Electronic Supplementary Material (ESI) for CrystEngComm.

Supplementary information

Studies on crystal structure and the arrangement of water in sitagliptin L-tartrate hydrates

Eszter Tieger[1, 2], Violetta Kiss[2], György Pokol[1], Zoltán Finta[2], Michal Dušek[3], Jan Rohlíček[3], Eliska Skořepová[2, 4]

[1]Department of Inorganic and Analytical Chemistry, Budapest University of Technology and Economics, Szt. Gellért tér 4, H-1111 Budapest, Hungary, tiegereszter@gmail.com

[2] Zentiva k.s., U kabelovny 130, 102 37, Prague, Czech Republic

[3]Institute of Physics ASCR, v. v. i., Na Slovance 2, 182 21 Prague, Czech Republic

[4]Department of Solid State Chemistry, University of Chemistry and Technology, Prague, Technicka 5, 166 28 Prague, Czech Republic

Table S1. Crystallographic data and details of structure refinement of SLT Phase 1

Chemical formula	$C_{40}H_{42}F_{12}N_{10}O_{14} \times 1\ H_2O$
Formula weight	557.4
Crystal system, Space group	monoclinic $P2_1$
Z	2
a, Å	7.2124 (10)
b, Å	29.0912 (10)
c, Å	11.686 (12)
α, °	90
β, °	106.181 (10)
γ, °	90
V, Å^3	2354.8
S	4.38
λ	1.5418 Å (CuK_α)

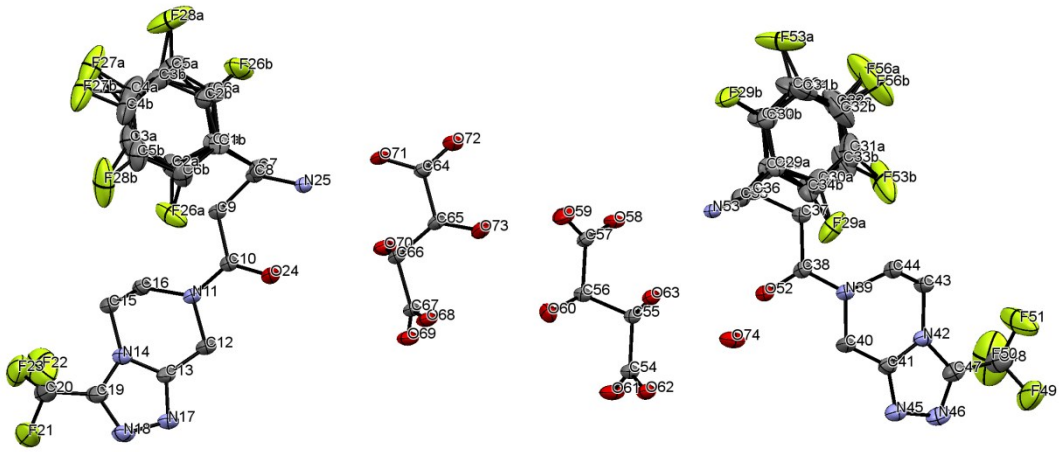

Figure S1. The molecular configuration and atom-numbering scheme for the sitagliptinium cations, the hydrogen l-tartrate anions and water molecule in SLT Phase 1.

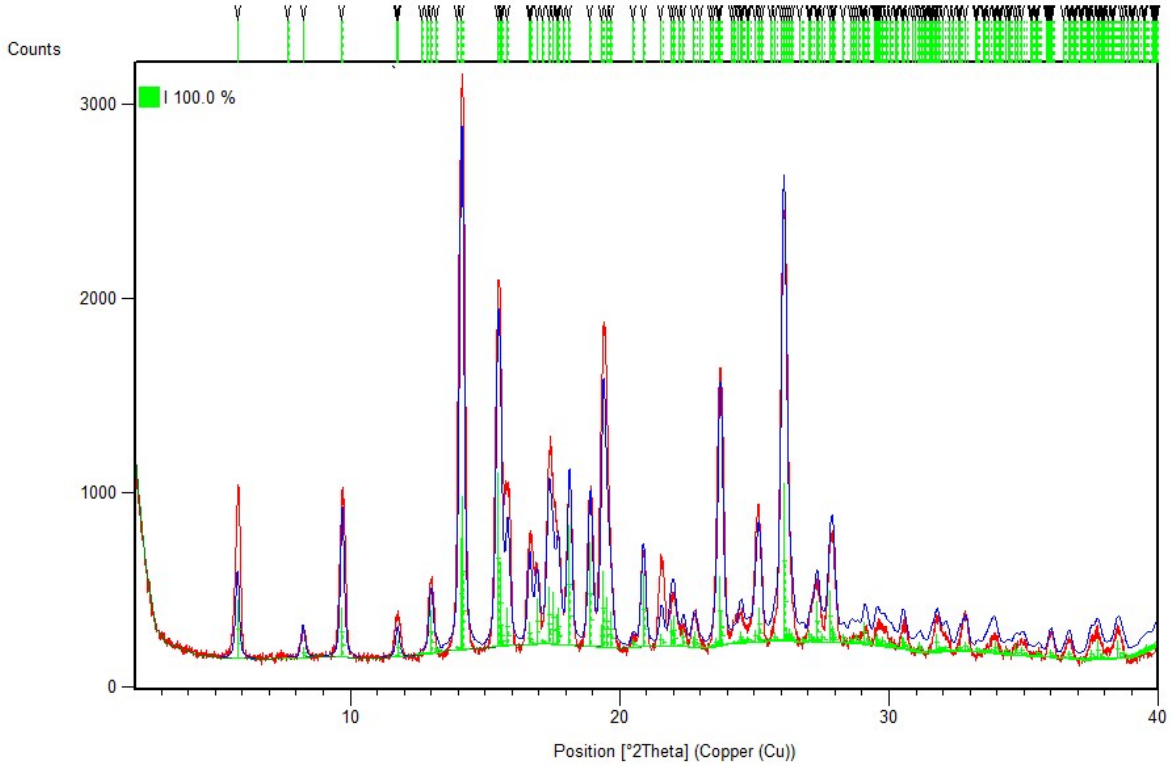

Figure S2. Final Rietveld plot of SLT Phase 1 showing the measured data (blue line), calculated data (red line). Calculated Bragg positions are shown by vertical bars.

