystals and a solution phase. In contrast, samples forming a racemic compound (Figure 1b) have an additional racemic solid phase present.

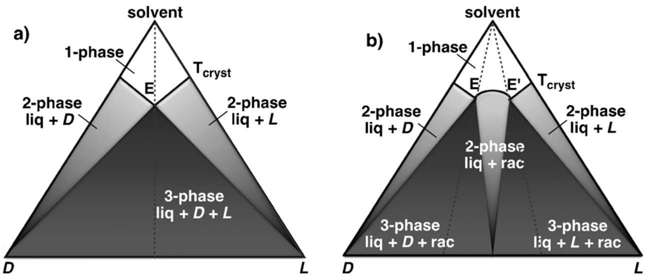

Figure 1. Representation of the ternary phase diagram of a chiral compound for conglomerate-forming (a) and racemate (b) crystals.

While classical preferential separation requires a stable enantiopure phase, new methods have exploited the difference in solubility of the enantiopure and racemic phases to achieve enantiopurification[9,10] by crystallization of the racemate, leaving an enantiopure solution. In either case, the optimization of crystallization techniques to effect separation of enantiomers requires knowledge of solubility equilibria for both the enantiopure and racemic solid phases. The relative solubility difference can be determined experimentally by measuring the enatiomeric excess (*ee*) in the saturated solution. This is possible because the solubility of both the D and L enantiopure crystals are equal; at equilibrium the solution phase of conglomerate-forming compounds will always contain equal concentrations of both enantiomers ($ee = 0$). Alternatively, saturated solutions in contact with both racemic and enantiopure crystals will contain an excess of one enantiomer because the racemic and enantiopure crystals have different solubilities. Therefore, measurement of the solution-phase enantiomeric excess allows researchers to categorize compounds in terms of the tie line shown on the phase diagram in Figure 1b.

While this measurement appears straightforward, accurate *ee* measurement requires an appropriate analytical technique, such as chiral stationary-phase HPLC, NMR in a chiral solvent, or polarimetry. Furthermore, sufficient

---

[*] Dr. A. Otero-de-la-Roza
National Institute for Nanotechnology, National Research Council of Canada
11421 Saskatchewan Drive, Edmonton, Alberta T6G 2 M9 (Canada)
B. H. Cao, I. K. Price, Prof. Dr. J. E. Hein, Prof. Dr. E. R. Johnson
Chemistry and Chemical Biology, School of Natural Sciences, University of California, Merced
5200 North Lake Road, Merced, CA 95343 (USA)
E-mail: ejohnson29@ucmerced.edu

[**] We thank the National Science Foundation, award number CHE-1300686 and the Spanish Malta/Consolider initiative (no. CSD2007-00045).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201403541.

samples of both the racemic and enantiopure crystal phases must be obtained to create saturated solutions at equilibrium. Thus the limitation of experimental analytics, coupled with the need for pure samples, can cause serious limitations. It would thus be extremely valuable if theory could be used to predict the relative solubility between the two crystalline forms of target molecules. Because the energies of the two crystalline lattices determine their relative solubilities, we thus need to be able to accurately calculate the small energy difference between the enantiopure and racemic crystals to predict the solution-phase *ee*.

Until recently, density-functional theory (DFT) methods could not provide accurate treatment of intermolecular interactions, owing to neglect of London dispersion,[11] the force responsible for long-range attractive interactions between non-polar chemical species.[12] There are now several successful DFT-based dispersion approaches in common use,[13–18] including the exchange–hole dipole moment (XDM) dispersion model of Becke and Johnson.[19,20] The XDM method has been shown to double the accuracy of other dispersion corrections for molecular crystals[21] giving heats of sublimation to within about 1 kcal mol^{-1} on average. In this work, we apply the XDM method to prediction of amino acid chiral eutectic points (points E and E′ in Figure 1), which is an even more sensitive test of theory owing to the very small inherent energy differences. We show that XDM predicts energy differences with an accuracy much better than 1 kcal mol^{-1}, probably thanks to error cancellation.

