

Serendipitous isolation of a disappearing conformational polymorph of succinic acid challenges computational polymorph prediction.

Supplementary Information

Paolo Lucaioli^a, Elisa Nauha^a, Ilaria Gimondi^b, Louise S Price^c, Rui Guo^c, Luca Iuzzolino^c, Ishwar Singh^a, Matteo Salvalaglio^b, Sarah L Price^c, Nicholas Blagden^a

^a*School of Pharmacy (PL, IS, NB), School of Chemistry (EN), University of Lincoln, Joseph Banks Laboratories, Green Lane, Lincoln, LN6 7DL, U.K.*

^b*Thomas Young Centre and Department of Chemical Engineering, University College London, Torrington Place, London, WC1E 7JE, U.K.*

^c*Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, U.K.*

Contents

1	Experimental production and characterization of γ-succinic acid	3
1.1	Peptide synthesis and L-Leu-L-Leu dipeptide	3
	Crystallographic data	4
1.2	Powder X-ray diffraction of starting materials.	5
1.3	Attempts to reproduce the crystallization of the γ polymorph	6
1.3.1	A) Investigating the role of trifluoroacetic acid (code: SATFA)	6
1.3.2	B) Investigating the influence of the Leu-Leu dipeptide: reproduction of the initial experiment	9
1.3.3	C) Investigating the influence of the mono-methyl hydrogen succinate	10
1.4	Analysis of crystal structures containing succinic acid.	13
	Neutral molecule CSD Python API analysis	13
1.4.1	Results	13
1.5	Comparison of crystal structures of succinic acid in all protonation states using dSNAP.	17
1.5.1	Analysis of non-ionic conformations.	17
1.5.2	Analysis of ion conformations.	19
2	Crystal Structure Prediction	21
2.1	Conformational analysis of Succinic Acid	21
2.2	Details of CSP methodology	21
2.3	Match of known and CSP structures	24
3	DFT-D relative energetics of experimental and CSP crystal structures of succinic acid	25
3.1	Method	25
3.2	Results	26
4	Molecular Dynamics Studies	28
4.1	Methods	28
4.1.1	MD conformers of succinic acid	28
4.1.2	Succinic acid crystals	28
4.1.3	Markov State Model (MSM)	29
4.1.4	Tools	29
4.2	Results: MD conformers of succinic acid in aqueous solution	29
4.2.1	Metadynamics Simulations	29
4.2.2	Analysis MD trajectories: MSM	32
4.2.3	Summary	35
4.3	Molecular simulations of bulk crystals	35
4.3.1	β Polymorph	35
4.3.2	γ Polymorph	36
4.3.3	Succinic acid conformational transitions in the bulk of the β phase.	37
4.3.4	Structural relaxation and melting of the γ polymorph.	37

1 Experimental production and characterization of γ -succinic acid

Paolo Lucaioli, Elisa Nauha and Nicholas Blagden

As part of wide co-crystallization screening, the L-Leu-L-Leu dipeptide has been crystallized along with succinic acid through slow solvent evaporation experiments. The analysis of different crystals obtained from the vial used for the experiment showed the presence of a new polymorphic phase of the acid (here defined as γ) along with the already known β form.

Further experimental work was undertaken, aiming to reproduce the formation of the γ polymorph, with an emphasis on identifying the specific aspects of the initial crystallization cocktail that promoted this novel phase. This required crystallization experiments to be conducted that take into account the possible influence of different chemical species involved in the initial crystallization.

1.1 Peptide synthesis and L-Leu-L-Leu dipeptide

Paolo Lucaioli and Ishwar Singh

The dipeptide used for the co-crystallization experiment has been synthesized through Fmoc Solid Phase Peptide Synthesis (SPPS), a well-established synthetic procedure based on the sequential addition of Fmoc-amino protected amino acids (aa) to a solid support resin. The final step of this method is removal of the peptidic molecule from the resin beads: this reaction is performed using a 'cleavage cocktail' containing trifluoroacetic acid (TFA). This harmful and corrosive compound ($pK_a=0.23$) binds to the positively charged N-terminal of the peptide (Figure S1) generating a strong ion pair (trifluoroacetate salts).

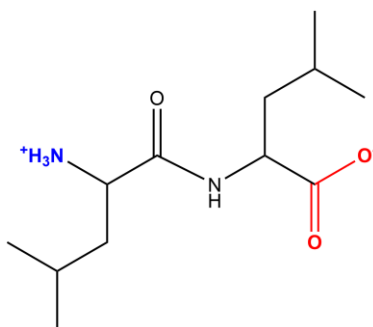


Figure S1. Molecular structure of the L-Leu-L-Leu dipeptide. The molecule is represented here in its usual zwitterionic state with a positively charged amino terminal (blue) and a negatively charged carboxy terminal (red).

The presence of such a chemical entity negatively affects both the physiochemical and biological properties of the peptide. For this reason, additional purification steps through chromatographic techniques followed by ion exchange reactions are routinely performed to purify the desired product.

For the co-crystallization experiment described in the present work, the L-Leu-L-Leu dipeptide was not purified: the TFA-contaminated product has been recovered from the synthesis liquor via evaporation (rotavapor) and lyophilization. Such treatments can remove most of the excess of TFA, but the final freeze-dried material is still affected by the presence of the contaminant.

Crystallographic data

Paolo Lucaioli, Elisa Nauha

Table S1. Crystal data and structure refinement for succinic acid γ form.

	γ succinic acid
Empirical formula	C ₄ H ₆ O ₄
Formula weight	118.09
Crystal system	Monoclinic
Space group	C2/c
a/Å	5.7015(5) Å
b/Å	8.4154(8) Å
c/Å	10.3538(8) Å
α	90°
β	90.374(3)°
γ	90°
Volume	496.77(7) Å ³
Z	4
Density (calculated)	1.579 Mg m ⁻³
Absorption coefficient	0.145 mm ⁻¹
F(000)	248
Crystal size	0.143 x 0.077 x 0.023 mm ³
Theta range	3.936 to 29.202°.
Index ranges	-7<=h<=7, -11<=k<=11, -13<=l<=14
Reflections collected	11388
Independent reflections	671
R(int)	0.1089
Completeness to theta = 25.242°	100.0 %
Max. and min. transmission	0.7458 and 0.6938
Data / restraints / parameters	671 / 0 / 40
Goodness-of-fit on F²	1.095
R1 [I>2sigma(I)]	0.0472
wR2 [I>2sigma(I)]	0.0818
Largest diff. peak and hole	0.317 and -0.220 e.Å ⁻³

The main feature and difference of the new γ form is represented by the position of the two carboxylic functionalities. In the α and β polymorphs, the carboxylic acid groups are pointing in opposite directions with respect to the plane on which C2 and C3 are positioned (Figure S2b and Figure S2b) while in the new γ form the two functionalities are placed on the same side of it (Figure S2a).

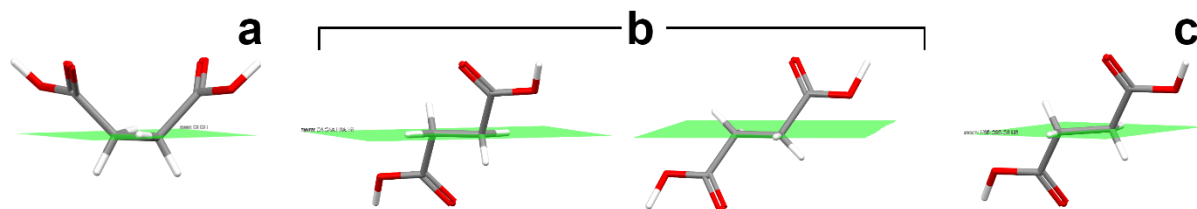


Figure S2. Spatial orientation of the two carboxylic groups in the succinic acid molecules for (a) the γ , (b) the α and (c) the β polymorphs. Cambridge Structural Database¹ structures with refcodes SUCACB07² and SUCACB11³ have been selected as representative of the α and β polymorphs respectively.

The relative orientation of the carboxylic groups can be described by measuring the absolute value of the $C_1-C_2\cdots C_3-C_4$ torsion angle (Table S2). When this parameter is equal to 180° , they are exactly in opposite directions while a torsion angle equal to zero describes two eclipsed groups pointing in the same direction.

Table S2. Torsion angles of the different polymorphic forms of succinic acid.

CSD Refcode	Polymorphic form	Torsion angle
SUCACB07	α polymorph	Molecule A: 180° Molecule B: 180°
SUCACB11	β polymorph	180°
CCDC 1836394	γ polymorph	75.43°

1.2 Powder X-ray diffraction of starting materials.

The powder diffractogram of the new γ polymorph was calculated from the solved structure using CSD Mercury. This was compared with the diffractograms of the already know forms (α and β) and that of the succinic acid supplied by Tokyo Chemical Industries & Co., used as starting material (Figure S3). Such comparison shows that the starting material contains the β polymorph and possibly some contaminant but there are no traces of γ .

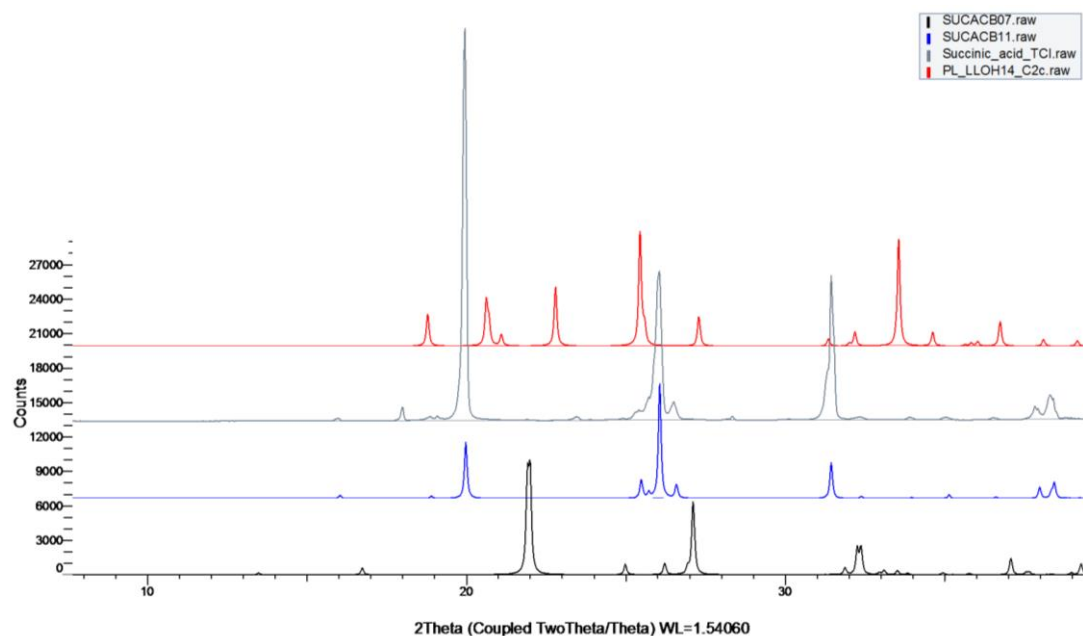


Figure S3. Powder X-ray diffractograms comparing the starting material (slate grey) and the three polymorphs (red for γ , blue for β , black for α)

1.3 Attempts to reproduce the crystallization of the γ polymorph

Paolo Lucaioli, Elisa Nauha and Nick Blagden

Co-crystallization of the peptide is a non-trivial operation due to the high flexibility of the molecular backbone (and side chains). This is demonstrated by the presence of only few solvated forms of Leu-Leu dipeptide in the CSD. The presence of an additional contaminating agent (trifluoroacetate anion) generating a strong ion pair with the positively charged terminal of the peptide represents a further complication for the co-crystallization process. The slow solvent evaporation process often results in the formation of oily residues that are not suitable for XRD analysis. Nevertheless, when solid material is obtained at the end of the solvent evaporation, small crystals can often be found on the walls or the bottom of the glass vial along with non-crystalline material (frequently represented by sticky/rubbery residual matter). The retrieval of crystals for X-ray analysis (by SCXRD) is not straightforward. When crystals with dimensions suitable for single crystal analysis are separated, frequently they do not diffract well (it is often difficult to just index them during the unit cell determination) and different crystals must be screened to test the quality of the diffraction spots.

The use of the material for PXRD is not convenient because the diffractograms obtained from such analyses are not qualitatively good and reliable due to the presence of amorphous excess combined with a large number of possible unknown phases present in the mixture.

Using the solid material for PXRD analysis might also lead to a loss of crystalline material. Some experiments have shown that the repetition of the same co-crystallization procedure (e.g.: same method, conditions, starting material, stoichiometric ratio, etc.) can lead to different crystalline adducts with different composition or result in a failed experiment with no crystalline material at all. Despite the operational difficulty, the best way to analyze the co-crystallization product is through single crystal X-ray diffraction and this is how we found (luckily!) the single crystal of the new γ polymorph of succinic acid.

Nevertheless, PXRD analysis has been possible when experiments have been carried out in the absence of peptide in the experimental attempts to obtain the γ polymorphic form. (We have easily run powder X-ray diffraction analysis when crystallization experiments have been carried out using commercial samples.)

1.3.1 A) Investigating the role of trifluoroacetic acid (code: SATFA)

Mixed solutions (with variable stoichiometric ratio) of succinic acid and trifluoroacetic acid were prepared using methanol as solvent (Table S3) for crystallization via slow solvent evaporation.

Table S3. Experimental plan for the crystallization experiments of succinic acid in the presence of different molar equivalents of TFA.

Sample name	Component A	Component B	Molar ratio
PL_SATFA_1:0.1_a	Succinic acid	Trifluoroacetic acid	1:0.1
PL_SATFA_1:0.1_b	Succinic acid	Trifluoroacetic acid	1:0.1
PL_SATFA_1:0.25_a	Succinic acid	Trifluoroacetic acid	1:0.25
PL_SATFA_1:0.25_b	Succinic acid	Trifluoroacetic acid	1:0.25
PL_SATFA_1:0.5_a	Succinic acid	Trifluoroacetic acid	1:0.5
PL_SATFA_1:0.5_b	Succinic acid	Trifluoroacetic acid	1:0.5
PL_SATFA_1:0.75_a	Succinic acid	Trifluoroacetic acid	1:0.75
PL_SATFA_1:0.75_b	Succinic acid	Trifluoroacetic acid	1:0.75
PL_SATFA_1:1_a	Succinic acid	Trifluoroacetic acid	1:1
PL_SATFA_1:1_b	Succinic acid	Trifluoroacetic acid	1:1

Crystalline material obtained at the end of the solvent evaporation in each vial was analysed by powder X-ray diffraction. Diffractograms of the different stoichiometric ratio mixtures were compared to check for any possible difference in peaks (Figure S4).

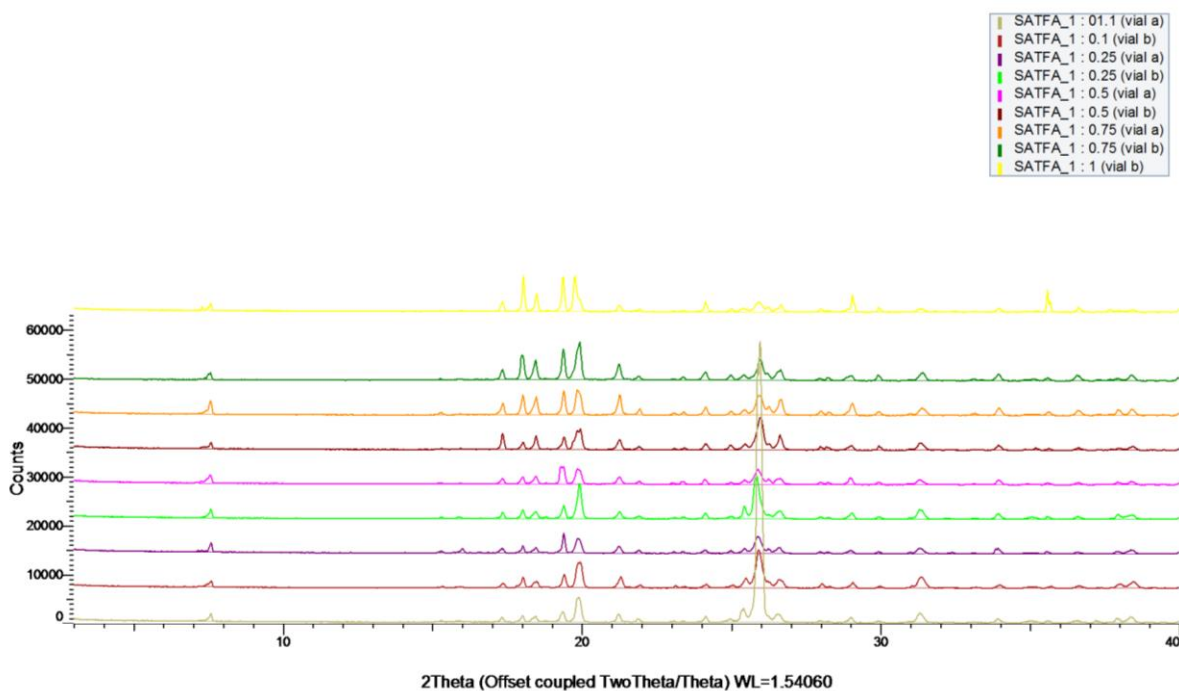


Figure S4. Comparison of the diffractograms obtained through PXRD from crystalline residues of the different SA:TFA molar ratios (after solvent evaporation).

The diffraction patterns of the different succinic acid : trifluoroacetic acid molar ratios are congruent. For this reason, one of the samples has been chosen as representative (PL_SATFA, green in Figure S5) and used for a comparison with the diffractograms of:

- α succinic acid (refcode: SUCACB07) – black in Figure S5
- β succinic acid (refcode: SUCACB11) – blue in Figure S5
- γ succinic acid (PL_LLOH14_C2c) – red in Figure S5

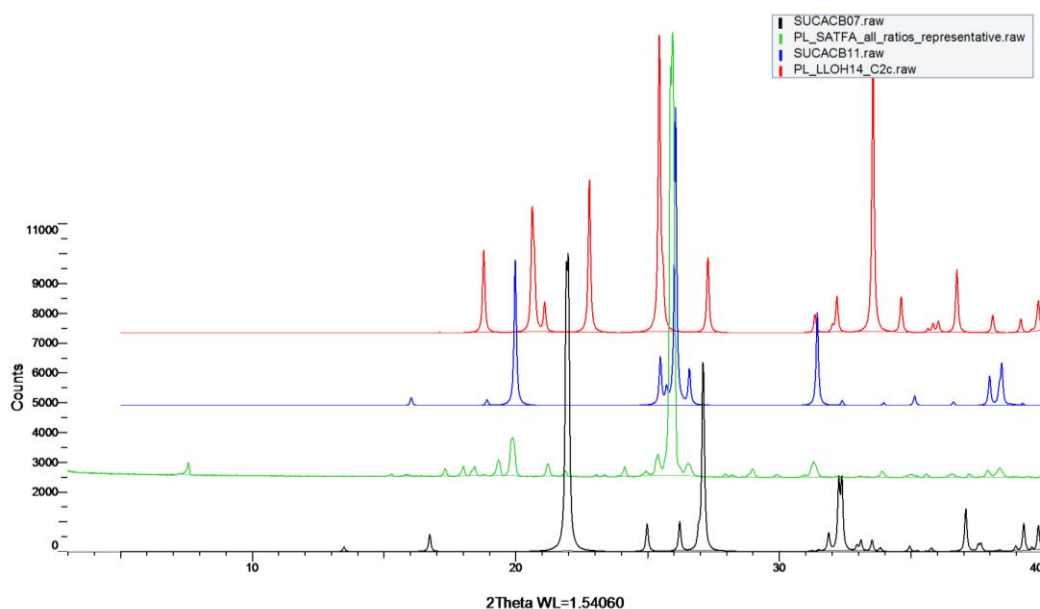


Figure S5. PXRD comparison of the representative sample of the SATFA experiment (green) and the three different polymorphs of succinic acid (black for α , blue for β , red for γ).