Table S2. Crystallographic data and details of structure refinement of SLT Phase 2

Chemical formula	$C_{20}H_{21}F_6N_5O_7 \times 0.5\ H_2O$
Formula weight	557.4
Crystal system, Space group	triclinic *P*1
Z	1
a, Å	7.21076(11)
b, Å	11.66322(14)
c, Å	15.64522(15)
α, °	99.3299(8)
β, °	98.8646(13)
γ, °	105.9897(12)
V, Å^3	1220.43(3)
R_p	0.0141
R_{wp}	0.0203
R_{exp}	0.0034
S	5.81
λ	1.5418 Å (CuK_α)
$2\theta_{min}$, $2\theta_{max}$, °	4.007, 79.992
Increment in 2θ, °	0.013
nb. restraints	192
nb. constraints	199
refined parameters	270
$\Delta\rho_{min}$, $\Delta\rho_{max}$ (eÅ^{-3})	-0.36, 0.25

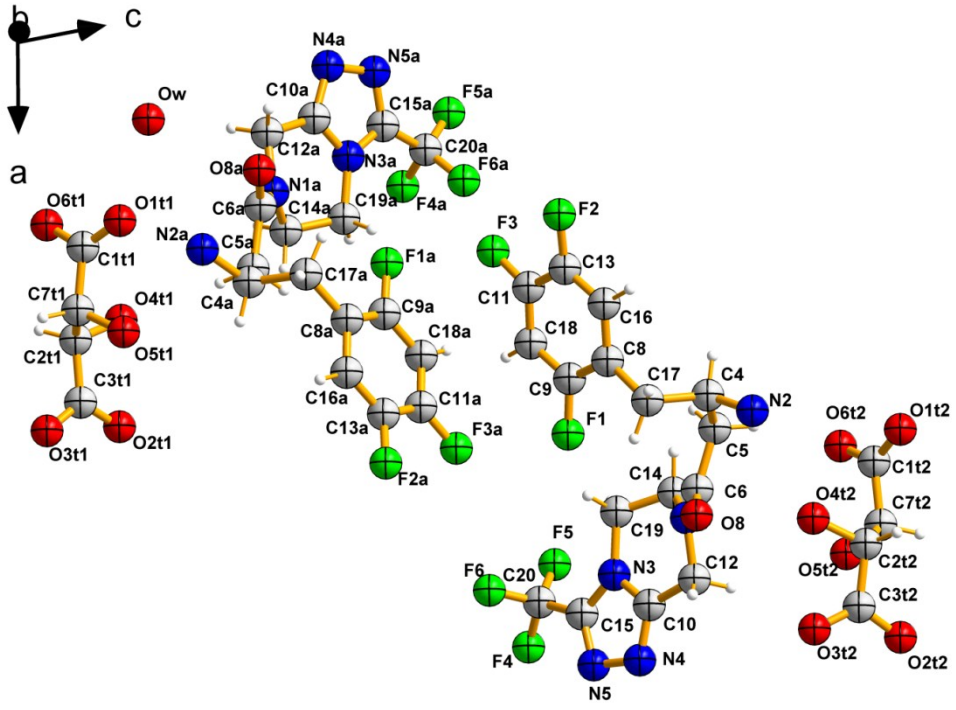

Figure S3. Molecular configuration and atom-numbering scheme for the sitagliptinium cation, the hydrogen l-tartrate anion and water molecule in SLT Phase 2

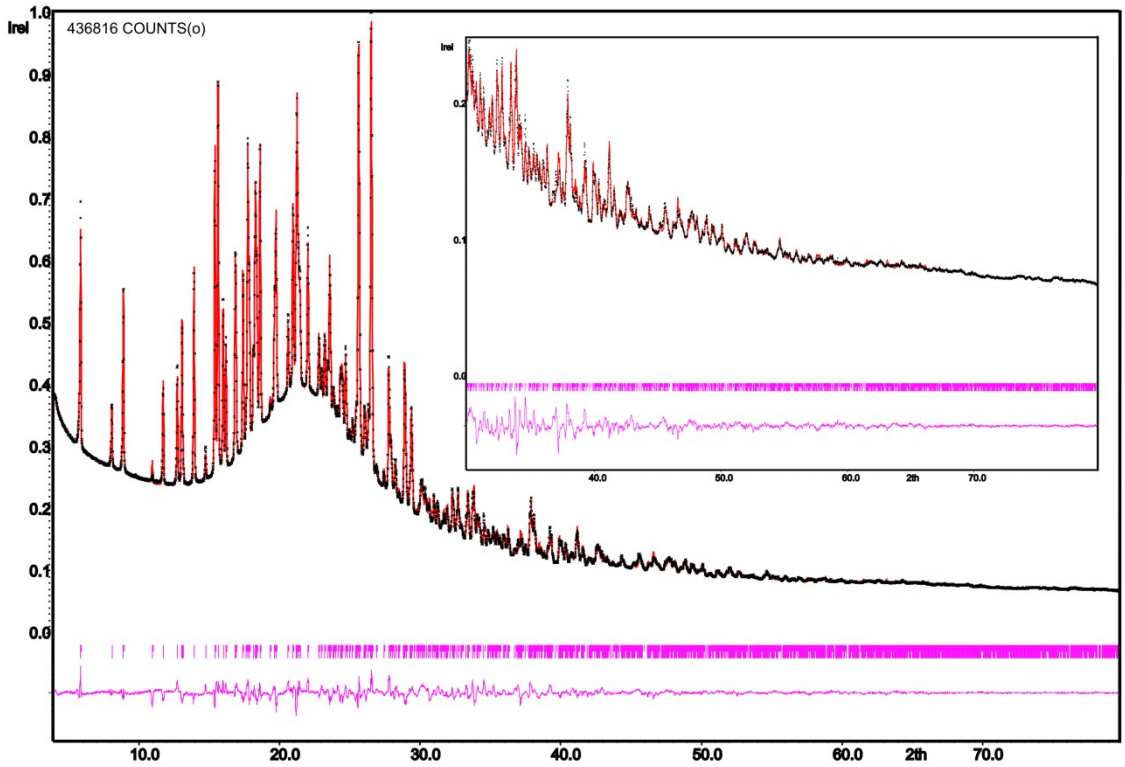

Figure S4. Final Rietveld plot of SLT Phase 2 showing the measured data (black thin-cross), calculated data (red line) and difference curve (pink line). Calculated Bragg positions are shown by vertical bars.

Table S3. Crystallographic data and details of structure refinement of SLT Phase 4

Chemical formula	$C_{20}H_{21}F_6N_5O_7 \times 4\ H_2O$
Formula weight	620.45
Crystal system, Space group	monoclinic, *C* 2
Z	4
a, Å	45.2368 (4)
b, Å	8.2403(4)
c, Å	7.1817(18)
α, °	90
β, °	95.326 (4)
γ, °	90
V, Å^3	2665.53
S	6.63
λ	1.5418 Å (CuK_α)

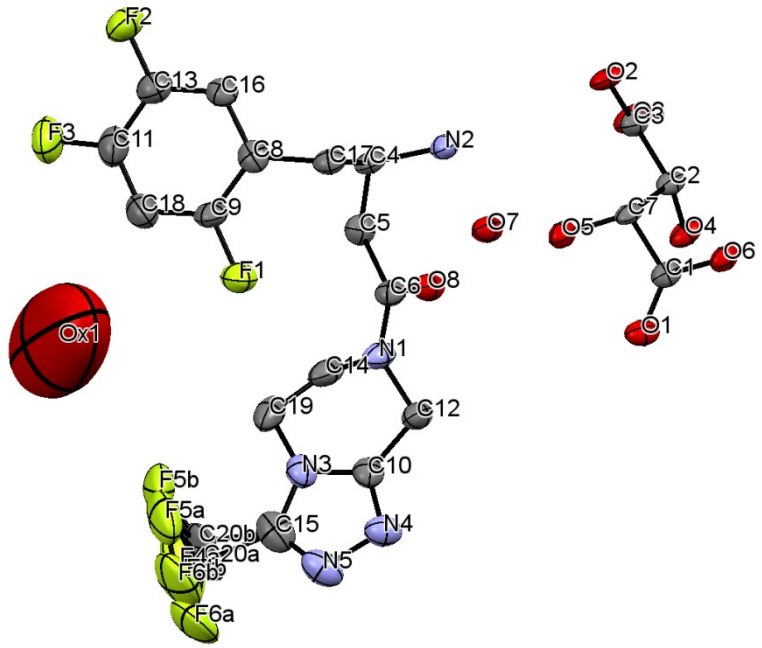

Figure S5. Molecular configuration and atom-numbering scheme for the sitagliptinium cation, the hydrogen l-tartrate anion and the water molecules in SLT Phase 4.

