Supplementary Information.

Systematic Finite-Temperature Reduction of Crystal Energy Landscapes.

Nicholas F. Francia,† Louise S. Price,‡ Jonas Nyman,¶§ Sarah L. Price,‡ and Matteo Salvalaglio*†

†Thomas Young Centre and Department of Chemical Engineering, University College London, London WC1E 7JE, UK.
‡Department of Chemistry, University College London, London WC1H OAJ, UK.
¶Small Molecule Design & Development, Eli Lilly and Company, Indianapolis, Indiana 46285, USA
§Holly Chemistry, Hangarvägen 13, S-69135 Karlskoga, Sweden

E-mail: m.salvalaglio@ucl.ac.uk
Contents

S1 CSP_0 study and comparison with the GAFF Lattice Energies 3

S2 Fingerprints and cluster analysis 6
  S2.1 Urea ......................................................... 6
  S2.2 Succinic Acid ............................................. 10

S3 Tables 12
  S3.1 Urea ......................................................... 12
  S3.2 Succinic Acid ............................................. 18

References 24
Figure S1: Lattice energy vs density plot of structures minimized at 0 K using the GAFF force field for the urea (A) and succinic acid (B) systems. (C-D) Comparison between the lattice energies of the different methods involved. Structures are ordered by the lattice energy ranking of $\Psi_{mol}$, enhancing the previously known crystal forms.

S1 CSP 0 study and comparison with the GAFF Lattice Energies

After the minimization with GAFF, we calculated the lattice energy of each structure (SI Tab. S4 for urea and SI Tab. S5-S6 for succinic acid) defined as the energy difference between a molecule in the lattice and a molecule in vacuum:

$$E^{\text{latt}} = \frac{E^{\text{crystal}}}{n_{\text{mols}}} - E^{\text{vacuum}}$$  \hspace{1cm} (1)

For urea, the energies obtained with GAFF$^1$ are comparable with the $\Psi_{mol}$ ones but the ranking is different. In Fig. S1C, we compare the two methods by showing the lattice energy difference relative to the $\Psi_{mol}$ global minimum of each structure. The known urea structures optimized with
Figure S2: (A) The dominant planar conformation (aAa) of succinic acid and (B) the top and side view of conformer gA-g. (C) The H-bonding networks with motifs $C_1^1(4)\bar{R}_2^2(6)$ of structures S1107, S18857 and S15665 and (D) $C_1^1(4)$ of structures NC1, NC2 and NC3, that contain molecules in the conformation shown in part B.

the GAFF force field, all appear among the lowest energy structures. Based on GAFF calculations the experimental forms I and IV rank among the 10 most stable urea structure, while Form A emerges as the most stable. Based on $\Psi_{mol}$ calculations experimental forms I and IV rank 60th and 41st - with an energy difference with respect to the global minimum of 9.0 kJ mol$^{-1}$ and 7.8 kJ mol$^{-1}$, respectively.

For succinic acid, the $\Psi_{mol}$ ranking$^2$ identifies form $\beta$ as the global minimum while using GAFF $\beta$ is in 6th position with a difference of 0.6 kJ mol$^{-1}$ with respect to the most stable structure. In the $\Psi_{crys}$ set, the global energy minimum is the labile polymorph $\gamma$, with $\alpha$ at +0.3 kJ mol$^{-1}$ and $\beta$ at +1.6 kJ mol$^{-1}$. Looking at the experimental structures, the GAFF force field is in good agreement with the $\Psi_{mol}$ CSP.0 study results showing the $\beta$ polymorph as the most stable, followed by $\gamma$ at +4.34 kJ mol$^{-1}$ and $\alpha$ at 6.91 kJ mol$^{-1}$. Fig. S1D shows that the lattice energies are very sensitive to the method used with variations of similar magnitude between the three methods. In general, differences are smaller compared to urea (Fig. S1C), with the exception of structures S1107, S18857 and S15665 which are the only ones that do not exhibit the $R_2^2(8)$ H-bonding motif. Structures
S1107 and S18857 have the H-bonding motif $R^2_2(6)C_1^1(4)$ (Fig. S2C) and the molecules in a planar conformation. In WTmetaD simulations, they convert to geometries that exhibit the succinic acid in the gA-g conformation, with the C-C-C-O dihedral angles around 90°. The resulting H-bonding network maintains the $C_1^1(4)$ chain, shown in Fig. S2D, but not the ring as the O-H hydrogen bond that forms the ring motif passes from 2.03 Å to 3.76 Å.

The structure closest to the high-temperature α polymorph presents a slightly different geometry with respect to the experimental structure SUCACB07 available in the CSD. The typical $R^2_2(8)$ H-bonding motif and chain orientation within each layer is maintained but the layer stacking changes from “ABCABC...” in the experimental form to “ABABAB...” in structure S9393 resulting in a $\text{RMSD}_{30} = 1.052$ Å. Despite this, after equilibration these two structures are grouped in the same cluster.

Table S1: Comparison between the experimental α structure from the CSD and the best match from the CSP.0 study, showing the unit cell, density, packing motif of the experimental structure and of its closest approximation identified by $\Psi_{mol}$. The the lattice energy obtained with GAFF forcefield for the two structures is reported in the last row.

| Structures | CSD - SUCACB07 | $\Psi_{mol}$ |
|------------|----------------|-------------|
| Unit cell  | ![Image](image1.png) | ![Image](image2.png) |
| Density / $g \cdot cm^{-3}$ | 1.531 | 1.534 |
| Layer stacking | ABCABC... | ABABAB... |
| $E_{\text{att}}$(GAFF) / $kJ\cdot mol^{-1}$ | -126.26 | -125.93 |
Figure S3: The two fingerprints considered for the cluster analysis of the urea set: the distributions of (A) the distances between each center of mass and (B) the intermolecular orientations obtained by computing the angles between vectors $\overrightarrow{CO}_i \cdot \overrightarrow{CO}_j$ and $\overrightarrow{NN}_i \cdot \overrightarrow{NN}_j$ of each pair of molecules, $i$ and $j$.

