# Polymorph-Specific Solubility Prediction using Constant Chemical Potential Molecular Dynamics

Neha[†], Manya Aggarwal[†], Aashutosh Soni, and Tarak Karmakar[*]

*Department of Chemistry, Indian Institute of Technology, Delhi, 110016 New Delhi, India*

E-mail: tkarmakar@chemistry.iitd.ac.in

Phone: +91 11 2654 8549

**Abstract**

Molecular Dynamics (MD) simulations offer a robust approach to understanding material properties within a system. Solubility is defined as the analytical composition of a saturated solution expressed as a proportion of designated solute in a designated solvent, according to IUPAC. It is a critical property of compounds and holds significance across numerous fields. Various computational techniques have been explored for determining solubility, including methods based on chemical potential determination, enhanced sampling simulation, and direct coexistence simulation, and lately, machine learning-based methods have shown promise. In this investigation, we aim to find the solubility values of a compound through Constant Chemical Potential Molecular Dynamics, a method rooted in direct coexistence simulation. The primary purpose of using this method is to overcome the limitation of the direct simulation method by maintaining a constant chemical potential for a sufficiently long time. Urea is chosen as a prototypical system for our study, with a particular focus on three of its polymorphs. Our approach effectively discriminates between the polymorphs of urea based on their respective solubility values; polymorph III is found to have the highest solubility followed by form IV and I.

# Introduction

Solubility is a sought-after physicochemical property that influences the behavior of substances in various physical and chemical environments. Solubility finds applications in many fields ranging from the chemical industry, where the difference between the solubility of product and impurities is exploited to purify the product, to the geochemical sciences where solubility is studied to investigate the contamination of soils by inorganic salts. Moreover, solubility has a pronounced effect on the pharmaceutical industry as it regulates the bioavailability of drugs in biological systems. Researchers have long been interested in determining the solubility of active pharmaceutical ingredients (API) and sought ways to improve their solubility through the reduction in particle size, the addition of solubilizing excipients, salt formation etc.[1–3] Most of the time, the solubility of drugs is predicted using experiments[4] which is a time-consuming and tedious task for some systems, for instance, moisture-free conditions need to be maintained for determining the solubility of a hygroscopic salt.[5] Molecular Dynamics (MD) simulations are a potential tool to predict solubility and thereby, help optimize time and experimental cost.

MD simulations[6,7] propose a simple yet powerful computational method for studying the behavior of systems at the atomistic level. Although the method has been around for half a century, the use of MD simulations is not popular for determining solubility. Ferrario *et al's*[8] study was the first of this kind in which they predicted the aqueous solubility of KF using MD simulations. In this pioneering study, they computed the chemical potential of the solute in solid and solution phase at different solution concentrations and the difference in chemical potential at each concentration was used to predict the solubility. Although, the density of the solution did not show a convergence with the experimental density at higher concentrations, a surprisingly reasonable agreement between the estimated and experimental solubility was obtained. Since then, plenty of work using different MD simulation approaches has been reported.[9–15] Direct Coexistence Simulation (DCS)[16] is one such method in which the crystal surface is exposed to its saturated solution and allowed to attain equilibrium. The

concentration of solution at equilibrium serves as the solubility, it is a brute-force method that gives a range of solubility, often close to the experimental solubility but it suffers from finite-size effect and also the longer time scales required for equilibration[17] pose a computational limitation. In a 2015 study, Manzanilla-Granados *et al.*[18] carried out the DCS using NaCl as the test case and got a notable correspondence with the experimental value given that the simulations ran for several microseconds. Carruthers *et al.* computed the upper and lower bounds of solubility of urea using two force field models[19] developed by Özpınar[20] and Hölzl[21] using DCS. Despite the long simulations of $\sim$1 $\mu$s the predicted aqueous solubility values were 2-6 times lower than the experimental values. Assuming the accuracy of the force field, this discrepancy is attributed to the finite-size effect present in DCS. To overcome the finite-size effect operational in DCS, we utilize the Constant Chemical Potential Molecular Dynamics (C$\mu$MD) approach developed by Perego *et al.*[22] to predict the solubility.

In this work, we propose a way to study the solubility of different polymorphs of a molecular crystal using C$\mu$MD simulation drawing upon urea as a prototypical system. Urea is taken as an example because it shows a relatively fast crystal growth, and its growth and dissolution have been widely studied using MD simulations.[10,23–26] To the best of our knowledge, so far, all studies related to urea crystal growth and dissolution are done on polymorph I. Urea shows a high degree of polymorphism and is known to exist in up to five polymorphic forms, among which Form I (tetragonal, space group $P42_1m$) is the most stable at ambient conditions followed by Form III (orthorhombic, $P2_12_12_1$), and Form IV (orthorhombic, $P2_12_12$) which are obtained at high pressures.[27–29] The crystallographic structures of polymorphic forms II and V have not been reported yet. In this work, we restricted ourselves to studying forms I, III, and IV, because of the obvious reason for the availability of the crystal structures. Our results unveil that the polymorph III and IV have a greater solubility value than polymorph I by a factor of three and two respectively. This trend in values can be understood by acknowledging the organization of molecules in these polymorphs, forms III and IV have less symmetry as compared to form I. In the latter, a square-shaped configu-

ration between four hydrogen-bonded molecules is realized, which is disrupted in the other two forms, causing the molecules to move in the solution phase easily.

# Computational Details

## Force Field

The success of MD simulations in predicting the material properties of the system depends largely on the force field used. To model the interactions of urea in crystal and solution, we have used Generalized AMBER Force Field (GAFF)[30] along with the TIP3P water model. Antechamber tool[31] was used to obtain the force field parameters, and the charges were assigned using the AM1-BCC charge model.