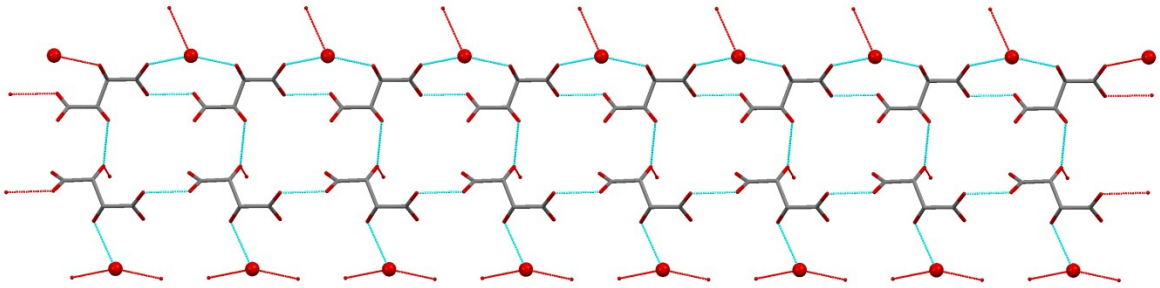

Figure S6. Projection of the hydrogen-bonded hydrogentartrate sheet in SLT Phase 1.

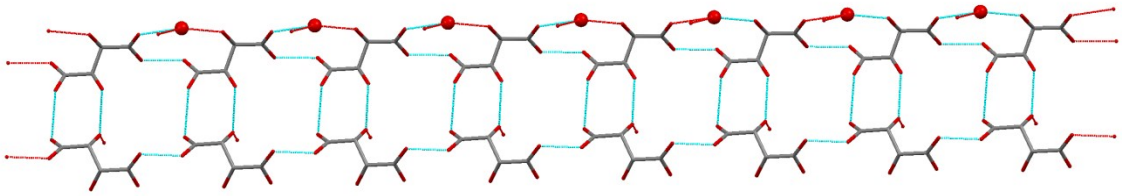

Figure S7. Projection of the hydrogen-bonded hydrogentartrate sheet in SLT Phase 2.

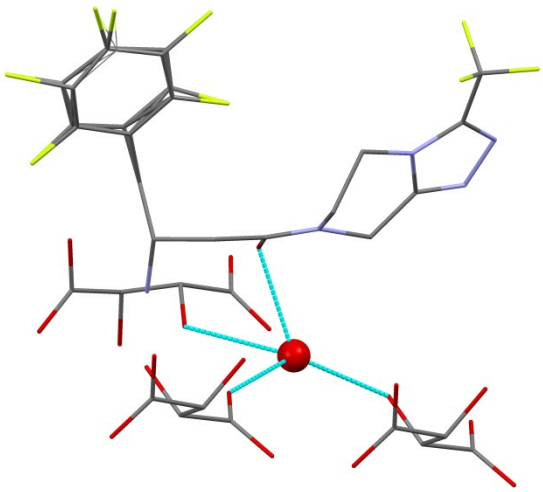

Figure S8. Hydrogen-bonded water molecule in SLT Phase 1.

Structure determination of the Phase 1, 2 and 4.

Refinement of Phase 1

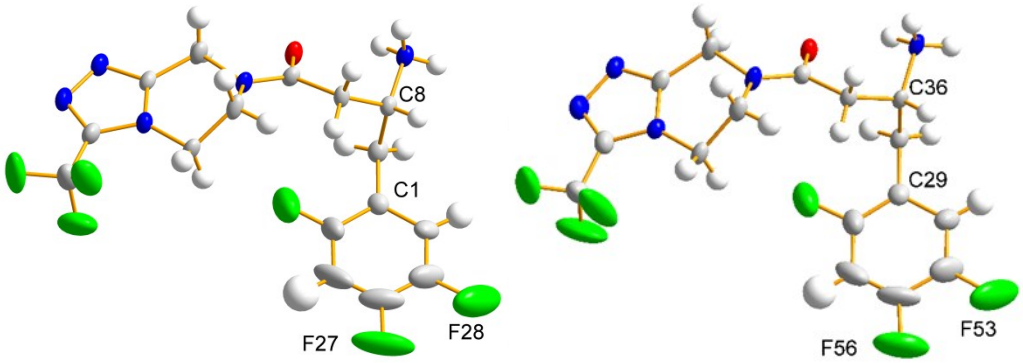

Figure S9. Sitagliptin molecules (A) – left – and (B) – right – in the structure of SLT Phase 1.

After initial refinement the difference Fourier map exhibited maxima in unexpected positions. Using Jana2006[S1] for the refinement, we found that occupancy of F26, F28, F53 and F29 is by ~20% lower than the full occupancy. At the same time, difference maxima appeared close to hydrogen atoms of these two rings.

The two molecules C1…C6,H31,H61,F26,F27,F28 (A) and C29,….,C34,H311,H341, F29,F53,F56 (B) have their rings in the following orientation (Figure S9): looking from F27 towards C1 such that C8 is above the ring, F28 is on the right (for molecule B, looking from F56 towards C29 such that C36 is above the ring, F53 is also on the right). Natural explanation of the decreased occupancy on some fluorine atoms seems to be that there is also some smaller portion of molecules where the ring is rotated by 180 degree around F27-C1 (for A) or F56-C29 (for B) axis. Such rotation brings positions of hydrogen to fluorine causing decrease of the occupancy, and position of fluorine to hydrogen, causing ghost maxima around the hydrogen atoms.

In order to prove this idea, we constructed two rigid bodies (Figure S10): A and B. Each rigid body was placed in two positions defined by the original placement of the molecule (Aa and Ba) and by the above-explained rotation (Ab and Bb). Occupancy of all positions was refined, keeping the sum for Aa, Ab and Ba, Bb to be the full occupancy. The distances C1a-C7, C1b-C7 were kept the same as well as C29a-C35, C29b-C35.

[S1] Petříček V., Dušek M., Palatinus L., Crystallographic Computing System JANA2006: General Features, Zeitschrift für Kristallographie, **2014**, 229, 5, 345-352

The disordered rigid body model significantly improved the refinement fit and decreased the extremes of the difference Fourier map: R value decreased from 6.03% to 4.38%, GOF from 1.73 to 1.22 (Jana2006 uses the true GOF based on experiment which is - unlike SHELX - sensible to the small structure changes), and the difference Fourier map was almost flat (Figure S9). The refined occupancies were 0.744(3) and 0.256(3) for the molecule Aa, Ab; 0.812(4) and 0.188(4) for the molecule Ba, Bb.