Vials have been also inspected optically under a microscope to double check the presence of crystals with different properties. Some of them have been randomly chosen and mounted on loops for unit cell determinations (and comparison with the α and β polymorphs). Results of such arbitrary screening are shown in Table S4.

Table S4. Unit cell measurements of crystals randomly selected from the different crystallization vials.

Sample	Vial	Position	Aspect	Result
PL_SATFA_1:0.1	a	bottom	plate	beta polymorph
	a	bottom	plate	beta polymorph
	a	wall	block	beta polymorph
	b	bottom	needle	N/A (bad quality reflections)
	b	bottom	elongated plate	beta polymorph
	b	wall	thin needle	N/A (bad quality reflections)
PL_SATFA_1:0.25	a	bottom	plate	beta polymorph
	a	bottom	needle	beta polymorph
	a	wall	elongated plate	beta polymorph
	a	wall	elongated plate	beta polymorph
	a	wall	plate	UNKNOWN CELL*
	b	bottom	plate	beta polymorph
	b	wall	thick plate	beta polymorph
PL_SATFA_1:0.5	a	bottom	thick plate	UNKNOWN CELL*
	a	bottom	plate	beta polymorph
	a	wall	thick plate	beta polymorph
	b	bottom	elongated plate	beta polymorph
	b	bottom	plate	beta polymorph
	b	wall	plate	N/A (bad quality reflections)
	b	wall	plate	UNKNOWN CELL*
	b	wall	plate	beta polymorph
PL_SATFA_1:0.75	a	bottom	elongated plate	UNKNOWN CELL*
	a	bottom	plate	UNKNOWN CELL*
	a	bottom	plate	beta polymorph
	b	bottom	block	UNKNOWN CELL*
	b	wall	plate	beta polymorph
	b	wall	plate	beta polymorph
PL_SATFA_1:1	a	bottom	plate	beta polymorph
	a	bottom	block	UNKNOWN CELL*
	a	bottom	elongated plate	beta polymorph
	a	wall	thin plate	beta polymorph
	a	wall	block	N/A (bad quality reflections)
	b	bottom	thin plate	beta polymorph
	b	wall	plate	beta polymorph

***UNKNOWN CELL:** the structure solution showed that the new crystal form is represented by a new polymorph (II) of mono-methyl hydrogen succinate (CCDC deposition number 1836683; the previously reported form has reffcode MESUCC), the succinic acid mono-methyl ester (Figure S6). The small extra peaks in the powder patterns of the SATFA samples are explained by the peaks from the mono-methyl hydrogen succinate. (We note that the mono-methyl hydrogen succinate is in the folded conformation in both forms.)

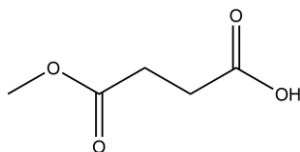


Figure S6. Mono-methyl hydrogen succinate molecular structure.

Figure S7 shows the reaction mechanism of generating the ester in the vial used for our crystallization experiments. The reaction is a simple acid catalysed esterification of a carboxylic acid in which the TFA acts as the catalyst while the solvent is the reacting alcohol.

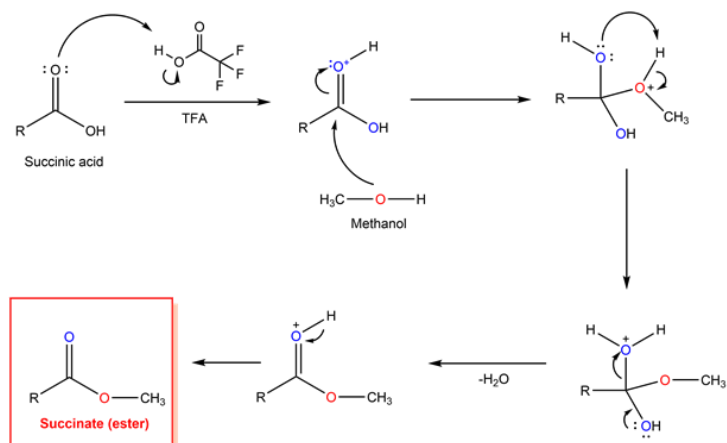


Figure S7. The formation of the mono-methyl hydrogen succinate is catalysed by the trifluoroacetic acid (strong acid) through a Fischer esterification

1.3.2 B) Investigating the influence of the Leu-Leu dipeptide: reproduction of the initial experiment

A reproduction of the experiment that resulted in the isolation of the new polymorphic form of succinic acid has been carried out.

An equimolar solution of TFA-contaminated L-Leu-L-Leu dipeptide and succinic acid in MeOH has been used for a co-crystallization experiment. After the solvent evaporation, the vial was screened under the microscope. The residue obtained was a sticky/rubbery material (frequently encountered when co-crystallization experiments with TFA-contaminated peptides are performed): some fragile needles and small blocks were recovered from the vial (with some difficulties) and used for unit cell determination. Results of such screening are shown in Table S5.

Table S5. Unit cell measurement of single crystals collected from the crystallization vial of the experimental reproduction of the initial co-crystallization attempt that resulted in the isolation of the γ polymorph.

Aspect	Result
needle	N/A (bad quality reflections)
block	beta polymorph
block	N/A (bad quality reflections)
block	N/A (bad quality reflections)
elongated plate	beta polymorph
elongated plate	beta polymorph
needle	N/A (bad quality reflections)
block	beta polymorph

The sticky material contained in the vial has been re-dissolved and re-crystallized using the same solvent (methanol) to attempt a better crystallization but the residue of the solvent evaporation was similar to the previous one. Crystals suitable for SCXRD analysis have been retrieved with difficulty from it and used for unit cell measurements. Results of this second attempt are shown in Table S6.

Table S6. Unit cell determinations of crystals retrieved and separated from the sticky material obtained after re-crystallization.

Aspect	Result
block	beta polymorph
block	N/A (bad quality reflections)
block/plate	N/A (bad quality reflections)
block	not crystalline

Note: PXRD analyses are not available for these samples since the sticky material obtained from the co-crystallization through slow solvent evaporation experiment could not be recovered from the vial and could not be placed in a suitable support for X-ray diffraction.

1.3.3 C) Investigating the influence of the mono-methyl hydrogen succinate

The presence of an additional chemical species in the crystallization environment, represented by the serendipitously synthesized ester, required an additional investigation to understand the possible influence of such a component in the formation of the γ polymorph of succinic acid.

Commercial mono-methyl hydrogen succinate (obtained from Sigma Aldrich) has been used to prepare mixed solutions (solvent: methanol) with succinic acid in different stoichiometric ratios (Table S7). Vials containing the mixed solutions were capped with perforated parafilm and placed at 20 °C in an incubator to allow slow solvent evaporation.

Table S7. Experimental plan for the SAE (succinic acid:ester) crystallization screening.

Sample name	Component A	Component B	Molar ratio
PL_SAE_1:0.1_a	Succinic acid	Mono-methyl hydrogen succinate	1:0.1
PL_SAE_1:0.1_b	Succinic acid	Mono-methyl hydrogen succinate	1:0.1
PL_SAE_1:0.25_a	Succinic acid	Mono-methyl hydrogen succinate	1:0.25
PL_SAE_1:0.25_b	Succinic acid	Mono-methyl hydrogen succinate	1:0.25
PL_SAE_1:0.5_a	Succinic acid	Mono-methyl hydrogen succinate	1:0.5
PL_SAE_1:0.5_b	Succinic acid	Mono-methyl hydrogen succinate	1:0.5
PL_SAE_1:0.75_a	Succinic acid	Mono-methyl hydrogen succinate	1:0.75
PL_SAE_1:0.75_b	Succinic acid	Mono-methyl hydrogen succinate	1:0.75
PL_SAE_1:1_a	Succinic acid	Mono-methyl hydrogen succinate	1:1
PL_SAE_1:1_b	Succinic acid	Mono-methyl hydrogen succinate	1:1

Solid material obtained at the end of the solvent evaporation in each vial was used for PXRD analysis of the different samples to investigate the presence of any possible difference in the diffractograms (Figure S8).

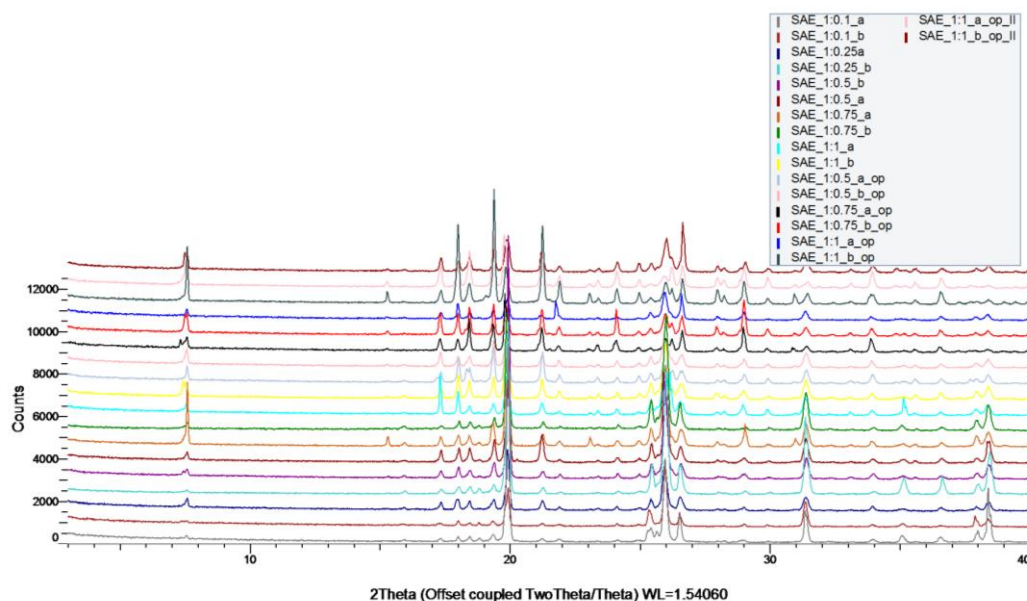


Figure S8. Comparison of the diffractograms obtained through PXRD from crystalline residues of the different SA:E molar ratios (after solvent evaporation). Some vials showed crystallized material outside the perforated parafilm used to cap them. Such external material has been collected and analyzed separately (samples labelled as 'op').

The comparison of the diffractograms of the different samples show that the samples are all similar. For this reason, one of these diffractograms (green in Figure S9) has been chosen as representative of the experimental plan and has been used for an overlay with the diffractograms of:

- mono-methyl hydrogen succinate known polymorph (refcode: MESUCC) – orange in Figure S9
- new polymorph of mono-methyl hydrogen succinate – magenta in Figure S9
- α succinic acid (refcode: SUCACB07) – black in Figure S9
- β succinic acid (refcode: SUCACB11) – blue in Figure S9
- γ succinic acid (PL_LLOH14_Csc) – red in Figure S9

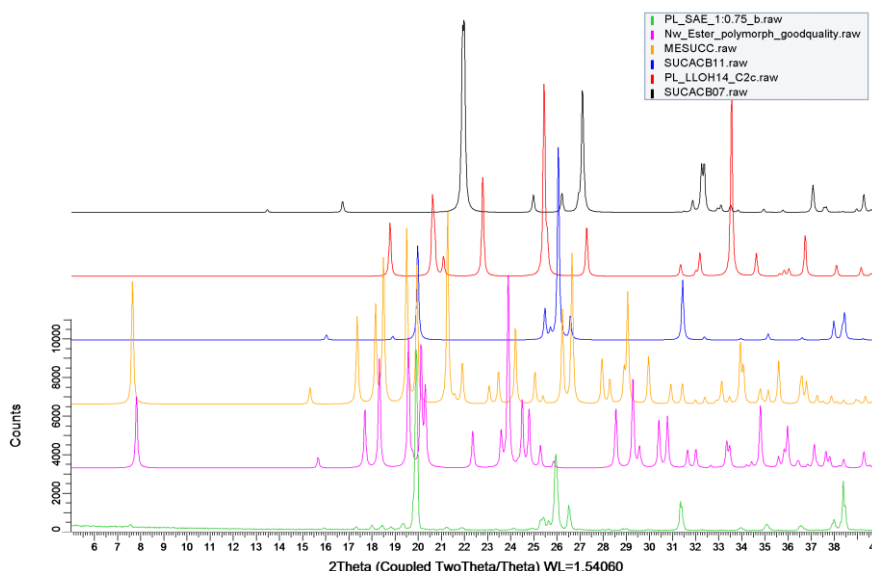


Figure S9. PXRD comparison of the representative sample of the SAE experiment (PL_SAE_1:0.75_b, Table S7, green) and the different polymorphs of both succinic acid and mono-methyl hydrogen succinate.

Vials have been inspected under the microscope and crystals selected randomly (from different positions inside the container) and mounted on loops for unit cell determination. Results of such screening are reported in Table S8.

Table S8. Unit cell determination of randomly selected crystals from the solid material in each crystallization vial.

Sample	Vial	Position	Aspect	Result
PL_SAE_1:0.1	a	bottom	plate	N/A (bad quality reflections)
	a	bottom	plate	succinic acid beta polymorph
	a	bottom	thick plate	succinic acid beta polymorph
	a	wall	plate	succinic acid beta polymorph
	a	wall	block	N/A (bad quality reflections)
	a	wall	block	succinic acid beta polymorph
	a	wall	block	succinic acid beta polymorph
	b	bottom	block	succinic acid beta polymorph
	b	bottom	elongated plate	succinic acid beta polymorph
	b	bottom	block	succinic acid beta polymorph
	b	wall	block	succinic acid beta polymorph
	b	wall	plate	succinic acid beta polymorph
	b	wall	block	succinic acid beta polymorph
	b	wall	plate	succinic acid beta polymorph
PL_SAE_1:0.25	a	bottom	thick plate	succinic acid beta polymorph
	a	bottom	block	N/A (bad quality reflections)
	a	bottom	block	succinic acid beta polymorph
	a	wall	plate	succinic acid beta polymorph
	a	wall	plate	succinic acid beta polymorph
	a	wall	block	succinic acid beta polymorph
	b	bottom	plate	succinic acid beta polymorph
	b	bottom	block	succinic acid beta polymorph
	b	bottom	block	succinic acid beta polymorph
	b	wall	plate	succinic acid beta polymorph
	b	wall	block	succinic acid beta polymorph

	b	wall	plate	succinic acid beta polymorph
	b	wall	block	succinic acid beta polymorph
PL_SAE_1:0.5	a	bottom	plate	succinic acid beta polymorph
	a	bottom	thick plate	succinic acid beta polymorph
	a	bottom	plate	succinic acid beta polymorph
	a	wall	plate	succinic acid beta polymorph
	a	wall	block	succinic acid beta polymorph
	a	wall	plate	N/A (bad quality reflections)
	b	bottom	block	N/A (bad quality reflections)
	b	wall	plate	succinic acid beta polymorph
	b	wall	block	succinic acid beta polymorph
	b	wall	thick plate	MONO-METHYL HYDROGEN SUCCINATE
	b	wall	block	MONO-METHYL HYDROGEN SUCCINATE
	b	wall	block	MONO-METHYL HYDROGEN SUCCINATE
PL_SAE_1:0.75	a	bottom	plate	succinic acid beta polymorph
	a	bottom	block	succinic acid beta polymorph
	a	bottom	plate	N/A (bad quality reflections)
	a	wall	block	succinic acid beta polymorph
	a	outside parafilm	block	N/A (bad quality reflections)
	a	outside parafilm	block	MONO-METHYL HYDROGEN SUCCINATE
	b	wall	block	succinic acid beta polymorph
	b	wall	plate	succinic acid beta polymorph
	b	bottom	block	succinic acid beta polymorph
	b	bottom	block	succinic acid beta polymorph
	b	bottom	plate	succinic acid beta polymorph
	b	wall	thick plate	succinic acid beta polymorph
	b	wall	block	succinic acid beta polymorph
	b	wall	plate	succinic acid beta polymorph
b	outside parafilm	plate	MONO-METHYL HYDROGEN SUCCINATE	
PL_SAE_1:1	a	bottom	thick plate	succinic acid beta polymorph
	a	bottom	elongated plate	succinic acid beta polymorph
	a	bottom	plate	succinic acid beta polymorph
	a	wall	plate	succinic acid beta polymorph
	a	wall	plate	succinic acid beta polymorph
	b	bottom	thick plate	MONO-METHYL HYDROGEN SUCCINATE
	b	bottom	block	N/A (bad quality reflections)
	b	bottom	block	MONO-METHYL HYDROGEN SUCCINATE
	b	bottom	plate	MONO-METHYL HYDROGEN SUCCINATE
	b	wall	thick plate	succinic acid beta polymorph
	b	wall	plate	succinic acid beta polymorph
	b	wall	elongated plate	succinic acid beta polymorph
	b	outside parafilm	block	MONO-METHYL HYDROGEN SUCCINATE
	b	outside parafilm	plate	MONO-METHYL HYDROGEN SUCCINATE

1.4 Analysis of crystal structures containing succinic acid.

Luca Iuzzolino and Paolo Lucaioli

The Cambridge Structural Database (CSD)¹ contains more than 850,000 experimentally-determined crystal structures, and it is an extremely valuable source of information on crystalline conformations. Two independent approaches were used to mine the data.

Neutral molecule CSD Python API analysis. Conquest 1.19 was used to retrieve all the crystal structures containing succinic acid, either as a single-component or in multi-component systems. The search was limited to non-disordered crystal structures with determined 3D coordinates, including hydrogen atoms, and with R factors smaller than 10%. The Crystal Packing Similarity Tool available through the CSD Python API 1.3.0 was used to remove redeterminations, defined by overlaying clusters of 30 molecules specifying a 20% distance tolerance and a 20° angle tolerance, and clustering structures with a 30/30 molecule match and an RMSD₃₀ < 0.3 Å. The retained structure had the lowest R-factor, except when a neutron study was available. The first criterion was sufficient to remove most of the 37 duplicates. When a crystal structure contained more than one independent molecule of succinic acid, if the central torsion angle did not vary between the independent molecules by more than 15°, it was only counted once. There was only one crystal structure (HOGFIU01) in which there were one folded and one planar molecule in the asymmetric unit. Hence, there are 142 individual succinic acid molecules in 141 unique crystal structures.

All the succinic acid molecules were either planar or folded, i.e. all the carboxylic acid groups were parallel to the central carbon chain, with no twisted conformations (such as aA-g or gGa on Figure S27, for example). A minority formed conformers with *cis* carboxylic acid groups. The value of the central torsion angle, as output by Conquest, was used to classify the molecules as planar if it was between 170° and 190°, and folded otherwise. The CSD Python API was finally used to calculate the packing coefficient and the percentage of void space, with the default setting of a probe size of 1.2 Å and a grid spacing of 0.7 Å, of each crystal structure.

1.4.1 Results

Out of the 142 individual succinic acid molecules that were considered, 126 crystallized as planar and 16 as folded. The refcodes and formulae, and crystal properties for the planar conformations are given in Table S9 and those for the non-planar in Table S10.

Table S9. List of the refcodes and chemical formulae of the deposited crystal structures containing at least one molecule of succinic acid with a planar conformation.

