Investigating the hydration and structural changes of molecular organic materials under high-pressure conditions

Dissertation

for the award of the degree

"Doctor of Philosophy"

Ph.D. Division of Mathematics and Natural Sciences
of the Georg-August-Universität Göttingen

within the doctoral program Geology / Geography
of the Georg-August University School of Sciences (GAUSS)

submitted by

Rubén Granero-García

from Zaragoza (Spain)

Göttingen, 2016

Thesis committee

Dr. Francesca. P. A. Fabbiani
Prof. Dr. Werner F. Kuhs
Abteilung Kristallographie,
Fakultät für Geowissenschaften und Geographie

Members of the examination board

Reviewer: Dr. Francesca. P. A. Fabbiani
Second reviewer: Dr. Werner F. Kuhs

Further members of the examination board:
Dr. Ben Corry
Dr. Ricardo Mata
Dr. Heidrun Sowa
Dr. Kai Tittmann

Date of the examination: June 9$^{\text{th}}$, 2016

Where the spirits fall
Where live beautiful lies
Where I'm outa control
And I just cannot die

—S.G.B.

Acknowledgements

After painstakingly writing and formatting the other 220 pages of this document, it feels good to finally come to a place where I can lay the words unconstrained by the rigorous formality of scientific writing, a discipline about which, I just realised, I know rather little. My awesome supervisor, Francesca Fabbiani—often referred as *Boss*, pronounced with a delicate whiff of Spanish accent—has done everything in her power to sculpt my writing into what can be found in this dissertation. I am immensely thankful to her as a supervisor, mentor, and friend. For letting my curiosity wander around, but never too far from the path of productivity. For having always a witty response to my illogical digressions. For believing in this work even when I am still not completely sure of what I have accomplished.

I shall show my gratitude towards the members of my department, under the careful watch of Prof. Werner F. Kuhs, those who still linger around and those whose paths have diverged far from here. Sofiane, who has patiently answered all my questions, no matter how foreign. Stefano, who lifted the spirits during the time he spent freezing around here. Julian, always ready to discuss refinement strategies and willing to go through nightmarish paperwork. Patrick, from whom I have learnt almost everything I know about statistical mechanics—nothing like some hefty equations to deter people from peering at my desk—and a consid-

erable bunch of approaches to computer simulations. Andrzej, who, now I come to realise, has been answering the same question about hydrate clathrates for the last three years, slightly rephrased every time. Lars, because being able to ask directly the person who built the diffractometer has proven a more efficient learning technique than trial and error. And Dr. Heidrun Sowa, whose input on space-group symmetry has saved a considerable amount of time in several occasions. I do not forget the technical staff who keeps this department running: Heiner, Jan Martin, Karin, Klaus, Nils, and Ulf. Thank you all.

I have also enjoyed the fabulous opportunity to jump around in Australian soil, as Francesca reckoned it was best to send me over 16000 km away if I were to learn about computer simulations, and Ben Corry gladly accepted the proposal and let me consume his computing time. I could not be more grateful to both for my time there; also to Natalie and Nishank, from whom I learned a fair amount.

I want to finish by mentioning a few friends, who have not contributed to the scientific dimension of this work, but without whom the results would have been different. Almudena, Jovana, and Aldo, thanks for you support, for—how to put it diplomatically—putting up with my utter lack of diplomacy. Claudia (Wandkte), my go-to girl to find things around the university. And the honourable mentions go to Claudia (Millán), who is yet to miss her usual summer in German lands, and to Sergio, who has earnestly kept up with the philosophical disquisition from the other side of the Internet since I changed the neck-wrecking winds of my hometown by the stiff coldness of the Northern regions. Thanks.