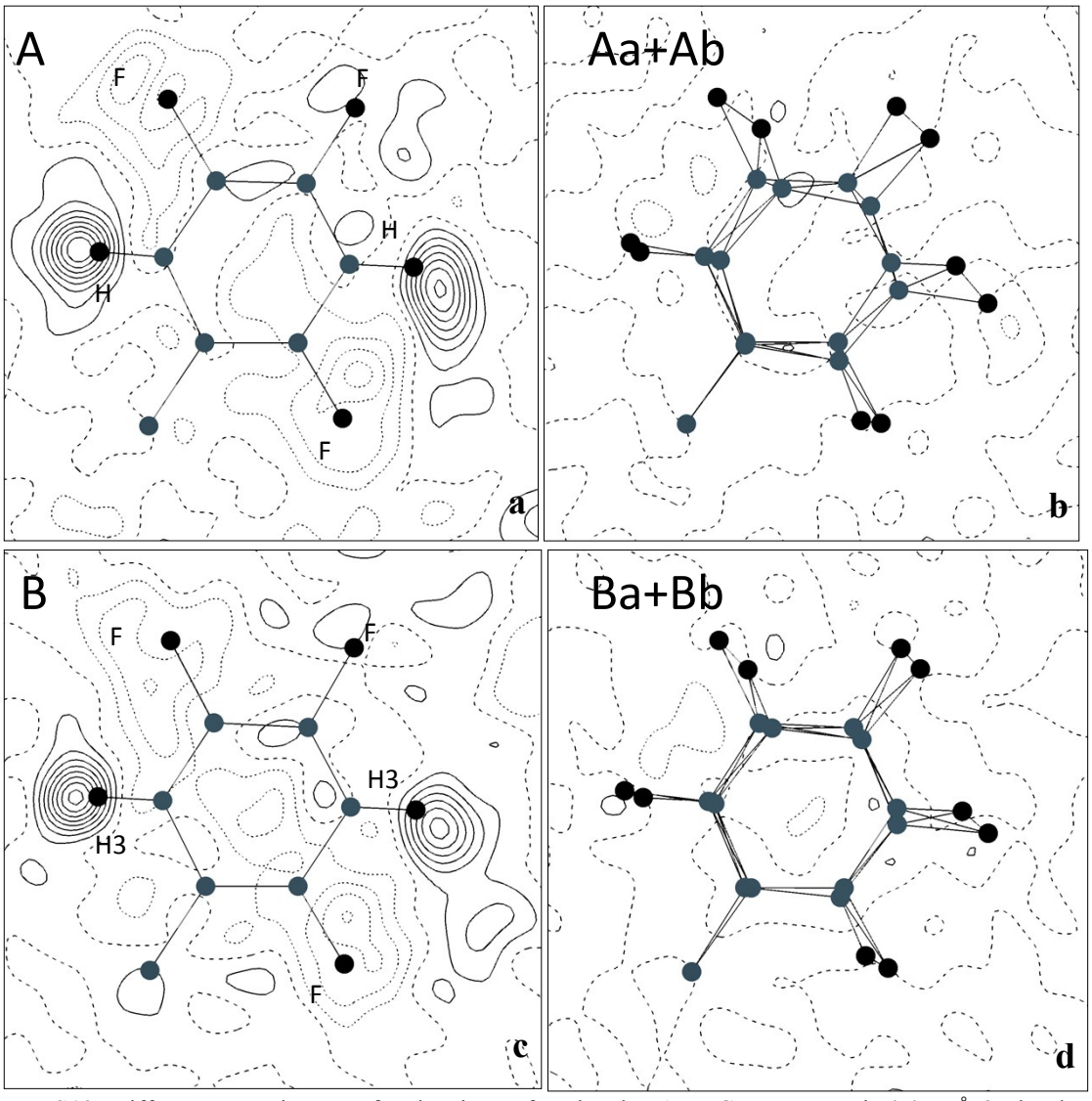

Figure S10. Difference Fourier map for the rings of molecules A,B. Contour step is 0.05e-Å-3, the density is merged at the distance 1 Å out of the plane of the rings. Density (a), (c) is calculated before introducing disorder; (b), (d) after refinement of the disordered model.

The disorder cannot be removed by changing sample of by repeating the crystallization. It was observed in several samples from different batches and seems to be a regular feature, including the ratio of the minor and major disordered part.

Refinement of Phase 2

Phase 2 could not be prepared in a form of high quality single crystals, however we succeeded to get partially distorted fragment of the crystal structure from single crystal data of very low quality. This fragment was used as a starting model for the Rietveld refinement with Jana2006 based on high quality powder data. The model was solved in the triclinic unit cell in space group *P*1 and it contained both units of SLT in the unit cell together with one water molecule.

After confirmation of the unit cell and Le Bail refinement using zero-shift, background, profile parameters and asymmetry correction, Rietveld refinement was done of the initial model kept as a rigid body. Only one overall isotropic ADP parameter was refined at this step to fit the scale parameter and to see the correctness of the model. In final Rietveld refinement we removed the rigid body constraint and we refined positions and one isotropic ADP parameter for all non-hydrogen atoms. Hydrogen atoms bonded to carbon were kept in their theoretical positions (U_{iso} of each hydrogen atom was kept as 1.2*U_{eq} of the parent atom). As expected with powder data, hydrogen atoms of the water molecule, -NH_3^+ and -OH groups could not be located and refined. Scale, background, March-Dollase parameter and unit cell parameters were refined together with the structure model. The final model was restrained with 78 bonds and 114 bond-angles restraints to get a result with reasonable geometry. Each restrained bond and bond-angle value was set individually according to the crystal structure of the Phase 4 tetrahydrate. The final agreement profile factors were Rp = 1.42%, Rwp = 2.04%, GOF = 5.93.

Refinement of Phase 4

The CF_3 groups exhibited strong disorder. In order to describe the disorder, a rigid body CF_3 was defined in two positions “a” and “b”, each defined by a translation vector and three rotations. The distances C15-C20a and C15-C20b were restrained to keep the same refined value and similar restriction was applied for distances C-F in the rigid body. Occupancies of the rigid body positions “a” and “b” were constrained to keep in the sum the full occupancy. The refined values of the occupancies were 0.559(11) and 0.441(11) for “a” and “b”, respectively. Hydrogen atoms attached to carbon were kept in the geometrically correct position, with C-H distance 0.96 Å. Positions of hydrogen atoms attached to nitrogen and oxygen were refined with N-H distance restrained to 0.86 Å and O-H distance restrained to 0.9 Å. U_{iso} of any hydrogen atoms was calculated as 1.2 multiple of U_{eq} of its parent atom.

Phase 4 is non-centrosymmetric. Due to the presence of only light atoms, Flack parameter was not refined in the final structure model. However, its value 0.2(3) obtained from other refinement indicated that we worked with the proper enantiomer.