S2 Fingerprints and cluster analysis

S2.1 Urea

In general, we can define a structural fingerprint for each structure, $i$, as a set of probability densities of the relative position, relative orientation and relevant conformational degrees of freedom:

$$F_i = \{ p_i(r) \ p_i(\theta) \ p_i(\Phi) \}$$  \hspace{1cm} (2)

We assume that urea is a rigid molecule so that the fingerprint depends only on the relative position and orientation. The different geometries are distinguished by using:

- The distribution of the distances between each pair of centers of mass, $p_i(r_{COM})$

- The 2D distribution of the intermolecular orientation given by the angles $\theta$ and $\psi$ between each pair of vectors $\overrightarrow{CO}$ and $\overrightarrow{NN}$, $p_i(\theta, \psi)$

Both these distributions are generated by post-processing the final 500 ps of a molecular dynamics trajectory using PLUMED2.\textsuperscript{3} However, some precautions must be considered. The distances between centers of mass are calculated without considering periodic boundary conditions. The result is that supercells of the same geometry but of different size will generate different distributions as the distance increases, as shown in Fig. S4A. For this reason, the distribution is truncated at
Figure S4: (A) Effect of using supercells of different sizes on the distribution of centers of mass for the orthogonal structure U98 with cell vectors \( \mathbf{a} = 7.5777 \), \( \mathbf{b} = 8.4116 \), \( \mathbf{c} = 7.5048 \). (B) In these three different geometries, we can see that the angle between the C-O vector of two molecules remains constant. On the other hand, using the N-N vector we can distinguish the geometries typical of form A (top), form IV (middle), and form I (bottom). (C) One and two dimensional distributions of form I (left) and structure U146 (right).

the length of half of the shortest box vector. When computing the dissimilarity between centers of mass distributions the Hellinger distance is computed on a domain bounded by the shortest half-box distance. In Fig. S4A, we can see the difference in the distribution between two supercells obtained by replicating the same unit cell of structure U98, two or four times in each direction. From \( \sim 0.75 \) nm (half of the smallest box vector) the two distributions differ from each other and cannot be used for the comparison.

The use of a bivariate distribution of relative orientations in the definition of the urea fingerprint is
motivated by the fact that the angle $\theta$ (between C-O vectors), used to identify melted structures, was unable to identify unequivocally changes in the relative orientation of urea molecules, as shown in Fig. S4B. The introduction of the $\psi$ intermolecular angle is crucial, to distinguish form I, form IV and form A which all share overlapping probability distributions of $\theta$. Furthermore, the use of two univariate distributions, $p_i(\theta)$ and $p_i(\psi)$, failed to distinguish some structures, as shown in Fig. S4C where the form I structure is compared with a similar structure that exhibits stacking disorder along one axis. In this case, the two-dimensional distributions corresponding to different packings are immediately distinguishable while the one-dimensional ones overlap.

After generating these fingerprints for every structure and normalizing the distributions, we need to define a distance between each pair of structures and generate a (dis)similarity matrix (here labeled with $\Delta$). To this aim, we use the Hellinger distance, $H_{ij}$, a measure of the overlap between two normalized distributions.$^4$ For each pair of structures, $i$ and $j$, we calculated $H_{ij}^{\text{COM}}$ and $H_{ij}^{\text{CO,NN}}$, defined as:

$$H_{ij}^{\text{COM}} = \sqrt{1 - \int_{0}^{r_{\text{min}}} \sqrt{p_i(r_{\text{COM}}) p_j(r_{\text{COM}})} \, dr_{\text{COM}}}$$

(3)
$$H_{i,j}^{CO,NN} = \sqrt{1 - \int_{0}^{\pi} \int_{0}^{\pi} \sqrt{p_i(\theta, \psi) \ p_j(\theta, \psi)} \ d\theta d\psi}$$  \hspace{1cm} (4)$$

where $p_i$ and $p_j$ are the previously generated probability distributions and the integral term corresponds to the Bhattacharyya coefficient ($BC(p_i, p_j)$). The latter term quantifies the similarity between two probability distributions and ranges from 0 (completely different) to 1 (identical). Therefore, a Hellinger distance equal to 0 refers to identical distributions while values from 0 to 1 indicate increasingly different ones. The Hellinger distance is inherently symmetric, and numerically well behaved when dealing with concentrated probability densities.

In order to equally weight the contribution of the Hellinger distances associated with the elements of the fingerprint in the calculation a dissimilarity between structure $i$ and $j$ we rescale $H_{i,j}^{COM}$ and $H_{i,j}^{CO,NN}$ by their respective maximum values computed over the entire set of structure pairs. The norm of the vector of rescaled Hellinger distances, computed over all the distributions included in the fingerprint and appropriately normalized, defines a distance between two structures $\Delta_{i,j}$:

$$\Delta_{i,j} = \left\| \left[ \frac{H_{i,j}^{COM}}{\max(H^{COM})}, \frac{H_{i,j}^{CO,NN}}{\max(H^{CO,NN})} \right] \right\| \left\| \left[ \frac{1}{\max(H^{COM})}, \frac{1}{\max(H^{CO,NN})} \right] \right\|$$  \hspace{1cm} (5)

The resulting distance matrix is shown in Fig. S5A. We use the dissimilarity matrix to perform a clustering via the Fast Search and Find of Density Peaks (FSFDP) method proposed by Rodriguez & Laio,\(^5\) which does not require prior knowledge of the number of clusters. Starting from the dissimilarity matrix, for each structure $i$, FSFDP calculates the density parameter, $\rho_i$, that indicates how many other structures are close to $i$, and the distance parameter, $\sigma_i$, that is the minimum distance between $i$ and another structure of higher density as:

$$\rho_i = \sum_j e^{-\left(\frac{d_{ij}}{\sigma_i}\right)}$$  \hspace{1cm} (6)

$$\sigma_i = \min_{j: \rho_j > \rho_i} d_{ij}$$  \hspace{1cm} (7)

By plotting $\rho$ vs $\sigma$, as in Fig. S5B, structures characterized by high values of both $\sigma$ and $\rho$ are identified as cluster centers. As we aim to maintain also unique structures, that exhibit $\rho = 0$, a
Table S2: Clusters identified following the finite T MD simulation.