Contents

Abbreviations and symbols

α, β, γ	unit-cell angles
$\bar{\alpha}$	average thermal expansion coefficient
β	compressibility, *or*
	reciprocal thermodynamic temperature
ϵ	Lennard-Jonnes depth well
ϵ_0	electrical permittivity
ζ	Nosé-Hoover thermostat friction
θ	deviation from ideal conformation, *or*
	diffraction angle
λ	trajectory coordinate, *or*
	wavelength
μ	average, *or*
	chemical potential
ν	state of a system
ρ	crystal density, *or*
	electron density
σ	Lennard-Jones equilibrium distance, *or*
	standard deviation
χ	order parameter
Ω	NVE partition function
A	Helmholtz free energy
ADP	atomic displacement parameter
AP	ambient-pressure
a, b, c	unit-cell axes
B_0	bulk modulus at ambient pressure
B_0'	pressure derivative of the bulk modulus

CD	cyclodextrin
α-CD	α-cyclodextrin
β-CD	β-cyclodextrin
$Corr$	correlation
Cov	covariance
CSD	Cambridge Structural Database
CSP	crystal structure prediction
C_V	constant-volume heat capacity
dabcoHI	1,4-diazabicyclo[2.2.2]octane hydroiodide
DAC	diamond anvil cell
dMBzIm	5,6-dimethylbenzimidazole
e	electron
E	energy
f	correction factor
F	force, *or*
	structure factor
FES	(Gibbs) free energy surface
FOM	figure of merit
G	Gibbs free energy
H	enthalpy
HP	high-pressure
H bond	hydrogen bond
h, k, l	Miller indices
I	intensity of a reflection
k_B	Boltzmann constant
K	kinetic energy

L	length of a simulation box		S	entropy, *or*
m	mass			goodness of fit
M	number of ideal particles		t	time
MD	molecular dynamics		T	temperature, *or*
$\boldsymbol{M}$	Parrinello-Rahman barostat coupling matrix			transmission factor
N	hydration number, *or*		^{t}BA	*tert*-butylamine
	number of particles		U	potential
NPT	constant N, P, and T		U_{C}	Coulomb potential
NVE	constant N, V, and E		U_{LJ}	Lennard-Jones potential
NVT	constant N, V, and T		v	velocity
P	pressure, *or*		V	unit-cell volume, *or*
	probability			volume
PBC	periodic boundary conditions		V_{excl}	volume per chemical formula excluding water
PME	particle mesh Ewald		V_{H}	volume within a Hirshfeld surface
PTM	pressure transmitting medium		$V_{H,ntt}$	unattributed volume per chemical formula
q	atomic charge		$V_{H,nwat}$	volume of a molecule other than water
Q	Cremer-Pople puckering amplitude, *or*		$V_{H,wat}$	volume of a water molecule
	mass of coupling bath, *or*		V_{M}	molecular volume
	NVT partition function		V_{Z}	volume per chemical formula
r	radius		Var	variance
R	R factor		x, y, z	coordinates
RMS	root mean square		Z	NPT partition function, *or*
RSS	residual sum of squares			number of molecules per unit cell
s	reduced coordinate			

Introduction

Molecular organic materials are ordered solids whose stability and structure is dictated by the intermolecular interactions of their constituent organic molecules.[1] These materials are at the centre of many technological applications—from electronics (organic semiconductors) and photonics (organic lasers) to medicine (targeted drug delivery)—thanks to the fact that individual molecules can be functionalised and their arrangements tailored to obtain specific properties, giving rise to the field of molecular engineering. Intermolecular interactions are a key concept in this field, and a better understanding of these is essential to improve crystal designing techniques.

Polymorphism is both a blessing and a curse for crystal engineers.[2] A blessing because different polymorphs might show different properties, proving that the final properties of a solid are not solely determined by its constituents, but also by their arrangement in the solid state. A curse because predicting which of the polymorphs found in a crystal's structural landscape—for instance, by means of crystal structure prediction (CSP)—will actually crystallise under certain experimental conditions, is not an easy task. The same can be said about solvate formation.