Two symmetry independent water molecules were found in the structure. One of them (O7) could be refined including hydrogen atoms while the other located in channels along c was strongly disordered and manifested in difference Fourier map as a cloud of electron density. For description of the corresponding electron density we used oxygen position labelled O1x and located in the general position within the most pronounce difference Fourier maximum. For this position, we refined occupancy and ADP tensors up to the 5th order in order to describe possible lattice water positions. The resulting cloud of the electron density is shown in Figure S11. Occupancy of O1x was refined as 3.23(2). Taking into the account hydrogen atoms, this is roughly three molecules of water.

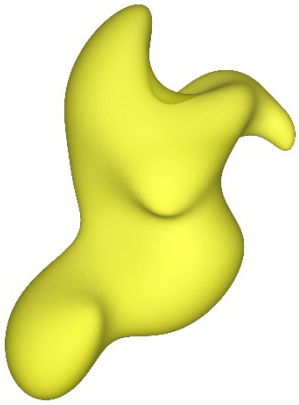

Figure S11. Electron density corresponding to the atom O1x described with the 5th order ADP tensors. Isosurfaces were visualized by the program VESTA[S2] at the level 0.069 electrons.

S2 Momma K. and Izumi F., VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data, Journal of Applied Crystallography, **2011**, 44, 1272-1276

Structure determination trials of Phase M

Transmission electron microscopy (TEM) was conducted on a Philips CM120 transmission electron microscope with a LaB_6 cathode operating at 120 kV. The microscope is equipped with a CCD Camera Olympus SIS Veleta with 14 bit dynamical range. The precession assisted electron diffraction tomography[S3] was measured in low-dose regime using low-temperature sample holder at temperatures about 100 K. The tilt step of 2° and precession angle of 1.5° were used during the measurement. Programs PETS and Jana2006 were used for processing the raw data.

The crystals are very fine needles (Figure S12), which degrade fast upon electron beam irradiation. Therefore, accumulation of data covering range up to about 80° was possible. It was possible to find the corresponding unit cell obtained from PXRD in most of the 3D reconstructions. The resolution of the obtained data was about 1.6 Å. Moreover, the data were frequently obscured by twining about c* axis, disorder along the same axis and bending of the crystals during the measurement. About forty crystals were measured and only two of them gave usable data. These two crystals had the same orientation towards the e-beam so the resulting coverage after combination of the datasets remained only 40%. Solving the structure using simulated annealing in Sir2011[S4] was unsuccessful.

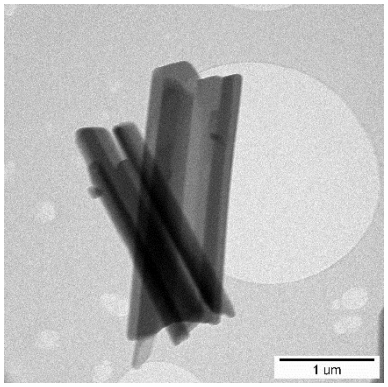

Figure S12. Morphology of SLT Phase M crystals.

S3 U. Kolb, T. Gorelik, C. Kuebel, M. T. Otten, and D. Hubert, Ultramicroscopy 107, **2007**, 507-513

S4 Burla, M. C., Caliandro, R., Camalli, M., Carrozzini, B., Cascarano, G.L., Giacovazzo, C., Mallamo, M., Mazzone, A., Polidori, G. & Spagna, R. , J. Appl. Cryst., **2012**, 45, 351-356.

Powder diffraction measurement was performed in capillary on laboratory diffractometer PANalytical Empyrean equipped with focussing mirror and with PIXCel^{3D} detector. The diffraction pattern was measured from 4° to 80°2θ with step size 0.013°2θ and counting time 18000 s per step.

The Le Bail refinement confirmed the triclinic unit cell with these newly refined parameters: a = 7.19853(17), b = 11.6983(3), c = 15.4294(3), α = 97.7099(11), β = 95.3209(18) and γ = 106.9252(16). The measured powder diffraction data was not sufficient for direct determination of the crystal structure from measured intensities. In addition, the high complexity of the crystal structure (the expected number of degrees of freedom is 45!) does not allow us to solve the crystal structure by using direct space methods. We have used the crystal structure of Phase 2 as a starting model in Rietveld refinement of Phase M. Fractional coordinates of Phase 2 were loaded to the unit cell of Phase M. Due to the differences between the unit cells of the phases, the starting model was slightly distorted. Nevertheless, the calculated powder diffraction pattern of the starting model was in a relatively good agreement with the measured data (R_p = 7.15% and R_w = 11.86%), see Fig. S13. After applying bond and bond-angle restraints calculated from the known crystal structures of sitagliptin, which keep the model geometrically correct and which also allow changing positions and shapes of all moieties during the refinement, we have achieved a better profile fit (R_p = 4.15% and R_w = 6.09%), see Fig. S14. Unfortunately, the final fit is not adequate. This indicates that the real structural model is different. The difference Fourier map shows several relatively weak maximums and minimums, which were almost randomly spread in the unit cell and which were impossible to interpret. In addition, several attempts to interpret these peaks did not lead to satisfactory result. Since the difference Fourier map does not show any reasonable maximum nor minimum, it could not be clarified, what is the difference between Phase 2 and Phase M. Based the good agreement of the theoretical and measured data, it is a reasonable assumption, that at least the main crystal motifs, such as molecular packing of sitagliptin molecules is similar in the structure of Phase 2 and Phases M.