The *ee* in solution can be predicted by considering the equilibria at the eutectic point (E and E′ in Figure 1 b). We assume an excess of the L enantiomer, corresponding to the right side of the racemate phase diagram. At the eutectic point, E′, three phases coexist in equilibrium: a saturated solution, enantiopure solid $\mathrm{L}_{(s)}$, and racemate solid, $\mathrm{DL}_{(s)}$. For the purpose of our theoretical treatment, we also consider the gas-phase molecules, which at a given temperature and volume are related to the solid and solvated species by the Helmholtz free energies of sublimation (ΔF_{sub}) and of solvation (ΔF_{solv}), respectively.

$$\mathrm{L}_{(s)} \xrightleftharpoons{\Delta F_{\mathrm{sub}}(\mathrm{L})} \mathrm{L}_{(g)} \xrightleftharpoons{\Delta F_{\mathrm{solv}}} \mathrm{L}_{(\mathrm{solv})} \tag{1}$$

$$\mathrm{DL}_{(s)} \xrightleftharpoons{2\Delta F_{\mathrm{sub}}(\mathrm{DL})} \mathrm{D}_{(g)} + \mathrm{L}_{(g)} \xrightleftharpoons{2\Delta F_{\mathrm{solv}}} \mathrm{D}_{(\mathrm{solv})} + \mathrm{L}_{(\mathrm{solv})} \tag{2}$$

Because the solution is saturated, the solubility products are:

$$K_{\mathrm{sp}}(\mathrm{L}) = [\mathrm{L}] = \exp\left(-\frac{\Delta F_{\mathrm{sub}}(\mathrm{L}) + \Delta F_{\mathrm{solv}}}{RT}\right) \tag{3}$$

$$K_{\mathrm{sp}}(\mathrm{DL}) = [\mathrm{D}][\mathrm{L}] = \exp\left(-2\frac{\Delta F_{\mathrm{sub}}(\mathrm{DL}) + \Delta F_{\mathrm{solv}}}{RT}\right) \tag{4}$$

The conglomerate-forming and racemate crystals will have different sublimation energies. However, in an achiral solvent, the two enantiomers have identical solvation energies. This fact is what ultimately enables the accurate calculation of the enantiomeric excess. After some algebra, it can be shown that the *ee* in solution is:

$$ee = \frac{[\mathrm{L}] - [\mathrm{D}]}{[\mathrm{L}] + [\mathrm{D}]} = \frac{\beta^2 - 1}{\beta^2 + 1} \tag{5}$$

with:

$$\begin{aligned}\beta &= \exp\left(-\frac{\Delta F_{\mathrm{sub}}(\mathrm{L}) - \Delta F_{\mathrm{sub}}(\mathrm{DL})}{RT}\right)\\ &= \exp\left(-\frac{E(\mathrm{L}) - E(\mathrm{DL})}{RT}\right)\end{aligned} \tag{6}$$

where E is the electronic energy (coming directly from the DFT calculation) and we assume that the vibrational contributions to the sublimation free energy are equal for both crystals. This assumption introduces an error that is on the order of a few kJ mol^{-1},[22,23] but is standard in the treatment of the related problems of polymorph ranking[11] and crystal structure prediction.[24] Successful application of the methods devised to tackle problems in these fields also rely on accurate molecular crystal energetics.

The accuracy of this simple model can be demonstrated by using available experimental heats of formation. For example, in the case of alanine, ΔH_f^0 is -563.63 kJ mol^{-1} for the DL phase and -561.24 kJ mol^{-1} for the L phase.[25–27] Using this data, the *ee* from Equations (5) and (6) is 74.6 %, to be compared with the experimental value of 60.4 %.[28] Similarly for leucine, ΔH_f^0 is -637.56 kJ mol^{-1} for the enantiopure phase and -640.70 kJ mol^{-1} for the DL phase.[25,27] This gives a calculated *ee* of 85.3 %, to be compared with the experimental value of 87.9 %.[28] Unfortunately, such careful calorimetric measurements of the heats of formation are not frequently available for the L and DL phases at the same time and, in general, we must turn to density-functional theory to predict the relative energetics of the two crystal phases.