Refcode	Chemical Formula	% Packing coefficient	% Void spacing	Hydrogen bond formed by succinic acid
SUCACB03 β	C4 H6 O4	80.37	0	R22(8) chain
SUCACB07 α	C4 H6 O4	73.777	0	R22(8) chains
ACESUY	C6 H6 N6 O2,0.5(C4 H6 O4)	75.136	0	with coformer
AVEKAN	2(C14 H12 Fe1 N2),C4 H6 O4	66.642	0	with coformer
AXUFIJ	C13 H9 N1,0.5(C4 H6 O4)	70.1	0	with coformer
BEPTIB	2(C4 H10 N1 O1 1+),C4 H6 O4,C4 H4 O4 2-	70.89	0	with coformer
BICQAH	C6 H7 N3 O1,C6 H6 N2 O1,C4 H6 O4	75.32	0	with coformer
BULGEU	C4 H6 O4,Cs1 1+,F1 1-	69.446	0	with F ⁻ ion
BZASUC	2(C7 H7 N1 O1),C4 H6 O4	69.379	0	with coformer
CAKZUL	C17 H13 Cl1 N4,0.5(C4 H6 O4),3(H2 O1)	66.076	0	with water
CEJXOF	(C10 H22 N2 O12 P4 Pb4) _n ,n(C4 H6 O4)	76.314	0	with coformer
CIWFUL	C20 H26 N4 O2,C4 H6 O4	72.373	0	with coformer
CUJMIE	C18 H22 N4 O2,C4 H6 O4	72.915	0	with coformer
CUVDUT	C4 H5 N3 O1,0.5(C4 H6 O4)	73.965	0	with coformer
DIKCIK	2(C15 H13 N3 O4 S1),C4 H6 O4	69.961	0	with coformer
DILVAX	0.5(C4 H6 O4),C4 H4 F1 N3 O1	74.807	0	with coformer
DOSZAO	C18 H16 N6,C4 H6 O4	71.758	0	with coformer
DUZPAQ	2(C6 H6 N2 O1),C4 H6 O4	73.379	0	with coformer
EFAPUY	C5 H6 Cl1 N3 O1,0.5(C4 H6 O4)	72.034	0	with coformer
EMAPIT	2(C12 H10 N4),1.5(C4 H6 O4)	67.835	0	with coformer
ENICOU	C14 H13 N3 O4 S2,0.5(C4 H6 O4)	73.032	0	with coformer
ENICOU01	2(C14 H13 N3 O4 S2),C4 H6 O4	70.05	0	with coformer

EWOZIZ	2(C5 H11 N1 O2),C4 H6 O4	68.722	0	with coformer
EXIPEH	C4 H6 O4,C4 H4 O4 2-,2(C2 H8 N1 O1 1+)	71.823	0	with coformer
EXIPEH01	C4 H6 O4,C4 H4 O4 2-,2(C2 H8 N1 O1 1+)	73.838	0	with coformer
FADGIC	2(C6 H7 N3 O1),C4 H6 O4	73.225	0	with coformer
FADGIC02	2(C6 H7 N3 O1),C4 H6 O4	71.958	0	with coformer
FAGXOB	(C16 H12 N2 O4 Zn1)n,n(C4 H6 O4)	69.395	0	with coformer
FIDCIG	2(C6 H7 N3 O1),C4 H6 O4	67.866	0	with coformer
FOTDAV	2(C7 H7 N1 O2),C4 H6 O4,C4 H4 N2	72.631	0	with coformer
FOTDID	C12 H8 N2,C7 H7 N1 O2,0.5(C4 H6 O4),H2 O1	71.618	0	with coformer
FOTFOL	C6 H4 N2 O4,2(C7 H6 Br1 N1 O1),C4 H6 O4	71.964	0	with coformer
GABZUF	C16 H14 N4,C4 H6 O4	71.072	0	with coformer
GALBIF	C10 H8 N4,C4 H6 O4	70.048	0	with coformer
GAWLOG	2(C12 H8 N4 O1),C4 H6 O4	71.032	0	with coformer
GESBAJ	C7 H7 N1 O2,0.5(C4 H6 O4)	73.827	0	with coformer
GOKBAL	C10 H8 N2,C4 H6 O4	72.299	0	with coformer
GOKBAL01	C10 H8 N2,C4 H6 O4	76.032	0	with coformer
HAGNEJ	C14 H18 N1 1+,0.5(C4 H4 O4 2-), C4 H6 O4, H2 O1	68.525	1.23	with water
HIQMAX	4(C7 H6 Br1 N1 O1),C4 H6 O4	69.869	0	with coformer
HIQSIL	2(C7 H6 Br1 N1 O1),C4 H6 O4	70.574	0	with coformer
HOGFIU01 (mol.1)	2(C6 H6 N2 O1),C4 H6 O4	69.64	0	with coformer
HOLNIG	2(C4 H7 N2 1+),C4 H4 O4 2-,C4 H6 O4	71.019	0	with coformer
HOLNIG01	C4 H6 O4,2(C4 H7 N2 1+),C4 H4 O4 2-	72.215	0	with coformer
HOQGIF	2(C4 H6 O4),C2 H3 N3 S1	73.676	0	with coformer
HUVGIQ	C12 H10 N4,C4 H6 O4	72.385	0	with coformer
IHESOD	2(C5 H7 N2 1+),C4 H4 O4 2-,C4 H6 O4	71.823	0	with coformer
IHESOD01	C4 H6 O4,2(C5 H7 N2 1+),C4 H4 O4 2-	73.822	0	with coformer
IKEQEU	C35 H38 Cl2 N8 O4,0.5(C4 H6 O4)	69.781	0	with coformer
IQIHEW01	C4 H6 O4,2(C2 H3 N3)	69.963	0	with coformer
IRAMEV	C13 H12 N2 O2,0.5(C4 H6 O4)	69.121	0	with coformer
IRAMEV01	2(C13 H12 N2 O2),C4 H6 O4	70.562	0	unknown
ISUTEV	C6 H18 N2 2+,C4 H4 O4 2-,C4 H6 O4	73.615	0	with coformer
JAZBES	C13 H14 N2,C4 H6 O4	68.436	0	with coformer
JUWHOA	C10 H8 N2 O1,0.5(C4 H6 O4)	72.906	0	with coformer
KAYGUO	C11 H9 N3 O1,C4 H6 O4	71.437	0	with coformer
KAYHOJ	C11 H9 N3 O1,C4 H6 O4	72.851	0	with coformer
KEFBED	2(C7 H8 N2 O1),C4 H6 O4	73.997	0	with coformer
KFSCCN01	C4 H4 D2 O4,K1 1+,F1 1-	87.96	0	with F ⁻ ion
KOHPOM	C6 H15 N2 O2 1+,0.5(C4 H4 O4 2-), 0.5(C4 H6 O4)	74.695	0	with coformer
LABZUJ	C5 H9 N1 O2,0.5(C4 H6 O4)	71.576	0	with coformer
LADDIF	2(C7 H7 Br1 N2 O1),C4 H6 O4	70.16	0	with coformer
LAPXEH	2(C20 H32 N5 O8 P1),C4 H6 O4	63.203	0	with coformer
LATJAT	C10 H8 N2,C4 H6 O4	72.638	0	with coformer
LATJEX	C15 H10 O1,0.5(C4 H6 O4)	71.224	0	with coformer
LATKEY	C9 H11 N3 O1,0.5(C4 H6 O4)	71.25	0	with coformer
LATLEY	C14 H14 N4,C4 H6 O4	71.524	0	with coformer
LATTOR	C5 H5 N3 O1,0.5(C4 H6 O4)	69.714	0	with coformer
LEGZEE	2(C3 H7 N6 1+),C4 H4 O4 2-,C4 H6 O4,2(H2 O1)	74.088	0	with coformer and water
LOXSUC11	2(C18 H19 Cl1 N3 O1 1+),C4 H6 O4, C4 H4 O4 2-,2(H2 O1)	69.351	0	with water
LUNNUD	2(C6 H6 N2 O1),C4 H6 O4	73.741	0	with coformer
LUNNUD01	C4 H6 O4,2(C6 H6 N2 O1)	74.061	0	with coformer
MADPEO	C12 H20 N4 Ni1 O6 2+,C4 H6 O4,C4 H4 O4 2-	71.623	0	with coformer
MELNIA	(C16 H14 Cu1 N2 O5)n,0.5n(C4 H6 O4)	70.95	0	with coformer
MOGZEP	2(C18 H15 Cl1 N2 O2 S1),C4 H6 O4	65.281	0	with coformer
MOSMIR	2(C3 H5 N2 S1 1+),C4 H4 O4 2-,C4 H6 O4	74.097	0	with coformer

MOXSOI	C20 H13 F3 N4 O2 S1,0.5(C4 H6 O4)	69.063	0	with coformer
MOXSIO1	C20 H13 F3 N4 O2 S1,0.5(C4 H6 O4)	70.364	1.18	with coformer
NIJGIY	C9 H9 N5,0.5(C4 H6 O4)	69.692	0	with coformer
NISTAK	2(C5 H5 N1 O1),C4 H6 O4	73.301	0	with coformer
NISTAK01	2(C5 H5 N1 O1),C4 H6 O4	71.345	0	with coformer
OCIPUM	C11 H11 N3 O1,0.5(C4 H6 O4)	71.854	0	with coformer
OGAHAF	C4 H6 O4,C4 H8 O2 S2	73.563	0	with coformer
ORIXUK	2(C7 H10 N1 1+),C4 H6 O4,C4 H4 O4 2-	69.694	0	with coformer
PILKOM	C7 H6 N2 O3,0.5(C4 H6 O4)	73.094	0	with coformer
PILSOU	C19 H23 F1 N3 O3 1+,0.5(C4 H6 O4), 0.5(C4 H4 O4 2-)	73.568	0	with coformer
PUTFAL	2(C26 H36 N3 O2 1+),C4 H4 O4 2-,C4 H6 O4	65.084	0	with coformer
QIPNAE	2(C12 H14 N2 O1),C4 H6 O4	67.031	0	with coformer
RAJFEO	2(C17 H19 F3 N1 O1 1+),2(C1 1-),C4 H6 O4	66.905	0	with coformer
RESGAY	2(C7 H9 N1),C4 H6 O4	69.003	0	with coformer
RESGIG	2(C7 H9 N1),C4 H6 O4	66.875	0	with coformer
RESGOM	2(C7 H9 N1),C4 H6 O4	68.636	0	with coformer
RESHAZ	2(C7 H9 N1),C4 H6 O4	67.982	0	with coformer
RESHIH	2(C7 H9 N1),C4 H6 O4	71.008	0	with coformer
RETZEW	2(C7 H9 N1),C4 H6 O4	68.61	0	with coformer
ROSTUQ	2(C13 H13 N3 O2 S1),C4 H6 O4	70.212	0	with coformer
SEQVIV	C5 H5 N5,C4 H6 O4	73.246	0	with coformer
SOCMII	C12 H9 N3 O5 S1,0.5(C4 H6 O4)	70.41	0	with coformer
SOSBAD	C18 H12 N4,0.5(C4 H6 O4)	65.444	0	with coformer
SOVQEZ	C4 H6 O4,C4 H7 N5,H2 O1	73.345	0	with coformer and water
SUCTAN04	2(C3 H8 N2 O1),C4 H6 O4	71.403	0	with coformer
SUDDEC04	C4 H6 O4,2(C2 H6 N2 O1)	70.253	0	with coformer
TACCIL	C12 H12 N2,C4 H6 O4	69.836	0	with coformer
TELDAR	C19 H24 N2 O2,C4 H6 O4	66.958	0	with coformer
TUMNAS	C20 H22 Cu1 N2 O4,0.5(C4 H6 O4),H2 O1	67.736	0	with water
TUMNEW	C20 H22 N2 Ni1 O4,0.5(C4 H6 O4),H2 O1	68.78	0	with water
UGOSOA	2(C4 H7 N2 S1 1+),C4 H6 O4,C4 H4 O4 2-	69.516	0	with coformer
UHACEM	C4 H6 O4,2(C4 H7 N1 O1)	71.532	0	with coformer
UHADEN	C4 H6 O4,2(C3 H6 N2 O1)	72.493	0	with coformer
UKOSIX	2(C6 H7 N3 O1),C4 H6 O4	75.253	0	with coformer
ULOYUR	C11 H10 N2 S1,C4 H6 O4	70.623	0	with coformer
UMINOT	C12 H10 N2,C4 H6 O4	71.644	0	with coformer
UNIRIT	C4 H6 O4,C1 H4 N2 O1	71.83	0	with coformer
VAXWAU01	C4 H6 O4,C4 H4 N2	75.43	0	with coformer
VEJXAJ06	C4 H6 O4,2(C1 H4 N2 O1)	69.206	0	with coformer
VIGDEV	C18 H16 N4,0.5(C4 H6 O4)	70.209	0	with coformer
VORCOV	C6 H6 N4,C4 H6 O4	69.996	0	with coformer
WEJPIL	C27 H21 Cl1 N2 O2,0.5(C4 H6 O4)	71.02	0	with coformer
WOJHEI	C10 H8 N2 O2,C4 H6 O4	74.625	0	with coformer
WOQBOT	C12 H8 N2,C4 H6 O4	70.341	0	with coformer
XAQPAI	2(C8 H12 N1 O1 1+),C4 H4 O4 2-,C4 H6 O4	66.373	0	with coformer
XEPBEB	C10 H8 N4,C4 H6 O4	71.391	0	with coformer
YOWDET	C6 H15 N4 O2 1+,0.5(C4 H4 O4 2-), 0.5(C4 H6 O4),H2 O1	71.214	0	with coformer and water
YUXTUI	(C10 H14 N2 Ni1 O4)n,n(C4 H6 O4), 4n(H2 O1)	74.144	0	with water
YUZDUT	C4 H6 O4,C12 H20 Co1 N4 O6 2+,C4 H4 O4 2-	73.885	0	with coformer
ZUKXIM	2(C6 H9 N3 O1),C4 H6 O4	72.276	0	with coformer

Table S10. List of the 16 refcodes and chemical formulae of the deposited crystal structures containing at least one molecule of succinic acid with a folded conformation.

Refcode	Chemical Formula	% Packing coefficient	% Void spacing	Hydrogen bond formed by succinic acid
New form γ	C4 H6 O4	75.649	0	R22(8) chain
CIRXAD	C2 H4 N4, C4 H6 O4	72.7	0	with coformer
DUWLAK	C17 H24 N8 O5 S1 V1, C4 H6 O4	65.808	1.11	with coformer
GADBEV	C5 H11 N1 O2 S1, 0.5(C4 H6 O4)	70.678	0	with coformer
HOGFIU01 (mol.2)	2(C6 H6 N2 O1), C4 H6 O4	72.747	0	with coformer
HUPPEP	2(C6 H14 N1 1+), C4 H6 O4, C4 H4 O4 2-	70.126	0	with coformer
JEDLAG01	C11 H10 N4 O1, C4 H6 O4	69.139	0.66	with coformer
KIJSEC	C42 H70 O35, C4 H6 O4, 7(H2 O1)	70.823	0	with coformer and water
KTHSUC	C4 H5 O4 1-, C4 H6 O4, K1 1+	77.939	0	with coformer
OLOFUQ	C20 H18 N1 O4 1+, C1 H1 O2 1-, C4 H6 O4	74.138	0	with formate
OMEKIC	C22 H23 N3 O4, C4 H6 O4, H2 O1	68.916	0	with coformer and water
PEKQOM	C27 H27 N5 O3, 1.5(C4 H6 O4)	67.635	0	with coformer
QEVMEJ	C7 H9 N1 O1, C4 H6 O4	67.101	0	with coformer
SERMOR	C4 H5 N3, C4 H6 O4	67.521	0	with coformer
SOVPOJ	C13 H12 Cu1 N2 O5, C4 H6 O4	72.339	0	with coformer and succinic acid, but no chain
XOBCIB	C15 H12 N2 O1, 0.5(C4 H6 O4)	69.713	0	unknown
XUBVEW	C10 H8 N2 S2, C4 H6 O4	68.53	0	with coformer

Table S11. Summary of the data shown in Table S9 and Table S10.

% Planar conformers	88.8
% Folded conformers	11.2
Average packing coefficient with planar conformers/%	71.4
Average packing coefficient with folded conformers/%	70.4
# structures with void space and a planar conformer	2/127
# structures with void space and a folded conformer	2/16

The planar conformers are clearly much more common in the CSD. On average, the crystal structures containing the planar conformer pack slightly better than those with the folded conformer, although the difference is only 1%, and probably not statistically significant. The polymorphs of single-component succinic acid have packing coefficients of ~74-80%, which are higher than those of most cocrystals, salts and solvates because of the dense packing of the carboxylic acid chain. Only four crystal structures, two with planar (HAGNEJ and MOXSOI01) and two with folded (DUWLAK and JEDLAG01) conformers of succinic acid, contain recognizable voids. Even in those four cases the void spacing is small, indicating that crystal structures containing succinic acid tend to pack well with themselves.

1.5 Comparison of crystal structures of succinic acid in all protonation states using dSNAP.

Paolo Lucaioli, Elisa Nauha and Nicholas Blagden

A conformational cluster analysis using dSNAP has been carried out in order to obtain a wide overview of the conformational families of succinic acid molecules and ions. The analysis has been performed doing CSD searches for the molecules shown in Figure S10.

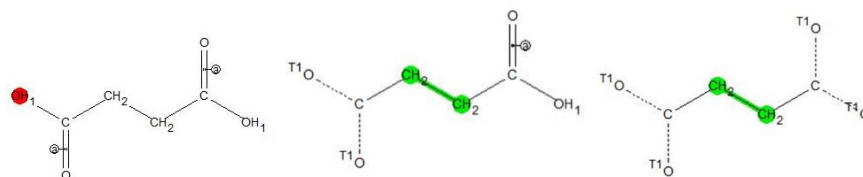


Figure S10. Succinic acid molecule/ions drawn in the Build query tab of CSD for the co-crystal, hemisuccinate and succinate searches

1.5.1 Analysis of non-ionic conformations.

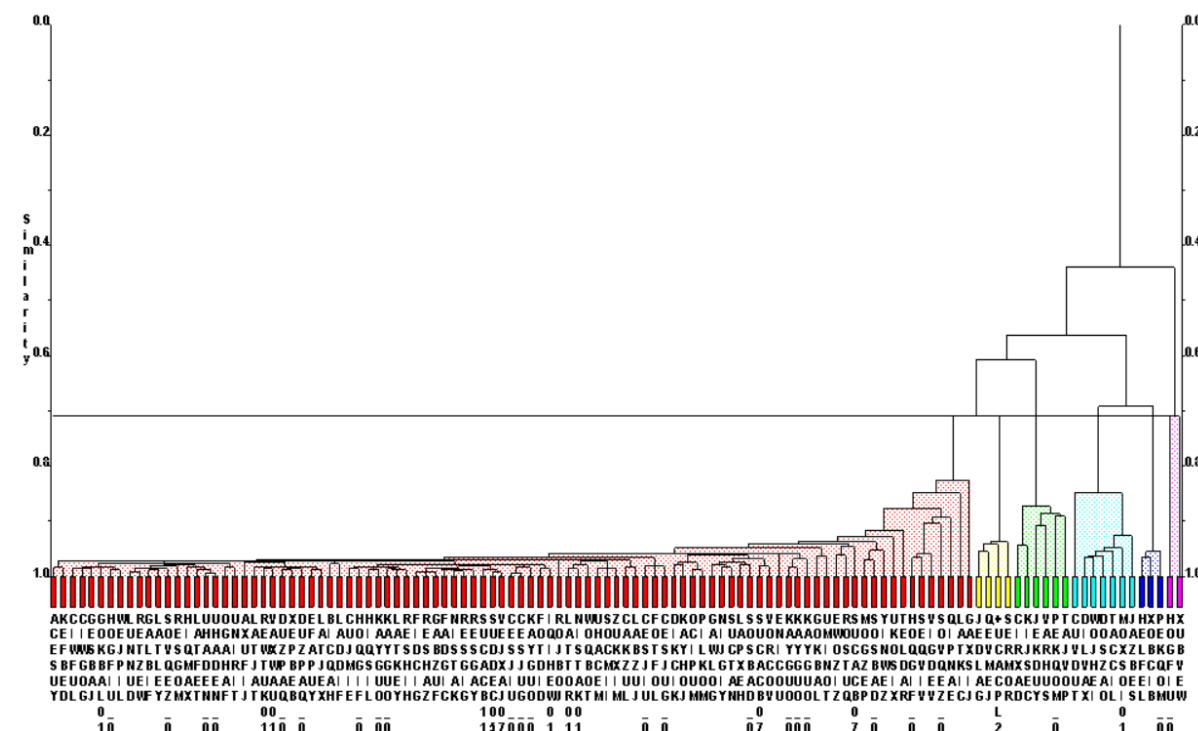


Figure S11. Dendrogram of the co-crystal (fully protonated acid) clustering in dSNAP. Each structure is denoted by its refcode on the horizontal axis, and the tie-lines denote the level of similarity. (The new γ form is denoted +UCACPL2.)