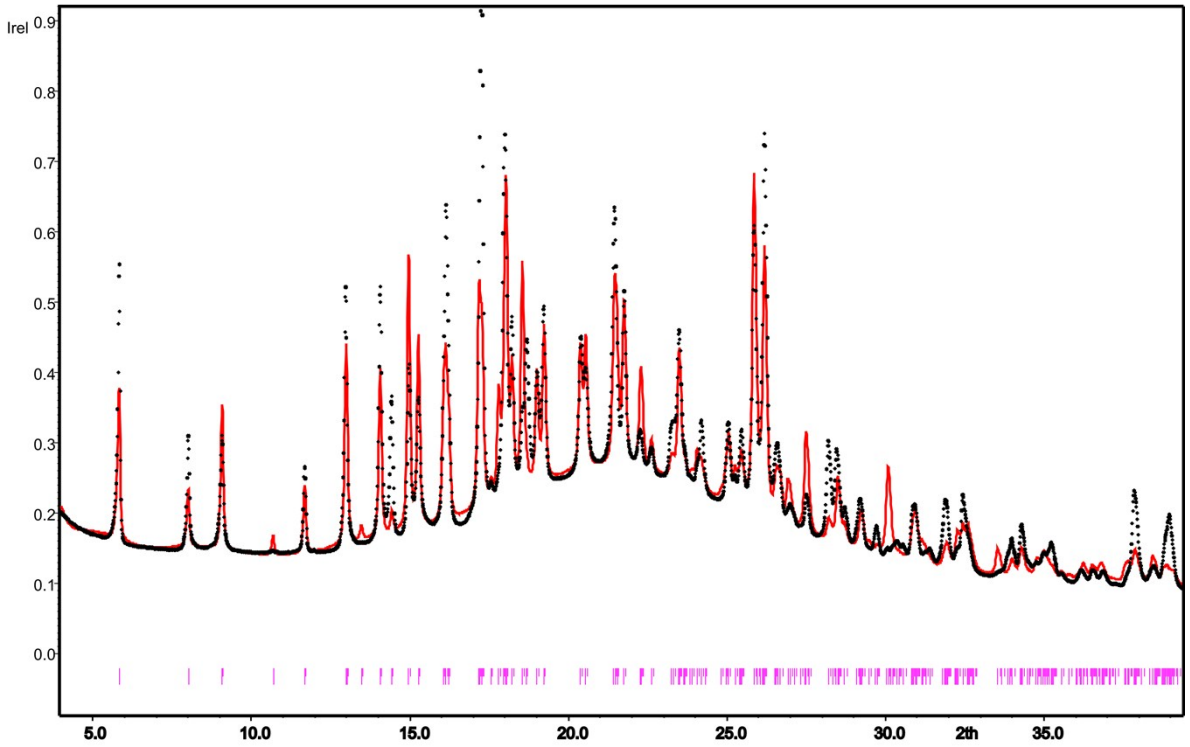

Figure S13. Comparison of theoretical (red) and measured (black dots) patterns of Phase M. The theoretical model is a starting model of the crystal structure of Phase 2 saved in the unit cell of the Phase M.

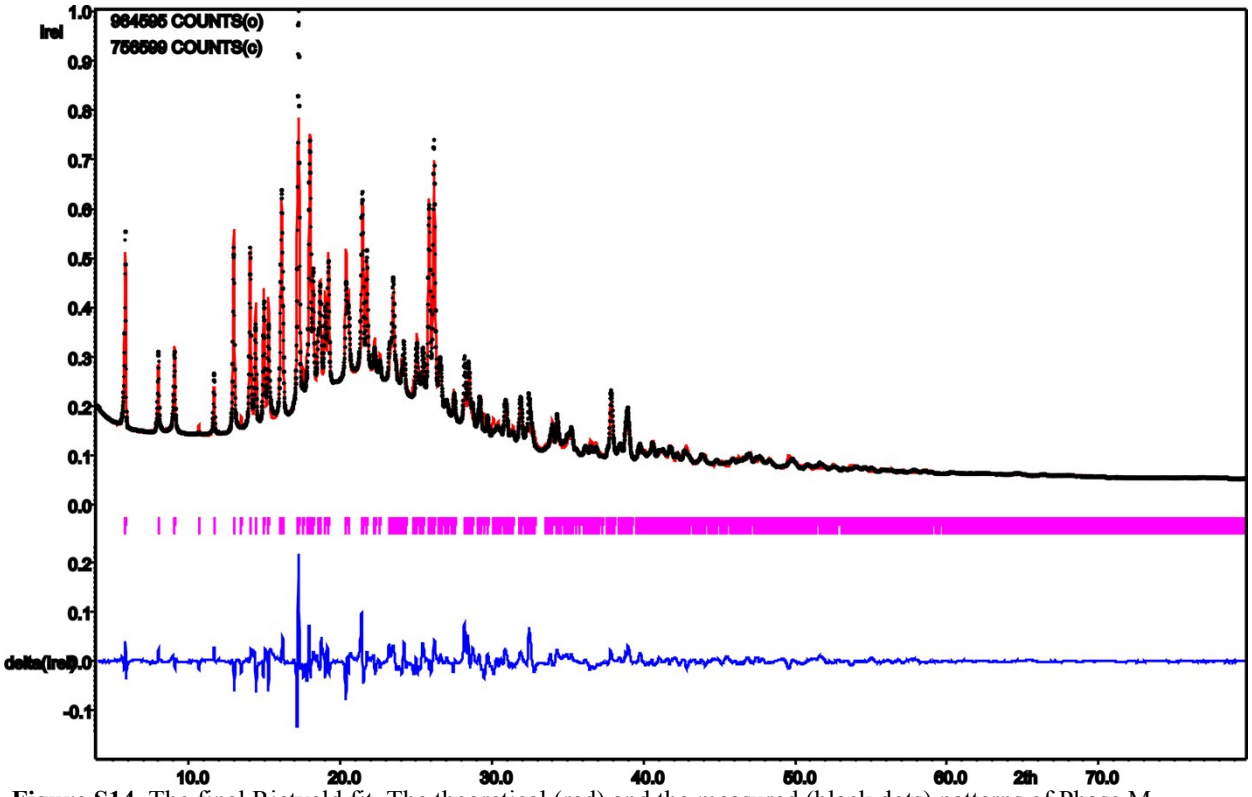

Figure S14. The final Rietveld fit. The theoretical (red) and the measured (black dots) patterns of Phase M.

Investigation of SLT Phase 4 tetrahydrate

Since only few crystals of Phase 4 were obtained, a thorough characterization was impossible. SC-XRD, Raman and infrared spectroscopy and hot-stage misroscopy could be applied to gain information of such small quantities.

FTIR spectra were recorded by Nicolet Thermo 6700 spectrometer.

General settings:

Number of sample scans: 45

Number of background scans: 45

Resolution: 4.000

Sample gain: 4.0

Optical velocity: 0.6329

Aperture: 100.00

Raman spectra were recorded by FT-Raman Bruker RFS 100/S Spectrometer

General settings:

Excitation source: Nd-YAG laser (1064 nm)

Applied spectral domain: 4000-200 cm^{-1}

Applied laser power: 250 mW

Detector: liquid nitrogen cooled Ge-diode detector (D418-T)

Resolution: 4 cm^{-1}

Number of accumulations: 128

Scattering geometry: 180° (back scattering)

Aperture: 3.5 mm

Hot stage microscopy

Microscope: Nikon Eclipse Ni

Camera ProgRes CT3 (Jenoptik)

Hot-stage: LTS 420 from LINKAM

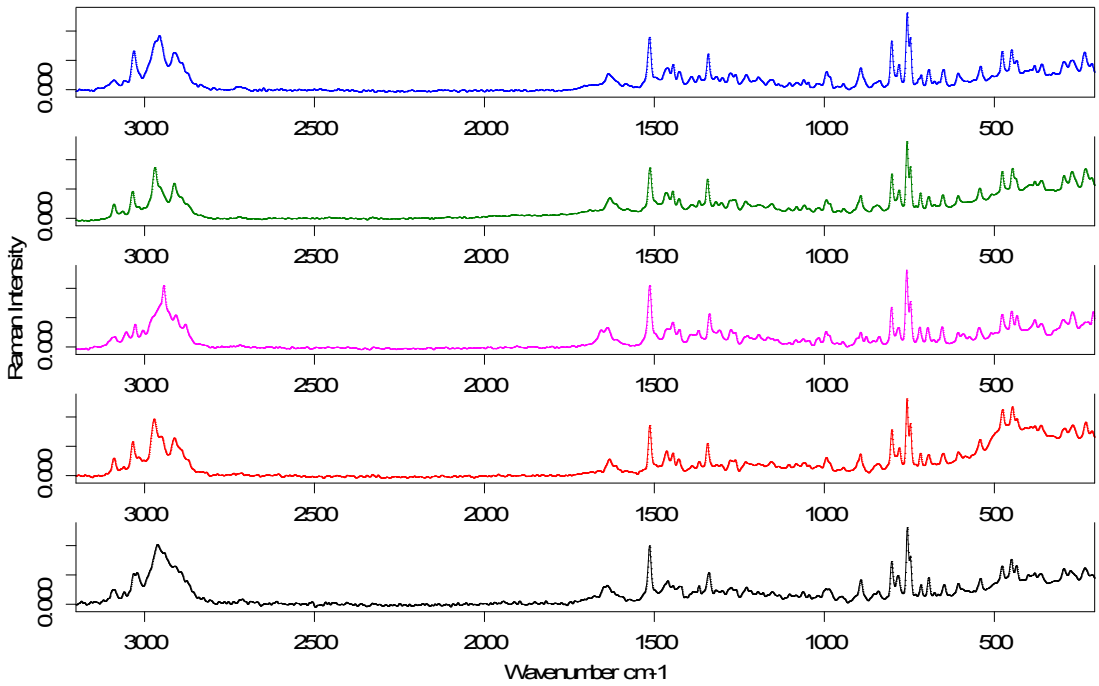

Figure S15. Raman spectra of the SLT hydrates: Phase 1, Phase 2, Phase 3, Phase M and Phase 4