Structural studies are a basic tool to broaden our understanding about intermolecular interactions, and provide very valuable information that can potentially be used to determine which crystalline phase will form and which properties it will have. Screening of crystalline phases of organic substances has been predominantly performed by means of temperature- and solvent-based techniques. However, the use of pressure, thanks to the development of the diamond anvil cell (DAC), has been attracting increasingly more interest. The past 15 years have seen an important rise in the number of studies of molecular materials by means of pressure. Compression and in-situ high-pressure crystallisation have proven a successful way to access phases (polymorphs and solvates, especially hydrates) which cannot be obtained by other means (see §1.1 for a brief review on the topic).

Reporting new phases of molecular organic materials is of great interest to expand our knowledge; recently the interest has been turning towards understanding *why* these phases actually form. The work of Parsons,[3–5] combining high-pressure crystallography with energy calculations performed with the PIXEL method,[6] is probably the best example. The advantage of the PIXEL method is that it breaks down the energy of intermolecular interactions in four contributions: Coulombic, polarisation, dispersion, and repulsion. Upon compression, a phase transition might occur when, for one or more interactions, the repulsion outweighs the other three contributions as pressure brings the molecules closer together. In such a case, the phase transition is a way to release the internal strain of the structure.[3] In other cases, the overall energy of the structure increases because of the PV work; a phase transition towards a denser phase can stabilise the structure by decreasing the energy associated to the PV term via smaller volumes.[4] In both situations, the new phase is more efficiently packed due to less volume being available at higher pressures. In some cases, zero-point energies and entropic contributions can become important to lower the overall energy of the high-pressure favoured phase.[5] Nevertheless, two polymorphs are related by a phase transition only if there is a suitable path for it.[7] Certain phases cannot be accessed by direct compression, because the transformation is kinet-

ically hindered, but can be accessed by in-situ crystallisation at high pressure. These studies benefit from the fact that pressure modifies the volume of a sample while keeping its thermal energy constant, which allows to directly correlate the changes observed in the sample with the applied pressure, something that is more complex when triggering crystal transformations by means of changes in temperature.[8]

The present work contains a collection of high-pressure crystallographic experiments combined with molecular dynamics (MD) simulations aimed at exploring *how*, rather than *why*, pressure modifies structures at the molecular level and molecular transformations can occur. Three organic materials have been studied: two cyclodextrins and *tert*-butylamine. While all these substances are interesting for their industrial applications, they have been selected as representatives of very different types of systems; the former are large molecules with a great ability to form hydrogen bonds (H-bond), while the latter is small and compact with limited capability of generating inter-molecular interactions. Studying such distinct systems allows to deepen our understanding of the behaviour of water in very different chemical environments. This is brought to light in the first chapter, where a database study including many different molecular hydrates is carried out and the differences between the substances are analysed.

Database study on hydrate formation at high pressure

Water is the most common solvent present in molecular organic materials,[9] and is easily incorporated in their structure due to its small size and H-bond forming capabilities.[10] Despite its ubiquitous presence and a considerable amount of research on the topic (Ch. 2), our understanding of the role of water in organic crystals is still limited, and prediction and control of hydrate formation remains an open question in crystal engineering.[11] A complete database study of hydrates formed at high pressure has been performed to bring to light trends on hydrate formation which can further contribute to systematise our overall knowledge on phase transformations. An equation of state for the compressibility of lattice water has been modelled as part of this study, providing an use-

ful tool to easily compare hydrates with their corresponding anhydrous forms at pressures below 5 GPa.

Cyclodextrins

Cyclodextrins (CDs) are cyclic derivatives of starch formed by several α-D-glucopyranose units linked together by (1→4) O-glycosidic bonds. Research on this type of compounds is mainly focused on those comprising six (α-CD), seven (β-CD) and eight (γ-CD) pyranose units. Both α- and β-CD have been studied in this work (Ch. 5 and Chs. 3 and 4, respectively). Topologically, CDs are described as hollow truncated cones in which the wider opening of the cavity is delimited by the secondary hydroxyl groups. The inner part of the cone is mainly apolar because of the alkyl groups of the glucose rings, while the outer part is polar thanks to the presence of several hydroxyl groups.[12]