## Simulation System setup

To study the dynamics at the interface, the (001) face of crystalline urea was exposed to solution in the three polymorphs. The given face is known to grow fast in the case of polymorph I,[24] hence the choice. The unit cell information was obtained from the Cambridge Structural Database (CSD) with the following deposition numbers 731958 (polymorph I), 731959 (polymorph III), and 731961 (polymorph IV).[28] The first step in preparing the system is to check the validity of the force field. To accomplish this a $2 \times 2 \times 2$ supercell was built by translating the unit cell in three dimensions using AVOGADRO[32] software. Then the supercell was minimized using the steepest descent algorithm. After that, the minimized configuration was equilibrated according to the NVT ensemble at 300K followed by NPT equilibration at the same temperature and 1 bar pressure for 10 ns. Velocity-rescale thermostat with a temperature coupling constant of 0.5 ps and Berendsen barostat with a pressure coupling constant of 10 ps were used during simulations. After this, the lattice cell parameters obtained from NPT simulation were compared with the experimental values. The lattice cell parameters obtained from NPT simulation were found to be in close agreement with the

experimental values, the lattice cell lengths deviated by 1.25%, 9.10%, 2.26% on average for polymorphic forms I, III, and IV respectively, and the lattice cell angles showed insignificant deviation.[33] The supercell was again replicated in 3D using the genconf utility, which resulted in a crystal slab of dimensions 2.22×2.22×1.89. The simulation box was extended along the z direction, thus generating a simulation box of size 2.22×2.22×18.16 approximately. The insert-molecules feature of GROMACS was used to evenly fill the simulation box with water and urea molecules according to different concentrations, chosen around the experimental solubility of polymorph I.[34] The above steps were followed for forms III and IV to obtain systems which were then simulated further.

## Constant Chemical Potential Molecular Dynamics(C$\mu$MD) Simulation

All simulations were carried out using the GROMACS[35] molecular dynamics simulation package version 2021.4 patched with Plumed[36] version 2.8.0. Each system was first minimized using the steepest descent algorithm until the force on each atom was at most 100 kJ/mol or when 50000 steps were reached. After energy minimization, the system was equilibrated according to the NVT ensemble at 300 K followed by NPT equilibration for 100 and 500 ps respectively. The NPT simulation was carried out in the isothermic-isobaric ensemble at 300 K and 1 bar. The time step was 1 fs and all bond lengths were constrained using the LINCS algorithm. The long-range electrostatic interactions were calculated using the Particle Mesh Ewald (PME) method with a Fourier spacing of 0.16 nm, a PME order of 4, and an Ewald tolerance of $1.0 \times 10^{-6}$. The van der Waals and real space components of the Coulombic interactions were truncated at 0.80 nm, with the van der Waals effect being moderated by a potential-switch function. The pressure was regulated using the Berendson barostat during equilibration with a compressibility of $4.5 \times 10^{-5}$ bar$^{-1}$. The first crystal layer at the interface was seen to dissolve during the equilibration of the system. Further, NPT simulations were carried out by maintaining a constant concentration of the solution

around the interface of the crystal slab in solution, a transition region of 1.5 nm, a control region of 3.0 nm, and a force region of 0.5 nm with the rest of the simulation box used as reservoir[22] was implemented using plumed driver package, these distances are calculated from the solid-solution interface on the fly. Parrinello-Rahman barostat with a pressure coupling constant of 10 ps was used during the NPT production run. The force constant was regulated to maintain the desirable concentration in the control region. Figure 1 depicts the schematic of a typical C$\mu$MD simulation setup. The number density of urea in the control region was plotted against time to see if the target number density was maintained during the simulation, the duration until which the concentration is maintained is of interest and is referred to as the stability time. Table 1 summarises the composition of the simulation box along with the stability time of C$\mu$MD simulations.

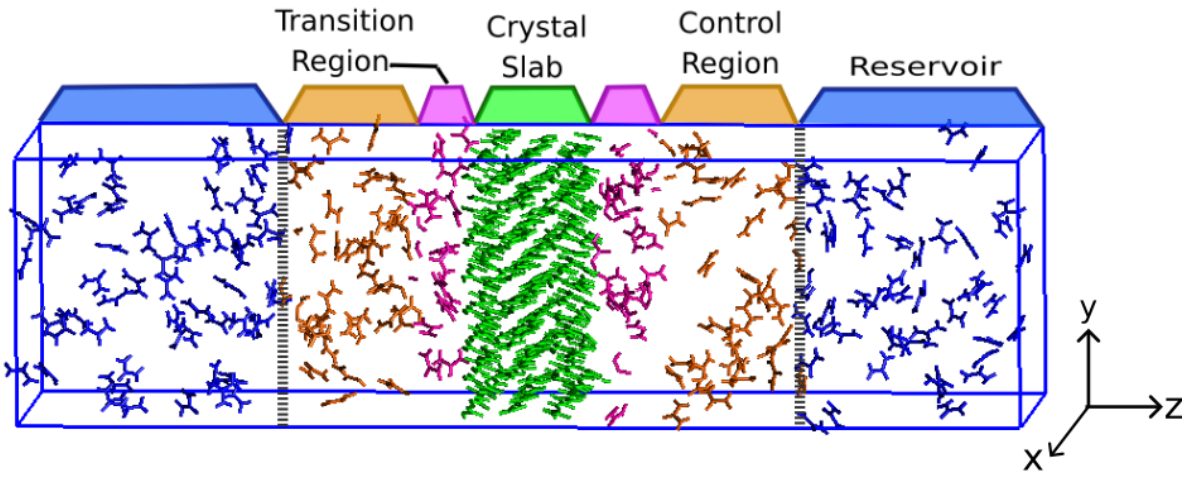



Figure 1: Schematic of C$\mu$MD simulation box having urea crystal in the middle of the box and urea solution on both sides. Different regions, as described in the text, are defined from the solid-solution interface. Urea molecules are colored to demarcate the region to which they belong, the grey-striped bands represent the force region. Water molecules are not shown for better visualization.