This analysis shows it is possible to identify 2 main conformations of succinic acid involved in co-crystals:

- Planar conformations represented by the red and the cyan (Figure S11 and Figure S12) clusters; the red one contains both previously known polymorphs.
- Folded conformations represented by the yellow, green and blue (Figure S11 and Figure S13) clusters; the new polymorphic form is contained in the yellow cluster.

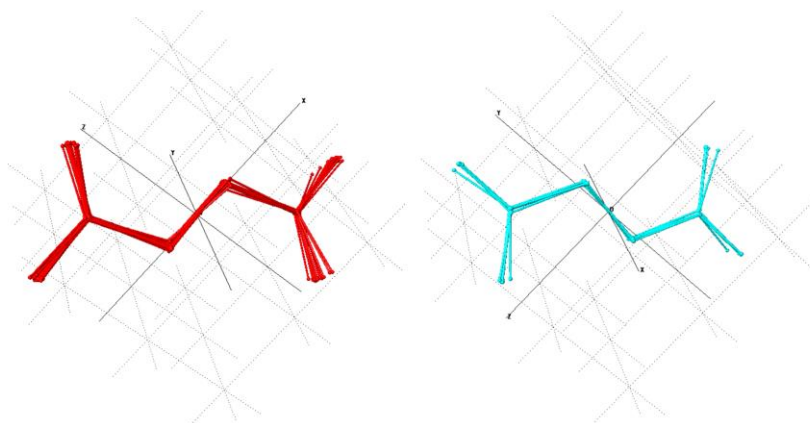


Figure S12. Planar conformations of succinic acid in red and cyan clusters

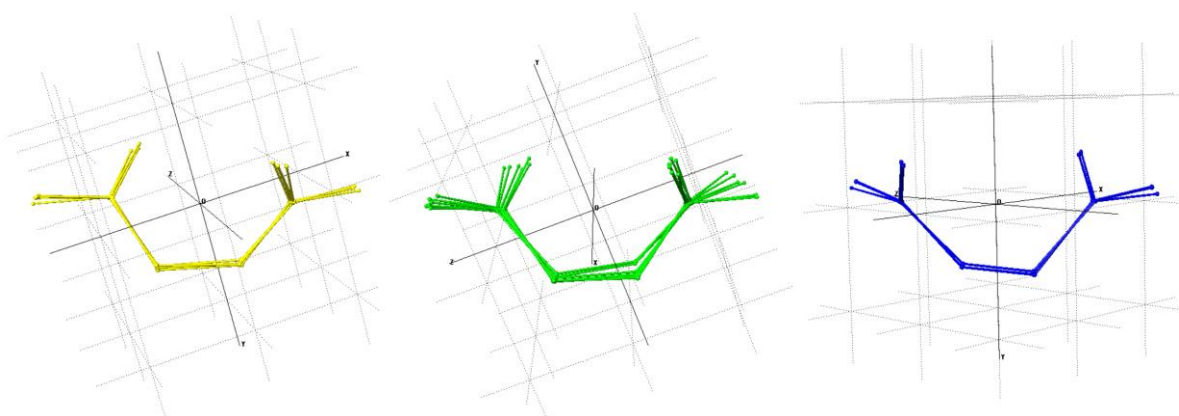


Figure S13. Non-planar (folded) conformations of succinic acid in yellow, green, and blue clusters

The folded non-planar conformations are grouped in three different clusters. The new polymorphic form (labelled +UCACPL2 on Figure S11) from the yellow cluster has been compared with KIJSEC (green cluster) and HELFEL (blue cluster) using the same method described above and the conformational differences are shown in (Figure S14). The main difference is the relative orientations of the carboxylic functionalities: to visualize it in a better way, it might be useful to consider the directions C=O bonds. Placing the carbon chain on the plan we can see three different situations (Figure S14):

- yellow cluster: the two C=O bonds are pointing in opposite directions on the two sides of the carbon chain
- green cluster: C=O bonds are pointing towards the same side of the carbon chain
- blue cluster: the two C=O bonds are pointing on opposite directions on the two sides of the carbon chain but the carboxylate groups have a completely different orientation if compared with those of the yellow group.

Hits grouped in the magenta cluster cannot be visualized using the Multiple Fragment Viewer in dSNAP. Structures of HOGFIU and XUBVEW (folded) were loaded into Mercury and a structure comparison was carried out using the Multiple Structure option and considering other conformations from the yellow, green and blue clusters (Figure S15).

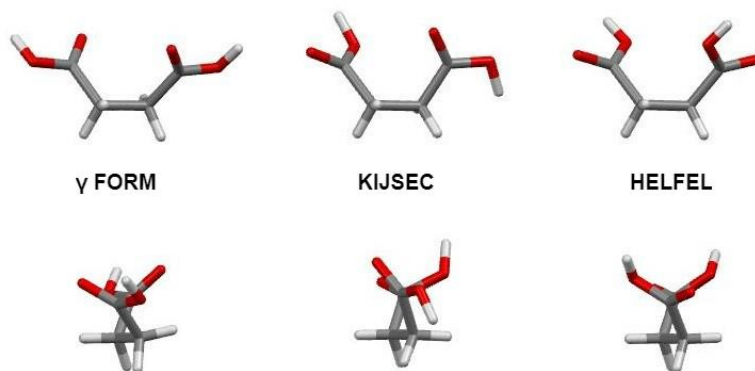


Figure S14. Conformational comparison of the γ form (yellow cluster), KIJSEC (green cluster) and HELFEL (blue cluster). Two different point of views are shown to help the evaluation and comparison of the conformations.

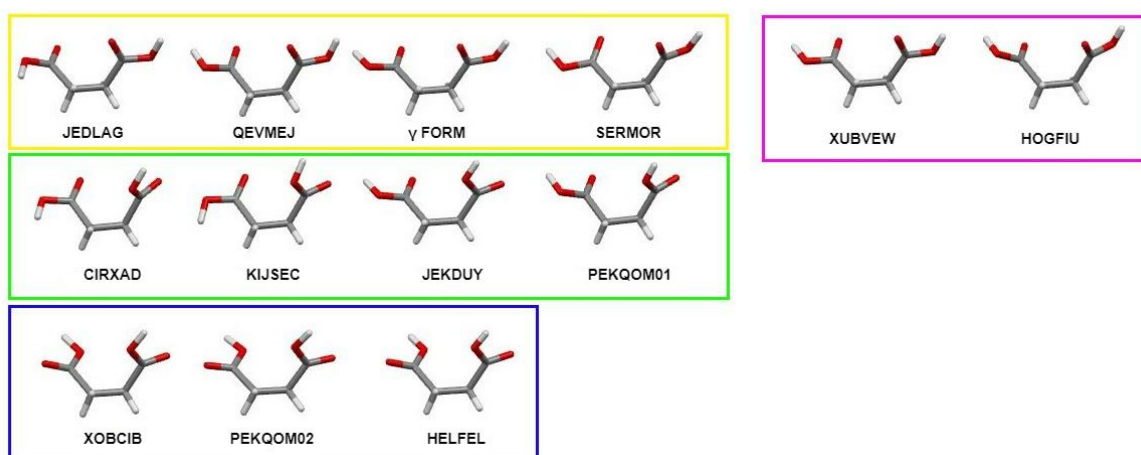


Figure S15. Folded molecules selected from different clusters and compared with hits from the magenta cluster.

1.5.2 Analysis of ion conformations.

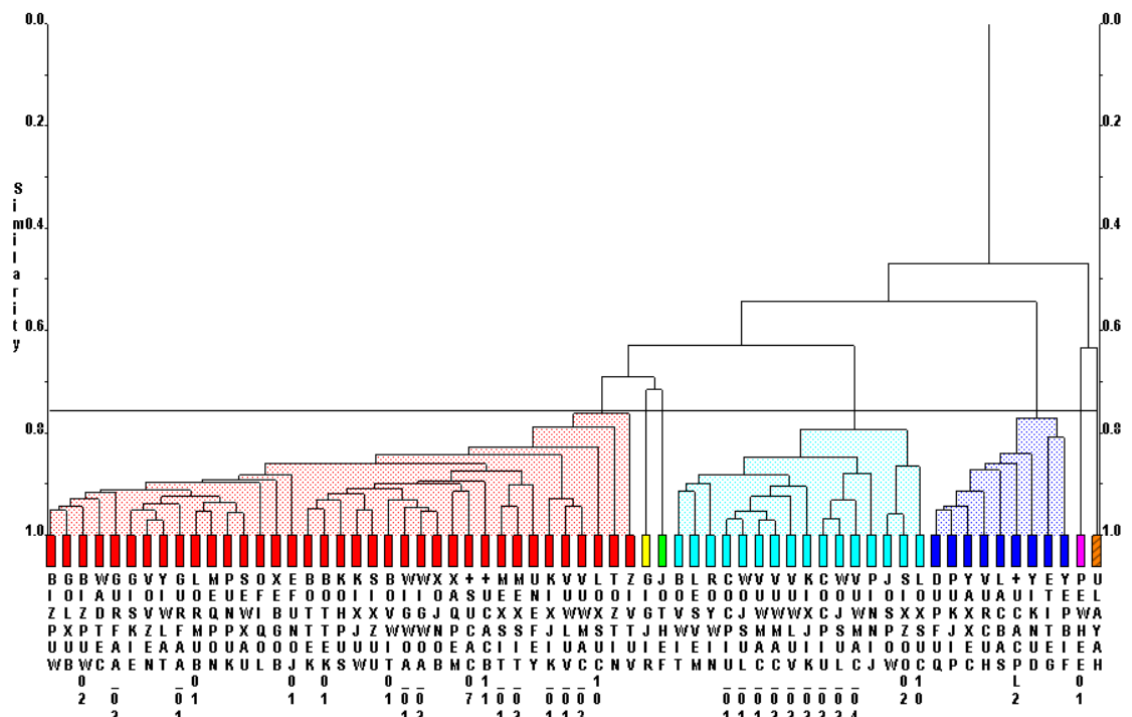


Figure S16. Dendrogram showing the different conformational families of hemisuccinates, including the three polymorphs of succinic acid (+SUCAC07 for α , +UCACB11 for β and +UCACPL2 for γ).

The dSNAP cluster analysis results in 8 different conformational families for the hemi-succinate structures obtained from the CSD search. Planar conformations are represented by the red, yellow, green, and cyan clusters. Folded conformations are represented by blue, magenta, and striped-orange clusters (Figure S16). The new γ form is in the blue cluster and the α and β forms are in the red cluster.

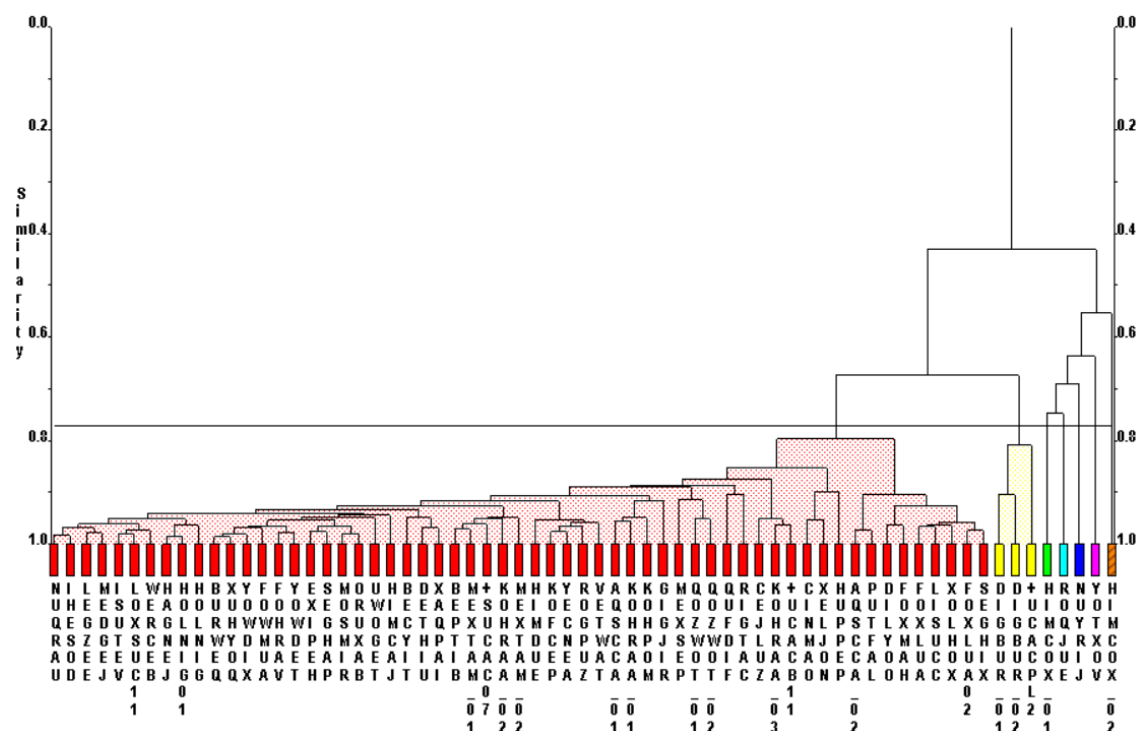


Figure S17. Dendrogram showing the different conformational families of succinates including the three polymorphs of succinic acid (+SUCAC07 for α , +UCACB11 for β and +UCACPL2 for γ).

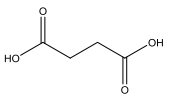
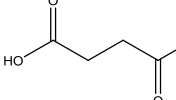
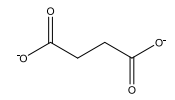
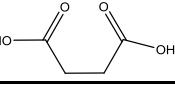
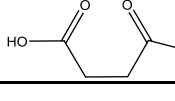
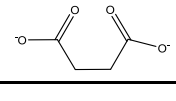
The results of the cluster analysis for the succinate ion identify two well-defined clusters (Figure S17):

- The red cluster representing the planar conformations; it also contains the two previously known polymorphs.
- The yellow cluster containing folded conformations of succinate; in this group, we can find the new γ polymorph.

The five clusters on the right hand side of the dendrogram (green, cyan, blue, magenta and striped-orange) show some error in the elaboration of the data.

Table S12 sums up the occurrence of the planar and folded conformations that can be found in the dSNAP analysis of structures in the CSD for the different protonation states.

Table S12. Percentage occurrence of planar and folded succinic acid molecules or ions in the CSD.

Conformation % in search	Neutral		Anion		Di-anion	
Planar		41.98 %		21.81 %		25.10 %
Folded		5.77 %		4.11 %		1.23 %

2 Crystal Structure Prediction

Louise Price and Sarah Price

2.1 Conformational analysis of Succinic Acid

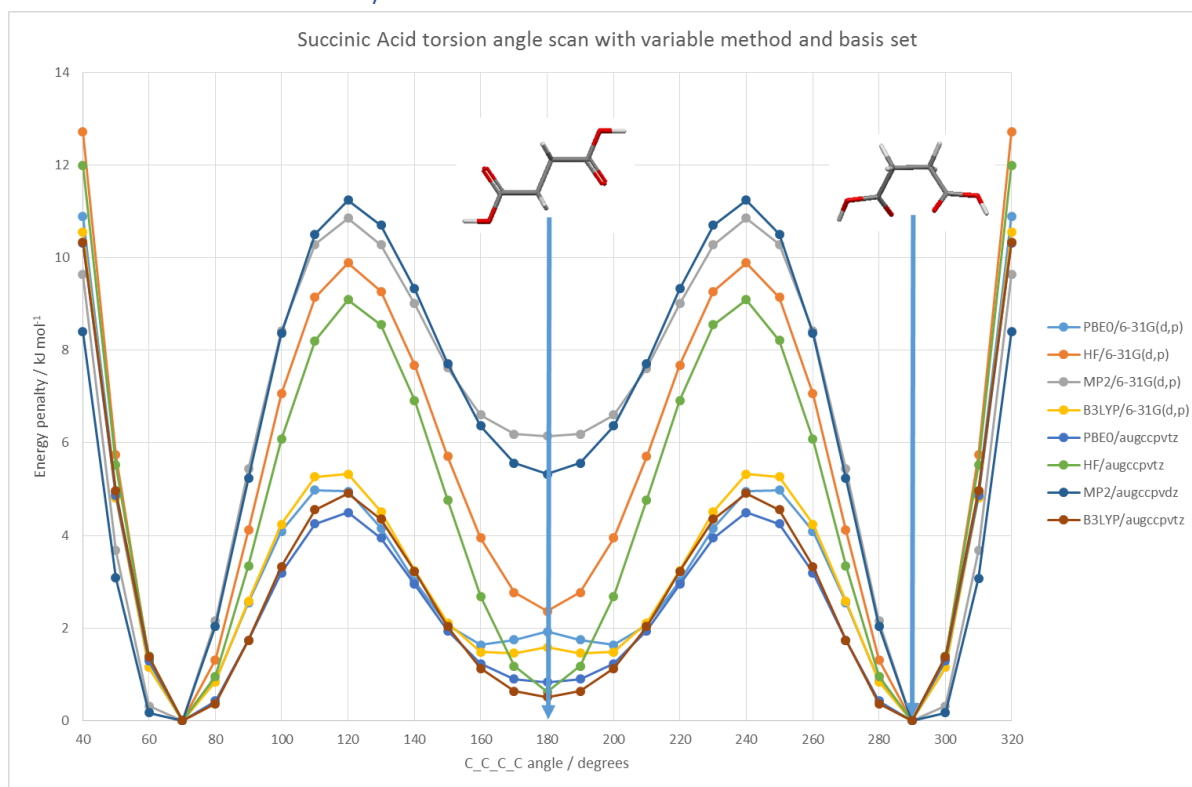


Figure S18. Relaxed scans of the C₁C₂C₃C₄ torsion angle, with different computational methods and basis sets.