If the structures of the enantiopure and racemic crystals are known and the solvent is achiral, it is straightforward to calculate the *ee* at the eutectic by computing the relative energy difference between the two crystals using DFT-XDM and applying Equations (5) and (6). We will show that, because of the exponential in Equation (6), the enantiomeric excess is extremely sensitive to the predicted energy difference between the two crystals and the range of 0–95 % *ee* is spanned by only a 4.5 kJ mol^{-1} energy difference. Therefore, accurate prediction of the *ee* is an extremely challenging test for dispersion-corrected DFT. Previous applications of dispersion-corrected DFT literature[29] have focused solely on determining which of the enantiopure and racemate phases are more stable.

To test our theoretical procedure, we considered a subset of the proteinogenic amino acids, for which solvent-free single-crystal X-ray crystal structures are available.[30] We assume (except where explicitly investigated) that the crystal structure obtained in the preferential crystallization is the same as that reported in the database. For all of the crystals, the cell geometries and atomic positions were relaxed using the B86b exchange functional[31] with PBE correlation[32] and XDM dispersion. The vibrational effects on the crystal geometries are neglected. Calculations were performed with the quantum ESPRESSO program[33,34] (available under a free license; XDM is implemented in the development version) using the projector augmented wave (PAW) method[35] and the plane-waves/pseudopotentials approach. All crystals were calculated using a $4\times4\times4$ **k**-point grid and a 80 Ry cutoff energy, as in previous work.[21]

The eutectic points for a series of amino acids were measured experimentally to provide reference data to examine the predictive capability of the XDM method. For each amino acid, three scalemic mixtures consisting of different initial enantiomeric excess (*ee*) were created by either mixing pure enantiomers (D + L amino acids) or a single enantiomer with a racemic material (L + rac amino acid). Each sample was treated with a small volume of solvent to produce saturated suspensions, which were stirred at 25 °C overnight. Aliquots of the suspension were withdrawn and filtered by centrifugation. The enantiomeric excess of the solution phase was measured by chiral-phase HPLC-MS on the underivatized amino acid samples directly from the filtration. This aliquoting process was repeated daily over four days to ensure the sample had reached equilibrium. The reported eutectic point is the average of the solution-phase *ee* for the three scalemic mixtures for each amino acid.

Theoretical and experimental *ee* and energies are collected in Table 1. Using the XDM dispersion model, the solution-phase chiral eutectics are predicted to be in very good agreement with the experimental data for all of amino acids except proline. The mean absolute error of the calculated *ee* results is 9.9 % relative to experiment.

Table 1: Solution-phase enantiomeric excess of selected amino acids obtained with the B86bPBE-XDM method and compared with experimental data.

Amino acid	DFT ΔE[c]	DFT *ee*	Expt. *ee*
serine	17.221	100.0	100.0[a]
histidine	4.209	93.5	93.7[a]
leucine	3.971	92.2	87.9[a]
alanine	2.017	67.1	60.4[a]
cysteine	2.113	69.2	58.4[b]
tyrosine	2.180	70.6	51.7[b]
valine	1.807	62.3	44.1[b]
proline	−0.791	0.0	44.4[b]
aspartic acid	−1.996	0.0	0.7[b]
glutamic acid	−8.665	0.0	0.7[b]

[a] Experimental data from previous work.[28] [b] Experimental data from the present work. [c] The calculated energy difference per molecule between the enantiopure and racemic crystals [kJ mol^{-1}].

