Supporting information for "Solubility prediction of organic molecules with molecular dynamics simulations"

Zoran Bjelobrk,[†] Dan Mendels,[‡] Tarak Karmakar,[¶] Michele Parrinello,[*,¶] and Marco Mazzotti[*,†]

†*Institute of Energy and Process Engineering, ETH Zürich, CH-8092, Switzerland*

‡*Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, United States*

¶*Istituto Italiano di Tecnologia (IIT), Via Morego, 30, 16163 Genova GE, Italy*

E-mail: michele.parrinello@iit.it; marco.mazzotti@ipe.mavt.ethz.ch

Simulation setup equilibration

For all equilibration simulations we used Gromacs 2016.5[1] with full atomistic description of all molecule types. We used the velocity rescaling thermostat,[2] periodic boundary conditions, the Ewald particle mesh approach[3] for the electrostatic interactions, and the LINCS algorithm[4,5] to constrain the covalent bonds involving hydrogens. The non-bonded cutoff was set to 1 nm.

The simulation box specifications including the box lengths L_x, L_y, and L_z, as well as number of solute molecules N_s and solvent molecules N_l are listed in Table 1 for all studied systems in this work, namely urea grown in acetonitrile (MeCN), ethanol (EtOH), and methanol (MeOH), and naphthalene grown in EtOH and toluene (MePh). For the urea-

MeCN system, a larger simulation box was used to reach the low solubility regime. The solubility does not depend on the simulation box sizes for the reported setups. However, simulation box sizes which are smaller than the reported ones do suffer from finite size effects, especially if the length of the liquid phase in z direction (perpendicular to the crystal surface) is not long enough. Too short liquid phase lengths can lead to a weak orientation pattern of the solvent which can cause a significant drop in solubility for the given solute compound.

Table 1: Simulation box specifications of all studied systems.

	urea			naphthalene	
	MeCN	EtOH	MeOH	EtOH	MePh
N_s [-]	504	225	250	260	320
N_l [-]	1400	390	525	706	350
L_x [nm]	3.78200	2.26842	2.26842	3.28663	3.28663
L_y [nm]	4.25456	2.83593	2.83593	2.98478	2.98478
L_z [nm]	10.46135	8.22554	8.06331	12.13055	12.96988
T [K]	300	300	300	280	280
p [bar]	1	1	1	1	1

It suffices to perform simulations under NVT instead of the computationally more expensive NPT conditions, since the growth of a single kink site does not noticeably alter the pressure of the system. To obtain the appropriate simulation box lengths we used the following equilibration protocol for all considered systems.

First, a seed crystal was constructed from XRD data (urea polymorph I[6] and naphthalene polymorph I[7]) with the face of interest perpendicular to the z-axis: for urea face $\{110\}$ and for naphthalene face $\{00\bar{1}\}$. The crystal system energy was then minimized with the conjugate gradient algorithm with a tolerance of the maximum force of 50 kJ mol^{-1} nm^{-1}, followed by a temperature equilibration at NVT conditions for 1 ns at an integration time step of 0.5 fs, to reach the targeted temperatures. The pressure equilibration was achieved by running the simulation setup for a further 25 ns at NPT conditions using the anisotropic Parrinello-Rahman barostat[8] with the same integration time step of 0.5 fs. The data of the last 20 ns of the simulation were used to calculate the average box lengths L_x, L_y, and their

average ratio L_x/L_y to identify the simulation box frame closest to the average values.

Second, we submerged each of the crystals in the corresponding solute-solvent mixture using the genbox utility of Gromacs.[5] The same energy minimization, and NVT equilibration were performed as for the crystal equilibration step. For the NPT equilibration we used the semi-isotropic barostat to allow expansion/contraction of the simulation box only along the z-axis while keeping the already averaged L_x and L_y constant. From the NPT equilibration run we used the last 20 of the 25 ns to compute the average box length L_z at the pressure of 1 bar. Again, the simulation frame with the simulation box length along z closest to L_z was chosen as initial configuration for the concentration profile equilibration step with CμMD.

Third, we used the CμMD algorithm[9] to obtain the targeted concentration profiles in the vicinity of the crystal surface. A simulation time of 25 ns ensures to reach the targeted concentration profile. During the concentration profile equilibration, a harmonic potential was used to push the molecules away from the crystal surface, which do not belong to the unfinished layer. The unfinished surface layer was prevented from dissolving using a potential acting through the surface structure CV.

Simulation box visualizations are shown in Figure 1 and the unfinished surface layer visualizations are shown in Figure 2. The surface layer was cut along the [001] direction for urea and along the [010] direction for naphthalene.

Constant chemical potential method

We briefly discuss the CμMD method.[9] The scheme shown in Figure 3 depicts the simulation setup, with the corresponding solute concentration profile, $c(z)$, along the z axis. Periodic boundary conditions are introduced for all spatial directions. The crystal surface, with an unfinished layer comprising the kink site, is exposed to the solution.

To keep the solution concentration constant the CμMD algorithm is introduced, which works as follows. The liquid phase of the simulation box is partitioned along the z axis into