2.2 Details of CSP methodology

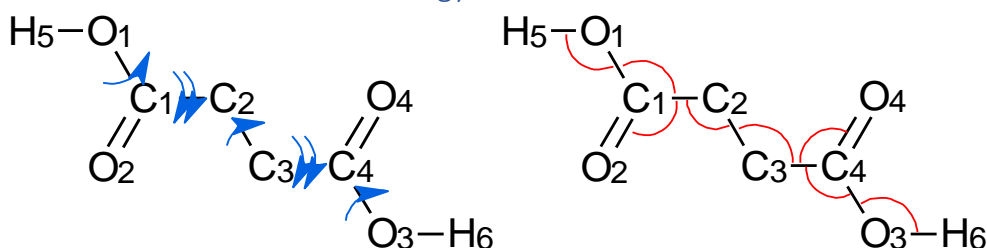


Figure S19. Molecular numbering of Succinic Acid used in this work, with the independent degrees of freedom optimized with CrystalOptimizer marked (blue for 7 torsions, red for 8 angles). Where a double arrow is marked, both torsions around the same bond (e.g. O₂-C₁-C₂-C₃ and O₁-C₁-C₂-C₃) were allowed to optimize independently.

A CSP search for succinic acid was carried out using CrystalPredictor version 2.2.⁴ This was done with a flexible molecule, where the main torsion angles were allowed to vary as described in Table S13. For each combination of independent degrees of freedom, the entire molecule was optimized with GAUSSIAN at the PBE0/6-31G(d,p) level. The Hessian matrix for each Local Approximate Model (LAM) was generated at the same level of theory, to enable accurate interpolation between points. This version of CrystalPredictor allowed the use of non-uniform LAMs, so the grid that was generated (Table S13) was inspected, and any areas with jumps in the energy of more than 10 kJ mol⁻¹ had an additional LAM point calculated. This was repeated with smaller maximum allowed discontinuity sizes until all parts of the LAM were smooth to within 1 kJ mol⁻¹.

1,000,000 crystal structures were generated by CrystalPredictor using any conformation of the molecule described in the database of LAM points. Point charges fixed to those calculated at each LAM conformation in the database were used in the search, which covered the 61 space groups P1, P-1, P21, P21/c, P21212, P212121, Pna21, Pca21, Pbca, Pbcn, C2/c, Cc, C2, Pc, Cm, P21/m, C2/m, P2/c, C2221, Pmn21, Cmc21, Aba2, Fdd2, Iba2, Pnna, Pccn, Pbcm,

Pnmm, Pmmn, Pnma, Cmcm, Cmca, Fddd, Ibam, P41, P43, I-4, P4/n, P42/n, I4/m, I41/a, P41212, P43212, P-421c, I-42d, P31, P32, R3, P-3, R-3, P3121, P3221, R3c, R-3c, P61, P63, P63/m, P213, PA-3, P2221, Pba2. A similar number of structures were produced in each spacegroup, as a small highly symmetrical molecule such as succinic acid will have a tendency to crystallize in higher symmetry space groups than is typical on the CSD. The FIT repulsion-dispersion potential was used.

Table S13. Independent degrees of freedom of succinic acid used in the CrystalPredictor search.

Description of angle	Min LAM point /°	Max LAM point	LAM validity
C1_C2_C3_C4	10	350	±10°
O1_C1_C2_C3	120	240	±20
O3_C4_C3_C2	120	240	±20°
H5_O1_C1_C2	160	200	±20
H6_O3_C4_C3	160	200	±20

The resulting crystal structures were compared by CrystalPredictor's internal clustering program, Analyse, and the 26,827 unique structures within 20 kJ mol⁻¹ were retained.

For each crystal structure, a rigid molecule charge density evaluation at the PBE0/6-31G(d,p) level was carried out, and the distributed multipole analysis of this carried out with GDMA2.2.⁵ The crystal structure was then minimized with DMACRYS, with this multipole description of the charge density and the FIT potential. At this stage of the calculations, some crystal structures were found to have very low density, as there were only very weak or no interactions between layers of the molecules. In these cases a small external pressure (0.2 GPa) was applied to the DMACRYS minimization and the crystal structure reminimized. The pressure was then removed and the structure minimized with DMACRYS again, to ensure that all final crystal structures had been evaluated with the same minimization method. The energy from the rigid molecule charge density evaluation calculation was used to calculate the intramolecular energy penalty for each individual conformation relative to the gas phase optimized conformation, which was added to the intermolecular energy calculated from the distributed multipoles and FIT exp-6 potential. A clustering algorithm based on powder pattern similarity and crystal structure overlay was used to remove duplicate structures, and there remained at this stage 1835 unique crystal structures within 20 kJ mol⁻¹ of the global minimum energy structure.

CrystalOptimizer2.4.4⁶ was then used to minimize the crystal structures, allowing the seven torsion angles and eight bond angles shown in Figure S19 to be optimized in response to the packing forces. The same database of LAMs as was used for the CrystalPredictor search was extended, with LAM validity restricted to any point within 5° for torsion angles and 1° for bond angles, and a second database of the distributed multipole analyses was created, and used to generate the DMA file required for the DMACRYS minimization of each structure. Any structures that corresponded to saddle points on the energy surface (negative eigenvalues are seen in the DMACRYS minimization) were reminimized at lower symmetry structures with the same computational model.

The CSP landscape was therefore calculated from the molecular PBE0 6-31G(d,p) charge densities to give the conformational energy and distributed multipoles, with the empirical FIT potential. To assess the possible effect of an average polarization of the charge density within the crystal, a separate estimate was made by performing the molecular wavefunction calculation in a polarizable continuum (PCM) of $\epsilon=3$.

Structures were inspected by PLATON to see if the symmetry could be increased. The search method uses a whole molecule, but the internal symmetry of the molecule can be included in the space group symmetry (as is the case for all three experimentally observed polymorphs, which contain only half molecules in the asymmetric units).

The results of the CSP search, covering the crystal structures up to -100 kJ mol⁻¹ in total lattice energy, are given in Table S14, and are available from the UCL Chemistry authors on request. The structures are labelled by their rank at the CrystalPredictor stage.

Table S14. CSP generated structures of Succinic Acid. The structures in bold correspond to the experimentally observed polymorphs, whose spacegroups and lattice parameters are included for comparison.

Label	Spacegroup [†]	a / Å	b / Å	c / Å	α / °	β / °	γ / °	Lattice energy / kJ mol ⁻¹	Conformation [*]
A23778	P21 (P21/c)	5.3798	8.7792	5.1469	90	93.8	90	-110.537	aAa
SUCACB03	P21/c	5.464(1)	8.766(3)	5.004(1)	90	93.29(3)	90		aAa
A8149	Pca21 (Pbcn)	8.1662	5.2878	11.4305	90	90	90	-109.149	aGa
A763	P21/c	5.2755	18.1732	5.1788	90	94.981	90	-107.666	aAa
A20629	Pa (P21/c)	6.2105	7.3703	5.4215	90	96.958	90	-106.546	aAa
A4158	P-1	4.7981	5.1337	5.3336	84.866	83.223	73.848	-106.264	aAa
A19217	P-1 (C2/c)	5.7798	8.2793	8.2793	89.408	118.526	90.839	-105.868	aGa
gamma	C2/c	5.7015(5)	8.4154(8)	10.3538(8)	90	90.374(3)	90		aGa
A6712	C1 (P-1)	2.5237	4.7917	10.7391	92.573	95.592	102.808	-105.679	aAa
A22937	Cc	7.0321	7.7117	24.2261	90	129.953	90	-105.532	aAa
A5862	Fd (Cc)	6.9877	37.6146	7.6892	90	91.501	90	-105.166	aAa
A11752	Pa (Pca21)	17.2354	5.5604	5.2402	90	90	90	-104.89	aAa
A2798	Pbca	9.5981	8.9601	11.982	90	90	90	-104.004	aGa
A10824	Pbca	5.2491	11.3415	17.0103	90	90	90	-103.95	aGa
A19123	P21/c	5.1301	18.8799	5.2776	90	84.552	90	-103.434	aAa
A16898	la	5.1642	19.3814	5.1913	90	84.222	90	-102.879	aAa
A1308	Pc	5.2294	17.6804	5.5968	90	89.717	90	-102.637	aAa
A1904	Pbcn	17.028	5.2864	11.4491	90	90	90	-102.498	aGa
A20462	P21/c	5.3874	19.5035	4.9549	90	96.293	90	-102.495	aAa
A4846	A1 (C2/c)	7.7198	6.8631	9.996	90	103.511	90	-102.385	aAa
A1488	P21/c	7.7036	12.7095	5.6266	90	69.035	90	-102.37	aAa
A2407	P212121	9.3048	7.6681	7.0567	90	90	90	-102.016	aAa
A4200	P21	7.7307	14.1747	4.8411	90	74.797	90	-101.968	aAa
A804	P21/c	7.7452	13.8764	5.2004	90	116.794	90	-101.62	aAa
A9393	P21	5.7367	6.4344	7.4512	90	109.591	90	-101.546	aAa
SUCACB07	P-1	6.867(3)	7.198(2)	5.727(2)	109.10(2)	97.18(3)	101.84(3)		aAa
A2779	P-1	8.1009	4.7786	7.973	69.591	74.485	106.635	-101.502	aAa
A14464	P-1	5.649	5.1223	10.0908	87.804	90.544	61.347	-101.469	aGa
A1161	P21/c	9.6778	7.7009	7.1521	90	77.433	90	-101.314	aAa
A4939	C2/c	14.6944	4.8964	14.1913	90	92.538	90	-101.305	aAa
A1667	P21/c	7.6855	9.5845	7.4948	90	69.198	90	-101.212	aAa
A1107	An	10.2049	4.7834	10.2751	90	85.76	90	-101.184	aAa
A2778	P21/c	4.7734	7.1727	16.0872	90	68.441	90	-101.106	aAa
A2396	P43 (P43212)	8.1463	8.1463	7.6656	90	90	90	-101.089	aAa
A2473	P21/c	7.6956	9.7533	7.2278	90	109.36	90	-101.077	aAa
A18857	Pca21	15.0057	4.7754	6.9762	90	90	90	-101.073	aAa
A2585	Pca21	7.7078	10.3596	13.0366	90	90	90	-101.029	aAa
A2526	Pna21	10.2756	7.7262	6.5519	90	90	90	-100.963	aAa
A4816	Pbca	16.8618	5.2928	11.5439	90	90	90	-100.926	aGa
A618	P21/c	5.4798	5.3262	17.8804	90	85.526	90	-100.86	aAa
A10263	P-1	6.8933	6.1187	7.6835	62.951	111.126	84.026	-100.782	aAa
A123	Pc (P21/c)	5.5266	5.2441	8.8198	90	96.796	90	-100.757	aAa
A23063	P-1	5.3905	8.2038	5.9262	95.576	87.658	85.438	-100.742	aAa
A3917	P21/c	7.0502	10.2598	7.708	90	67.494	90	-100.69	aAa
A16988	P-1	5.2804	5.2485	9.6746	104.951	91.706	94.453	-100.612	aAa
A4947	Cn	4.7663	14.7748	14.5334	90	84.999	90	-100.591	aAa
A9203	P21/c	9.99	7.6735	14.6736	90	68.668	90	-100.571	aAa
A11173	P21/c	10.8176	7.7012	14.3429	90	59.496	90	-100.341	aAa
A3882	P21/c	7.6906	6.9587	10.4855	90	67.935	90	-100.315	aAa
A2910	P-1	5.1724	10.8902	4.9019	90.149	70.817	86.886	-100.266	aAa
A15665	Pna21 (P21212)	4.6843	10.6533	10.5247	90	90	90	-100.235	aGa
A3841	Fdd2	29.8541	13.7215	4.9553	90	90	90	-100.2	aAa
A2478	C2/c	12.7098	5.6708	14.3828	90	84.227	90	-100.153	aAa

[†] The space group of the search generated structure is given. If a difference space group symmetry is identified by Platon, this is given in parentheses. ^{*} Conformation as defined in SI Section 4.2.

2.3 Match of known and CSP structures

Structures A23778, A9393 and A19217 generated in the CSP study corresponded to the low energy β form, the high energy α form and the new conformational γ form, respectively (Figure S20). It is apparent that the high temperature form is less well reproduced with the energy model chosen for this work, which is at a notional 0 K, although some temperature effects are included in the FIT repulsion-dispersion potential parameterization of room temperature crystals structure data.

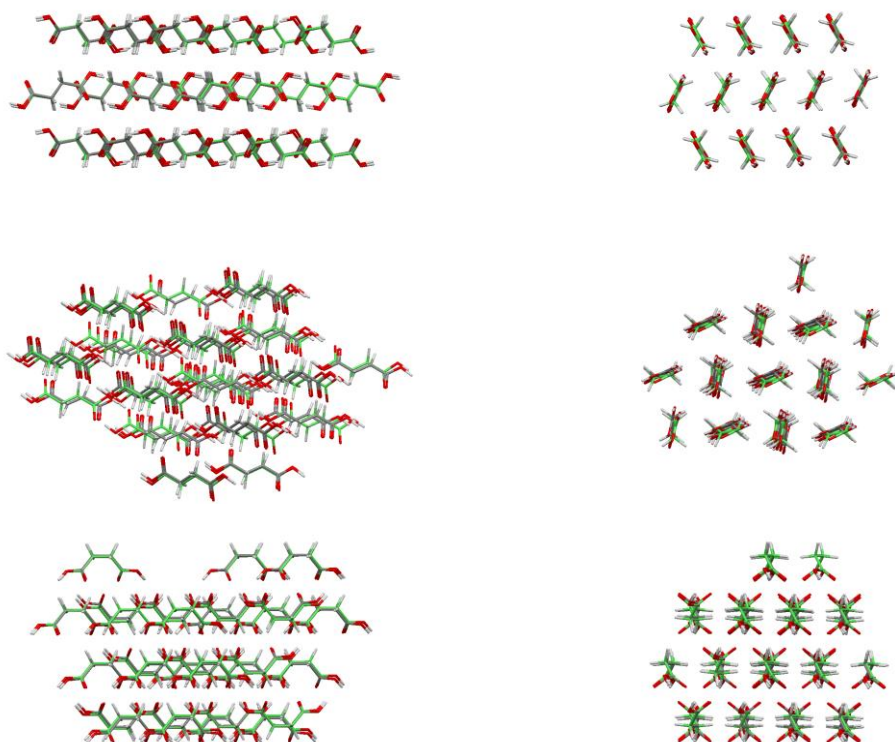


Figure S20. Overlays of experimental forms (coloured by element) with search structures (green). Top: monoclinic β form SUCACB03 with A23778 ($\text{RMSD}_{30} = 0.1854 \text{ \AA}$), Middle: triclinic α form SUCACB07 with A9393 ($\text{RMSD}_{30} = 1.052 \text{ \AA}$), Bottom: new γ form with A19217 ($\text{RMSD}_{30} = 0.170 \text{ \AA}$). Left: side view of the chains, Right: end view of the chains.

3 DFT-D relative energetics of experimental and CSP crystal structures of succinic acid

Rui Guo and Sarah Price

The input structures of succinic acid used in the periodic DFT-D calculations are listed in Table S15.

Table S15. Initial cell parameters of experimental and 3 low-lying CSP crystal structures as input for DFT-D optimization.

Name	Space Group	a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)
SUCACB03 β	P21/c	5.464	8.766	5.004	90	93.29	90
SUCACB07 α	P-1	6.867	7.198	5.727	109.10	97.18	101.84
γ (this work)	C2/c	5.7015	8.4154	10.3538	90	90.37	90
dA8149	Pbcn	8.1662	5.2878	11.4305	90	90	90
dA763	P21/c	5.2755	18.1732	5.1788	90	94.98	90
dA4158	P-1	4.7981	5.1337	5.3336	84.866	83.223	73.848

3.1 Method

Full DFT-D crystal structure optimizations were carried out with CASTEP ver. 18.1 for the three experimental forms, and three competitive CSP structures, using the pure PBE functional and Tkatchenko-Scheffler (TS)⁷ or Grimme's D02⁸ dispersion correction scheme, with on-the-fly ultrasoft pseudopotentials. A plane wave cutoff energy of 1100 eV and k-point grid spacing of 0.10 Å⁻¹ were used after extensive convergence testing. Structural optimizations were carried out using the BFGS algorithm with an SCF electronic energy tolerance of 10⁻¹⁰ eV, force convergence tolerance of 0.001 eV/Å, and fine grid scale of 4. The PBE-TS and PBE-D02 optimized structures are listed in Table S16 and Table S17.

After DFT-D optimizations, other dispersion correction schemes beyond pairwise dipole-dipole interactions, such as Tkatchenko-Scheffler's Many-Body Dispersion scheme (MBD*)⁹ and Grimme's D03 scheme,^{10, 11} were used for single-point energy calculations on the PBE-TS and PBE-D02 optimized crystal structures, with the results shown in Table S18.

For the β and γ forms, phonon calculations were performed on PBE-TS optimized structures, using the finite displacement method with 2x6x2 and 2x2x2 supercells for β and γ respectively to make sure the phonon Brillouin zones were sufficiently sampled to converge the energy difference. Zero-point energies and vibrational free energy corrections were estimated from the phonon density of states calculations within the harmonic oscillator approximation from 10 K to 460 K and are given in Figure S21 and Figure S22.

3.2 Results

Table S16. PBE-TS optimized cell parameters of experimental structures and selected CSP crystal structures of succinic acid.

Name	Space Group	a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)	Cell volume (Å ³)	RMSD ₁₅ (Å)*
SUCACB03 (β form)	P2 ₁ /c	5.4864	8.6848	5.0896	90	91.57	90	242.42	0.078
SUCACB07 (α form)	P-1	6.7152	7.1321	5.6680	108.64	96.47	101.38	247.60	0.101
γ (this work)	C2/c	5.5690	8.2549	10.5923	90	90.09	90	486.95	0.142
dA8149	Pbcn	7.7532	5.2346	12.0771	90	90	90	490.15	0.386
dA763	P2 ₁ /c	5.4213	17.5912	5.1163	90	91.87	90	487.67	0.280
dA4158	P-1	4.7651	5.3850	5.1553	88.18	83.54	68.72	122.48	0.367

* RMSD₁₅ was calculated with MERCURY using default setting compared to the input structures.

Table S17. PBE-D02 optimized cell parameters of experimental structures and selected CSP crystal structures of succinic acid.

Name	Space Group	a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)	Cell volume (Å ³)	RMSD ₁₅ (Å)*
SUCACB03 (β form)	P2 ₁ /c	5.4876	8.5526	4.9190	90	93.35	90	230.47	0.099
SUCACB07 (α form)	P-1	6.6014	7.0616	5.7123	108.33	97.64	104.71	237.83	0.173
γ (this work)	C2/c	5.6339	8.0667	10.4289	90	89.54	90	473.95	0.129
dA8149	Pbcn	7.5512	5.1705	12.0188	90	90	90	469.26	0.402
dA763	P2 ₁ /c	5.3701	17.5635	4.9682	90	94.04	90	467.42	0.255
dA4158	P-1	4.7050	5.2711	5.0549	85.80	83.09	71.79	118.13	0.299

* RMSD₁₅ was calculated with MERCURY using default setting compared to the input structures.

The relative energies of all structures in Table S16 and Table S17, along with their single-point energies calculated with MBD* and D03 dispersion schemes, are listed in Table S18. For comparison, CrystOpt (PCM) relative energies are also shown.

Table S18. Relative energies of experimental structures and selected CSP crystal structures of succinic acids.

Name	PBE-TS optimized				PBE-D02 optimized				CrystOpt (PCM)
	TS	MBD*	D02	D03	TS	MBD*	D02	D03	
SUCACB03 (β form)	0	0	0	0	0	0	0	0	0
SUCACB07 (α form)	1.71	-0.56	-0.44	-0.93	1.10	-0.93	-0.53	-1.45	8.28
γ (this work)	-0.42	-0.10	-1.02	-1.08	-1.63	-0.53	-0.59	-2.14	3.69
dA8149	3.54	1.95	1.08	1.56	2.24	1.43	1.26	0.85	-0.44
dA763	0.21	0.40	0.32	-0.07	0.50	0.53	0.47	-0.19	3.08
dA4158	0.44	0.42	0.68	0.01	0.44	0.16	0.52	-0.63	4.40

The results in Table S18 were also shown in manuscript Figure 3, and compared to the energies in the search (CrystOpt) and CrystOpt(PCM).