Table 1: Simulation box dimensions, the composition of the system (Number of urea molecules in crystal ($N_u^c$), in solution ($N_u^s$), number of water molecules ($N_w$)) and stability time for C$\mu$MD simulations at different concentrations of urea solution for the three polymorphs.

| | Polymorph I | | | | Polymorph III | | | | Polymorph IV | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Concentration(g/L) | 545 | 400 | 300 | 200 | 1100 | 1060 | 1040 | 600 | 700 | 600 | 500 | 240 |
| Simulation box dimension (a×b×c) | 2.22× 2.22× 18.16 | 2.22× 2.22× 18.16 | 2.22× 2.22× 18.16 | 2.22× 2.22× 18.16 | 2.04× 3.17× 18.27 | 2.04× 3.17× 18.27 | 2.04× 3.17× 18.27 | 2.05× 3.18× 20.95 | 2.84× 3.07× 14.58 | 2.84× 3.07× 14.58 | 2.84× 3.07× 14.58 | 2.84× 3.07× 10.66 |
| $N_u^c$ | 128 | 128 | 128 | 128 | 384 | 384 | 384 | 384 | 256 | 256 | 256 | 256 |
| $N_u^s$ | 328 | 328 | 328 | 328 | 1100 | 1100 | 1100 | 1000 | 632 | 632 | 632 | 150 |
| $N_w$ | 1954 | 1954 | 1954 | 1954 | 514 | 514 | 514 | 1314 | 2199 | 2199 | 2199 | 2193 |
| Stability Time(ns) | 60 | 250 | 1000* | 108 | 490* | 1400* | 1000 | 1500 | 310 | 1200* | 1100* | 1000* |

* denotes the time of simulation in addition to stability time

# Results and Discussion

The simulations were extended to several microseconds and monitored closely. To get a clear picture of dynamics at the interface of the crystal slab in solution, the number of crystalline molecules ($N_c$) were plotted as a function of time.[24] The degree of crystallinity, $\Gamma_i(t)$ was measured as the product of two functions describing the local density, $\rho(n_i(t))$ and local order, $\phi_i(n_i(t), \vec{\theta}_i(t))$ around a given molecule. Local density is calculated by summing the number of molecules within a radial distance, $r$, around a given molecule, and local orientation is dependent on the angle between two molecules along a bond vector, present within radius $r$. $N_c$ is not necessarily an integer, since it is calculated using the contribution of local order and local density for a given molecule, a switching function having the Gaussian form is used to assign the contribution. A radial cutoff ($r$) of 0.6 nm was chosen and optimum local density ($\bar{n}$) was found to be 9, 17, and 16 for polymorphic forms I, III, and IV respectively. The angles between the -CO bond vector of two nearby crystal molecules were considered to find local orientation in the case of forms I and IV while, the vector along two N atoms

was examined for form III, as shown in figure 2.

$$\rho(n_i(t)) = \left\{ \begin{array}{ll} \exp\left(-\frac{(n_i(t)-\bar{n})^2}{2\sigma_n^2}\right) & , n_i(t) \leq \bar{n} \\ 1 & , n_i(t) \geq \bar{n} \end{array} \right\} \tag{1}$$

$$\phi_i(n_i(t), \vec{\theta}_i(t)) = \frac{1}{n_i(t)} \sum_{j=1}^{n_i(t)} \left\{ \exp\left(-\frac{(\theta_{i,j}(t) - \bar{\theta_0})^2}{2\sigma_\theta(\vec{\theta_0})^2}\right) + \exp\left(-\frac{(\theta_{i,j}(t) - \bar{\theta_1})^2}{2\sigma_\theta(\vec{\theta_1})^2}\right) \right\} \tag{2}$$

$$\Gamma_i(t) = \rho(n_i(t))\phi_i(n_i(t), \vec{\theta}_i(t)) \tag{3}$$

where $n_i(t)$ is the number of neighbors of the $i^{th}$ molecule, $\sigma_n$ is the standard deviation of the Gaussian distribution of $\bar{n}$ in the solid; $\vec{\theta}_i$ is the vector of the $n_i$ mutual orientations $\theta_{i,j}$ between the internal vectors associated with the $i^{th}$ molecule and its $j^{th}$ neighbor. $\Gamma_i(t)$ values were computed for all the urea molecules in the system. Then the number of molecules belonging to the crystalline phase at each time step t, can be obtained as:

$$N_c(t) = \sum_{i=1}^{N_{tot}} \Gamma_i(t) \tag{4}$$

where $N_{tot}$ is the total number of urea molecules in the simulation box. Table 2 contains the values of these parameters for the three polymorphs.

Table 2: Parameters used for the calculation of $N_c$ in urea polymorphs

| parameter | Polymorph I | Polymorph III | Polymorph IV |
|---|---|---|---|
| cutoff | 0.6 nm | 0.6 nm | 0.6 nm |
| $\bar{n}$ | 9 | 17 | 16 |
| $\sigma_n$ | 2.2 | 7 | 9 |
| $\vec{\theta}_0$ | 14.33° | 11.5° | 11.57° |
| $\vec{\theta}_1$ | 120.4° | 171.9° | 171.9° |
| $\sigma_\theta(\vec{\theta}_0)$ | 13.7° | 14.3° | 22.9° |
| $\sigma_\theta(\vec{\theta}_1)$ | 27.5° | 8.0° | 22.9° |