Interest on CDs arises from their ability to form inclusion compounds with guests that fit in the cavity and have some capability of H-bond formation.[13,14] Thanks to this property, CDs have found a vast number of applications. To name a few, CDs are used as agents to improve solubility, stability, and bioavailability of pharmaceutical substances,[15–18] as molecular reactors,[19–23] in the constructions of molecular machines,[24–26] and as enzyme models.[27–32]

In the framework of this work, CDs provide a challenge due to their large size and difficulty to crystallise in-situ at high pressure, but also a helpful insight in the high-pressure behaviour of large systems, especially in relation to their water content. In the case of α-CD, application of pressure does not promote a phase transition; instead, the structure undergoes a conformational change in order to increase its water content and reduce the overall volume of the system. The mechanism of this process has been explored via MD simulations. β-CD shows a very different behaviour, and two different super-hydrated phases have been obtained at high pressure. The formation of these phases is discussed in relation to the overall behaviour of hydrate formation discussed in Ch. 2.

tert-Butylamine

tert-Butylamine (^{t}BA) is a very small molecule and exhibits a completely different behaviour than CDs in terms of hydrate formation, as it forms semi-clathrates. Water inclusion compounds are crystalline solids consisting of a guest (a small molecule) trapped within water cages, formed by H-bonded water molecules.[33] Water inclusion compounds can be classified into three different groups: gas hydrates, alkylamine hydrates, and quaternary ammonium salt hydrates.[34] Only gas hydrates are true clathrates, and as such their water cages are characterised by faces in the shape of regular polyhedra. The other two categories are termed semi-clathrates because, although their structure generally resembles that of a true clathrate, they exhibit interactions between the guest molecule and the host water, which causes the shape of the cages to deviate from regular polyhedra. Clathrates hydrates, particularly gas hydrates, are relevant as gas storage materials, appearing naturally in the seafloor, and due to the problems they originate in production pipelines.[33]

Together with the two structures presented in this work, a total of nine different crystalline ^{t}BA hydrates have been reported. Although semi-clathrates are structurally closer to ice than to the organic hydrates presented in Ch. 2, and a comparison with the latter seems inappropriate, a system with so many different hydrates is a perfect example of how the use of high-pressure crystallisation provides access to the densest phases in the system. The study on ^{t}BA hydrates reveals that a thorough screen of hydration capabilities necessitates the exploration of pressure space, which can help rationalising trends observed at ambient pressure.

As part of this study a new structure of pure ^{t}BA, which can crystallise as two different polymorphs depending on the pressure, has also been reported. The high-pressure phase has been also studied by MD to better understand its behaviour, and on the basis of the simulation results it has been classified as an orientational glass.[35,36]

References

[1] A. I. Kitaigorodskii, *Molecular Crystals and Molecules*, Academic Press, 1973.

[2] D. Braga, L. Brammer and N. R. Champness, *CrystEngComm*, 2005, **7**, 1–19.

[3] P. A. Wood, R. S. Forgan, D. Henderson, S. Parsons, E. Pidcock, P. A. Tasker and J. E. Warren, *Acta Cryst.*, 2006, **B62**, 1099–1111.

[4] R. D. L. Johnstone, D. Francis, A. R. Lennie, W. G. Marshall, S. A. Moggach, S. Parsons, E. Pidcock and J. E. Warren, *CrystEngComm*, 2008, **10**, 1758–1769.

[5] R. D. L. Johnstone, A. R. Lennie, S. F. Parker, S. Parsons, E. Pidcock, P. R. Richardson, J. E. Warren and P. A. Wood, *CrystEngComm*, 2010, **12**, 1065–1078.

[6] A. Gavezzotti, *J. Phys. Chem. A*, 2003, **107**, 2344–2353.

[7] E. Boldyreva, *Cryst. Growth Des.*, 2007, **7**, 1662–1668.

[8] R. Winter and J. Jonas, *High-Pressure Chemistry, Biochemistry and Materials Science*, Springer Science & Bussiness Media, 1993.