| Cluster Centers | Structures               |
|-----------------|--------------------------|
| Form A          | U172 U177 U205 U2386 U289 U381 U730 U871 U1023 U2444 U449 U242 U374 U645 U680 U682 U325 U445 U600 U2363 |
| Form IV         | U1326 U341 U998 U63 U275 U2405 U1527 U411 |
| Form B          | U933 U939 U165 U836 U849 U899 U146 U146 U2012 U1816 U77 U951 |
| U115            | U2383 U2404 U690 U115 |
| U2223           | U2223 U609 U273 |
| Form I          | U472 U98 |
| U100            | U1011 U100 |
| U802            | U215 U802 |
| U2403           | U589 U2403 |
| Unique Structures | U1140 U1179 U2449 U302 U321 U499 U562 U628 U771 U792 U81 U1831 U1955 U49 U638 U721 U785 U1540 U1960 U2015 |

density cut-off is not applied and the only parameter to optimize is the $\sigma$ cut-off. This parameter discriminates between cluster centers, and all other structures that will be later assigned to a cluster. In this particular application, an ideal cutoff choice should allow to identify as separate structures all the high-density cluster centers as well as all isolated structures that have survived after the finite-temperature simulation steps. In the cases examined in this work the chosen cutoff is of the order of $\sigma \simeq 0.1$. For urea, a distance cut-off equal to 0.05 was found to be a suitable value to group crystal structures. In Tab. S2, we show the resulting 30 cluster centers with the structures assigned to them.

S2.2 Succinic Acid

Flexible molecules need a further set of fingerprints to describe their different conformations. For succinic acid, a new distribution of the torsional angles of the carbon chain, $p_i(\phi)$, is implemented to distinguish the planar and folded geometries. As discussed in the previous case we consider the distribution of the distances between centers of mass, $p_i(r_{\text{COM}})$. Two vectors have been used to define the molecular orientation: one connecting the two central carbon atoms ($\theta$) and the other connecting the terminal ones ($\psi$). The angle $\theta$ can distinguish how the central carbon atoms of a molecule are located with respect to molecules in adjacent chains or layers, as shown in Fig. S6A. The angle $\psi$ helps to define how different layers are oriented, as shown in Fig. S6B, and identifies the packing motifs of each structure.
The resulting distance matrix and decision graph are shown in Fig. S6C and S6D. The latter shows many cluster centers located around \( \rho = 1 \), as a consequence of the structures that the \( \Psi_{\text{crys}} \) and \( \Psi_{\text{mol}} \) sets have in common. For succinic acid, the distance cut-off is taken as \( \sigma = 0.115 \), which allow to retain for further analysis all the unique structures (\( \rho = 0 \)). The FSFDP method produced 32 clusters, ordered in Tab. S3 according to their respective number of members.

![Figure S6](image)

Figure S6: (A) Possible relative orientations of two succinic acid molecules belonging to different chains. (B) Examples of the different orientation between layers recognized analyzing the \( \Psi \) angle. (C) Distance matrix used for the FSFDP clustering method and resulting decision graph (D). In the latter, we show with different colors the most populated cluster centers, in grey other small clusters and with a black circle unique structures. All other structures, shown with a red circle, will be assigned to a cluster by the clustering algorithm.
Table S3: Clusters identified following the finite T MD simulation, divided by the two sets. Experimental forms $\beta$, $\alpha$, and $\gamma$ are labelled E1, E2 and E3, respectively.

| Clusters | $\Psi_{mol}$ Set | $\Psi_{crys}$ Set |
|----------|------------------|------------------|
| Form $\alpha$ | S10263 S1488 S23063 S2779 S3882 S4158 S9393 S16988 S4846 | S10263 S1488 S23063 S2779 S3882 S4158 S9393 S2778 E2 |
| S763 | S19123 S20462 S763 S804 | S1161 S19123 S20462 S22937 S763 |
| S2526 | S2396 S2407 S2526 S3917 | S2396 S2407 S2526 S3917 |
| S10824($\Psi_{crys}$) | S10824 S1904 S4816 | S10824 S1904 S4816 |
| Form $\beta$ | S10263 S1488 S23063 S2779 S3882 S4158 S9393 S16988 S4846 | S10263 S1488 S23063 S2779 S3882 S4158 S9393 S2778 E2 |
| S2585 | S11173 S1161 S2585 | S11173 S2585 |
| Form $\gamma$ | S14464 S19217 | S19217 E3 |
| S4200 | S2778 S4200 | S4200 |
| S123 | S2585 | S1161 |
| S1308 | S1308 | S1308 |
| S1667 | S1667 | S1667 |
| S20629 | S20629 | S20629 |
| S3841 | S3841 | S3841 |
| S8149 | S8149 | S8149 |
| S18857($\Psi_{crys}$) | S18857 | S18857 |
| S2473($\Psi_{crys}$) | S2473 | S2473 |
| S2478($\Psi_{crys}$) | S2478 | S2478 |
| S2798($\Psi_{crys}$) | S2798 | S2798 |
| S4939($\Psi_{crys}$) | S4939 | S4939 |
| Unique Structures | S1107 S11752 S22937 S2910 S4947 S618 | S11752 S14464 S16988 S4846 S4947 S618 S804 |

S3 Tables

S3.1 Urea

In Tab. S4, we show the evolution of the putative polymorphs of urea through the different steps. Structures are ordered as their rank in the CSP,0 stage. The energies are expressed in kJ mol$^{-1}$ but while in the CSP and energy minimization (EM) with GAFF steps they are lattice energies, in the NVT and NPT ones they are an average of the potential energy difference between a molecule in the crystal and a molecule in vacuum.
Table S4: The energies of the CSP.0-generated structures at different steps and the cluster centers assigned by the FSFDP clustering algorithm. Melted structures are shown in bright red.