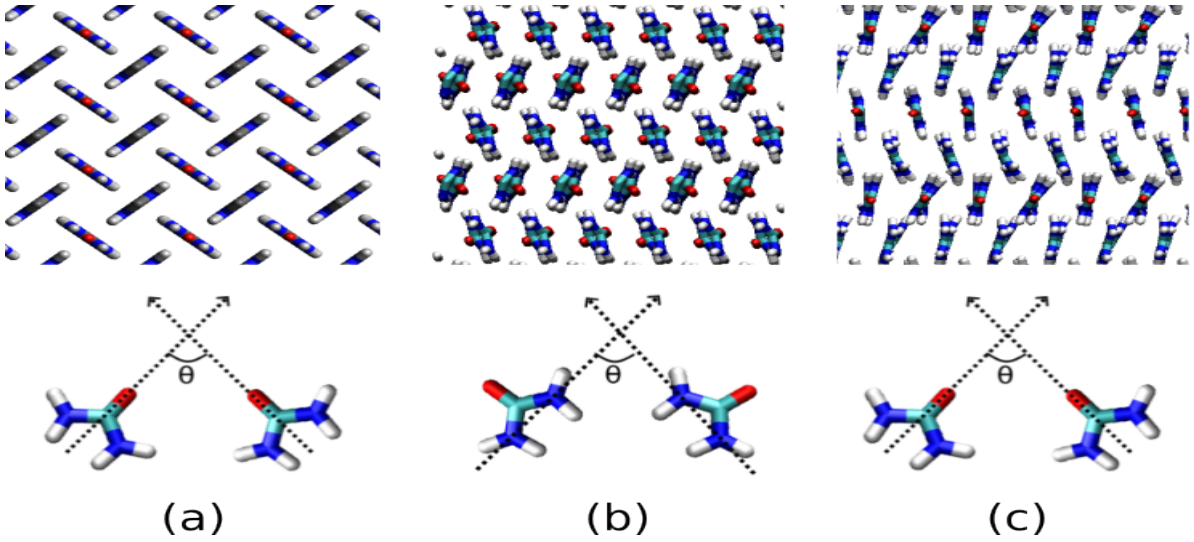
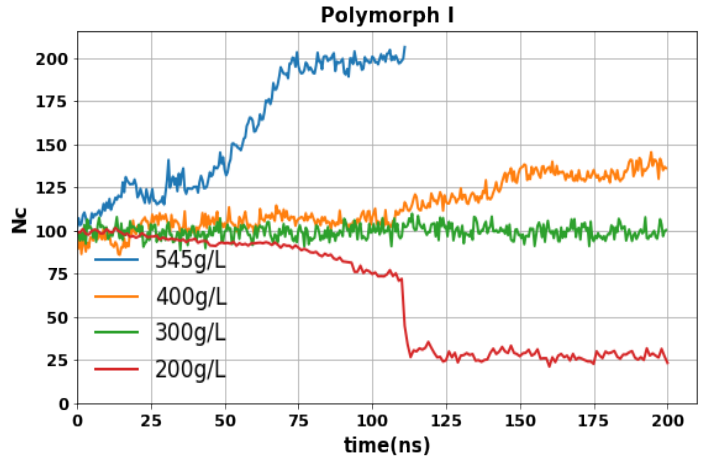


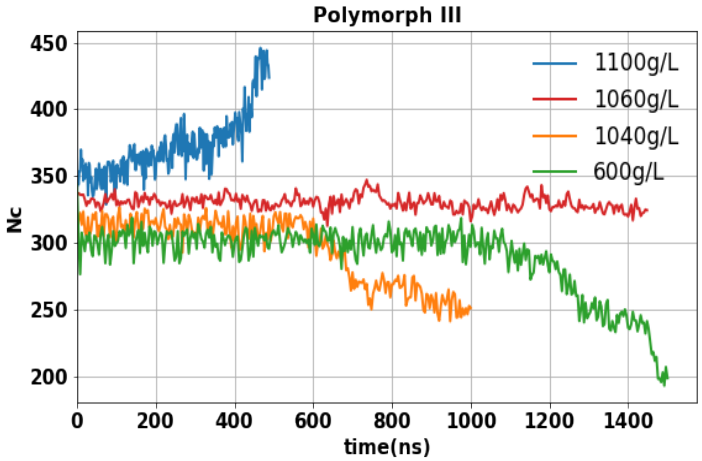

Figure 2: Energy minimized structures of (a) form I (b) form III and (c) form IV are shown along with the angles between the vectors, considered for finding local orientation in each structure.

## Polymorph 1

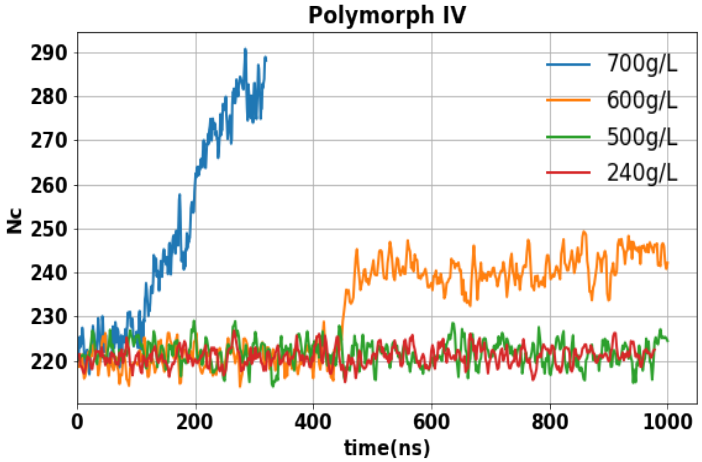
Initially, the crystal slab of urea had four layers composed of 128 molecules but, during the NPT equilibration, one layer at the interface dissolved partially owing to the initial encounter of the crystal with the solution. The system thus obtained was simulated using C$\mu$MD. Figure 3a shows the evolution of the number of crystalline molecules as a function of time for four different concentrations of urea solution. The experimental solubility of urea at 298 K lies in the range of 350-544g/L.[34] When the crystal was placed in the solution with a concentration closer to the solubility upper limit, the crystal grew with the molecules approaching in the proper orientation and sticking to the surface, as can be seen in figure 4. Relatively slower growth is observed when the crystal is surrounded with 400g/L urea solution. The crystal shows no appreciable growth or dissolution in 300g/L urea solution for as long as 1$\mu$s, and so, we predict the value of solubility to be in the range of 300-400 g/L. At a concentration lower than this value, the crystal dissolves.