Compared to CrystOpt which has the energy difference between α and β forms greater than 8 kJ mol⁻¹, DFT-D relative energies for the three known polymorphs, no matter which dispersion correction was used, all lie within a much narrower range of 2 kJ mol⁻¹. Frequently, the three known forms of succinic acids can be found within an energy range of less than 1 kJ mol⁻¹, which is within the current limit of accuracy of available computational methods for relative energies of different polymorphs of organic molecular crystals. With such a small energy difference, other contributions to the relative energies, which are routinely omitted, become more important, such as zero-point energy, free energy corrections and thermal expansion.

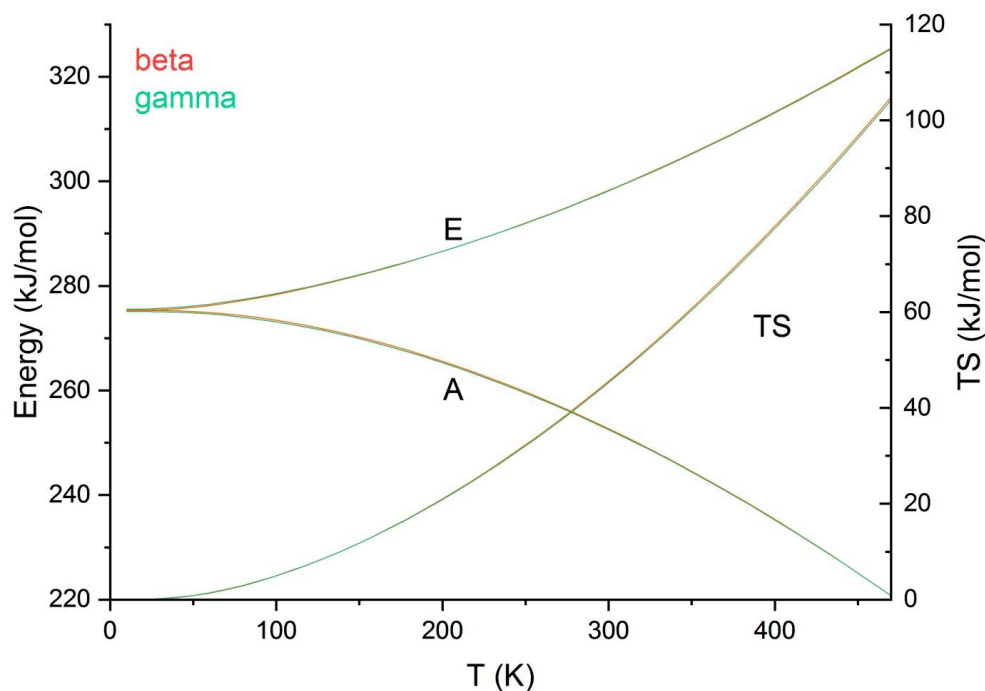


Figure S21. The Helmholtz free energy (A), internal energy (E , which includes lattice energy, zero-point energy and thermal correction), and the vibrational TS terms of β and γ forms of succinic acid, calculated using the PBE-TS method within the harmonic approximation. The difference is too small to be visible.

Although the Helmholtz free energy depends strongly on temperature (Figure S21) the difference between the two polymorphs is small. However, as shown in Figure S22, the relative stability of the γ form decreases with increasing temperature, and the Helmholtz free energies of the two forms cross each other around ambient temperature. This is mainly because of the larger vibrational entropy of the β form.

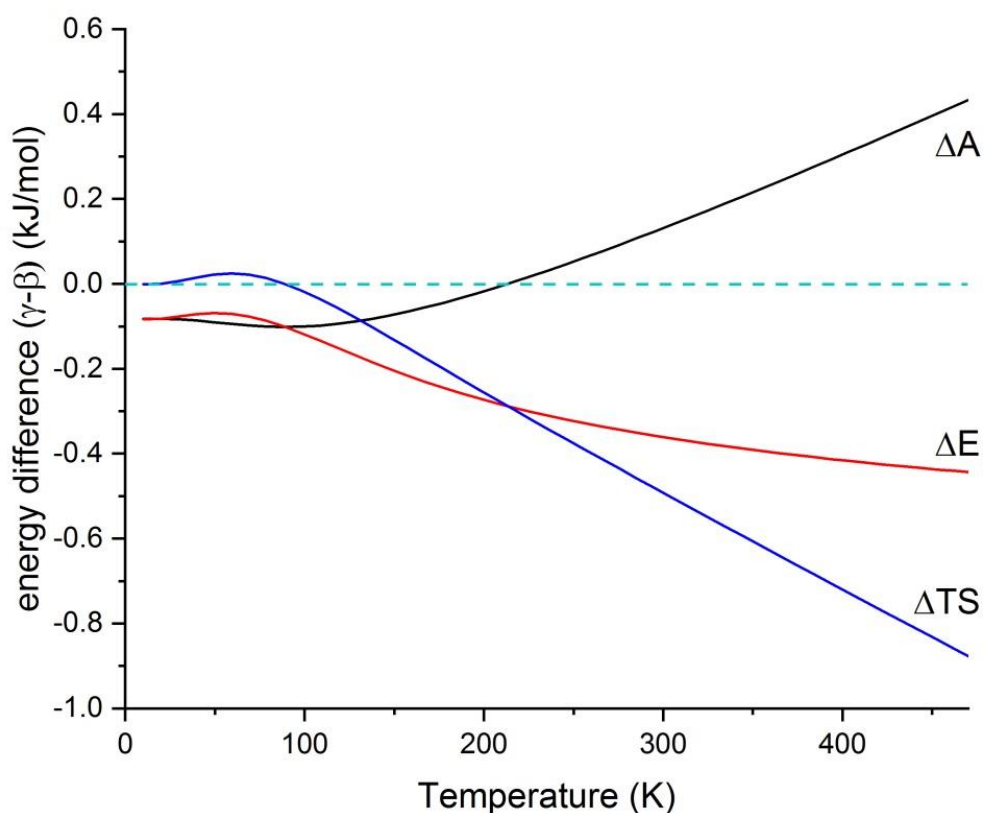


Figure S22. Relative Helmholtz free energies (ΔA), internal energy (ΔE , E includes lattice energy, zero-point energy and thermal correction) and vibrational TS terms (ΔTS) of the β and γ forms of succinic acid as calculated using PBE-TS lattice energy and phonons calculated within the harmonic approximation.

4 Molecular Dynamics Studies

Ilaria Gimondi and Matteo Salvalaglio

4.1 Methods

For the analysis of the conformational behaviour of succinic acid in water, we employ molecular dynamics (MD), well-tempered metadynamics (WTMetaD) and Markov State Model (MSM), which allow identification and distinction of succinic acid's conformers, evaluation of their free energy, and study of their equilibrium distribution and characteristic time for such equilibration. Crystal structure stability is investigated with MD.

Molecular dynamics (MD) is a powerful tool to access the molecular scale of a system and uncover the dynamics of its evolution; however, it presents limitations that prevent sampling rare events, as these processes take place on timescales that cannot be accessed by standard MD.¹²⁻¹⁶ To circumvent such limitations, enhanced sampling techniques are developed to accelerate rare events and enable their sampling. Among these techniques, we employ well-tempered metadynamics¹⁷ (WTMetaD), a variant of metadynamics¹⁸ (MetaD). In short, metadynamics methods are based on the introduction of an external history-dependent bias potential along with a well-chosen and low-dimensional set of collective variables (CVs), which are descriptors of the configurational state of the system. The introduction of such bias forces the system to leave the initial basin and explore other stable and metastable configurations in the CV-space. It is thus possible to sample rare events. Another advantage of MetaD/WTMetaD is the possibility of recovering the free energy of the system as a direct output of the simulation; such outputs provide the free energy as a function of the CVs (free energy surface, FES). Interesting, little *a priori* knowledge of the CV-space is required. As a comprehensive introduction to MetaD and WTMetaD is beyond the aim of this introduction, we refer the interested reader to insightful reviews on the technique and its applications.¹⁹⁻²¹ The model of succinic acid is the General Amber Force Field (GAFF)²² and that for water is TIP3P.

4.1.1 MD conformers of succinic acid

We investigate the evolution of one succinic acid molecule in a solution of 2157 water molecules. MD simulations are integrated with a 0.001 ps step, using dispersion corrections to evaluate van der Waals interactions, and particle mesh Ewald (pme) for the electrostatics, with 1 nm cutoff. The NPT ensemble is obtained by employing the Bussi-Donadio-Parrinello thermostat²³ and the Berendsen barostat²⁴ at 300 K and 1 bar, respectively. We perform 18 independent 100 ns-long MD simulations, starting from a different conformer of succinic acid (see Table S19).

As far as WTMetaD is concerned, we employ the same set up as standard MD. An additional concern here is the selection and definition of the collective variables, whose choice is pivotal to run effective simulations. As our focus is on succinic acid conformers, we employ as CVs the three main structural dihedrals of the molecule, hereafter referred to as t_0 , t_1 and t_2 (highlighted in Figure S23a, t_0 corresponds to the torsion of the carbon atoms (C1-C2-C3-C4), while t_1 and t_2 consider the rotation of the hydroxyl oxygen of the acid groups (O1-C1-C2-C3 and C2-C3-C4-O3, respectively). The bias is introduced along these CVs as a sum of Gaussians with initial height 1.2 kJ mol⁻¹, width 0.2 rad for each dihedral; the bias factor is 5 and the deposition takes place every 5 ps. The same setup is used to compute the free energy associated with conformational changes of a single succinic acid molecule in the bulk of the β phase. For simulations in water, WTMetaD started from conformer aAa (see Table S19) and ran for about 458 ns.

4.1.2 Succinic acid crystals

MD simulations of succinic acid molecular crystals employ the same set up discussed above. Supercells of the β and γ polymorphs are obtained by replicating 64 unit cells, i.e. structure SUCACB12 (refcode 929783) of the Cambridge Crystallographic Data Centre (CCDC) for β and the one reported in this work for γ . The so-built cells firstly undergo energy minimization, followed by 50 ns NVT equilibration. Then, for each polymorph, the equilibrated configuration is simulated in the NPT ensemble; in particular, three different pressure coupling conditions are employed to investigate the stability of the crystals: isotropic, anisotropic with and without possibility for the cell angles to fluctuate (hereafter referred to as fully anisotropic and anisotropic, respectively). Metadynamics simulations have been performed instead on larger supercells (512 molecules, i.e. 128 unit cells) of the γ polymorph in order to gain

insight into its structural relaxation and its destabilization due to fluctuations in the β cell angle. Details on these simulations are reported in the results section.

4.1.3 Markov State Model (MSM)

Markov State Models (MSM) are kinetics models employed to analyze MD trajectories through the construction of a network of macrostates.²⁵⁻²⁸ The statistical approach of this model allows the processing of long simulations in a more effective way, adding a new understanding of the underlying mechanism of a process defined by macrostates interactions, rather than a mere visualization of a high-dimensional space. In addition, MSM enables evaluation of equilibrium and kinetics properties observable on timescales not accessible from MD (as discussed above) from a combination of shorter MD simulations. Thanks to these advantages, MSM theory is commonly used in the study of macromolecule conformational changes, such as protein folding/unfolding, protein binding with ligands etc.

Briefly, the model construction begins with the definition of distinct macrostates, which consists of grouping structures in a kinetically relevant way. It is inevitable that a first approximation is based on the geometry of the system. Once the macrostates are identified, the transition matrix $\mathbf{C}(\tau)$ is built, where τ is the lag time and each element C_{ij} counts the observed transitions between the i -th and j -th states. From $\mathbf{C}(\tau)$ it is possible to obtain the probability matrix, whose eigenvalues and eigenvectors are central output of the model: indeed, adequately chosen eigenvalue and eigenvector represent respectively the characteristic relaxation time and equilibrium distribution. This brief overview of MSM aims only to introduce the statistical analysis conducted on our MD simulations, while details on the formulation and framework can be found in the cited literature.

4.1.4 Tools

MD and WTMetaD simulations were carried out using Gromacs 5.2.1¹⁴ patched with Plumed 2.3;²⁹ post-processing of the outputs employs Python,³⁰ Visual Molecular Dynamics (VMD)³¹ supplemented with GISMO.³²

4.2 Results: MD conformers of succinic acid in aqueous solution

4.2.1 Metadynamics Simulations

In order to uncover the conformational space of succinic acid we performed WTMetaD at 300 K and 1 bar biasing the three main structural dihedrals as CVs, as described in Section 4.1. Dihedrals t_0 , t_1 , and t_2 are defined in Figure S23a; we note that the carboxylic acid groups remain planar and the description of the dihedral angle space does not depend on the choice of a specific oxygen atom. The output FES, as a function of the three dihedrals, shows a wide conformational sampling and, due to its high dimensionality, the phase space appears difficult to visualize; to better present the results, Figure S23 plots the projection of the FES on the t_0 - t_1 plane (Figure S23b) (t_0 - t_2 omitted for symmetry) and on the t_1 - t_2 plane (Figure S23c), while in the 3D representations in Figure S23d and Figure S23e only volumes up to ~ 15 kJ mol⁻¹ above the global free energy minimum are represented. The study of the symmetry of succinic acid enables reduction of such a complicated configurational space to just one quarter of the total volume: the explicative quarter shown in Figure S23e corresponds to the volume for which $t_0 \geq 0$, and $t_1 \geq t_2$.

On this selected volume we located the free energy minima and identified the corresponding conformers, represented as black spheres in Figure S23e with volume proportional to their stability. First of all, we notice that as far as their carbon skeleton (t_0) is concerned the conformers are grouped in two families: anti ($t_0 \sim 180^\circ$) and gauche ($t_0 \sim 58-67^\circ$). Once the structural symmetry is taken into account, we identify 5 different gauche conformers and four anti, by studying the values of dihedrals t_1 and t_2 . To distinguish the configurations, we assign each a nomenclature composed of three letters that represent the values of the torsions in the t_1 - t_0 - t_2 sequence: a stands for anti, and g for gauche. To better identify the t_0 torsion, this is assigned a capital letter. As an example, the conformer with all the three dihedrals planar is aAa. One structure is identified as HE, as it does not belong to any of the two main families, but it has the highest free energy. The summary of this classification is reported in Table S19, while Figure S27a (and Figure 4 of the manuscript) shows a 3D ball-and-stick representation of the A and G conformers and the associated nomenclature. A detailed discussion of the relative stability of these configurations and its implications will follow with the comparison with MSM results.

Table S19. Succinic acid conformers in water at 300 K found after symmetry was considered, defined in the volume presented in Figure S23e. The name attributed to each structure is made of three letters based on the $t_{1,2}$ - t_0 - $t_{1,2}$ sequence, and are either g for gauche or a for anti. Structure HE does not belong to any of the two main families, and it is found only through WTMetaD. The probabilities of conformers in solution are reported as obtained from 18 100 ns-long unbiased molecular dynamics simulations through a Markov State Model (MSM) and from WTMetaD.

conformer	t_0 [°]	$t_{1,2}$ [°]	$t_{1,2}$ [°]	Free energy at 300 K [kJ mol ⁻¹]	probability MSM [%]	probability WTMetaD [%]
aAa	175.955 (A)	-175.955	-175.955	2.530	7.257 ± 1.306	8.571 ± 2.085
aA-g	175.955 (A)	175.955	86.975	5.407	6.288 ± 1.176	6.875 ± 1.560
-gA-g	175.955 (A)	86.975	86.975	8.983	0.747 ± 0.334	0.684 ± 0.225
gA-g	175.955 (A)	86.975	-90.986	8.548	0.838 ± 0.235	0.934 ± 0.228
aGa	62.682 (G)	-167.877	-171.887	0	67.576 ± 3.321	69.181 ± 3.556
gGa	66.750 (G)	70.760	-175.955	4.372	9.637 ± 1.585	8.574 ± 1.332
gG-g	58.671 (G)	82.907	-103.132	4.731	4.129 ± 0.999	2.751 ± 0.959
gGg	58.671 (G)	78.896	74.828	5.482	2.073 ± 0.476	1.500 ± 0.274
-gG-g	62.739 (G)	-111.211	-111.211	6.806	1.455 ± 0.706	0.884 ± 0.668
HE	74.828	54.603	-90.986	11.47	–	0.044 ± 0.035

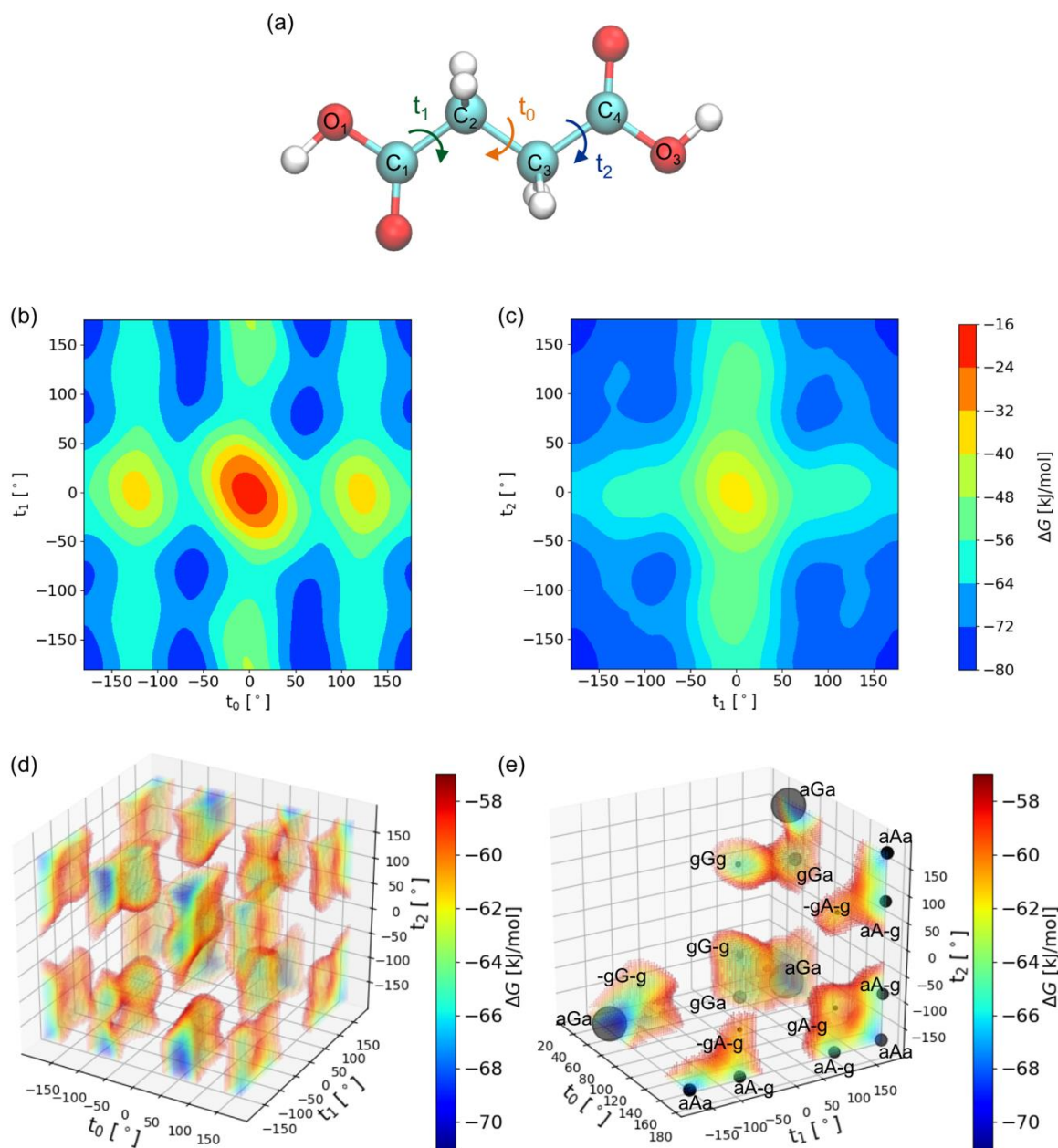


Figure S23. Results of WTMetaD on a molecule of succinic acid in water at 300 K biasing 3 CVs, namely t_0 , t_1 and t_2 . The molecular structure in (a) shows the mentioned dihedrals; in particular, t_0 (orange) represents the torsion of the four carbons C1-C2-C3-C4, while t_1 (green) and t_2 (blue) includes the oxygens of the hydroxyl group (O1-C1-C2-C3 and C2-C3-C4-O3 respectively). For an immediate visualization of the free energy of the system, we show in (b) and (c) two 2D FESs, projected on t_0 - t_1 and on t_1 - t_2 , respectively. The t_0 - t_2 FES is not reported for symmetry reasons. The 3D representation of the free energy as a function of all three dihedrals biased is reported in (d), while (e) shows the quarter of the space under analysis, together with the local minima, i.e. distinct conformers. Such conformers are represented as black spheres, centred in the minimum and with a volume proportional to its equilibrium probability, evaluated as a function of the free energy as: $p(t_0, t_1, t_2) = \exp\left(-\frac{\Delta G(t_0, t_1, t_2)}{kT}\right)$.