| Rank | IDs    | CSP($\Psi_{mol}$) kJ mol$^{-1}$ | EM(GAFF) kJ mol$^{-1}$ | NVT kJ mol$^{-1}$ | NPT kJ mol$^{-1}$ | Clusters   |
|------|--------|----------------------------------|------------------------|------------------|------------------|------------|
| 1    | U242   | -95.83                           | -107.50                | -76.60           | -75.51           | Form A     |
| 2    | U374   | -95.67                           | -106.54                | -76.12           | -75.27           | Form A     |
| 3    | Form B | -94.96                           | -104.09                | -74.00           | -71.27           | Form B     |
| 4    | U2145  | -94.8                             | -104.71                | -74.39           | melted           |
| 5    | U680   | -94.73                           | -100.31                | -70.74           | -75.61           | Form A     |
| 6    | U682   | -94.73                           | -100.29                | -70.75           | -75.47           | Form A     |
| 7    | U2147  | -94.67                           | -101.20                | -70.51           | melted           |
| 8    | Form A | -94.19                           | -108.18                | -78.10           | -75.62           | Form A     |
| 9    | U804   | -93.89                           | -99.57                 | -68.51           | melted           |
| 10   | U2015  | -93.88                           | -104.49                | -73.86           | -70.70           | U2015      |
| 11   | U899   | -93.72                           | -101.76                | -72.52           | -71.25           | Form B     |
| 12   | U836   | -93.28                           | -100.48                | -73.12           | -71.22           | Form B     |
| 13   | U1540  | -92.98                           | -99.61                 | -68.59           | -69.02           | U1540      |
| 14   | U1960  | -92.87                           | -101.43                | -70.95           | -65.24           | U1960      |
| 15   | U273   | -92.76                           | -100.89                | -70.29           | -68.20           | U2223      |
| 16   | U645   | -92.47                           | -102.92                | -72.56           | -75.62           | Form A     |
| 17   | U165   | -92.46                           | -103.54                | -73.23           | -71.24           | Form A     |
| 18   | U2403  | -91.25                           | -97.65                 | -69.04           | -68.98           | U2403      |
| 19   | U275   | -91.03                           | -99.65                 | -72.14           | -74.25           | Form IV    |
| 20   | U100   | -90.98                           | -100.34                | -72.70           | -70.72           | U100       |
| 21   | U1816  | -90.76                           | -102.52                | -72.38           | -71.39           | U146       |
| 22   | U77    | -90.7                            | -101.96                | -71.89           | -70.84           | U146       |
| 23   | U207   | -90.25                           | -96.73                 | -66.83           | melted           |
| 24   | U104   | -90.23                           | -97.97                 | -67.78           | melted           |
| 25   | U2062  | -90.19                           | -94.69                 | -66.45           | melted           |
Table S4: The energies of the CSP.0-generated structures at different steps and the cluster centers assigned by the FSFDP clustering algorithm. Melted structures are shown in bright red.

| Rank | IDs   | CSP($\Psi_{mol}$) kJ mol$^{-1}$ | EM(GAFF) kJ mol$^{-1}$ | NVT kJ mol$^{-1}$ | NPT kJ mol$^{-1}$ | Clusters       |
|------|-------|----------------------------------|------------------------|------------------|------------------|----------------|
| 26   | U1527 | -90.14                           | -98.58                 | -67.82           | -72.84           | Form IV        |
| 27   | U445  | -90.03                           | -102.36                | -72.27           | -75.29           | Form A         |
| 28   | U951  | -89.96                           | -102.28                | -72.11           | -71.10           | U146           |
| 29   | U1873 | -89.9                            | -99.08                 | -69.16           | melted           |
| 30   | U1428 | -89.84                           | -99.63                 | -70.54           | melted           |
| 31   | U2405 | -89.82                           | -98.04                 | -73.67           | -74.17           | Form IV        |
| 32   | U325  | -89.8                            | -97.17                 | -66.10           | -75.40           | Form A         |
| 33   | U650  | -89.41                           | -97.04                 | -67.05           | melted           |
| 34   | U34   | -89.29                           | -101.65                | -70.96           | melted           |
| 35   | U63   | -89.27                           | -99.65                 | -73.47           | -74.04           | Form IV        |
| 36   | U147  | -88.49                           | -92.44                 | melted           |
| 37   | U115  | -88.49                           | -99.01                 | -67.56           | -71.76           | U115           |
| 38   | U122  | -88.37                           | -93.22                 | melted           |
| 39   | U410  | -88.3                            | -98.91                 | -68.67           | melted           |
| 40   | U98   | -88.09                           | -98.55                 | -68.63           | -71.80           | Form I         |
| 41   | Form IV| -88.02                         | -106.14                | -76.01           | -74.32           | Form IV        |
| 42   | U1955 | -88.01                           | -98.06                 | -67.28           | -74.61           | U1955          |
| 43   | U449  | -87.94                           | -92.12                 | -68.95           | -71.66           | Form A         |
| 44   | U600  | -87.93                           | -91.48                 | -71.10           | -73.19           | Form A         |
| 45   | U1831 | -87.91                           | -92.98                 | -67.87           | -70.16           | U1831          |
| 46   | U49   | -87.9                            | -100.49                | -70.23           | -68.13           | U49            |
| 47   | U721  | -87.89                           | -92.06                 | -60.11           | -68.45           | U721           |
| 48   | U638  | -87.84                           | -96.98                 | -65.81           | -71.11           | U638           |
| 49   | U532  | -87.76                           | -92.09                 | -68.97           | melted           |
| 50   | U221  | -87.44                           | -96.26                 | melted           |
Table S4: The energies of the CSP\_0-generated structures at different steps and the cluster centers assigned by the FSFDP clustering algorithm. Melted structures are shown in bright red.