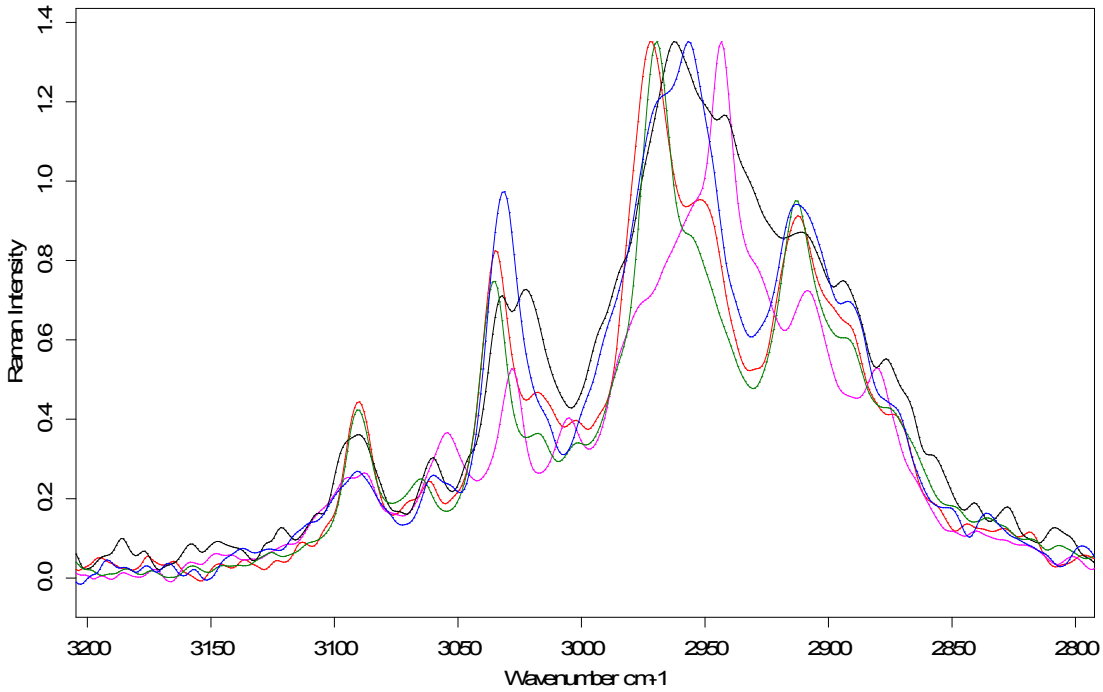

Figure S16. Overlay of the Raman spectra of the SLT hydrates – CH region
Phase 1, Phase 2, Phase 3, Phase M and Phase 4

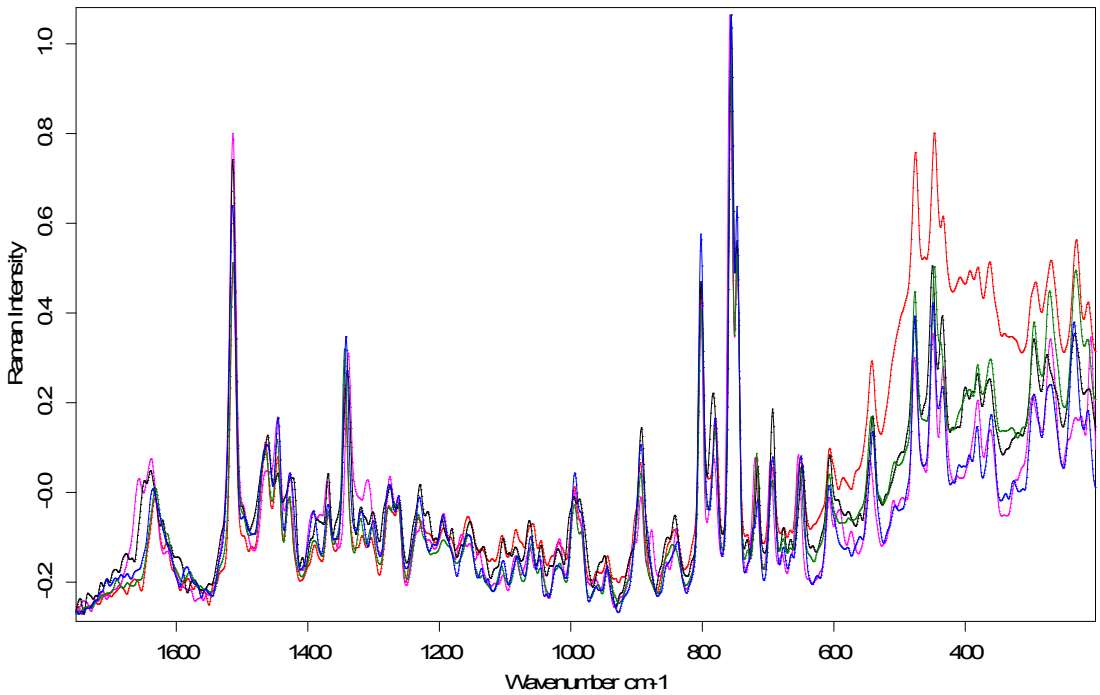

Figure S17. Overlay of the Raman spectra of the SLT hydrates – fingerprint region Phase 1, Phase 2, Phase 3, Phase M and Phase 4