Table S20. Potential energy of succinic acid conformers in solution (from 17 100 ns-long UB simulations) and from energy minimization in vacuum. For the values in solution, we present the breakdown of the final overall average into the contributions of the interactions within the succinic acid molecule and the ones between the molecule and the solvent (water); this allows comparison of the results of the energy minimization in vacuum with the potential energy due to only the interaction within the succinic acid molecule in solution, i.e. once the effect of the solvating shell is removed. No value is reported for aA-g, -gA-g and gA-g, as during the optimization the molecule rearranges to aAa. The optimized value for aGa is employed as reference for the lattice energy in Table S22 and Table S24, as this is the minimum.

conformer	E [kJ mol ⁻¹] in solution	E [kJ mol ⁻¹] in vacuum
aAa	-573.086 ± 20.550	-348.016
aA-g	-573.556 ± 20.635	–
-gA-g	-572.970 ± 19.723	–
gA-g	-574.853 ± 20.444	–
aGa	-575.769 ± 20.754	-350.507
gGa	-576.932 ± 20.425	-341.269
gG-g	-576.433 ± 20.314	-343.843
gGg	-577.014 ± 20.670	-343.125
-gG-g	-573.841 ± 19.417	-344.078
total average	-576.789 ± 21.206	–
SUC-SUC interactions	-366.260 ± 25.015	–
SUC-rest interactions	-210.529 ± 34.755	–

4.2.2 Analysis MD trajectories: MSM

Preliminary metadynamics exploration enables a clear and univocal definition of succinic acid conformers, which is a pivotal step towards the implementation of the Markov State Model. Indeed, it is key to determine kinetically relevant macrostates, often starting from geometrical considerations. Here, we consider each stable configuration as a state and identify the boundaries of the corresponding basin in the t_0 - t_1 - t_2 space (Figure S24); only the 9 A and G conformers are included in this model as the free energy of HE is much higher than the others, thus reducing its probability to close to zero.

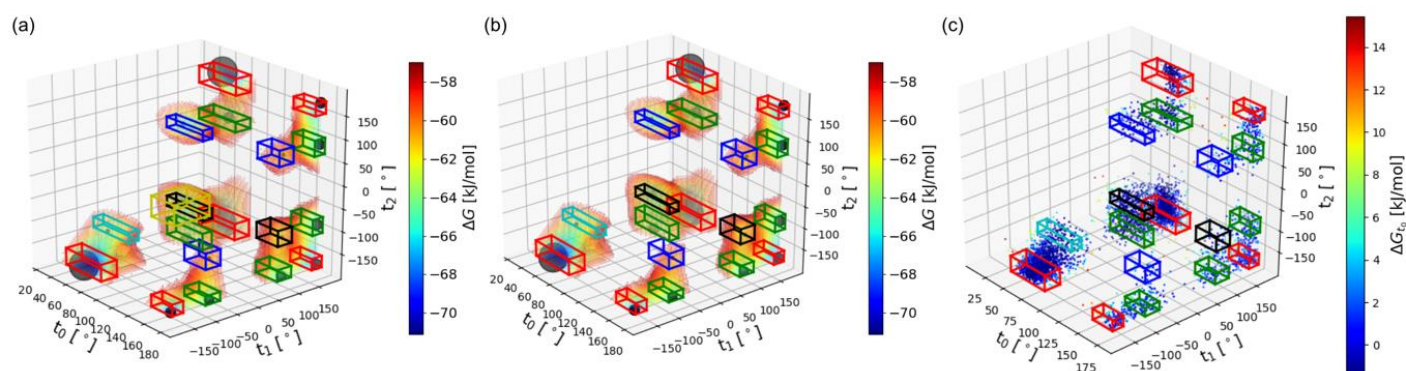


Figure S24. Identification of the conformational isomer macrostates in CV space. The colour code adopted for the states highlights regions of the CV space that belong to the same macrostate. It should be noted however that boxes with the same colour but different t_0 represent different conformers. (a) All the local minima in the 3D FES, corresponding to conformers possessing a finite lifetime. The yellow box identifies a high-energy state (HE) that has been deemed irrelevant in the construction of the MSM. (b) All the macrostates considered for the construction of the MSM, represented in CV space. (c) Projection of the conformational states sampled during an unbiased MD simulation. The colour represents the free energy associated only with the dihedral angle t_0 , quantifying the relative probability of the two main families of conformers (A, G).

As previously described, we ran 18 MD simulations 100 ns-long, starting from different conformers; the value of the dihedrals is reported with a lag time of 1 ps. Interestingly, the entire network of macrostates can be mapped within the first few nanoseconds of the runs, providing a large data set for MSM; this behaviour is in agreement with the low energy barriers found on the WTMetaD FES ($\Delta G_{A \leftrightarrow G} \sim 17$ kJ mol⁻¹, $\Delta G_{a \rightarrow g} \leq 4.3$ kJ mol⁻¹, $\Delta G_{g \rightarrow a} \leq 0.5$ kJ mol⁻¹, see Figure S25).

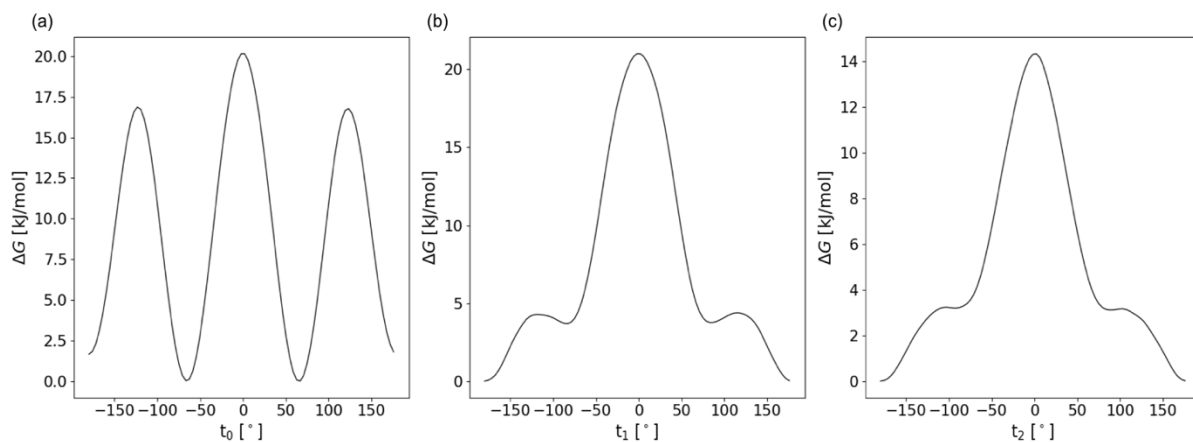


Figure S25. Monodimensional free energy surfaces in t_0 (a), t_1 (b), and t_2 (c).

To build the Markov State Model, the obtained MD trajectories need to be translated from the high dimensional space of coordinates to a sequence of macrostates (Figure S26). To carry out this classification we exploit the definition of states obtained from the analysis of the free energy surface computed with metadynamics. The next task consists of building the transition matrix, \mathbf{C} . The systematic analysis of the macrostates trajectories allows us to fill in C_{ij} elements by counting the $i \rightarrow j$ transitions; we highlight that the microscopic reversibility²⁷ is respected, as $C_{i \rightarrow j} = C_{j \rightarrow i}$. Moreover, we evaluate the overall residence time in each state representing a conformer to obtain the matrix of transition rates, \mathbf{K} . With this procedure we have thus built a kinetic network for the succinic acid conformers in water (Figure S27a).

Finally, with the aim of retrieving information on the equilibrium distribution of conformers and the relaxation time, we estimate the eigenvalues and left eigenvectors of the \mathbf{K} matrix. The stationary probability of each state is expressed by the eigenvector corresponding to eigenvalue zero, while the characteristic time of the process is expressed by the inverse of the smallest eigenvalue different from zero. The results are reported in Figure S27b and Figure S27c. For the system under investigation, we confirm that the conformational rearrangement is a fast process, with a relaxation time to equilibrium distribution of ~ 182 ps (Figure S27b).

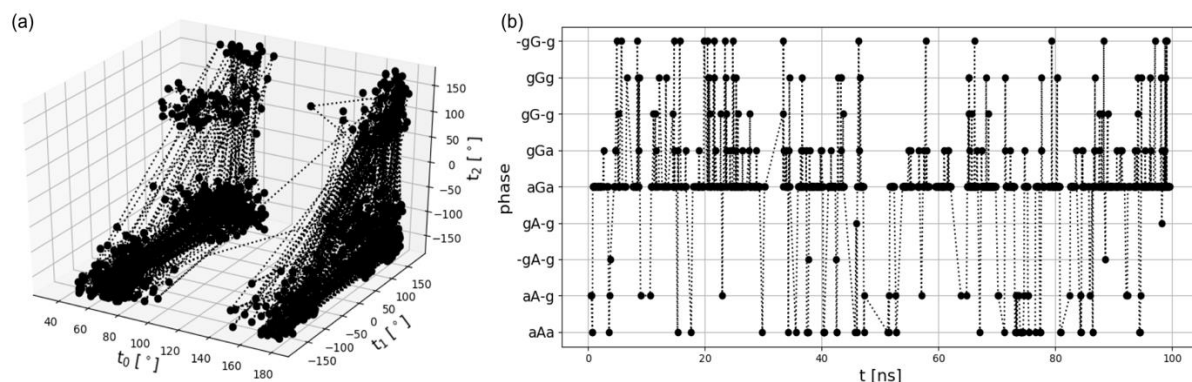


Figure S26. Representation of the state-to-state dynamics observed during an unbiased MD simulation. (a) 1 ns trajectory in CV space, highlighting transitions between the set of states reported in Figure S27, and (b) State to state dynamics during a 100 ns unbiased MD run.

We then validate the stationary composition obtained via MSM with the probabilities obtained from the WTMetaD free energy as $p(t_0, t_1, t_2) = \exp\left(-\frac{\Delta G(t_0, t_1, t_2)}{kT}\right)$. As shown in the histogram in Figure S27c, the results from the two

techniques are in good agreement, well within the error bars. The most abundant conformer in water is aGa, with a probability of $69 \pm 4\%$ (WTMetaD, MSM $67 \pm 3\%$), followed by gGa, aAa, aA-g, and gG-g, all below 10%, while the remaining conformers are in smaller and almost negligible amounts. This distribution is particularly interesting in light of the conformational polymorphism of succinic acid, in particular the well-known β and the newly discovered γ , built respectively from conformer aAa and aGa. The β crystal is well-known for being the only one easily nucleating from solution, despite the fact that its constituent conformer aAa in water has a probability of $8.6 \pm 2\%$ (WTMetaD, $7.3 \pm 1\%$ MSM) approximately one eighth the probability of aGa. On the other hand, despite the abundance of aGa conformers, it does not readily nucleate into either the new γ form or other polymorphs. This observation raises interest about the investigation of the mechanism of nucleation and the competing role of the conformers.

Finally, Figure S27a shows a well interconnected network, where each conformer can and does interconvert with all the others, regardless of the number of dihedrals that are switched at a time. This is again a sign of fast torsions.

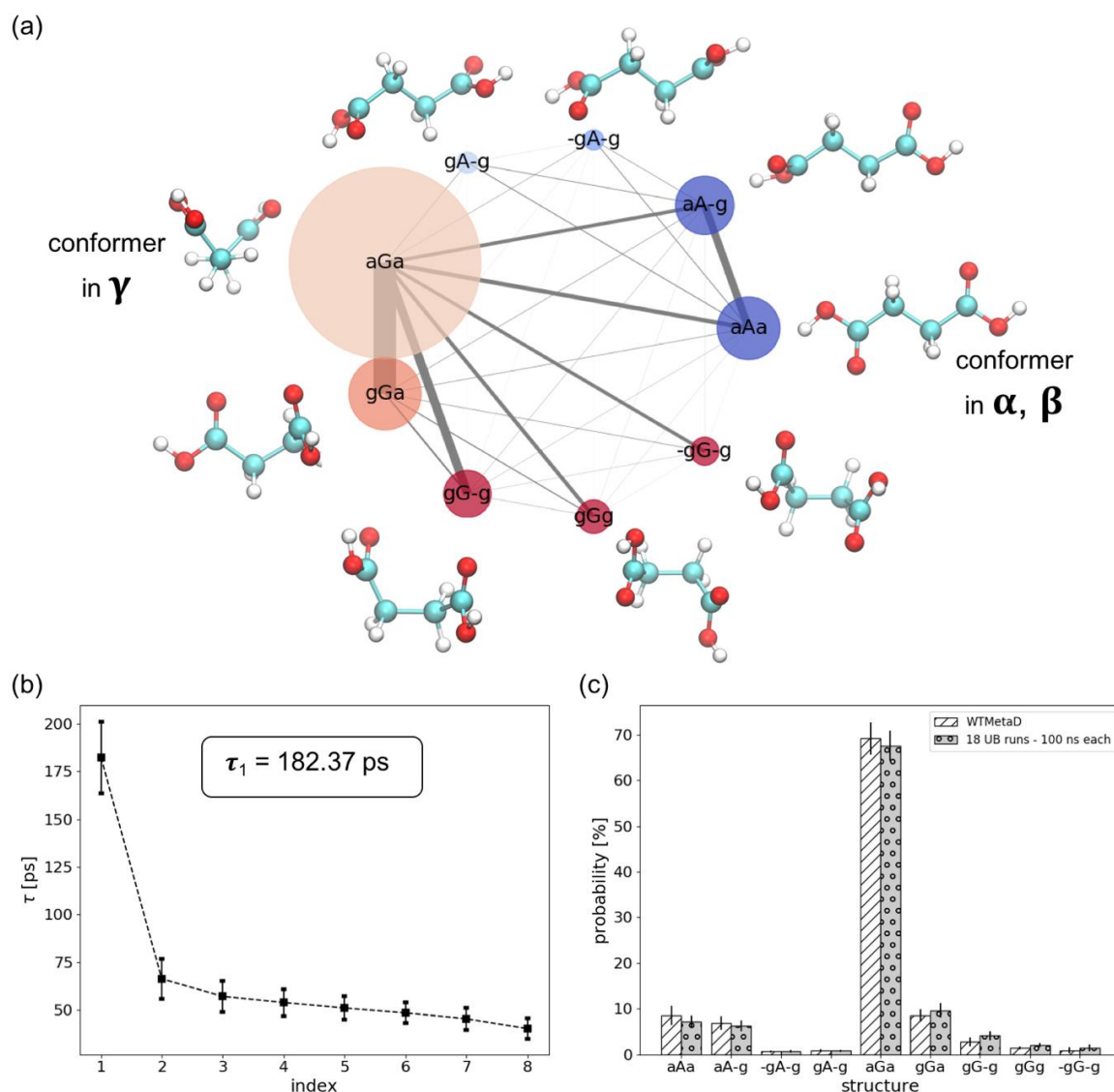


Figure S27. Results of MSM on unbiased MD on a molecule of succinic acid in water. (a) Network graph for the 9 gauche (shades of red) and anti (shades of blue) conformers. The size of the bubble corresponding to each graph is proportional to the overall time that the system spent with the specified conformational arrangement, whose molecular structure is reported outside the graph, and the width of the connections is proportional to the total number of conversions between the respective conformers. These results come from the overall values of 18 unbiased MD simulations run for 100 ns. (b) Characteristic times for the system under investigation, evaluated from the eigenvectors of the K matrix; τ_1 is the relaxation time in ps. (c) Equilibrium distribution of the conformers, expressed as percentage probability: the white striped columns report the result of WTMetaD, where the average and the error bar are weighted over simulation time, and the dotted grey bars refer to the unbiased simulations. For these last results and the plot in (b), the average and the errors are obtained from 18 100 ns unbiased MD simulations.

4.2.3 Summary

We investigated the conformational space for succinic acid with two different approaches: WTMetaD and MSM. The equilibrium ensemble emerging from these methods is consistent, with differences between the two within the error bars. The most stable conformer is aGa, with a probability of 69±4% (WTMetaD, MSM 67±3%), while aAa follows 3rd in the ranking, with a probability smaller than 10%, corresponding to a free energy ~3 kJ mol⁻¹ larger than aGa. This highlight is important as aGa, the dominant conformer in water solution, is the configuration found in the newly discovered γ polymorph, while the β crystal is built with the less abundant aAa. However, the global relaxation time of the network of conformational states in solution is rather fast, of the order of 182 ps. These observations encourage further investigation into the nucleation mechanism in water of β succinic acid rather than γ succinic acid, despite the latter having the most favoured conformer.