(a)




(b)




(c)

Figure 3: The number of crystalline molecules($N_c$) as a function of time for (a)Polymorph I (b)Polymorph III (c)Polymorph IV

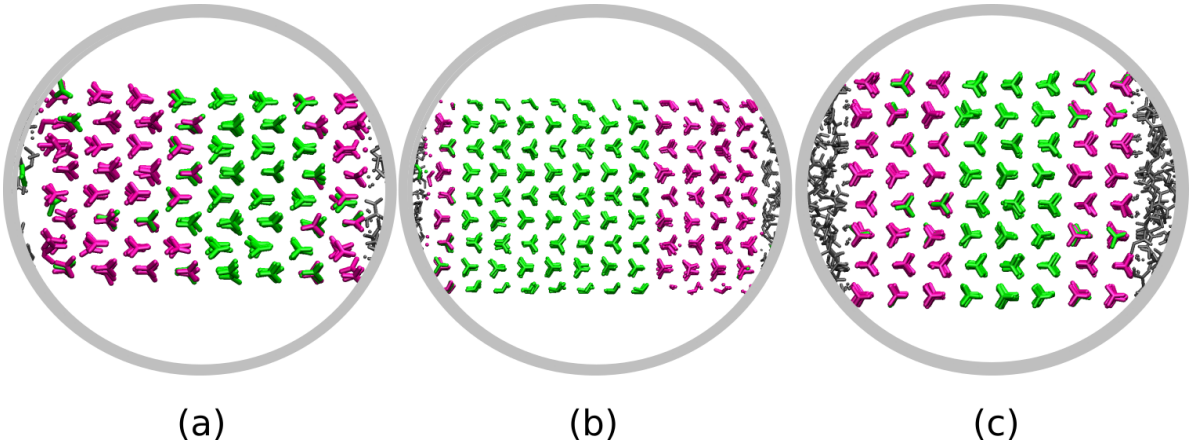



Figure 4: Snapshots of the simulation box taken after (a) 100 ns, (b) 550 ns and (c) 300 ns, showing growth in the three polymorphs. Urea molecules belonging to crystal are represented in green, urea molecules from solution that have oriented in the form of crystal are colored pink and the urea molecules in solution are grey. (a) In form I, the crystal had three layers initially, but five additional layers grew after 60 ns of simulation in 545g/L solution. (b) In form III, five more layers grew at the surface after 500 ns simulation in 1100 g/L solution. (c) Form IV also showed five new layers at the surface after 280 ns in 700 g/L solution.

## Polymorph III

The dynamics at the crystal interface in polymorph III (Figure 3b) show a trend similar to polymorph I. The solubility range is 600g/L with 1040g/L and 1100g/L serving as lower and upper limits to this range, as opposing events - growth and dissolution occur at these concentrations. The 1060g/L seems to be an approximate value of solubility, such a high value for this polymorph can be attributed to the weaker intermolecular forces particularly, hydrogen bonds in this crystal as compared to polymorph I.[28,37] In the latter, the carbonyl bond of the urea molecule forms four H-bonds, NH$\cdots$O with three different neighboring molecules, whereas, in polymorph III, the number of neighboring molecules making H-bonds to a given molecule reduces to two with the collapse of square tunnels formed from the four hydrogen-bonded molecules in form I. The structural compactness is reduced in phase III which encourages urea molecules to go into the solution phase thereby, increasing the solubility.

### Polymorph IV

Growth of the polymorph IV crystal occurs in 600g/L and 700g/L urea solutions as can be seen in Figure 3c. However, crystal dissolution was not observed even in urea solutions having a concentration as low as 240g/L for over 1$\mu$s. Neither growth nor dissolution of the crystal is observed at lower concentrations than 600g/L, the lower bound on solubility is hence not obtained. This highlights the large energy barrier for the urea molecule in the crystal to come out and pass into the solution. It is possible that (001) face may not be kinetically feasible for the study of this polymorph, further investigation into additional facets is required to establish the lower solubility limit, but that endeavor is deferred for future consideration. The crystal structure of phase IV is similar to form III in the sense that both have distortion of the square tunnel, which should increase the solubility of polymorph IV over form I. Compared to form III, polymorph IV exhibits greater symmetry and an additional NH$\cdots$O hydrogen bond.[28,37] This suggests that the solubility of the latter should be lesser than the solubility of the former, which indeed is found to be the case.

## Conclusions

Solubility values were found for the three polymorphs of urea using molecular dynamics simulations. This study effectively distinguishes between the polymorphs and yields solubility values consistent with our prior comprehension of their molecular arrangements. C$\mu$MD is an upgrade over unbiased MD simulations, which carefully looks after the shortcomings of DCS. Experimental solubility values for forms III and IV are not yet reported. The predicted solubility of form I, 300g/L, lies on the lower side of the experimentally determined value, 448g/L (average of the solubility range), and is only 67% accurate. The solubility uncertainty can be attributed to many factors, including the force field parameters. Recent advancements in machine learning-based potentials[38–46] will alleviate this issue to a great extent.