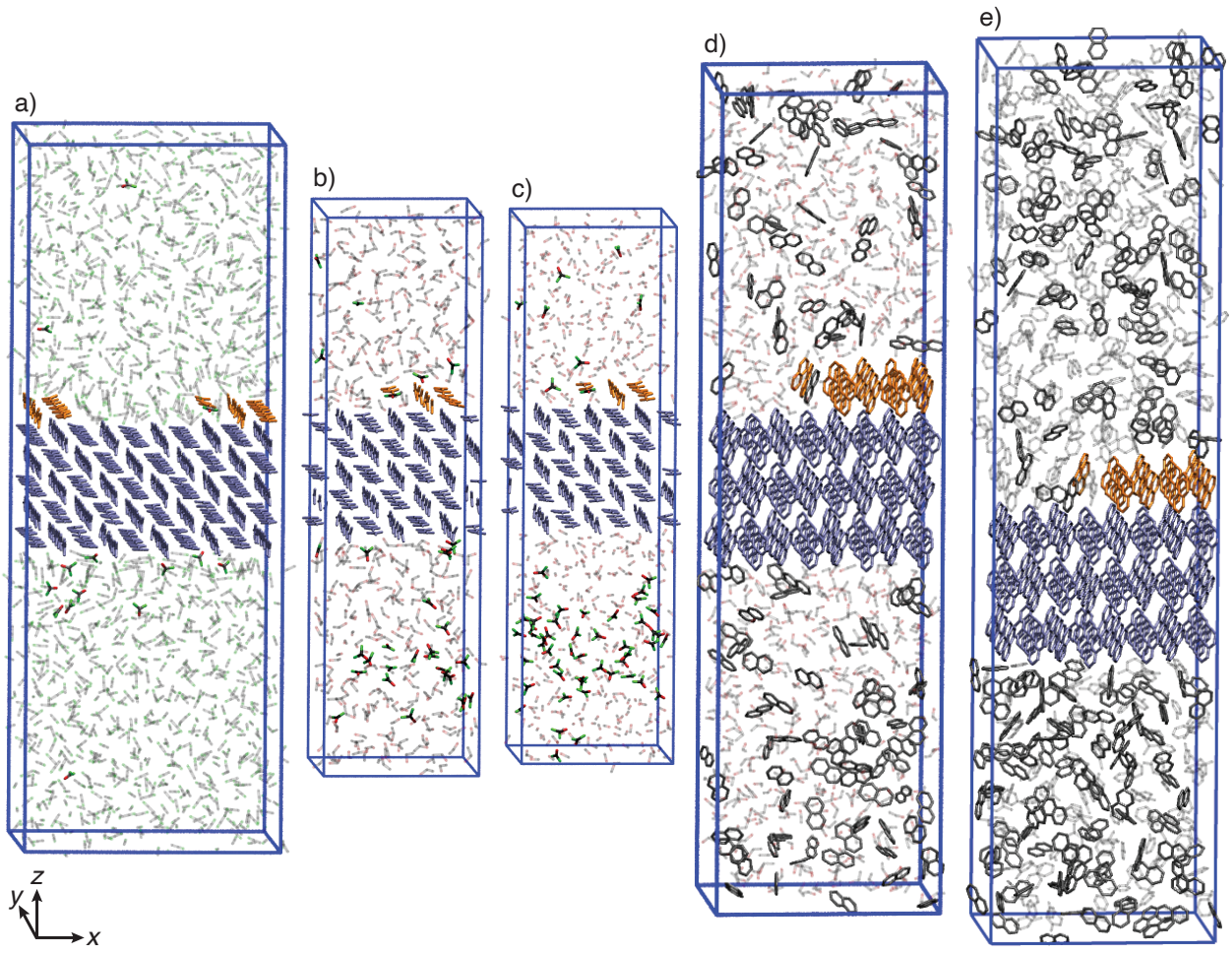

Figure 1: Simulation box visualizations; a) urea in acetonitrile, b) urea in ethanol, c) urea in methanol, d) naphthalene in ethanol, e) naphthalene in toluene. The bulk crystal molecules are colored in blue, the unfinished surface layer molecules are colored in orange. The atoms of the molecules in solution are colored in black for carbon, red for oxygen and green for nitrogen. Hydrogens are omitted for clarity. Solvent molecules are shown in faded colors.

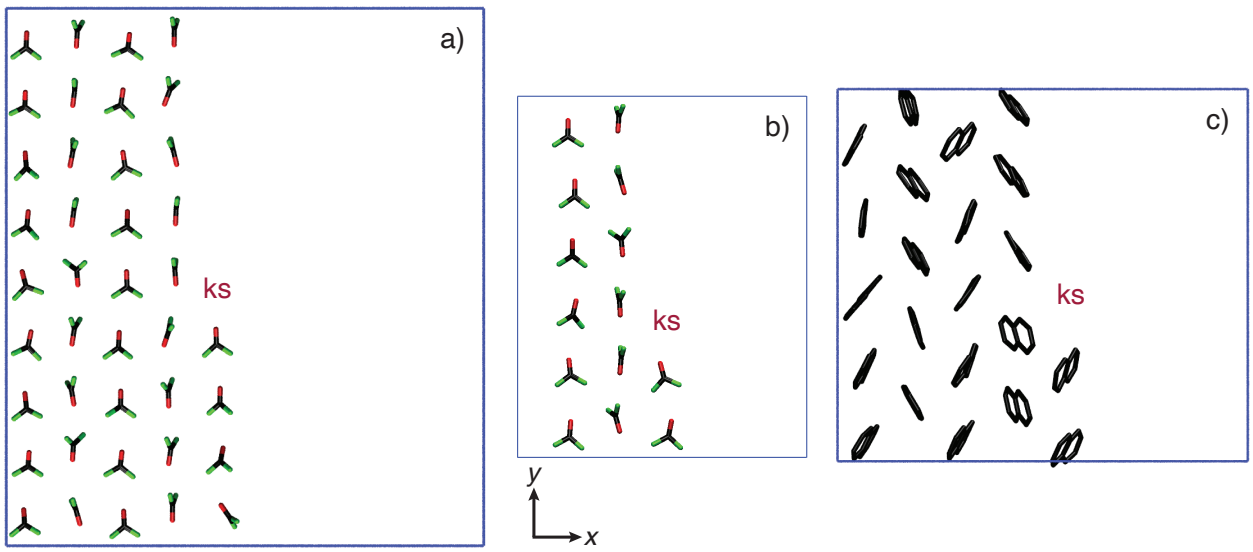

Figure 2: Visualizations of the unfinished surface layer molecules with the location of the biased kink site indicated by 'ks'. a) Urea in acetonitrile; b) urea in ethanol and methanol; c) naphthalene in ethanol and toluene. The atoms of the molecules are colored in black for carbon, red for oxygen and green for nitrogen. Hydrogens are omitted for clarity.

following segments: a transition region, control region and reservoir. An external force F_i^{μ} is introduced to control the flux of solutes i between the control region and reservoir at position z_{F}. F_i^{μ} is defined as follows:

$$F_i^{\mu} = k^{\mu}(c_{\mathrm{CR}}(t) - c_0)G_{\omega}(z_i, z_{\mathrm{F}}). \tag{1}$$

k^{μ} is a force constant, $c_{\mathrm{CR}}(t)$ is the concentration of the control region at time t, and c_0 is the predefined target concentration. $G_{\omega}(z_i, z_{\mathrm{F}})$ is a bell shaped function:

$$G_{\omega}(z_i, z_{\mathrm{F}}) = \frac{1}{4\omega}\left[1 + \cosh\left(\frac{z_i - z_{\mathrm{F}}}{\omega}\right)\right]^{-1}, \tag{2}$$

where z_i corresponds to the z position of solute molecule i and ω defines the height and width of the bell curve.

If $c(t)$ is at a given time step below c_0, then F_i^{μ} will accelerate the solute molecules from the reservoir towards the control region and vice versa. This creates a constant concentration

profile in the control region, as shown in Figure 3, and enables the simulation of kink growth at constant chemical potential. See ref. 9 for further details on the CμMD method.