| Rank | IDs   | CSP($\Psi_{mol}$) kJ mol\(^{-1}\) | EM(GAFF) kJ mol\(^{-1}\) | NVT kJ mol\(^{-1}\) | NPT kJ mol\(^{-1}\) | Clusters   |
|------|-------|----------------------------------|------------------------|----------------|----------------|-------------|
| 51   | U82   | -87.41                           | -98.97                 | -68.66         |               | melted      |
| 52   | U467  | -87.37                           | -98.01                 | -69.01         |               | melted      |
| 53   | U16   | -87.32                           | -101.51                | -71.37         |               | melted      |
| 54   | U2444 | -87.13                           | -99.84                 | -74.40         | -75.52        | Form A      |
| 55   | U603  | -87.04                           | -91.92                 |               |               | melted      |
| 56   | U1023 | -87.01                           | -93.54                 | -70.79         | -73.72        | Form A      |
| 57   | U785  | -86.93                           | -88.70                 | -70.46         | -72.10        | U785        |
| 58   | U976  | -86.87                           | -96.97                 | -66.81         |               | melted      |
| 59   | U238  | -86.85                           | -95.53                 |               |               | melted      |
| 60   | Form I| -86.83                           | -104.33                | -73.62         | -71.74        | Form I      |
| 61   | U730  | -86.76                           | -102.17                | -71.10         | -74.80        | Form A      |
| 62   | U589  | -86.69                           | -94.30                 | -64.81         | -69.16        | U2403       |
| 63   | U939  | -86.6                            | -96.43                 | -69.81         | -71.23        | Form B      |
| 64   | U172  | -86.48                           | -99.49                 | -72.07         | -71.98        | Form A      |
| 65   | U1011 | -86.44                           | -98.13                 | -73.63         | -70.88        | U100        |
| 66   | U871  | -86.43                           | -92.63                 | -70.70         | -71.87        | Form A      |
| 67   | U1326 | -86.34                           | -91.66                 | -67.06         | -72.76        | Form IV     |
| 68   | U2386 | -86.27                           | -93.85                 | -73.33         | -75.66        | Form A      |
| 69   | U628  | -86.15                           | -91.17                 | -69.10         | -71.55        | U628        |
| 70   | U1461 | -86.06                           | -96.07                 |               |               | melted      |
| 71   | U1541 | -85.94                           | -104.12                | -73.72         |               | melted      |
| 72   | U315  | -85.87                           | -94.25                 |               |               | melted      |
| 73   | U177  | -85.85                           | -98.50                 | -67.03         | -74.85        | Form A      |
| 74   | U499  | -85.84                           | -104.18                | -73.86         | -71.80        | U499        |
Table S4: The energies of the CSP.0-generated structures at different steps and the cluster centers assigned by the FSFDP clustering algorithm. Melted structures are shown in bright red.

| Rank | IDs   | CSP($\Psi_{mol}$) kJ mol$^{-1}$ | EM(GAFF) kJ mol$^{-1}$ | NVT kJ mol$^{-1}$ | NPT kJ mol$^{-1}$ | Clusters  |
|------|-------|----------------------------------|------------------------|-------------------|-------------------|-----------|
| 76   | U925  | -85.83                           | -92.36                 |                  |                   |           |
| 77   | U388  | -85.8                             | -95.48                 |                  |                   |           |
| 78   | U609  | -85.75                           | -96.44                 | -65.95           | -68.18            | U2223     |
| 79   | U206  | -85.72                           | -96.62                 | -66.31           |                  | melted    |
| 80   | U2325 | -85.69                           | -89.82                 |                  |                   |           |
| 81   | U690  | -85.66                           | -98.41                 | -67.28           | -71.73            | U115      |
| 82   | U341  | -85.59                           | -93.74                 | -68.78           | -74.17            | Form IV   |
| 83   | U562  | -85.51                           | -91.11                 |                  |                   |           |
| 84   | U577  | -85.51                           | -91.11                 |                  |                   |           |
| 85   | U2449 | -85.48                           | -96.70                 | -65.84           | -66.97            | U2449     |
| 86   | U1328 | -85.46                           | -97.33                 | -70.10           |                  | melted    |
| 87   | U302  | -85.36                           | -99.91                 | -69.81           | -68.61            | U302      |
| 88   | U2681 | -85.32                           | -95.63                 | -65.54           |                  |           |
| 89   | U933  | -85.28                           | -91.93                 | -66.82           | -71.25            | Form B    |
| 90   | U533  | -85.25                           | -103.35                | -72.90           |                  |           |
| 91   | U285  | -85.23                           | -92.22                 |                  |                   |           |
| 92   | U205  | -85.18                           | -94.96                 | -71.23           | -74.02            | Form A    |
| 93   | U1568 | -85.17                           | -89.51                 |                  |                   |           |
| 94   | U321  | -85.15                           | -98.52                 | -67.17           | -67.84            | U321      |
| 95   | U1140 | -85.1                             | -91.51                 | -59.51           | -58.33            | U1140     |
| 96   | U218  | -85.09                           | -95.19                 | -67.52           |                  |           |
| 97   | U2345 | -85.08                           | -94.44                 |                  |                   |           |
| 98   | U998  | -85.01                           | -91.61                 | -68.07           | -72.85            | Form IV   |
| 99   | U1504 | -85.0                            | -96.67                 | -66.76           |                  |           |
| 100  | U81   | -84.98                           | -98.47                 | -68.34           | -69.93            | U81       |
Table S4: The energies of the CSP_0-generated structures at different steps and the cluster centers assigned by the FSFDP clustering algorithm. Melted structures are shown in bright red.

| Rank | IDs  | CSP($\Psi_{mol}$) kJ mol$^{-1}$ | EM(GAFF) kJ mol$^{-1}$ | NVT kJ mol$^{-1}$ | NPT kJ mol$^{-1}$ | Clusters |
|------|------|---------------------------------|------------------------|------------------|------------------|-----------|
| 101  | U792 | -84.93                          | -91.51                 | -58.96           | -64.86           | U792     |
| 102  | U802 | -84.93                          | -91.39                 | -58.84           | -67.23           | U802     |
| 103  | U1340| -84.89                          | -90.94                 | -70.08           | melted           |          |
| 104  | U771 | -84.88                          | -91.40                 | -58.92           | -67.59           | U771     |
| 105  | U146 | -84.88                          | -99.42                 | -67.24           | -71.15           | U146     |
| 106  | U2455| -84.87                          | -96.51                 | -65.33           | melted           |          |
| 107  | U1179| -84.86                          | -96.64                 | -65.86           | -64.70           | U1179    |
| 108  | U1132| -84.85                          | -87.29                 | -66.05           | melted           |          |
| 109  | U1801| -84.78                          | -92.35                 |                  | melted           |          |
| 110  | U408 | -84.71                          | -98.76                 | -68.73           | melted           |          |
| 111  | U46  | -84.65                          | -102.72                | -72.02           | melted           | U115     |
| 112  | U245 | -84.65                          | -95.74                 |                  | melted           |          |
| 113  | U2223| -84.65                          | -95.59                 | -65.69           | -68.22           | U2223    |
| 114  | U2012| -84.6                           | -98.93                 | -72.00           | -71.50           | U146     |
| 115  | U1189| -84.6                           | -91.09                 |                  | melted           |          |
| 116  | U2383| -84.57                          | -94.36                 | -67.94           | -70.83           | U115     |
| 117  | U288 | -84.57                          | -95.88                 | -66.32           | melted           |          |
| 118  | U914 | -84.53                          | -87.97                 |                  | melted           |          |
| 119  | U2404| -84.51                          | -94.28                 | -66.25           | -71.48           | U115     |
| 120  | U381 | -84.46                          | -96.14                 | -71.57           | -75.62           | Form A   |
| 121  | U1573| -84.36                          | -95.61                 | -65.92           | melted           |          |
| 122  | U297 | -84.35                          | -94.35                 |                  | melted           |          |
| 123  | U289 | -84.35                          | -93.78                 | -68.84           | -75.30           | Form A   |
| 124  | U215 | -84.33                          | -95.45                 | -67.07           | -66.82           | U802     |
### S3.2 Succinic Acid