[9] C. H. Görbitz and H.-P. Hersleth, *Acta Cryst.*, 2000, **B56**, 526–534.

[10] A. L. Gillon, N. Feeder, R. J. Davey and R. Storey, *Cryst. Growth Des.*, 2003, **3**, 663–673.

[11] H. D. Clarke, K. K. Arora, H. Bass, P. Kavuru, T. T. Ong, T. Pujari, L. Wojtas and M. J. Zaworotko, *Cryst. Growth Des.*, 2010, **10**, 2152–2167.

[12] G. Wenz, *Angew. Chem. Int. Ed.*, 1994, **33**, 803–822.

[13] W. Saenger, *Angew. Chem. Int. Ed.*, 1980, **19**, 344–362.

[14] K. K. Chacko and W. Saenger, *J. Am. Chem. Soc.*, 1981, **103**, 1708–1715.

[15] F. Castelli, G. Puglisi, G. Giammona and C. A. Ventura, *Int. J. Pharm.*, 1992, **88**, 1–8.

[16] A. Paulidou, D. Maffeo, K. Yannakopoulou and I. M. Mavridis, *CrystEngComm*, 2010, **12**, 517–525.

[17] K. T. Arun, D. T. Jayaram, R. R. Avirah and D. Ramaiah, *J. Phys. Chem. B*, 2011, **115**, 7122–7128.

[18] S. M. Bruno, J. A. Fernandes, J. Marques, S. C. Neto, P. J. Ribeiro-Claro, M. Pillinger, F. A. Almeida Paz, M. P. M. Marques, S. S. Braga and I. S. Gonçalves, *Eur. J. Inorg. Chem.*, 2011, 4955–4963.

[19] K. Takahashi, *Chem. Rev.*, 1998, **98**, 2013–2033.

[20] T. J. Brett, J. M. Alexander, J. L. Clark, C. R. Ross II, G. S. Harbison and J. J. Stezowski, *Chem. Commun.*, 1999, 1275–1276.

[21] T. J. Brett, J. M. Alexander and J. J. Stezowski, *J. Chem. Soc., Perkin Trans. 2*, 2000, 1105–1111.

[22] Y. Nishioka, T. Yamaguchi, M. Yoshizawa and M. Fujita, *J. Am. Chem. Soc.*, 2007, **129**, 7000–7001.

[23] Z. Yang and H. Ji, *ACS Sustainable Chem. Eng.*, 2013, **1**, 1172–1179.

[24] A. Harada, *Acc. Chem. Res.*, 2001, **34**, 456–464.

[25] S. A. Nepogodiev and J. F. Stoddart, *Chem. Rev.*, 1998, **98**, 1959–1976.

[26] X.-J. Shen, H.-L. Chen, F. Yu, Y.-C. Zhang, X.-H. Yang and Y.-Z. Li, *Tetrahedron Lett.*, 2004, **45**, 6813–6817.

[27] R. Breslow and S. D. Dong, *Chem. Rev.*, 1998, **98**, 1997–2012.

[28] N. M. Milović, J. D. Badjić and N. M. Kostić, *J. Am. Chem. Soc.*, 2004, **126**, 696–697.

[29] F. Ortega-Caballero, C. Rousseau, B. Christensen, T. E. Petersen and M. Bols, *J. Am. Chem. Soc.*, 2005, **127**, 3238–3239.

[30] K. Kano, H. Kitagishi, C. Dagallier, M. Kodera, T. Matsuo, T. Hayashi, Y. Hisaeda and S. Hirota, *Inorg. Chem.*, 2006, **45**, 4448–4460.

[31] L. G. Marinescu and M. Bols, *Angew. Chem. Int. Ed.*, 2006, **45**, 4590–4593.

[32] M. L. Singleton, J. H. Reibenspies and M. Y. Darensbourg, *J. Am. Chem. Soc.*, 2010, **132**, 8870–8871.

[33] E. D. Sloan and C. A. Koh, *Clathrate hydrates of natural gases*, CRC Press, 3rd edn, 2007.