While the predicted eutectic points for most amino acids investigated are in good agreement with the observed values, two cases stand out and merit further discussion. First, DFT-XDM predicts that the enantiopure crystal phase of proline is more stable (zero *ee*). Initially, we hypothesized that a solvate crystal may have been forming in situ,[28] but the X-ray diffraction powder pattern of the solids isolated from the DMSO solution matched the reported values for the racemic and enantiopure crystals. The probable cause for the discrepancy is that the enantiomeric excess is extremely sensitive to the predicted energy difference between the two crystals in that particular region, as shown in Figure 2. The errors in the calculated *ee* values increase for the racemates going down the table and are largest for the steepest portion of the concentration versus energy curve. Thus, while the proline result is in error by 40 %, this represents an energy error of only 1.84 kJ mol^{-1}, demonstrating that XDM is actually giving a relatively accurate energy difference between the enantiopure and racemic crystal phases.

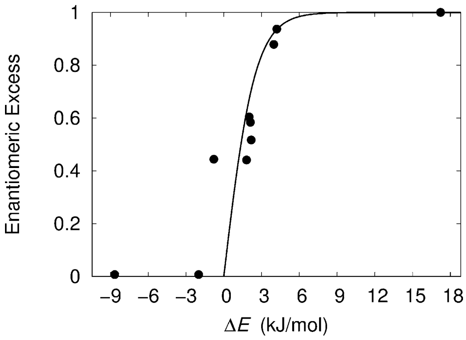

Figure 2. Dependence of the enantiomeric excess on the energy difference between racemate and enantiopure phases ($\Delta E = E(\text{L}) - E(\text{DL})$) according to our model (black line). The experimental *ee* graphed against our DFT energy differences are shown as data points.

The second case where there was a large discrepancy between theory and experiment is isoleucine, where DFT calculations predicted an *ee* value of 93.2 % compared with the experimental value of 51.7 %. Careful investigation by HPLC-MS (the chromatogram is shown in the Supporting Information) revealed that our sample contained both the natural isoleucine as well as a minor amount of the allo-isoleuceine diastereomer, which can co-crystallize with the isoleucine enantiomers. The solution phase enantiomeric excess does not represent the true relative solubility, as the presence of the minor diastereomer alters the solid-solution equilibrium. Thus, the theoretical and experimental data should not be directly compared (and the isoleucine point is not shown in Figure 2) since the effect of the allo diastereomer cannot be captured by the calculations. This example illustrates how theory can help direct experiment by identifying situations where the equilibrium is complicated by factors for which our simple two-state thermodynamic model does not account.

In summary, we have shown that a simple theoretical model combined with state-of-the-art dispersion-corrected density-functional calculations can be used to predict, using exclusively the knowledge of the crystal structures, the enantiomeric excess at the chiral eutectic of a scalemic solution in contact with the enantiopure and racemate crystal phases. This work validates the use of DFT-XDM to guide the choice of compounds for enantiomeric enrichment by selective crystallization.[9,10] Furthermore, the calculation of the enantiomeric excess at the chiral eutectic point is shown to be an exquisitely sensitive test of theoretical methods for intermolecular interactions in the solid phase. The XDM method achieves good agreement with experiment (to within 10 % on average), despite the highly non-linear energetic dependence of the enantiomeric excess.

Received: March 20, 2014
Revised: April 16, 2014

Keywords: density functional calculations · enantiomeric excess · racemic mixtures · solubility