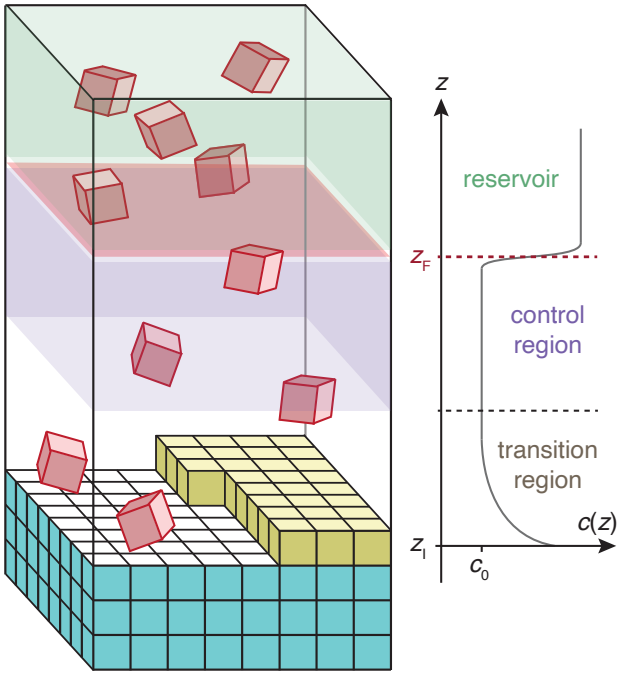

Figure 3: Scheme of the simulation setup with corresponding concentration profile, $c(z)$, along the z axis. The solute molecules are shown as cubes where the bulk crystal molecules are colored in green, the molecules of the unfinished layer, which comprises the kink site, are colored in yellow and the dissolved molecules are colored in red. Depiction of solvent molecules is omitted for clarity. The control region is shaded in violet and the reservoir is shaded in green. The position, z_F, where the external force acts is colored in red.

The CμMD parameters used in the simulations are listed in Table 2.

Table 2: Values of the CμMD parameters.

	urea			naphthalene	
	MeCN	EtOH	MeOH	EtOH	MePh
ω/L_z [-]	0.02	0.02	0.02	0.02	0.02
z_{TR}/L_z [-]	0.16	0.13	0.22	0.14	0.24
z_{CR}/L_z [-]	0.50	0.20	0.22	0.18	0.20
z_F/L_z [-]	0.68	0.37	0.50	0.34	0.48
$\Delta z/L_z$ [-]	1/120	1/120	1/120	1/120	1/120

Collective variables (CVs)

Biased CV

The values used for the biased CV s_b are listed in Table 3.

Table 3: Values of the biased CV used in the simulation.

		urea			naphthalene	
		MeCN	EtOH	MeOH	EtOH	MePh
s_s	$r_\mathrm{s}^{(x)}$ [nm]	2.8366	1.1364	1.1384	1.8129	3.0464
	$r_\mathrm{s}^{(y)}$ [nm]	2.2038	1.1096	1.1091	1.3490	1.3464
	$r_\mathrm{s}^{(z)}$ [nm]	1.1404	1.3231	1.4044	1.4945	0.1121
	σ_s [-]	0.30	0.20	0.20	0.4	0.22
	$\chi_{\mathrm{s}1}$ [-]	0.3	0.3	0.3	0.3	0.3
	$\chi_{\mathrm{s}2}$ [-]	3	3	3	3	3
	w_s [-]	0.5	0.5	0.5	0.5	0.5
s_l	$r_\mathrm{l}^{(x)}$ [nm]	2.8366	1.1364	1.1384	1.8129	3.0464
	$r_\mathrm{l}^{(y)}$ [nm]	2.2038	1.1096	1.1091	1.3490	1.3464
	$r_\mathrm{l}^{(z)}$ [nm]	1.1404	1.3231	1.4044	1.4945	0.1121
	σ_l [-]	0.1	0.1	0.1	0.2	0.2
	χ_l [-]	0.3	0.3	0.3	0.5	0.5
	w_l [-]	-0.25	-0.25	-0.25	-0.25	-0.5

To improve the sampling performance of the WTMetaD simulations, lower and upper wall potentials were used for the biased CV:

$$V_\mathrm{s} = \begin{cases} k_\mathrm{s,l}((s_\mathrm{s}^{\chi_{\mathrm{s}1}} + s_\mathrm{s}^{\chi_{\mathrm{s}2}}) - s_\mathrm{s,l})^2, & \text{if } (s_\mathrm{s}^{\chi_{\mathrm{s}1}} + s_\mathrm{s}^{\chi_{\mathrm{s}2}}) < s_\mathrm{s,l}, \\ k_\mathrm{s,u}((s_\mathrm{s}^{\chi_{\mathrm{s}1}} + s_\mathrm{s}^{\chi_{\mathrm{s}2}}) - s_\mathrm{s,u})^2, & \text{if } (s_\mathrm{s}^{\chi_{\mathrm{s}1}} + s_\mathrm{s}^{\chi_{\mathrm{s}2}}) > s_\mathrm{s,u}, \\ 0, & \text{else,} \end{cases} \tag{3}$$

and

$$V_\mathrm{l} = \begin{cases} k_\mathrm{l,l}(s_\mathrm{l}^{\chi_\mathrm{l}} - s_\mathrm{l,l})^2, & \text{if } s_\mathrm{l}^{\chi_\mathrm{l}} < s_\mathrm{l,l}, \\ k_\mathrm{l,u}(s_\mathrm{l}^{\chi_\mathrm{l}} - s_\mathrm{l,u})^2, & \text{if } s_\mathrm{l}^{\chi_\mathrm{l}} > s_\mathrm{l,u}, \\ 0, & \text{else.} \end{cases} \tag{4}$$

where $k_\mathrm{s,l}$, $k_\mathrm{s,u}$, $k_\mathrm{l,l}$, and $k_\mathrm{l,u}$ are the force constants and $s_\mathrm{s,l}$, $s_\mathrm{s,u}$, $s_\mathrm{l,l}$, and $s_\mathrm{l,u}$ are the thresholds below and above which the potentials are acting. The lower walls inhibit the biased simulations from getting stuck at s_s and s_l values of zero and the higher walls inhibit excessive agglomeration of solute or solvent molecules at the kink site, which is not relevant

for the kink growth process. The values of the potentials used in this work can be found in Table 4.

The WTMetaD[10] parameter values are presented in Table 5. W and σ_W are the height and width of the Gaussians, γ is the bias factor, τ the bias deposition stride, and Δs_b is the bin length of the grid on which the bias is stored.