[34] G. A. Jeffrey, *J. Inclusion Phenom.*, 1984, **1**, 211–222.

[35] S. C. Abrahams and W. N. Lipscomb, *Acta Cryst.*, 1952, **5**, 93–99.

[36] J. C. W. Folmer, R. L. Withers, T. R. Welberry and J. D. Martin, *Phys. Rev. B*, 2008, **77**, 1–10.

Experimental and theoretical background

1.1 High-pressure crystallography

1.1.1 Diamond anvil cells

The first design of a DAC was reported in 1959 by Weir et al.,[1] based on the previous design of an opposed anvil press by Bridgman.[2] The major breakthrough of the DAC was the use of diamonds as anvils, which not only allowed to reach very high pressures, but also made it possible to visually inspect the samples during pressure application, thanks to the transparent windows. This first design became very popular, as measurement techniques such as microscopy, IR spectroscopy, or X-ray diffraction could be applied to compressed samples.[3]

DACs evolved over fifteen years until Merrill and Bassett presented their own design,[4] later improved by Hazen and Finger,[5] on which most modern cells are based (Fig. 1.1). This type of cell, which is small enough

Figure 1.1: DAC based on the design by Merrill and Bassett.

to fit in the palm of a hand and can be easily mounted on a conventional diffractometer and many other instruments, has been at the core of high-pressure research since its creation.[6] In the Merrill-Bassett design, each diamond is supported on a backing plate, which in turn is mounted on a metal body. At the beginning, diamonds used as anvils had the brilliant cut, or a modification of this, which is characterised by a large flat table—the top part of the diamond (Fig. 1.2)—that is directly supported on the baking plate. In order to offer enough support, this design requires backing plates with relatively narrow window openings. Experimental results benefited enormously from the introduction of the Boehler-Almax diamond cut,[7] which is supported on the crown rather than on the table (Fig. 1.2), hence enough support can be provided even with wider window openings ($\sim$90°) and higher pressures can be attained.[8] Ahsbahs introduced another type of cut in which the table is not flat but approximately spherical, which adds to the previous advantages the fact that the optic path trough the diamond is equally long in all orientations of the DAC, removing the need to correct for the absorption of the diamonds.[9] In both cases, the support also attempts to facilitate the alignment of the diamonds.

Proper placement of the diamonds on the backing plates and alignment of the diamonds relative to each other is critical to ensure a homogeneous distribution of pressure and to minimise the risk of gasket failure, which can lead to diamond fracture. A modification of the alignment system was key to increase the pressure range of the DAC up to 25 GPa.[10] Backing plates were originally made out of beryllium; its transparency to X-rays was advantageous when only narrow windows were available. Steel or tungsten carbide baking plates, which block X-ray radiation, in combination with different wide window geometries, are preferred nowadays.[9,11] Synthetic diamond backing plates, which allow to reach pressures up to 223 GPa, are also available.[12]

The sample chamber is formed by a drilled gasket which is placed in between the two diamonds, forming a sealed compartment when pressure is applied (Fig. 1.2). Inside the chamber there should be a pressure calibrant, a pressure transmitting medium (PTM), and the sample itself

10
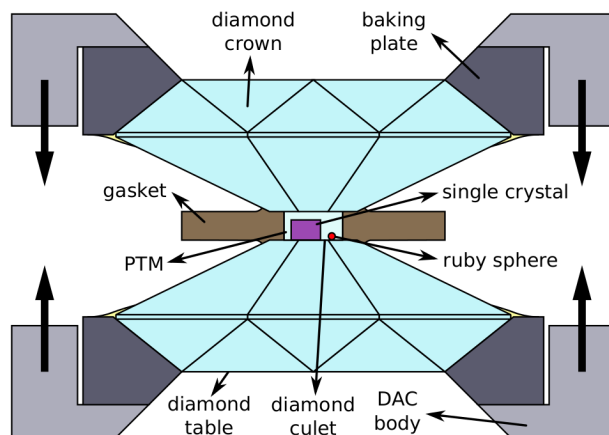
Figure 1.2: Diagram (not to scale) of a DAC with Boehler-Almax diamonds. The thick arrows indicate the direction in which pressure is applied via the pressing screws.