[1] L. C. R. Pasteur, *C. R. Hebd. Seances Acad. Sci.* **1848**, *26*, 535–538.
[2] J. E. Hein, D. Gherase, D. G. Blackmond, *Top. Curr. Chem.* **2012**, *333*, 83–108.
[3] D. G. Blackmond, *Cold Spring Harbor Perspect. Biol.* **2010**, *2*, a002147.
[4] P. Herdewijn, M. V. Kisaku, *Origin of Life: Chemical Approach*, Helvetica Chimica Acta, Wiley-VCH, **2008**.
[5] R. C. Wende, P. Schreiner, *Green Chem.* **2012**, *14*, 1821–1849.
[6] M. Kaspereit, S. Swernath, A. Kienle, *Org. Process Res. Dev.* **2012**, *16*, 353–363.
[7] G. Coquerel, *Top. Curr. Chem.* **2006**, *269*, 1–51.
[8] G. Levilain, G. Coquerel, *CrystEngComm* **2010**, *12*, 1983–1992.
[9] J. E. Hein, B. Huynh Cao, C. Viedma, R. M. Kellogg, D. G. Blackmond, *J. Am. Chem. Soc.* **2012**, *134*, 12629–12636.
[10] J. E. Hein, B. Huynh Cao, M. van der Meijden, M. Leeman, R. M. Kellogg, *Org. Process Res. Dev.* **2013**, *17*, 946–950.
[11] C. Rovira, J. J. Novoa, *J. Chem. Phys.* **2000**, *113*, 9208–9216.
[12] F. London, *Z. Phys.* **1930**, *63*, 245–279.
[13] S. Grimme, *J. Comput. Chem.* **2006**, *27*, 1787–1799.
[14] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
[15] A. Tkatchenko, M. Scheffler, *Phys. Rev. Lett.* **2009**, *102*, 73005.
[16] K. Lee, E. D. Murray, L. Kong, B. I. Lundqvist, D. C. Langreth, *Phys. Rev. B* **2010**, *82*, 081101.
[17] O. A. Vydrov, T. V. Voorhis, *J. Chem. Phys.* **2010**, *133*, 244103.
[18] E. Torres, G. A. DiLabio, *J. Phys. Chem. Lett.* **2012**, *3*, 1738–1744.
[19] A. D. Becke, E. R. Johnson, *J. Chem. Phys.* **2007**, *127*, 154108.
[20] A. Otero-de-la-Roza, E. R. Johnson, *J. Chem. Phys.* **2013**, *138*, 204109.
[21] A. Otero-de-la-Roza, E. R. Johnson, *J. Chem. Phys.* **2012**, *137*, 054103.
[22] A. J. Stone, *The Theory of Intermolecular Forces*, International Series of Monographs on Chemistry, Clarendon, Oxford, **1997**.
[23] A. Gavezzotti, G. Filippini, *J. Am. Chem. Soc.* **1995**, *117*, 12299–12305.
[24] D. A. Bardwell, et al., *Acta Crystallogr. Sect. B* **2011**, *67*, 535–551.
[25] H. Y. Afeefy, J. F. Liebman, S. E. Stein, *NIST Standard Reference Database, NIST Chemistry WebBook*, *Vol. 69*, National Institute of Standards and Technology, Gaithersburg MD, 20899, **2014**, http://webbook.nist.gov.
[26] H. M. Huffman, E. L. Ellis, *J. Am. Chem. Soc.* **1936**, *58*, 1728–1733.
[27] H. M. Huffman, E. L. Ellis, *J. Am. Chem. Soc.* **1937**, *59*, 2150–2152.
[28] M. Klussmann, A. J. P. White, A. Armstrong, D. G. Blackmond, *Angew. Chem.* **2006**, *118*, 8153–8157; *Angew. Chem. Int. Ed.* **2006**, *45*, 7985–7989.
[29] J. Kendrick, M. D. Gourlay, M. A. Neumann, F. J. J. Leusen, *CrystEngComm* **2009**, *11*, 2391–2399.
[30] F. H. Allen, *Acta Crystallogr. Sect. B* **2002**, *58*, 380–388.
[31] A. D. Becke, *J. Chem. Phys.* **1986**, *85*, 7184.
[32] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
[33] P. Giannozzi, et al., *J. Phys. Condens. Matter* **2009**, *21*, 395502.
[34] A. Otero-de-la-Roza, E. R. Johnson, *J. Chem. Phys.* **2012**, *136*, 174109.
[35] P. E. Blöchl, *Phys. Rev. B* **1994**, *50*, 17953.
[36] The term scalemic refers to any non-racemic mixture of enantiomers, and it is generally used to describe non-enantiopure samples (that is, $0 < ee < 100$).