## 6. AUTHOR INFORMATION


### Corresponding Author*

**Tarak Karmakar** - Department of Chemistry, Indian Institute of Technology, Delhi, Hauz Khas, New Delhi - 110016; ORCID: 0000-0002-8721-6247;

E-mail: tkarmakar@chemistry.iitd.ac.in

### Author

**Neha** - Department of Chemistry, Indian Institute of Technology, Delhi, Hauz Khas, New Delhi - 110016

**Manya Aggarwal** - Department of Chemistry, Indian Institute of Technology, Delhi, Hauz Khas, New Delhi - 110016

**Ashutosh Soni** - Department of Chemistry, Indian Institute of Technology, Delhi, Hauz Khas, New Delhi - 110016


### Notes:

† The authors have contributed equally. The authors declare no competing financial interest.

### Data Availability

Input files for the three polymorphs and the crystallinity calculation code can be found at https://github.com/Manya120101

## 7. ACKNOWLEDGEMENTS


Neha thanks IIT Delhi for the institute Ph.D. fellowship. T.K. acknowledges the Science and Engineering Research Board (SERB), New Delhi, India for the Start-up Research Grant (File No. SRG/2022/000969). We also acknowledge IIT Delhi for the Seed Grant. We thank

the IIT Delhi HPC facility for computational resources.

# References

(1) Bhalani, D. V.; Nutan, B.; Kumar, A.; Singh Chandel, A. K. Bioavailability Enhancement Techniques for Poorly Aqueous Soluble Drugs and Therapeutics. *Biomedicines* **2022**, *10*, 2055.

(2) Kumar, R.; Thakur, A. K.; Chaudhari, P.; Banerjee, N. Particle Size Reduction Techniques of Pharmaceutical Compounds for the Enhancement of Their Dissolution Rate and Bioavailability. *Journal of Pharmaceutical Innovation* **2021**, *17*, 333–352.

(3) Sareen, S.; Joseph, L.; Mathew, G. Improvement in solubility of poor water-soluble drugs by solid dispersion. *International Journal of Pharmaceutical Investigation* **2012**, *2*, 12.

(4) Wolthuis, E.; Pruiksma, A. B.; Heerema, R. P. Determination of solubility: A laboratory experiment. *Journal of Chemical Education* **1960**, *37*, 137.

(5) Bjelobrk, Z.; Rajagopalan, A. K.; Mendels, D.; Karmakar, T.; Parrinello, M.; Mazzotti, M. Solubility of Organic Salts in Solvent–Antisolvent Mixtures: A Combined Experimental and Molecular Dynamics Simulations Approach. *Journal of Chemical Theory and Computation* **2022**, *18*, 4952–4959.

(6) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*, 1129–1143.

(7) Badar, M. S.; Shamsi, S.; Ahmed, J.; Alam, M. A. *Transdisciplinarity*; Springer, 2022; pp 131–151.

(8) Ferrario, M.; Ciccotti, G.; Spohr, E.; Cartailler, T.; Turq, P. Solubility of KF in water by

molecular dynamics using the Kirkwood integration method. *The Journal of Chemical Physics* **2002**, *117*, 4947–4953.

(9) Gupta, J.; Nunes, C.; Vyas, S.; Jonnalagadda, S. Prediction of Solubility Parameters and Miscibility of Pharmaceutical Compounds by Molecular Dynamics Simulations. *The Journal of Physical Chemistry B* **2011**, *115*, 2014–2023.

(10) Bjelobrk, Z.; Mendels, D.; Karmakar, T.; Parrinello, M.; Mazzotti, M. Solubility Prediction of Organic Molecules with Molecular Dynamics Simulations. *Crystal Growth & Design* **2021**, *21*, 5198–5205.

(11) Bjelobrk, Z.; Rajagopalan, A. K.; Mendels, D.; Karmakar, T.; Parrinello, M.; Mazzotti, M. Solubility of Organic Salts in Solvent–Antisolvent Mixtures: A Combined Experimental and Molecular Dynamics Simulations Approach. *Journal of Chemical Theory and Computation* **2022**, *18*, 4952–4959.

(12) Pyrhönen, J.; Bansal, K. K.; Bhadane, R.; Wilén, C.-E.; Salo-Ahen, O. M. H.; Rosenholm, J. M. Molecular Dynamics Prediction Verified by Experimental Evaluation of the Solubility of Different Drugs in Poly(decalactone) for the Fabrication of Polymeric Nanoemulsions. *Advanced NanoBiomed Research* **2021**, *2*.

(13) Liu, S.; Cao, S.; Hoang, K.; Young, K. L.; Paluch, A. S.; Mobley, D. L. Using MD Simulations To Calculate How Solvents Modulate Solubility. *Journal of Chemical Theory and Computation* **2016**, *12*, 1930–1941.

(14) Hassanzadeh, A.; Sabzi, F. Prediction of CO2 and H2 solubility, diffusion, and permeability in MFI zeolite by molecular dynamics simulation. *Structural Chemistry* **2021**, *32*, 1641–1650.

(15) Salo-Ahen, O. M. H. et al. Molecular Dynamics Simulations in Drug Discovery and Pharmaceutical Development. *Processes* **2020**, *9*, 71.

(16) Espinosa, J. R.; Sanz, E.; Valeriani, C.; Vega, C. On fluid-solid direct coexistence simulations: The pseudo-hard sphere model. *The Journal of Chemical Physics* **2013**, *139*.

(17) Espinosa, J. R.; Young, J. M.; Jiang, H.; Gupta, D.; Vega, C.; Sanz, E.; Debenedetti, P. G.; Panagiotopoulos, A. Z. On the calculation of solubilities via direct coexistence simulations: Investigation of NaCl aqueous solutions and Lennard-Jones binary mixtures. *The Journal of Chemical Physics* **2016**, *145*.