(liquid samples and samples in solution can act as PTM on their own), as described in the following section. A variety of gasket materials are available, and in this work Inconel 718, steel, and brass are used, with thickness around 150 μm, pre-indented according to the guidelines from Dunstan,[13] depending on each specific DAC used. In all the DACs employed in this work (Appx. I), pressure is applied by means of screws connecting the two steel bodies which house the diamonds on their backing plates. The load is transmitted from a relatively large surface on the crown of the diamonds to a much smaller one on the culet, with the subsequent amplification. Load transmission can be further controlled by the volume of the sample chamber (diameter of the hole and thickness of the gasket); for the same position of the pressing screws, the smaller the sample chamber the higher the pressure. Pressure is applied by the diamonds to the PTM and also to the gasket, which through plastic deformation allows to recreate hydrostatic conditions in the interior of the sample chamber, up to the solidification point of the PTM ($\sim$1 GPa in the case of aqueous solutions). As most of the work here presented is limited to low pressures, below the formation of ice VI, relatively large sample chambers are used whenever possible, with the intention to grow larger crystals to facilitate the collection of X-ray data. Pressure inside the sample chamber is monitored by the shift of the R_1 fluorescence line of Cr^{3+}, from a small ruby sphere loaded along with the sample.[14,15] For all the experiments here reported, a 532 nm laser was used as an excitation source and the shift was measured on an OceanOptics USB4000

spectrometer (the uncertainty of pressure measured by this method is 0.05 GPa). For further information on the design and operation of the DAC, refer to the work of Miletich et al.[16] For a technical description of the DACs used in this work, see Appx. I.

The development of the DAC has not been idle in the recent years, but has steered towards the increase of the range of applications of the device. A DAC especially designed to work at the lower end of the pressure scale, with improved viewing windows, has been designed and used to monitor (by means of microscopy and spectroscopy) the behaviour of microbes under pressure.[17] At this range of pressures, calibration with the ruby method is not sensitive enough, and a fluorescent dye embedded in a polymeric matrix is used instead. Such sensors can be used in aqueous solution up to 323 K in the 3-12 pH range. For the purpose of magnetic measurements at cryogenic temperatures, pressure cells constructed of metal are impractical, as the magnetic field will induce current in the metallic parts, heating the sample. A plastic DAC has been built for such studies, substituting the gasket by a polymer containing diamond powder.[18] The pressing screws have been substituted by a turnbuckle mechanism, which allows for further miniaturisation. A similar turnbuckle pressing mechanism has been implemented in metal to create a miniature DAC which can be fully immersed in a conventional cryostream of a diffractometer, allowing for easy low-temperature high-pressure diffraction studies.[19] In neutron diffraction, the Paris-Edinburgh cell[20] has been the preferred choice for high-pressure, because of its larger sample chamber, although recently Binns et al. reported being successful on the use of a miniaturised Merril-Bassett DAC constructed of a copper-beryllium alloy for the collection of neutron diffraction data.[21] Further development of DACs is oriented towards creating wider windows while being able to attain very high pressures. A DAC with an opening angle of 100° on one side and 90° on the other, equipped with flat diamonds and the sample chamber drilled in polymer (supported on a traditional metallic gasket), has been used in this work. See Appx. I.3 for a more detailed description.

1.1.2 Experiments in a diamond anvil cell

There are three distinct types of experiments that can be performed with DACs: in-situ crystallisation of liquids, in-situ crystallisations of solids from solution, and compression studies.[22] The present work contains examples of all three cases.