As before, we show the evolution of the putative polymorphs of succinic acid through different steps. The CSP\((\Psi_{\text{crys}})\) column of Tab. S6 shows the relative lattice energies as a result of the \(\Psi_{\text{crys}}\) study. We show the experimental forms (optimized in the \(\Psi_{\text{crys}}\) set) as Form \(\beta\), Form \(\alpha\) and Form \(\gamma\).

Table S5: The energies of the \(\Psi_{\text{mol}}\) set of structures of succinic acid at different steps and the cluster centers assigned by the FSFDP clustering algorithm. The clustering method has been applied to the two sets together. Cluster centers are then distinguished by including \(\Psi_{\text{crys}}\) in parentheses. Melted structures are shown in bright red.

| Rank | IDs   | CSP\((\Psi_{\text{mol}})\) kJ mol\(^{-1}\) | EM(GAFF) kJ mol\(^{-1}\) | NVT kJ mol\(^{-1}\) | NPT kJ mol\(^{-1}\) | Clusters          |
|------|-------|-------------------------------------------|-------------------------|----------------|----------------|----------------|
| 1    | S23778| -110.54                                   | -133.17                 | -81.31         | -93.39         | Form \(\beta\)  |
| 2    | S8149 | -109.15                                   | -132.43                 | -80.46         | -91.75         | S8149          |
| 3    | S763  | -107.67                                   | -132.82                 | -81.20         | -91.91         | S763           |
| 4    | S20629| -106.55                                   | -130.54                 | -77.13         | -90.12         | S20629         |
| 5    | S4158 | -106.26                                   | -131.22                 | -79.18         | -93.94         | Form \(\alpha\) |
| 6    | S19217| -105.87                                   | -128.83                 | -76.84         | -88.75         | Form \(\gamma\) |
| 7    | S22937| -105.53                                   | -127.31                 | -75.23         | -86.74         | S22937         |
| 8    | S5862 | -105.17                                   | -129.31                 | -77.21         | -89.70         | S5862          |
| 9    | S11752| -104.89                                   | -133.02                 | -80.98         | -91.33         | S11752         |
| 10   | S2798 | -104.00                                   | -126.44                 | -74.29         | -85.85         | S2798\((\Psi_{\text{crys}})\) |
| 11   | S10824| -103.95                                   | -130.92                 | -78.68         | -88.40         | S10824\((\Psi_{\text{crys}})\) |
| 12   | S19123| -103.43                                   | -127.58                 | -80.38         | -92.32         | S763           |
| 13   | S16898| -102.88                                   | -126.03                 | -74.04         | -88.04         | S5862          |
| 14   | S1308 | -102.64                                   | -129.29                 | -77.47         | -90.95         | S1308          |
| 15   | S1904 | -102.50                                   | -129.61                 | -77.64         | -88.64         | S10824\((\Psi_{\text{crys}})\) |
| 16   | S20462| -102.50                                   | -124.57                 | -77.96         | -91.90         | S763           |
| 17   | S4846 | -102.39                                   | -125.93                 | -77.57         | -92.34         | Form \(\alpha\) |
| 18   | S1488 | -102.37                                   | -129.29                 | -79.27         | -92.15         | Form \(\alpha\) |
Table S5: The energies of the $\Psi_{mol}$ set of structures of succinic acid at different steps and the cluster centers assigned by the FSFDP clustering algorithm. The clustering method has been applied to the two sets together. Cluster centers are then distinguished by including $\Psi_{crys}$ in parentheses. Melted structures are shown in bright red.

| Rank | IDs    | CSP($\Psi_{mol}$) kJ mol$^{-1}$ | EM(GAFF) kJ mol$^{-1}$ | NVT kJ mol$^{-1}$ | NPT kJ mol$^{-1}$ | Clusters |
|------|--------|---------------------------------|------------------------|------------------|------------------|----------|
| 19   | S2407  | -102.02                         | -129.43                | -77.51           | -90.26           | S2526    |
| 20   | S4200  | -101.97                         | -126.90                | -74.63           | -88.19           | S4200    |
| 21   | S804   | -101.62                         | -130.50                | -78.50           | -93.67           | S763     |
| 22   | S9393  | -101.55                         | -125.93                | -73.45           | -92.24           |          |
| 23   | S2779  | -101.50                         | -120.05                | -68.53           | -93.48           |          |
| 24   | S14464 | -101.47                         | -126.29                | -74.45           | -86.44           |          |
| 25   | S1161  | -101.31                         | -127.76                | -75.46           | -87.77           | S2585    |
| 26   | S4939  | -101.31                         | -128.88                | -76.48           | -86.21           | S4939($\Psi_{crys}$) |
| 27   | S1667  | -101.21                         | -125.91                | -73.59           | -85.33           | S1667    |
| 28   | S1107  | -101.18                         | -112.56                | -62.40           | -75.90           | S1107    |
| 29   | S2778  | -101.11                         | -121.80                | -75.52           | -90.07           | S4200    |
| 30   | S2396  | -101.09                         | -125.19                | -75.70           | -88.00           | S2526    |
| 31   | S2473  | -101.08                         | -124.72                | -72.49           | -86.30           | S2473($\Psi_{crys}$) |
| 32   | S18857 | -101.07                         | -113.72                | -63.19           | -75.81           | S18857($\Psi_{crys}$) |
| 33   | S2585  | -101.03                         | -125.64                | -72.91           | -86.61           | S2585    |
| 34   | S2526  | -100.96                         | -125.15                | -76.69           | -90.23           | S2526    |
| 35   | S4816  | -100.93                         | -124.86                | -72.53           | -87.12           | S10824($\Psi_{crys}$) |
| 36   | S618   | -100.86                         | -132.04                | -79.88           | -89.98           | S25263   |
| 37   | S10263 | -100.78                         | -125.65                | -73.47           | -93.93           |          |
| 38   | S123   | -100.76                         | -133.18                | -80.99           | -93.73           | S123     |
| 39   | S23063 | -100.74                         | -131.41                | -79.48           | -93.89           |          |
| 40   | S3917  | -100.69                         | -122.85                | -70.79           | -90.26           | S2526    |
| 41   | S16988 | -100.61                         | -126.85                | -78.06           | -89.09           |          |
Table S5: The energies of the $\Psi_{\text{mol}}$ set of structures of succinic acid at different steps and the cluster centers assigned by the FSFDP clustering algorithm. The clustering method has been applied to the two sets together. Cluster centers are then distinguished by including $\Psi_{\text{crys}}$ in parentheses. Melted structures are shown in bright red.