Table 4: Values of the wall potentials parameters used for the biased CV.

		urea			naphthalene	
		MeCN	EtOH	MeOH	EtOH	MePh
s_s	$k_\mathrm{s,l}$ [kJ/mol]	15	15	15	15	15
	$k_\mathrm{s,u}$ [kJ/mol]	15	15	15	15	15
	$s_\mathrm{s,l}$ [-]	0.03	0.03	0.03	0.05	0.04
	$s_\mathrm{s,u}$ [-]	2.08	2.03	2.03	2.22	2.40
s_l	$k_\mathrm{l,l}$ [kJ/mol]	15	15	15	15	15
	$k_\mathrm{l,u}$ [kJ/mol]	15	15	15	15	15
	$s_\mathrm{l,l}$ [-]	0.025	0.025	0.025	0.04	0.08
	$s_\mathrm{l,u}$ [-]	0.995	0.995	0.995	1.10	1.10

Table 5: Values of the well-tempered Metadynamics parameters.

	urea			naphthalene	
	MeCN	EtOH	MeOH	EtOH	MePh
W [kJ/mol]	0.2	0.2	0.2	0.2	0.4
σ_W [-]	0.06	0.03	0.03	0.03	0.03
γ [-]	4	3	3	2	4
τ [ps]	1	1	1	1	1
Δs_b [-]	0.02	0.01	0.01	0.01	0.01

Surface structure CV

To prevent the dissolution of the unfinished layer, a harmonic potential wall is introduced through the surface structure CV, s_st. s_st is defined as the logistic function:

$$s_\mathrm{st} = \frac{1}{1 + \exp\left(-\sigma_\mathrm{st}(\tilde{s}_\mathrm{st} - \tilde{s}_\mathrm{st,0})\right)},$$

with step position $\tilde{s}_{\mathrm{st},0}$ and steepness σ_{st}, of following function:

$$\tilde{s}_{\mathrm{st}} = \sum_i \left(\sum_k \left[\cos^{\eta_x} \left(\frac{\nu_x \pi}{L_x} (x_i - \bar{x}_k) \right) \cos^{\eta_y} \left(\frac{\nu_y \pi}{L_y} (y_i - \bar{y}_k) \right) \right] \exp \left\{ - \frac{(z_i - \bar{z})^2}{2\sigma_z^2} \right\} \right).$$

$\nu_{x/y}$ corresponds to the number of unit cells along the x and y axes, $L_{x/y}$ is the length of the simulation box in x/y direction, $\bar{x}_k$ and $\bar{y}_k$ are the coordinates of the k-th molecule center position in x and y direction within the crystal unit cell. The exponent $\nu_{x/y}$ is a positive even integer and defines the width of the sinusoid peaks. For the z part of s_{st}, $\bar{z}_i$ defines the position in z direction, and σ_z is the width of the Gaussian like curve. The expression is summed over all solute molecules i in the unfinished surface layer.

The form of s_{st} is such that its value is 1 if the center of the solute molecule is at its adsorption site and otherwise zero. This is achieved by setting the steepness, σ_{st}, and position, $\tilde{s}_{\mathrm{st},0}$, of the logistic function step of s_{st} accordingly. Figure 4 shows the contour lines of s_{st} together with the histogram of the urea carbon atom positions of the unfinished layer.

A harmonic wall potential, V_{st}, is introduced to the system through s_{st}:

$$V_{\mathrm{st}} = \begin{cases} k_{\mathrm{st}} (s_{\mathrm{st}} - s_{\mathrm{st},0})^2, & \text{if } s_{\mathrm{st}} < s_{\mathrm{st},0}, \\ 0, & \text{else,} \end{cases}$$

where k_{st} is the force constant and $s_{\mathrm{st},0}$ is the threshold below which the harmonic potential acts. V_{st} prevents the surface molecules from dissolving while at the same time it does not interfere with their thermal lattice vibrations as shown in Figure 4. The parameters used in the simulations are shown in Table 6.

While the naphthalene face $\{00\bar{1}\}$ is stable enough that no dissolution of the surface on the opposite site of the crystal or the layer below the unfinished layer is observed within the simulation time spans of ~ 1 μs, these layers can dissolve for urea face $\{110\}$ grown in ethanol and methanol. We introduced a harmonic potential through the surface structure

CV also for the urea layer on the opposite crystal surface (layer 1) and for the layer below the unfinished surface layer (layer 5). The used parameter values are reported in Table 7.

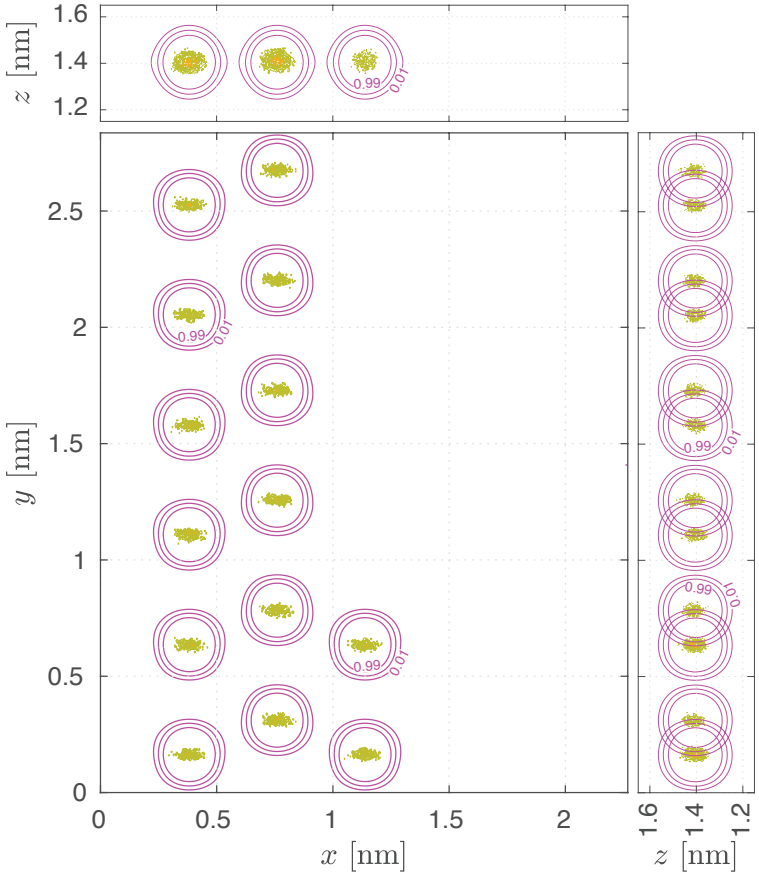

Figure 4: Contour lines of the surface structure CV (pink curves) and histogram of urea center positions (yellow dots) of the unfinished surface layer projected along all three spatial directions.