4.3 Molecular simulations of bulk crystals

4.3.1 β Polymorph

Table S21. Results for the unit cell of the β -polymorph. NPT MD simulations at 300 K - 1 bar with a supercell of 128 molecules (equivalent to 64 unit cells) are run for 50 ns. Temperature and pressure controls are achieved through Bussi-Donadio-Parrinello thermostat and Berendsen barostat, respectively. In particular, the results are presented for different types of pressure coupling employed, which we refer to as: isotropic, anisotropic and fully anisotropic. Here, we base the distinction between fully anisotropic and anisotropic on whether or not the off-diagonal elements of the cell matrix are allowed to fluctuate. The initial structure for each run is minimized beforehand and equilibrated NVT for 50 ns. The experimental values for the unit cell are at 298 K, obtained from structure SUCACB12 (deposition number 929783) of the Cambridge Structural Database (CSD).

conditions	a [Å]	b [Å]	c [Å]	α [°]	β [°]	γ [°]
isotropic NPT	5.107 ± 0.002	8.889 ± 0.004	5.528 ± 0.002	90	91.510 ± 6.8e-4	90
anisotropic NPT	5.193 ± 0.004	9.248 ± 0.0148	5.348 ± 0.004	90	91.587 ± 0.001	90
fully anisotropic	5.261 ± 0.006	9.038 ± 0.015	5.426 ± 0.007	89.998 ± 0.093	89.791 ± 0.124	90.000 ± 0.0932
fully anisotropic	5.260 ± 0.003	9.037 ± 0.007	5.425 ± 0.004	90.001 ± 0.037	89.826 ± 0.063	90.000 ± 0.034
anisotropic NPT PR	5.274 ± 0.036	9.084 ± 0.3052	5.423 ± 0.066	90	89.629 ± 0.004	90
experimental	5.0993	8.8763	5.5198	90.00	91.508	90.00

Table S22. Results for the energy of the β -polymorph. NPT MD simulations at 300 K - 1 bar with a supercell of 128 molecules (equivalent to 64 unit cells) are run for 50 ns. Temperature and pressure controls are achieved through Bussi-Donadio-Parrinello thermostat and Berendsen barostat, respectively. In particular, the results are presented for different type of pressure coupling employed, which we refer to as: isotropic, anisotropic and fully anisotropic. Here, we base the distinction between fully anisotropic and anisotropic on whether or not the off-diagonal elements of the cell matrix are allowed to fluctuate. The initial structure for each run is minimized beforehand and equilibrated NVT for 50 ns. The experimental values for the unit cell are at 298 K, obtained from structure SUCACB12 (deposition number 929783) of the Cambridge Structural Database (CSD). The reference used for the lattice energy is the aGa conformer.

conditions	potential energy per molecule [kJ mol ⁻¹]	minimized potential energy per molecule [kJ mol ⁻¹]	lattice energy [kJ mol ⁻¹]
isotropic NPT	-430.987 ± 1.026	-482.668	-132.161
anisotropic NPT	-430.891 ± 1.021	-482.348	-131.841
fully anisotropic NPT	-430.492 ± 1.026	-482.136	-131.629
fully anisotropic NPT - $\tau = 10$ ps	-430.456 ± 1.013	-	-
anisotropic NPT PR	-429.557 ± 2.0217	-	-
experimental	-	-483.002	-132.495

4.3.2 γ Polymorph

Table S23. Results for the unit cell of the γ -polymorph. NPT MD simulations at 300 K - 1 bar with a supercell of 256 molecules (equivalent to 64 unit cells) are run for 50 ns. Temperature and pressure controls are achieved through Bussi-Donadio-Parrinello thermostat and Berendsen (B) or Parrinello-Rahman (PR) barostat, respectively; for PR simulations, the initial configuration is the one equilibrated from B. In particular, the results are presented for different type of pressure coupling employed, which we refer to as: isotropic, anisotropic and fully anisotropic. Here, we base the distinction between fully anisotropic and anisotropic on whether or not the off-diagonal elements of the cell matrix are allowed to fluctuate. The initial structure for each run is minimized beforehand and equilibrated NVT for 50 ns.

conditions	a [Å]	b [Å]	c [Å]	α [°]	β [°]	γ [°]
isotropic NPT, B, $\tau = 5$ ps	5.803 \pm 0.002	8.566 \pm 0.003	10.539 \pm 0.003	90	90.370 \pm 1 e-4	90
isotropic NPT, B, $\tau = 1$ ps	5.803 \pm 0.003	8.566 \pm 0.004	10.538 \pm 0.005	90	90.370 \pm 1 e-4	90
anisotropic NPT, B, $\tau = 5$ ps	6.011 \pm 0.003	8.789 \pm 0.006	9.975 \pm 0.005	90	90.405 \pm 2e-4	90
anisotropic NPT, B, $\tau = 1$ ps	6.011 \pm 0.004	8.790 \pm 0.009	9.973 \pm 0.006	90	90.405 \pm 3e-4	90
anisotropic NPT, PR	6.009 \pm 0.018	8.790 \pm 0.024	9.974 \pm 0.023	90	90.405 \pm 1e-3	90
fully anisotropic NPT, B, $\tau = 5$ ps	5.681 \pm 0.003	9.134 \pm 0.007	10.293 \pm 0.005	90 \pm 0.03	99.274 \pm 0.065	90.044 \pm 0.026
fully anisotropic NPT, B, $\tau = 1$ ps	5.681 \pm 0.004	9.134 \pm 0.008	10.292 \pm 0.009	90 \pm 0.07	99.274 \pm 0.065	90.045 \pm 0.058
Experimental	5.7015	8.4154	10.3539	90	90.374	90

Table S24 Results for the energy of the γ -polymorph. NPT MD simulations at 300 K - 1 bar with a supercell of 256 molecules (equivalent to 64 unit cells) are run for 50 ns. Temperature and pressure controls are achieved through Bussi-Donadio-Parrinello thermostat and Berendsen (B) or Parrinello-Rahman (PR) barostat, respectively; for PR simulations, the initial configuration is the one equilibrated from B. In particular, the results are presented for different type of pressure coupling employed, which we refer to as: isotropic, anisotropic and fully anisotropic. Here, we base the distinction between fully anisotropic and anisotropic on whether or not the off-diagonal elements of the cell matrix are allowed to fluctuate. The initial structure for each run is minimized beforehand and equilibrated NVT for 50 ns. The reference employed for the lattice energy is the aGa conformer.

conditions	potential energy per molecule [kJ mol ⁻¹]	minimized potential energy per molecule [kJ mol ⁻¹]	lattice energy [kJ mol ⁻¹]
isotropic NPT, B, = 5 ps	-423.670 \pm 0.721	-475.586	-125.079
isotropic NPT, B, = 1 ps	-423.677 \pm 0.728	-	-
anisotropic NPT, B, = 5 ps	-426.304 \pm 0.717	-478.146	-127.639
anisotropic NPT, B, = 1 ps	-426.335 \pm 0.722	-	-
anisotropic NPT, PR	-426.319 \pm 0.743	-	-
fully anisotropic NPT, B, = 5 ps	-429.854 \pm 0.715	-481.589	-131.082
fully anisotropic NPT, B, = 1 ps	-429.889 \pm 0.723	-	-
experimental	-	-477.396	-126.889

4.3.3 Succinic acid conformational transitions in the bulk of the β phase.

During the analysis of the MD trajectories of the β phase, some molecules in the bulk displayed an interesting conformational behaviour, i.e. the reversible transition from conformer aAa to gA-g (Figure S28a). This rearrangement takes place as the carbon skeleton of succinic acid is free to rotate, in particular the central carbon atoms (C2 and C3, as defined in Figure S23a); importantly, such rearrangement does not modify the H-bond chain. To gain a better understanding of this phenomenon, we perform WTMetaD on a molecule in bulk biasing t_0 , t_1 and t_2 , as previously done for a single molecule in solution. Results are reported in Figure S28b-c. As expected, the conformer aAa is the most stable, since it is the conformer found in the β crystal. Moreover, significant differences emerge in comparison with the behaviour of the single molecule in solution presented in Figure S23. First of all, the conformations with the lowest free energy present planar t_0 , thus they belong to the anti (A) family (Figure S28b); in addition, within the A family we observe interesting modifications, such as a change in location or disappearing of some free energy minima, and the improved ranking of conformer gA-g, now second (~ 20 kJ mol $^{-1}$ above aAa). This analysis confirms what is empirically observed in MD: the aAa \leftrightarrow gA-g transition in the solid bulk of β is possible, but always reverts back to the more stable aAa conformer.

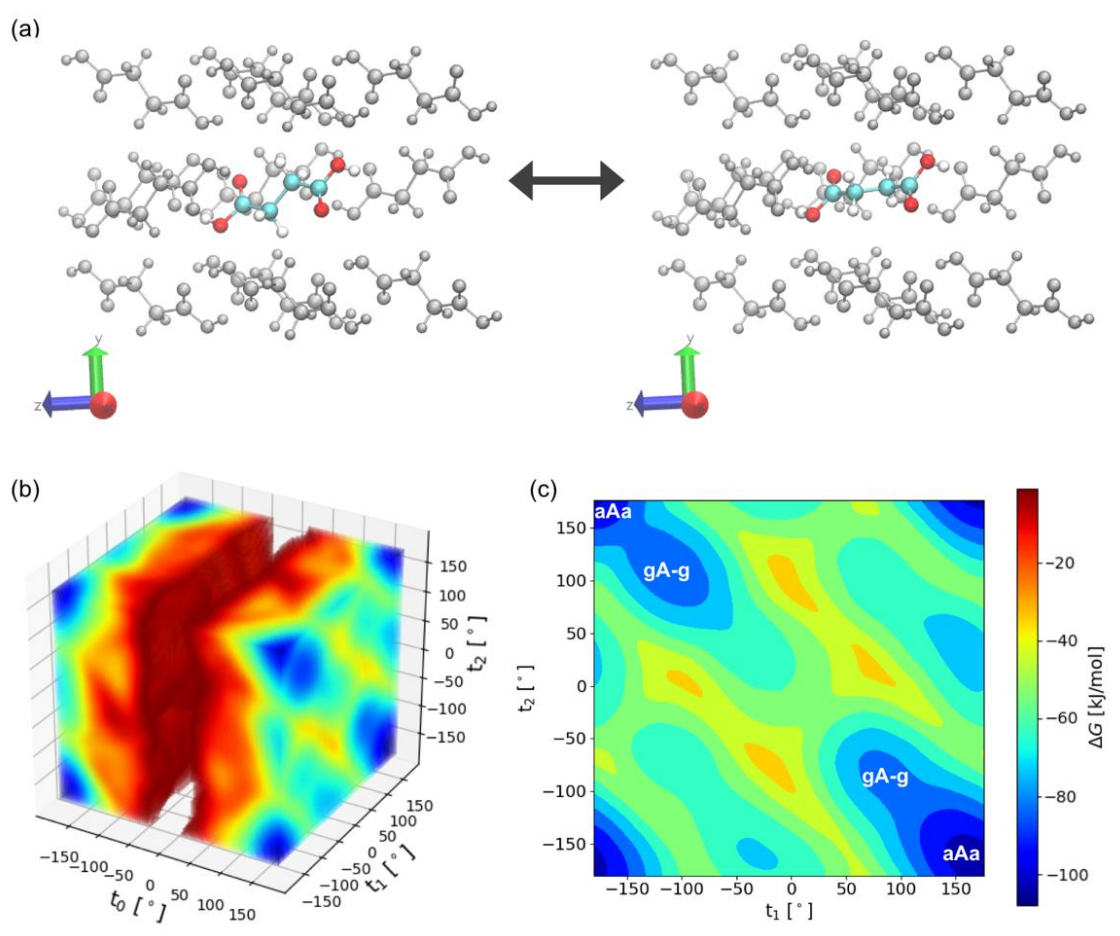


Figure S28. Analysis of the conformational behavior of a molecule of succinic acid in the bulk of the β phase. (a) Snapshots from MD simulations displaying the aAa \leftrightarrow gA-g transition (aAa on the left, gA-g on the right) for the highlighted molecule; it can be observed that the H-bonds are not affected by the rotation of the central carbons. (b-c) Free energy surfaces from WTMetaD: (b) as a function of the three biased dihedrals (t_0 , t_1 , t_2), while (c) projected on t_1 and t_2 . On (c) the locations of the most important conformers aAa and gA-g are shown.

4.3.4 Structural relaxation and melting of the γ polymorph.

As discussed in the main paper and reported in Table S23, NPT simulations of the γ polymorph highlight that its structure undergoes a relaxation that brings the β angle to 99° . To further characterize this transformation and to investigate whether it may be involved in the destabilization mechanism of the γ polymorph we have carried out metadynamics simulations of a γ supercell including 512 succinic acid molecules with the aim of enhancing structural

fluctuations along the β angle in order to recover the free energy surface associated with the distortion of the relaxed β configuration (reported in Figure 5b in the main manuscript), and to explore the mechanism of melting associated with this distortion.

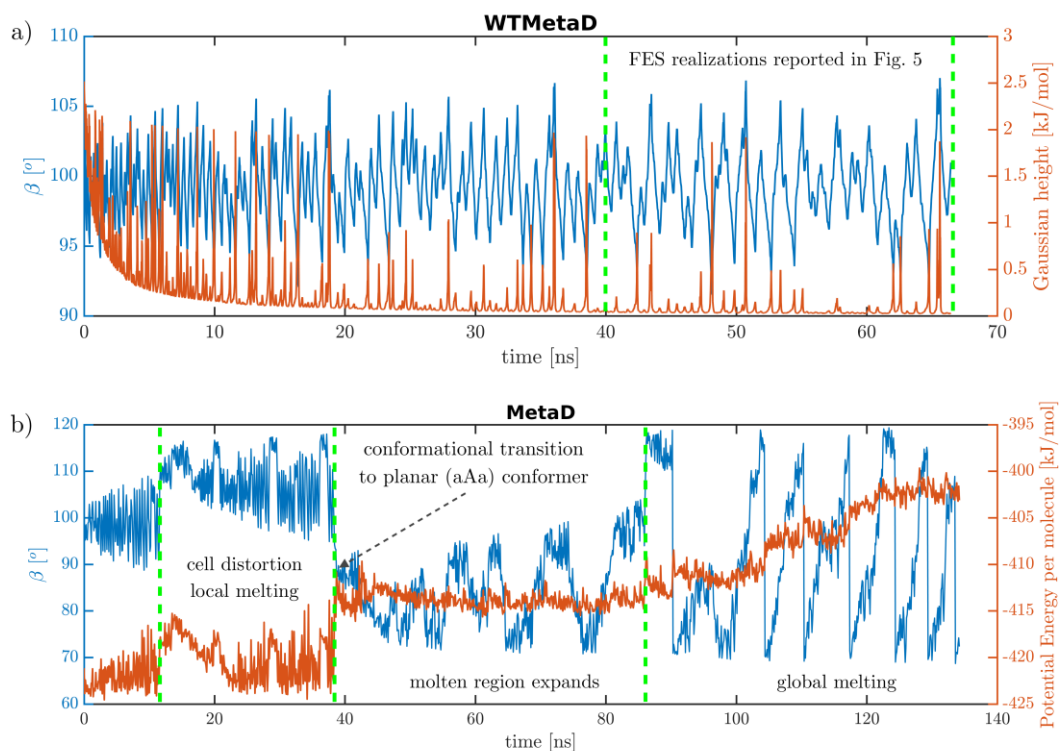


Figure S29. (a) Trajectory in CV space and Gaussian height of the WTMetaD simulations performed to compute the free energy surface reported in Figure 5 of the main paper. The time interval from which FES realizations have been extracted has been highlighted. (b) CV trajectory and potential energy per molecule during the explorative MetaD simulation carried out to investigate the destabilization of the γ polymorph due to fluctuations in the β angle.

To accomplish our first task, we have carried out a Well-Tempered metadynamics (WTMetaD) simulation, using cell angle β as a collective variable, with a bias factor of 150, a Gaussian height of 2.5 kJ mol⁻¹, σ of 5E-3 rad, and a deposition pace of 500 steps. With this setup the free energy surface can be converged in the vicinity of the relaxed configuration and the free energy gain per molecule associated with the relaxation observed in unbiased simulations is estimated to be ~ 3 kJ mol⁻¹.

To tackle the second task, we have carried out standard metadynamics simulations with a Gaussian height of 2.5 kJ mol⁻¹, σ of 5E-3 rad, and a deposition pace of 500 steps. In this simulation we observed that fluctuations along the β angle induce the local melting of succinic acid crystal layers that in turn induce irreversible conformational transitions and lead to the destabilization of the γ polymorph, which undergoes irreversible melting.

In Figure S29 we report the CV dynamics during the simulation time for both the WTMetaD and the MetaD simulations. For the WTMetaD simulation we also report on a secondary axis the height of the Gaussian as a function of time, to highlight the exhaustive exploration of the relevant interval of β . In the case of the MetaD simulation we report the potential energy per molecule of the system, showing how the fluctuations in the angle are actually leading to a destabilization of the system.

1. C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Crystallographica Section B-Structural Science Crystal Engineering and Materials*, 2016, **72**, 171-179.
2. I. M. Dodd, S. J. Maginn, M. M. Harding and R. J. Davey, CSD personal communication.
3. V. R. Thalladi, M. Nüsse and R. Boese, *Journal of the American Chemical Society*, 2000, **122**, 9227-9236.
4. I. Sugden, C. S. Adjiman and C. C. Pantelides, *Acta Crystallographica Section B-Structural Science Crystal Engineering and Materials*, 2016, **72**, 864-874.
5. A. J. Stone, GDMA: A Program for Performing Distributed Multipole Analysis of Wave Functions Calculated Using the Gaussian Program System, 2010.
6. A. V. Kazantsev, P. G. Karamertzanis, C. S. Adjiman and C. C. Pantelides, in *Molecular System Engineering*, eds. C. S. Adjiman and A. Galindo, WILEY-VCH Verlag GmbH & Co., Weinheim, 2010, vol. 6, pp. 1-42.
7. A. Tkatchenko and M. Scheffler, *Physical Review Letters*, 2009, **102**, 073005.
8. S. Grimme, *Journal of Computational Chemistry*, 2006, **27**, 1787-1799.
9. A. Ambrosetti, A. M. Reilly, R. A. DiStasio and A. Tkatchenko, *J Chem Phys*, 2014, **140**, 18A508.
10. S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J Chem Phys*, 2010, **132**, 154104.
11. S. Grimme, S. Ehrlich and L. Goerigk, *Journal of Computational Chemistry*, 2011, **32**, 1456-1465.
12. M. E. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation*, Oxford University Press, Oxford, UK, 2010.
13. D. Frenkel and B. Smit, *Understanding Molecular Simulations: from Algorithms to Applications*, Elsevier, 2nd edn., 2002.
14. M. J. Abraham, D. van der Spoel, E. Lindahl and B. Hess, GROMACS User Manual version 5.1.1, www.gromacs.org.
15. J. Anwar and D. Zahn, *Angewandte Chemie-International Edition*, 2011, **50**, 1996-2013.
16. J. C. Palmer and P. G. Debenedetti, *Aiche Journal*, 2015, **61**, 370-383.
17. A. Barducci, G. Bussi and M. Parrinello, *Physical Review Letters*, 2008, **100**, 020603.
18. A. Laio and M. Parrinello, *Proceedings of the National Academy of Sciences of the United States of America*, 2002, **99**, 12562-12566.
19. A. Barducci, M. Bonomi and M. Parrinello, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2011, **1**, 826-843.
20. O. Valsjö, P. Tiwary and M. Parrinello, *Annual Review of Physical Chemistry*, 2016, **67**, 159-184.
21. F. Giberti, M. Salvalaglio and M. Parrinello, *Journal*, 2015, **2**, 256-266.
22. J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *Journal of Computational Chemistry*, 2004, **25**, 1157-1174.
23. G. Bussi, D. Donadio and M. Parrinello, *Journal of Chemical Physics*, 2007, **126**, 014101.
24. H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola and J. R. Haak, *Journal of Chemical Physics*, 1984, **81**, 3684-3690.
25. J. D. Chodera and F. Noe, *Current Opinion in Structural Biology*, 2014, **25**, 135-144.
26. V. S. Pande, K. Beauchamp and G. R. Bowman, *Methods*, 2010, **52**, 99-105.
27. G. R. Bowman, V. S. Pande and F. Noe, *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, Springer, Netherlands, 2014.
28. J. H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte and F. Noe, *Journal of Chemical Physics*, 2011, **134**, 174105.
29. G. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Computer Physics Communications*, 2014, **185**, 604-613.
30. G. van Rossum, Python: a computer language, 1998.
31. W. Humphrey, A. Dalke and K. Schulten, *Journal of Molecular Graphics & Modelling*, 1996, **14**, 33-38.
32. M. Ceriotti, G. Tribello and M. Parrinello, *Proceedings of the National Academy of Sciences of the United States of America*, 2011, **108**, 13023-13028.