| Rank | IDs   | CSP($\Psi_{\text{mol}}$) kJ mol$^{-1}$ | EM(GAFF) kJ mol$^{-1}$ | NVT kJ mol$^{-1}$ | NPT kJ mol$^{-1}$ | Clusters |
|------|-------|--------------------------------------|------------------------|------------------|------------------|-----------|
| 42   | S4947 | -100.59                               | -123.79                | -71.59           | -85.34           | S4947     |
| 43   | S9203 | -100.57                               | -120.96                | -71.01           |                  | melted    |
| 44   | S11173| -100.34                               | -126.09                | -73.92           | -85.65           | S2585     |
| 45   | S3882 | -100.31                               | -125.27                | -73.04           | -93.87           | Form $\alpha$ |
| 46   | S2910 | -100.27                               | -116.81                | -76.82           | -89.34           | S2910     |
| 47   | S15665| -100.23                               | -37.46                 |                  |                  | melted    |
| 48   | S3841 | -100.20                               | -120.92                | -68.91           | -81.23           | S3841     |
| 49   | S2478 | -100.15                               | -129.33                | -77.38           | -90.03           | S2478($\Psi_{\text{crys}}$) |
| -    | NC1   | -87.13                                | -132.19                | -93.61           | -93.35           | NC1       |
| -    | NC2   | -                                      | -110.21                | -85.66           | -86.93           | Disordered |
| -    | NC3   | -82.05                                | -126.38                | -88.35           | -89.28           | NC3       |

Table S6: The energies of the $\Psi_{\text{crys}}$ set of structures of succinic acid at different steps and the cluster centers assigned by the FSFDP clustering algorithm. The clustering method has been applied to the two sets together. Cluster centers are then distinguished by including $\Psi_{\text{crys}}$ in parentheses. Melted structures are shown in bright red.

| Rank | IDs   | CSP($\Psi_{\text{crys}}$) kJ mol$^{-1}$ | EM(GAFF) kJ mol$^{-1}$ | NVT kJ mol$^{-1}$ | NPT kJ mol$^{-1}$ | Clusters |
|------|-------|--------------------------------------|------------------------|------------------|------------------|-----------|
| 1    | Form $\gamma$ | 0.00                                  | -125.94                | -73.94           | -91.79           | Form $\gamma$ |
| 2    | S9393 | 0.13                                  | -125.78                | -73.57           | -93.08           | Form $\alpha$ |
| 3    | S19217| 0.13                                  | -125.62                | -73.57           | -92.11           | Form $\gamma$ |
| 4    | S14464| 0.20                                  | -125.61                | -73.59           | -86.65           | S14464($\Psi_{\text{crys}}$) |
| 5    | Form $\alpha$ | 0.27                                  | -126.26                | -73.67           | -92.54           | Form $\alpha$ |
| 6    | S4158 | 1.02                                  | -131.50                | -79.18           | -92.98           | Form $\alpha$ |
Table S6: The energies of the $\Psi_{\text{crys}}$ set of structures of succinic acid at different steps and the cluster centers assigned by the FSFDP clustering algorithm. The clustering method has been applied to the two sets together. Cluster centers are then distinguished by including $\Psi_{\text{crys}}$ in parentheses. Melted structures are shown in bright red.

| Rank | IDs       | CSP($\Psi_{\text{crys}}$) kJ mol$^{-1}$ | EM(GAFF) kJ mol$^{-1}$ | NVT kJ mol$^{-1}$ | NPT kJ mol$^{-1}$ | Clusters          |
|------|-----------|----------------------------------------|------------------------|------------------|------------------|-------------------|
| 7    | S23063    | 1.16                                   | -131.88                | -79.62           | -92.92           | Form $\alpha$     |
| 8    | S2526     | 1.22                                   | -124.53                | -79.02           | -90.22           | S2526             |
| 9    | S763      | 1.45                                   | -133.29                | -82.00           | -91.67           | S763              |
| 10   | S23778    | 1.57                                   | -133.04                | -81.17           | -92.64           | Form $\beta$      |
| 11   | Form $\beta$ | 1.58                                | -133.06                | -81.13           | -92.67           | Form $\beta$      |
| 12   | S20629    | 1.61                                   | -131.45                | -78.80           | -90.09           | S20629            |
| 13   | S8149     | 2.90                                   | -129.25                | -77.29           | -91.93           | S8149             |
| 14   | S618      | 3.02                                   | -133.53                | -81.36           | -92.58           | S618($\Psi_{\text{crys}}$) |
| 15   | S5862     | 3.51                                   | -129.14                | -76.96           | -93.41           | Form $\beta$      |
| 16   | S2585     | 3.51                                   | -125.46                | -73.14           | -86.65           | S2585             |
| 17   | S1488     | 3.54                                   | -127.33                | -76.98           | -91.45           | Form $\alpha$     |
| 18   | S22937    | 3.65                                   | -126.45                | -76.42           | -90.71           | S763              |
| 19   | S804      | 3.68                                   | -130.68                | -78.60           | -91.49           | S804($\Psi_{\text{crys}}$) |
| 20   | S11752    | 3.77                                   | -133.60                | -81.61           | -93.58           | S11752($\Psi_{\text{crys}}$) |
| 21   | S3882     | 3.78                                   | -125.93                | -77.03           | -92.83           | Form $\alpha$     |
| 22   | S4200     | 3.80                                   | -128.63                | -76.34           | -88.37           | S4200             |
| 23   | S16898    | 4.01                                   | -123.32                | -81.73           | -93.36           | Form $\beta$      |
| 24   | S2407     | 4.33                                   | -131.15                | -79.11           | -90.22           | S2526             |
| 25   | S20462    | 4.34                                   | -124.96                | -81.39           | -91.73           | S763              |
| 26   | S2798     | 4.42                                   | -125.09                | -72.95           | -85.85           | S2798($\Psi_{\text{crys}}$) |
| 27   | S16988    | 4.59                                   | -129.40                | -81.06           | -90.13           | S16988($\Psi_{\text{crys}}$) |
| 28   | S1161     | 4.60                                   | -128.11                | -75.69           | -83.94           | S763              |
| 29   | S2478     | 4.98                                   | -130.91                | -78.99           | -89.20           | S2478($\Psi_{\text{crys}}$) |
Table S6: The energies of the $\Psi_{\text{crys}}$ set of structures of succinic acid at different steps and the cluster centers assigned by the FSFDP clustering algorithm. The clustering method has been applied to the two sets together. Cluster centers are then distinguished by including $\Psi_{\text{crys}}$ in parentheses. Melted structures are shown in bright red.