Adsorption site CVs

The crystal surface with the unfinished layer is comprised of other sites than only the biased kink site: the kink site opposite to the biased kink site and edges (see Figure 5). For the relatively fast growing crystal systems urea and naphthalene, it can happen that the edges and kink sites are growing on other sites and not only on the biased kink site.

Similarly as we did with the surface structure CV to prevent molecules from dissolving, wall potentials are introduced along the edges through adsorption site CVs to prevent the

Table 6: Values of the surface structure CV parameters.

	urea			naphthalene	
	MeCN	EtOH	MeOH	EtOH	MePh
ν_x [-]	5	3	3	4	4
L_x [nm]	3.78200	2.26842	2.26842	3.28663	3.28663
η_x [-]	16	16	16	16	16
$\bar{x}_1$ [nm]	-1.3213	0.38002	0.38002	1.6210	1.6210
$\bar{x}_2$ [nm]	-0.9434	0.75827	0.75827	0.3875	0.3875
ν_y [-]	9	6	6	5	5
L_y [nm]	4.25456	2.83593	2.83593	2.98478	2.98478
η_y [-]	8	8	8	8	8
$\bar{y}_1$ [nm]	-1.4902	0.6391	0.6391	0.4540	0.4540
$\bar{y}_2$ [nm]	-1.3425	0.7874	0.7874	0.1521	0.1521
$\bar{z}$ [nm]	0.7538	1.2873	1.4044	1.4920	0.1187
σ_z [nm]	0.065	0.065	0.065	0.65	0.65
σ_{st} [-]	150	150	150	150	150
$\tilde{s}_{\mathrm{st},0}$ [-]	0.1	0.1	0.1	0.1	0.1
k_{st} [kJ/mol]	15	15	15	15	15
$s_{\mathrm{st},0}$ [-]	40	14	14	22	22

Table 7: Values of the surface structure CV parameters for the urea bulk crystal layers.

	urea layer 1		urea layer 5	
	EtOH	MeOH	EtOH	MeOH
ν_x [-]	3	3	3	3
L_x [nm]	2.26842	2.26842	2.26842	2.26842
η_x [-]	16	16	16	16
$\bar{x}_1$ [nm]	0.3800	0.3799	0.3800	0.3801
$\bar{x}_2$ [nm]	0.7583	0.7583	0.7583	0.7582
ν_y [-]	6	6	6	6
L_y [nm]	2.83593	2.83593	2.83593	2.83593
η_y [-]	8	8	8	8
$\bar{y}_1$ [nm]	0.6391	0.6376	0.6391	0.6376
$\bar{y}_2$ [nm]	0.7874	0.7862	0.7874	0.7875
$\bar{z}$ [nm]	-0.6009	-0.4842	0.9091	1.0259
σ_z [nm]	0.065	0.065	0.065	0.065
σ_{st} [-]	150	150	150	150
$\tilde{s}_{\mathrm{st},0}$ [-]	0.1	0.1	0.1	0.1
k_{st} [kJ/mol]	10	10	10	10
$s_{\mathrm{st},0}$ [-]	36	36	36	36

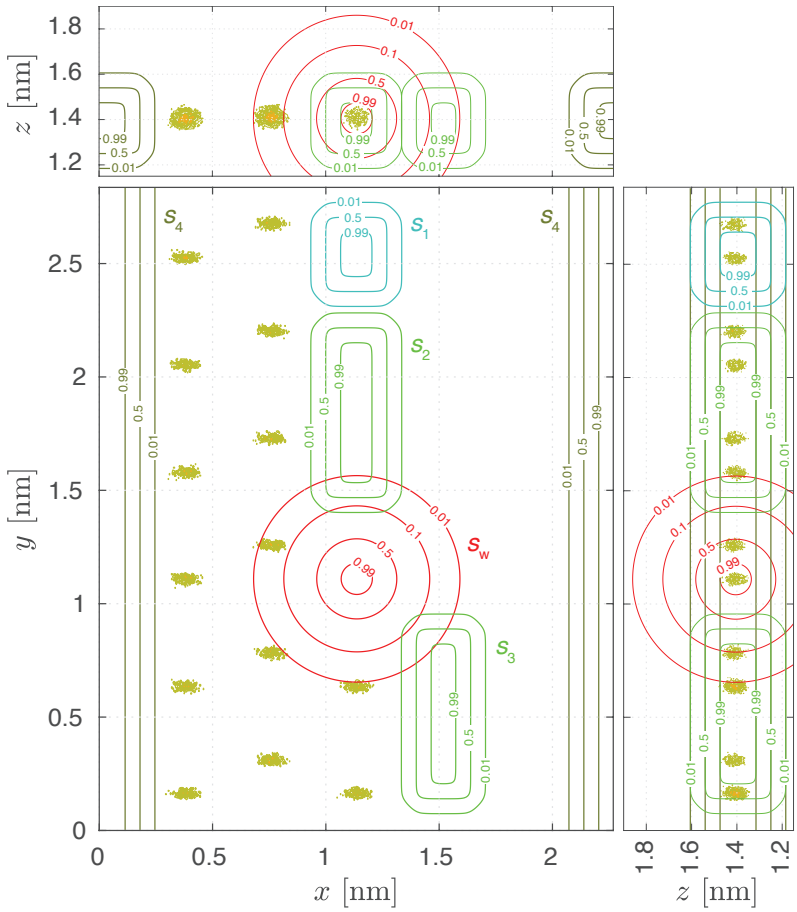

Figure 5: Contour lines of adsorption site CVs and histogram of urea carbon atom positions of the unfinished surface layer.