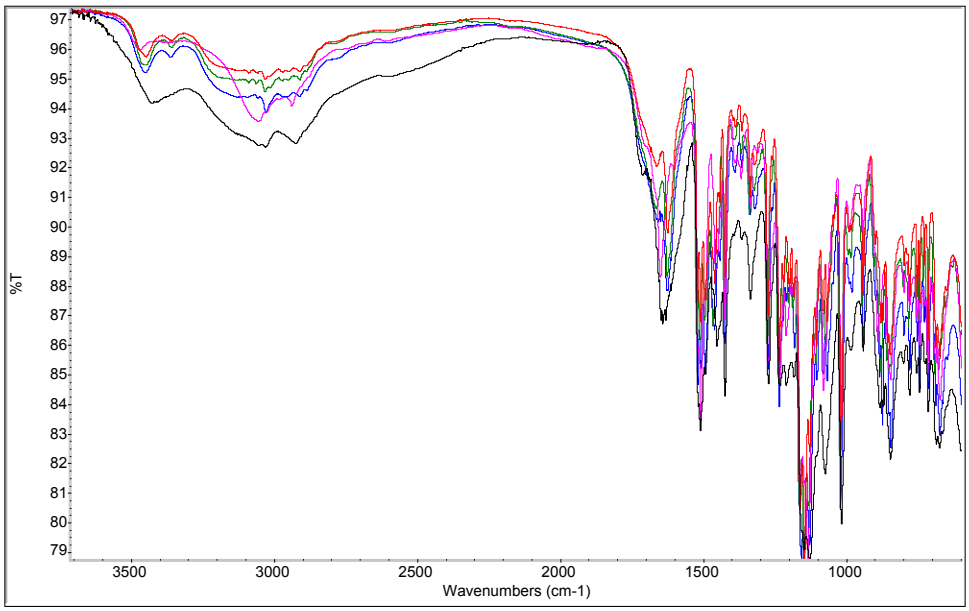

Figure S18. Overlay of the IR spectra of the SLT hydrates: Phase 1, Phase 2, Phase 3, Phase M and Phase 4

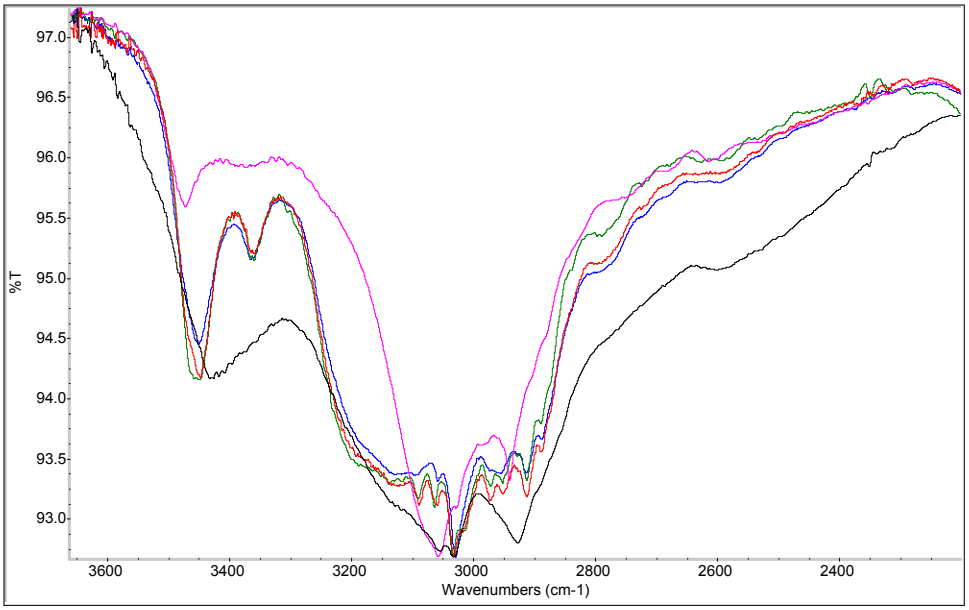

Figure S19. Overlay of the IR spectra of the SLT hydrates – CH region
Phase 1, Phase 2, Phase 3, Phase M and Phase 4

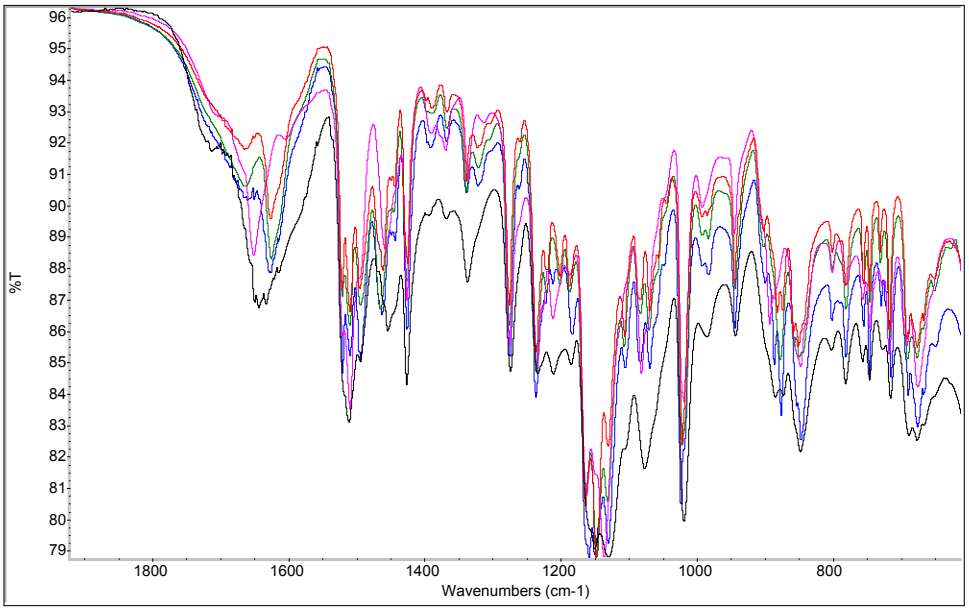

Figure S20. Overlay of the IR spectra of the SLT hydrates – fingerprint region
Phase 1, Phase 2, Phase 3, Phase M and Phase 4

The broad absorption band in the IR spectra between 3500 and 2600 cm^{-1} is due to the O-H, C-H and N-H stretching vibrations and the intermolecular H-bonding in the crystals. The sharp, characteristic absorption peaks at 3450 and 3320 cm^{-1} are assigned to NH asymmetric and symmetric stretching vibrations. The position, relative intensity and width of the peaks depend on the presence and strength of the intermolecular H-bonds. The spectra of Phase 1, 2 and M are very similar, and Phase 3 and 4 show similarity as well. The main difference between Phase 4 and the other structures is that the more extensive H-bonded network is

manifested in more intense and broader peaks. The most prominent differences in the Raman spectra are observed in this region as well. The peaks around 3030 and 2975 cm^{-1} are due to asymmetric and symmetric aromatic CH stretching vibrations.

The asymmetric COO^- stretching vibration gives a strong band at 1635 cm^{-1} in the IR spectrum, but a weak one in the Raman spectrum. The position of this peak is slightly shifted in different forms. There is a broad absorption band in the IR spectra in this region, since there are 3 sources of C=O strecthing vibration. In decreasing order of frequency: CO (sitagliptin), COOH (tartrate) and COO^- (tartrate). Water also has an absorption at around 1640 cm^{-1} which explains the even broader absorption observed in the case of Phase 4.

The CF_3 group absorbs strongly at 1275 and 1020 cm^{-1} in the IR spectrum and 756 cm^{-1} in the Raman spectrum.

Interestingly, the structure of Phase 3 and Phase 4 is similar comparing the XRPD patterns and the vibrational spectra (Figure S15-20).

Hot stage microscopy was applied to gain information with regards to the melting range and other potential transformations of Phase 4 during heating. No visible change occured during heating and the crystals melted in the range of 182-187°C (applying 1°C/min heating rate).