| Rank | IDs   | CSP($\Psi_{\text{crys}}$) kJ mol$^{-1}$ | EM(GAFF) kJ mol$^{-1}$ | NVT kJ mol$^{-1}$ | NPT kJ mol$^{-1}$ | Clusters       |
|------|-------|----------------------------------------|------------------------|-------------------|-------------------|----------------|
| 30   | S3917 | 5.01                                   | -122.44                | -71.90            | -90.21            | S2526          |
| 31   | S123  | 5.14                                   | -131.54                | -79.23            | -93.72            | S123           |
| 32   | S9203 | 5.25                                   | -120.09                | -75.41            | melted            |                |
| 33   | S19123| 5.26                                   | -133.80                | -82.17            | -90.16            | S763           |
| 34   | S2779 | 5.36                                   | -127.81                | -75.75            | -92.70            | Form $\alpha$  |
| 35   | S11173| 5.77                                   | -126.36                | -74.13            | -85.74            | S2585          |
| 36   | S10263| 5.78                                   | -124.01                | -71.66            | -90.96            | Form $\alpha$  |
| 37   | S10824| 5.83                                   | -128.38                | -76.05            | -88.30            | S10824($\Psi_{\text{crys}}$) |
| 38   | S1308 | 5.87                                   | -130.33                | -79.08            | -90.91            | S1308          |
| 39   | S4947 | 6.05                                   | -124.63                | -73.65            | -85.06            | S4947($\Psi_{\text{crys}}$) |
| 40   | S1904 | 6.45                                   | -127.29                | -75.26            | -88.83            | S10824($\Psi_{\text{crys}}$) |
| 41   | S1667 | 6.57                                   | -125.55                | -73.32            | -85.29            | S1667          |
| 42   | S4939 | 6.93                                   | -129.96                | -77.55            | -86.53            | S4939($\Psi_{\text{crys}}$) |
| 43   | S2473 | 7.15                                   | -124.78                | -72.78            | -86.26            | S2473($\Psi_{\text{crys}}$) |
| 44   | S4846 | 7.78                                   | -128.37                | -76.99            | -87.62            | S4846($\Psi_{\text{crys}}$) |
| 45   | S2910 | 8.01                                   | -98.27                 | melted            |                  |                |
| 46   | S4816 | 8.19                                   | -123.11                | -70.75            | -87.45            | S10824($\Psi_{\text{crys}}$) |
| 47   | S2396 | 11.31                                  | -125.96                | -76.41            | -87.98            | S2526          |
| 48   | S2778 | 11.41                                  | -122.95                | -75.98            | -93.36            | Form $\alpha$  |
| 49   | S3841 | 12.99                                  | -121.25                | -69.17            | -81.86            | S3841          |
| 50   | S15665| 29.19                                  | -96.88                 | melted            |                  |                |
| 51   | S1107 | 37.57                                  | -103.56                | melted            |                  |                |
| 52   | S18857| 38.28                                  | -114.41                | -64.05            | -75.76            | S18857($\Psi_{\text{crys}}$) |
Table S6: The energies of the $\Psi_{crys}$ set of structures of succinic acid at different steps and the cluster centers assigned by the FSFDP clustering algorithm. The clustering method has been applied to the two sets together. Cluster centers are then distinguished by including $\Psi_{crys}$ in parentheses. Melted structures are shown in bright red.

| Rank | IDs | CSP($\Psi_{crys}$) kJ mol$^{-1}$ | EM(GAFF) kJ mol$^{-1}$ | NVT kJ mol$^{-1}$ | NPT kJ mol$^{-1}$ | Clusters |
|------|-----|---------------------------------|------------------------|------------------|------------------|----------|
| -    | NC1 | 19.60                           | -132.19                | -93.61           | -93.35           | NC1      |
| -    | NC2 | -                               | -110.21                | -85.66           | -86.93           | Disordered |
| -    | NC3 | 24.80                           | -126.38                | -88.35           | -89.28           | NC3      |
References

(1) Case, D. et al. AMBER 2018. University of California, San Francisco 2018,

(2) Lucaioli, P.; Nauha, E.; Gimondi, I.; Price, L. S.; Guo, R.; Iuzzolino, L.; Singh, I.; Salvalaglio, M.; Price, S. L.; Blagden, N. Serendipitous Isolation of a Disappearing Conformational Polymorph of Succinic Acid Challenges Computational Polymorph Prediction. CrystEngComm 2018, 20, 3971–3977.

(3) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. Computer Physics Communications 2014, 185, 604–613.

(4) Nikulin, M. N. Hellinger Distance. Encyclopedia of Mathematics. 
   URL: http://www.encyclopediaofmath.org/index.php?title=Hellinger_distance&oldid=16453.

(5) Rodriguez, A.; Laio, A. Clustering by Fast Search and Find of Density Peaks. Science 2014, 344, 1492–1496.