growth of these sites. In this work these CVs, s_κ, are comprised of simple logistic switching functions in all three spatial directions. For solute molecule i s_κ is defined as:

$$s_\kappa = \sum_k \Bigg[\frac{1}{1+\exp(-\sigma_\kappa(x_k - x_{\mathrm{l}\kappa}))} \left(1 - \frac{1}{1+\exp(-\sigma_\kappa(x_k - x_{\mathrm{u}\kappa}))}\right) \tag{5}$$

$$\cdot \frac{1}{1+\exp(-\sigma_\kappa(y_k - y_{\mathrm{l}\kappa}))} \left(1 - \frac{1}{1+\exp(-\sigma_\kappa(y_k - y_{\mathrm{u}\kappa}))}\right) \tag{6}$$

$$\cdot \frac{1}{1+\exp(-\sigma_\kappa(z_k - z_{\mathrm{l}\kappa}))} \left(1 - \frac{1}{1+\exp(-\sigma_\kappa(z_k - z_{\mathrm{u}\kappa}))}\right) \Bigg], \tag{7}$$

where σ_κ is the steepness of the logistic functions. $x_{\mathrm{l},\kappa}$, $y_{\mathrm{l},\kappa}$, $z_{\mathrm{l},\kappa}$ define the lower bounds and $x_{\mathrm{u},\kappa}$, $y_{\mathrm{u},\kappa}$, $z_{\mathrm{u},\kappa}$ the upper bounds of the intervals (in each spatial direction) in which the adsorption site CV should act on the molecule center position x_i, y_i, z_i. s_κ is obtained by summing $s_{\kappa,i}$ over all solute molecules (excluding bulk- and unfinished surface layer molecules).

The harmonic wall potential is defined as:

$$V_\kappa = \begin{cases} 0, & \text{if } s_\kappa < s_{\kappa,0}, \\ k_\kappa (s_\kappa - s_{\kappa,0})^2, & \text{else}, \end{cases} \tag{8}$$

with force constant k_κ and threshold s_κ above which V_κ is active.

Figure 5 shows the contour lines of the adsorption site CVs (violet and green lines). The parameters used for the different systems are summarized in Table 8.

For the reweighting on the crystallinity CVs, only V_2 and V_3 (shown in green in Figure 5) were considered. V_1 and V_4 were not included because they are distant enough from the kink site and are addressing sites caused by the PBC and are not relevant for the kink growth process. Reweighting on V_2 and V_3 changes ΔF less than 0.5 kJ/mol (which is in the order of the overall accuracy of the WTMetaD sampling). The use of V_2 and V_3 is not necessary for simulations performed at undersaturated or around saturated conditions. Most APIs will not need these walls because their growth is kinetically hindered such that kink growth

Table 8: Values of adsorption site CVs. The coordinates origin is set to the box center.

		urea			naphthalene	
		MeCN	EtOH	MeOH	EtOH	MePh
	σ_1 [-]	80	80	80	80	80
	$x_{\mathrm{l}1}$ [nm]	0.8356	2.0742	2.0742	0.06	1.2931
	$x_{\mathrm{u}1}$ [nm]	1.1356	0.1558	0.1558	0.28	1.5131
	$y_{\mathrm{l}1}$ [nm]	1.9727	0.9620	0.9620	1.10	1.1205
s_1	$y_{\mathrm{u}1}$ [nm]	2.4789	1.3420	1.3420	1.45	1.4924
	$z_{\mathrm{l}1}$ [nm]	0.9405	1.0847	1.0847	1.31	−0.2679
	$z_{\mathrm{u}1}$ [nm]	1.2705	1.4147	1.4147	1.64	0.2521
	k_1 [kJ/mol]	30	30	30	30	30
	$s_{1,0}$ [nm]	0.2	0.2	0.2	0.2	0.2
	σ_2 [-]	80	80	80	80	80
	$x_{\mathrm{l}2}$ [nm]	0.8356	2.1642	2.1642	0.06	1.2931
	$x_{\mathrm{u}2}$ [nm]	1.1356	0.0958	0.0958	0.28	1.5131
	$y_{\mathrm{l}2}$ [nm]	2.4789	0.0520	0.0520	0.35	0.3176
s_2	$y_{\mathrm{u}2}$ [nm]	3.9973	0.8620	0.8620	1.06	1.0592
	$z_{\mathrm{l}2}$ [nm]	0.9405	1.0847	1.0847	1.31	−0.2679
	$z_{\mathrm{u}2}$ [nm]	1.2705	1.4147	1.4147	1.64	0.2521
	k_2 [kJ/mol]	5	5	5	5	5
	$s_{2,0}$ [nm]	1	1	1	1	1
	σ_3 [-]	80	80	80	80	80
	$x_{\mathrm{l}3}$ [nm]	0.4582	0.2801	0.2801	0.45	1.6804
	$x_{\mathrm{u}3}$ [nm]	0.6782	0.4801	0.4801	0.71	1.9404
	$y_{\mathrm{l}3}$ [nm]	0.2489	1.4980	1.4980	1.80	1.8020
s_3	$y_{\mathrm{u}3}$ [nm]	2.4789	2.2980	2.2980	2.65	2.6415
	$z_{\mathrm{l}3}$ [nm]	0.9405	1.0847	1.0847	1.31	−0.2679
	$z_{\mathrm{u}3}$ [nm]	1.2705	1.4147	1.4147	1.64	0.2521
	k_3 [kJ/mol]	5	5	5	5	5
	$s_{3,0}$ [nm]	1	1	1	1	1
	σ_4 [-]	80	80	80	80	80
	$x_{\mathrm{l}4}$ [nm]	2.5552	0.6374	0.6374	1.2092	2.4833
	$x_{\mathrm{u}4}$ [nm]	2.9552	0.8774	0.8774	1.6200	2.9933
	$y_{\mathrm{l}4}$ [nm]	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
s_4	$y_{\mathrm{u}4}$ [nm]	∞	∞	∞	∞	∞
	$z_{\mathrm{l}4}$ [nm]	0.9405	1.0847	1.0847	1.120	−0.2679
	$z_{\mathrm{u}4}$ [nm]	1.2705	1.4147	1.4147	1.760	0.2521
	k_4 [kJ/mol]	30	30	30	30	30
	$s_{4,0}$ [nm]	1	1	1	1	1
	σ_4 [-]	80	80	80	-	-
	$x_{\mathrm{l}4}$ [nm]	$-\infty$	$-\infty$	$-\infty$	-	-
	$x_{\mathrm{u}4}$ [nm]	∞	∞	∞	-	-
	$y_{\mathrm{l}4}$ [nm]	$-\infty$	$-\infty$	$-\infty$	-	-
s_5	$y_{\mathrm{u}4}$ [nm]	∞	∞	∞	-	-
	$z_{\mathrm{l}4}$ [nm]	−1.35	1.0847	1.0847	-	-
	$z_{\mathrm{u}4}$ [nm]	−0.90	1.4147	1.4147	-	-
	k_5 [kJ/mol]	40	40	40	-	-
	$s_{5,0}$ [nm]	0	0	0	-	-

events are rare within the simulation time span of ~ 1 μs.