(18) Manzanilla-Granados, H. M.; Saint-Martín, H.; Fuentes-Azcatl, R.; Alejandre, J. Direct Coexistence Methods to Determine the Solubility of Salts in Water from Numerical Simulations. Test Case NaCl. *The Journal of Physical Chemistry B* **2015**, *119*, 8389–8396.

(19) Carruthers, J.; Ferrario, M.; Anwar, J. Prediction of aqueous solubility of a strongly soluble solute from molecular simulation. *The Journal of Chemical Physics* **2023**, *159*.

(20) Özpınar, G. A.; Peukert, W.; Clark, T. An improved generalized AMBER force field (GAFF) for urea. *Journal of Molecular Modeling* **2010**, *16*, 1427–1440.

(21) Hölzl, C.; Kibies, P.; Imoto, S.; Noetzel, J.; Knierbein, M.; Salmen, P.; Paulus, M.; Nase, J.; Held, C.; Sadowski, G.; Marx, D.; Kast, S. M.; Horinek, D. Structure and thermodynamics of aqueous urea solutions from ambient to kilobar pressures: From thermodynamic modeling, experiments, and first principles simulations to an accurate force field description. *Biophysical Chemistry* **2019**, *254*, 106260.

(22) Perego, C.; Salvalaglio, M.; Parrinello, M. Molecular dynamics simulations of solutions at constant chemical potential. *The Journal of Chemical Physics* **2015**, *142*.

(23) Karmakar, T.; Piaggi, P. M.; Parrinello, M. Molecular Dynamics Simulations of Crystal Nucleation from Solution at Constant Chemical Potential. *Journal of Chemical Theory and Computation* **2019**, *15*, 6923–6930.

(24) Salvalaglio, M.; Vetter, T.; Giberti, F.; Mazzotti, M.; Parrinello, M. Uncovering Molecular Details of Urea Crystal Growth in the Presence of Additives. *Journal of the American Chemical Society* **2012**, *134*, 17221–17233.

(25) Salvalaglio, M.; Mazzotti, M.; Parrinello, M. Urea homogeneous nucleation mechanism is solvent dependent. *Faraday Discussions* **2015**, *179*, 291–307.

(26) Salvalaglio, M.; Perego, C.; Giberti, F.; Mazzotti, M.; Parrinello, M. Molecular-dynamics simulations of urea nucleation from aqueous solution. *Proceedings of the National Academy of Sciences* **2014**, *112*.

(27) Lamelas, F. J.; Dreger, Z. A.; Gupta, Y. M. Raman and X-Ray Scattering Studies of High-Pressure Phases of Urea. *The Journal of Physical Chemistry B* **2005**, *109*, 8206–8215.

(28) Olejniczak, A.; Ostrowska, K.; Katrusiak, A. H-Bond Breaking in High-Pressure Urea. *The Journal of Physical Chemistry C* **2009**, *113*, 15761–15767.

(29) Piaggi, P. M.; Parrinello, M. Predicting polymorphism in molecular crystals using orientational entropy. *Proceedings of the National Academy of Sciences* **2018**, *115*, 10251–10256.

(30) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.

(31) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling* **2006**, *25*, 247–260.

(32) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchi-

son, G. R. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of cheminformatics* **2012**, *4*, 1–17.

(33) Bekker, H.; Berendsen, H.; Dijkstra, E.; Achterop, S.; Vondrumen, R. v.; Vanderspoel, D.; Sijbers, A.; Keegstra, H.; Renardus, M. Gromacs-a parallel computer for molecular-dynamics simulations. 4th international conference on computational physics (PC 92). 1993; pp 252–256.

(34) Yalkowsky, S. H.; He, Y.; Jain, P. *Handbook of Aqueous Solubility Data*; CRC Press, 2016.

(35) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **2008**, *4*, 435–447.

(36) Promoting transparency and reproducibility in enhanced molecular simulations. *Nature Methods* **2019**, *16*, 670–673.

(37) Dziubek, K.; Citroni, M.; Fanetti, S.; Cairns, A. B.; Bini, R. High-Pressure High-Temperature Structural Properties of Urea. *The Journal of Physical Chemistry C* **2017**, *121*, 2380–2387.

(38) Ye, Z.; Ouyang, D. Prediction of small-molecule compound solubility in organic solvents by machine learning algorithms. *Journal of cheminformatics* **2021**, *13*, 98.

(39) Vermeire, F.; Chung, Y.; Green, W. Predicting Solubility Limits of Organic Solutes for a Wide Range of Solvents and Temperatures Using Machine Learning and Thermodynamics. 2022 AIChE Annual Meeting. 2022.

(40) Kurotani, A.; Kakiuchi, T.; Kikuchi, J. Solubility prediction from molecular properties and analytical data using an in-phase deep neural network (Ip-DNN). *ACS omega* **2021**, *6*, 14278–14287.

(41) Boobier, S.; Hose, D. R.; Blacker, A. J.; Nguyen, B. N. Machine learning with physicochemical relationships: solubility prediction in organic solvents and water. *Nature communications* **2020**, *11*, 5753.

(42) Panapitiya, G.; Girard, M.; Hollas, A.; Sepulveda, J.; Murugesan, V.; Wang, W.; Saldanha, E. Evaluation of deep learning architectures for aqueous solubility prediction. *ACS omega* **2022**, *7*, 15695–15710.