Crystallinity CVs

The reference atoms used for the crystallinity CVs, s_{c1} and s_{c2}, of urea and naphthalene are shown in Figure 6. The parameter values of s_{c1} and s_{c2} are listed in Table 9 and are chosen such that their values close to 1 correspond to a fully crystalline molecule at the biased kink site and values around 0 correspond to a fully dissolved biased kink site. Figure 7 shows the contour lines of the crystallinity CVs for the case of urea. The graph shows the histogram of the carbon atom positions (left) and the oxygen atom positions (right) at the crystal surface together with the contour lines of s_{c1} and s_{c2}, which take the crystalline carbon atom position and oxygen atom position respectively at the biased kink site as references. The values of s_{c1} and s_{c2} which are between 0 and 1 correspond roughly to the urea atom positions within the region of transition, which exhibit the lowest density in the histogram in the surroundings of the biased kink site. The region of transition coincides approximately with the space between the nearest neighbors.

It suffices to take only the solute positions at the biased kink site into consideration for the crystallinity CVs, while neglecting the solvent, since the states of the biased kink site containing vacuum are very short lived and immediately refilled either with solvent or solute.

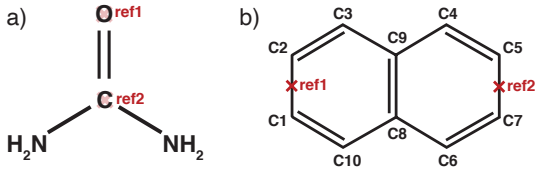

Figure 6: Reference positions of the crystallinity CVs, s_{c1} and s_{c2}. a) Urea: oxygen and carbon atom positions. b) Naphthalene: center of mass of carbon atom pairs C1-C2 and C5-C7 (which are interchangeable due to symmetry).

Table 9: Values of crystallinity CVs used in the reweighting.

		urea			naphthalene	
		MeCN	EtOH	MeOH	EtOH	MePh
s_{c1}	$r_{c1}^{(x)}$ [nm]	2.8366	1.1369	1.1384	1.8996	3.1273
	$r_{c1}^{(y)}$ [nm]	2.2038	1.1095	1.1091	1.3968	1.3919
	$r_{c1}^{(z)}$ [nm]	1.1404	1.2844	1.4044	1.2767	−0.0964
	σ_{c1} [-]	70	70	70	70	70
	d_{c1} [nm]	0.15	0.15	0.15	0.20	0.20
s_{c2}	$r_{c2}^{(x)}$ [nm]	2.8366	1.1372	1.1383	1.7266	2.9597
	$r_{c2}^{(y)}$ [nm]	2.2038	1.2304	1.2296	1.3119	1.3049
	$r_{c2}^{(z)}$ [nm]	1.1404	1.2827	1.4027	1.7121	0.3338
	σ_{c2} [-]	70	70	70	70	70
	d_{c2} [nm]	0.15	0.15	0.15	0.20	0.20

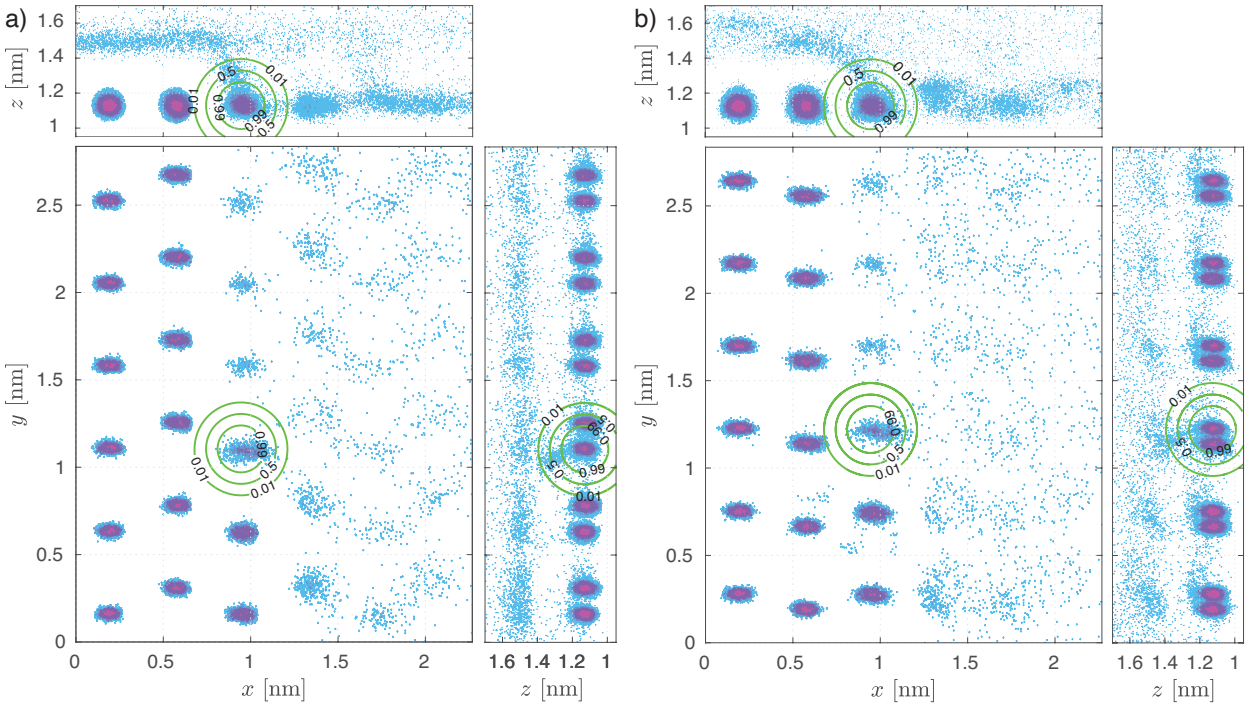

Figure 7: Histogram of urea atom positions of the crystal surface layer together with the contour lines of the crystallinity CVs (green lines) projected along all three spatial directions. a) Carbon atoms and s_{c1} contours. b) Oxygen atoms and s_{c2} contours.

Sampling of solubility with chemically distinct kink sites

ΔF should not depend on the kink site, as long as the growth unit corresponds to a single molecule along the edge of interest. To quantitatively verify this assumption we have run kink growth simulations of the four chemically distinct kink sites of urea face {110} with the unfinished surface layer cut along edge [001] and grown in ethanol at a solute mole fraction of $x = 0.020$. All four kink sites were biased simultaneously. The visualization of the crystal surface is shown in Figure 8a) and the unfinished surface layer with labeled kink sites ks1-4 is shown along the z-axis in Figure 8b).