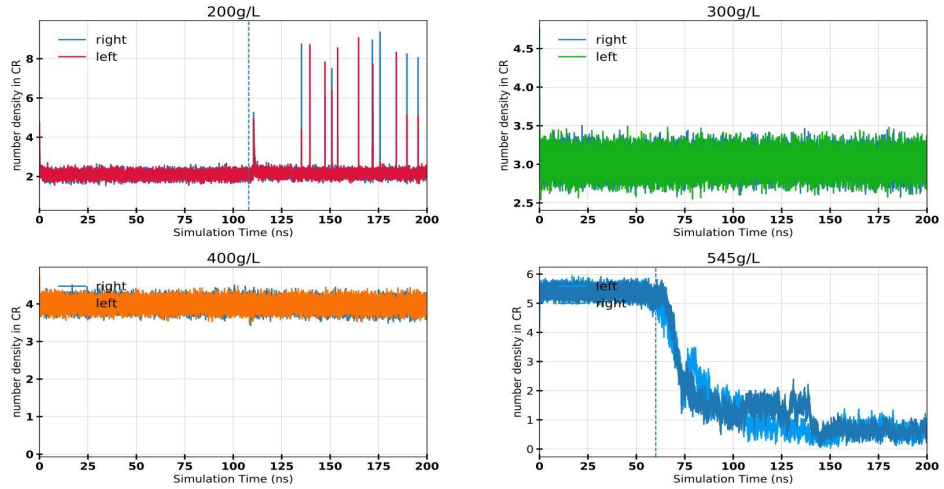
(43) An, F.; Sayed, B. T.; Parra, R. M. R.; Hamad, M. H.; Sivaraman, R.; Foumani, Z. Z.; Rushchitc, A. A.; El-Maghawry, E.; Alzhrani, R. M.; Alshehri, S., et al. Machine learning model for prediction of drug solubility in supercritical solvent: Modeling and experimental validation. *Journal of Molecular Liquids* **2022**, *363*, 119901.

(44) Lee, S.; Lee, M.; Gyak, K.-W.; Kim, S. D.; Kim, M.-J.; Min, K. Novel solubility prediction models: Molecular fingerprints and physicochemical features vs graph convolutional neural networks. *ACS omega* **2022**, *7*, 12268–12277.

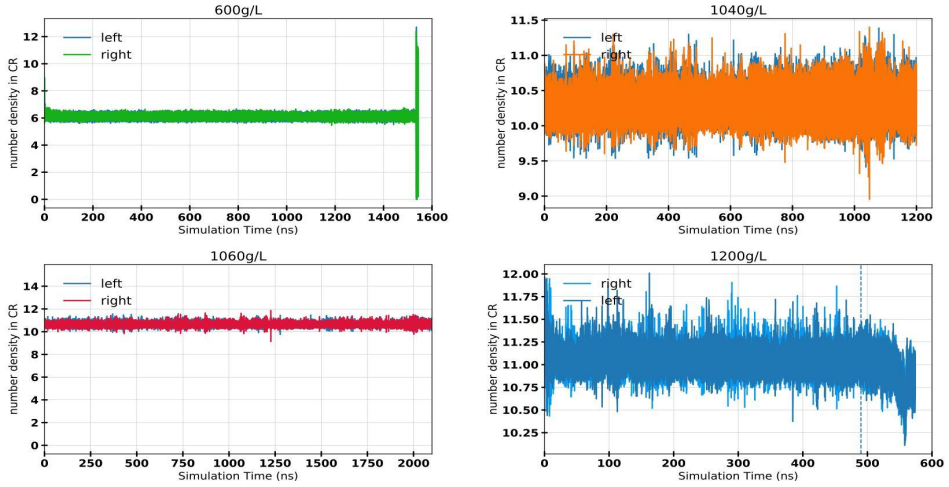
(45) Kim, B.; Manchuri, A. R.; Oh, G.-T.; Lim, Y.; Son, Y.; Choi, S.; Kang, M.; Jang, J.; Ha, J.; Cho, C.-H., et al. Experimental Analysis and Prediction of Radionuclide Solubility Using Machine Learning Models: Effects of Organic Complexing Agents. *Journal of Hazardous Materials* **2024**, 134012.

(46) GİDER, V.; BUDAK, C. Drug Solubility Prediction: A Comparative Analysis of GNN, MLP, and Traditional Machine Learning Algorithms. *Gazi University Journal of Science Part C: Design and Technology* 1–1.

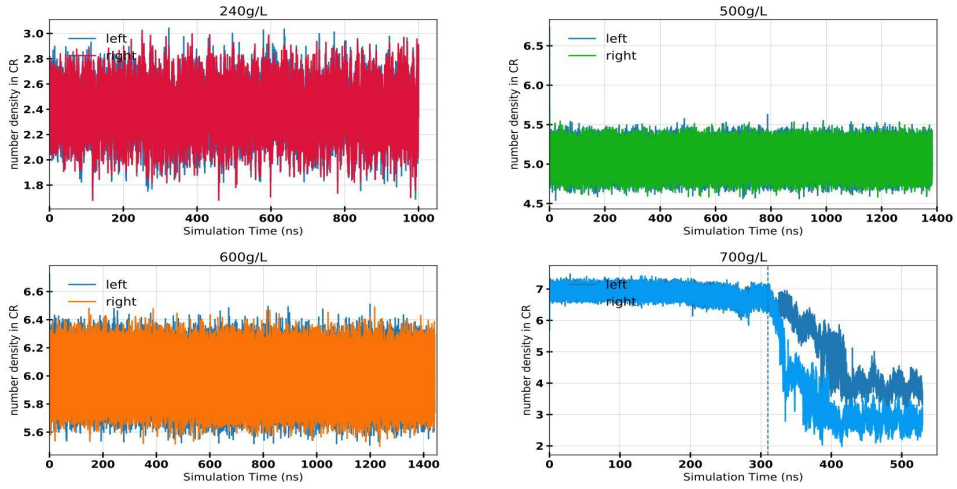
# Supporting Information Available




(a)




(b)




(c)

Figure 5: Number density in the control region is plotted as a function of time for C$\mu$MD simulations in (a) Polymorph I, (b) Polymorph III, and (c) Polymorph IV