The simulation convergence of growth and dissolution of the kink sites is presented in Figure 8c), which shows the time evolution of the energy difference of the grown and dissolved states $\Delta F(s_{\text{b1-4}})$ reweighted over the biased CVs, $s_{\text{b1-4}}$. The corresponding $F(s_{\text{b1-4}})$ averaged over the last 200 ns are shown in Figure 8d), which clearly show that the solubility is the same for all 4 kink sites (the energy differences are within the accuracy of the method of ~ 0.5 kJ/mol). The values of ΔF, obtained with reweighting on the crystallinity CVs, are: $\Delta F_{\text{ks1}} = 0.05$ kJ/mol, $\Delta F_{\text{ks1}} =$ -0.03 kJ/mol, $\Delta F_{\text{ks1}} = 0.04$ kJ/mol, and $\Delta F_{\text{ks1}} = 0.33$ kJ/mol. These values are in agreement with the ones reported in the main manuscript, which were obtained with a smaller simulation box setup.

It is interesting to note, that the activation energy barriers of $F(s_{\text{b1-4}})$ in Figure 8d) are smaller for kink sites, which face an oxygen atom of the unfinished row (ks1 and ks4) while the activation energy barriers of the kink sites facing amine groups of the unfinished row are slightly larger (ks2 and ks3).

References

(1) Abraham, M. J.; Murtola, T.; Schulz, R.; Pall, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.

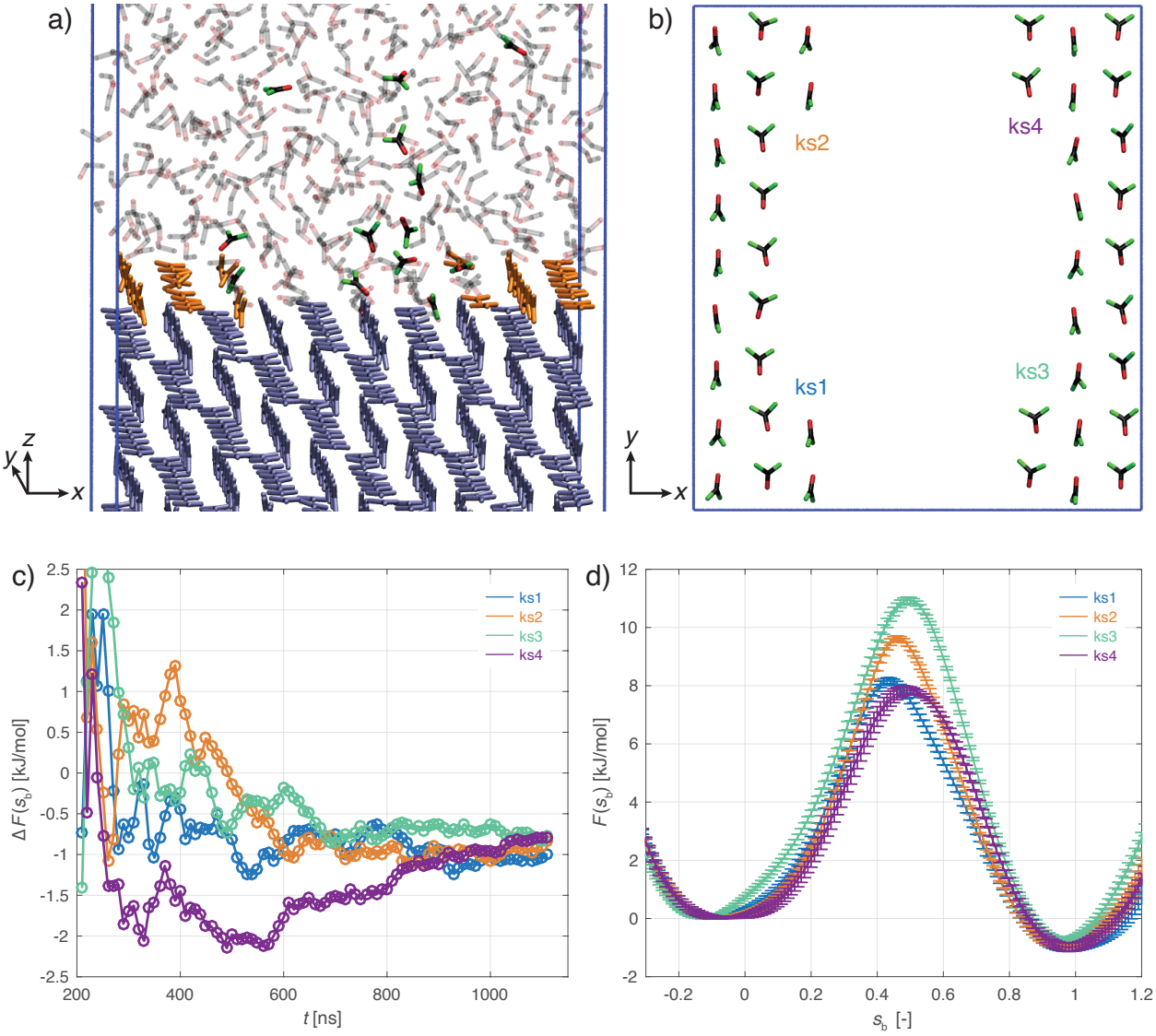

Figure 8: Biased simulations of urea grown in ethanol for four chemically distinct kink sites: a) Visualization of the crystal surface of the simulation box. b) Unfinished surface layer with labeled kink sites ks1-4. c) Time evolution of the energy difference of the grown and dissolved states $\Delta F(s_{\text{b1-4}})$ reweighted over the biased CVs, $s_{\text{b1-4}}$. d) Corresponding $F(s_{\text{b1-4}})$ averaged over the last 200 ns.

(2) Bussi, G.; Zykova-Timan, T.; Parrinello, M. Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *J. Chem. Phys.* **2009**, *130*, 074101.

(3) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(4) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.

(5) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(6) Sklar, N.; Senko, M. E.; Post, B. Thermal effects in urea: the crystal structure at –140°C and at room temperature. *Acta Crystallogr.* **1961**, *14*, 716–720.

(7) Cruickshank, D. W. J. A detailed refinement of the crystal and molecular structure of naphthalene. *Acta Crystallogr.* **1957**, *10*, 504–508.

(8) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(9) Perego, C.; Salvalaglio, M.; Parrinello, M. Molecular dynamics simulations of solutions at constant chemical potential. *J. Chem. Phys.* **2015**, *142*, 144113.

(10) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603.