In an in-situ crystallisation experiment, a crystal is directly grown under pressure inside the DAC. The simplest implementation consists of loading a liquid; however, the vast majority of molecular organic materials are solid at ambient pressure. Crystallisation of liquids and solids from solution are formally identical, as discussed below, although liquids have the advantage that the size of the crystals obtained is not a limitation, as the entire content of the sample chamber usually crystallises. For solid materials, a DAC is loaded with a saturated solution of the material in an appropriate solvent. The solution acts as the PTM, that is, the substance which receives the pressure from the diamonds and then transfers it to the crystal once it is formed. Under compression, high density phases are favoured; this promotes crystallisation, as solids tend to be denser than liquids and solutions. In the case of materials which can crystallise in different phases, either polymorphs, solvates, or cocrystals, denser phases are usually favoured during high-pressure crystallisation (see for instance the work by Fabbiani[23–28] and Katrusiak[29–31]). There are, nevertheless, exceptions to this behaviour, CDs being a good example (Chs. 3 and 5).

When crystallisation first occurs on increasing pressure, a polycrystalline sample is usually obtained, because at this stage the pressurisation rate tends to be fast and the sample is driven past the metastable zone for crystallisation. At this point, isobaric temperature cycling can be used to produce a single crystal. Cycling consists in heating to reduce the number of crystals in the sample chamber by dissolving them and then cooling to let the remaining crystals grow. The process is repeated until a single seed remains in the chamber, which is subsequently let to grow until it reaches a suitable size for X-ray diffraction. Once a single crystal is formed, pressure can be increased to ensure that the crystal will not further dissolve (for instance, under the heat of the X-ray

beam) or to explore possible phase transitions upon compression. Although having a unique single crystal inside the DAC is desirable, with the availability of micro-focussed synchrotron beams it is now possible to collect data on a specific small crystal in the presence of others.

One of the main limitations of in-situ crystallisation is the solubility of the sample in the solvent used as PTM. The amount of solution that can be loaded in a DAC is below 0.02 mm^3, which results in a very small amount of sample if said sample is poorly soluble, which does not allow to grow crystals of a suitable size for X-ray diffraction. To overcome this problem, crystals grown at ambient pressure can be loaded along with the saturated solution; subsequently the DAC is heated to dissolve the crystals in situ and effectively increase the concentration of the solution. In the particular case of CDs, heating is not needed, as the crystals dissolve, increasing the concentration of the solution, upon compression.

In a compression study, a crystalline sample is submitted to a gradual increase in pressure; data are collected at different pressure steps to identify and characterise possible phase transitions and changes at the molecular level (see for instance the work by Parsons,[32–37] Boldyreva,[38–40] and Lee[41,42]). X-ray diffraction and Raman spectroscopy are commonly used to analyse the compressed samples. To eliminate effects originating from the solvent, inert PTMs are usually employed (for instance, perfluorinated alkenes or pentane/cyclopentane mixtures in the case of polar substances). In other instances, an interacting PTM is desired, when it is the combination of an appropriate solvent and pressure which causes changes in the material under study, as it is the case of pressure-induced hydration (Ch. 5) and phase transitions that may be otherwise kinetically hindered.[22]

1.1.3 X-ray data collection and processing in a diamond anvil cell

High-pressure X-ray diffraction using a DAC suffers from three main problems:[43] small size of the sample, reduced access to the reciprocal space, and reflections with untrue intensities.

The small size of the sample can potentially be a problem, as indicated in the previous sections, but the availability of high-intensity micro-focused sources allows to measure small organic crystals in the home laboratory routinely. For the most difficult cases, or when rapid data collection is required, synchrotron sources offer a great alternative, with brighter beams which produce intense reflections even from poorly-diffracting crystals, and micro-focussed beams which allow to collect data on very small crystals, even in the presence of other specimens inside the sample chamber.

The access to the reciprocal space is limited by the body of the cell.[16] Only reflections coming out of the sample at an angle smaller than the opening of the window of the DAC can be collected. This translates into a cone of reflections in direct space—two cones joined by their vertices when the data is collected in both sides of the DAC—or a torus in reciprocal space (Fig. 1.3). This is seen in the diffraction frames in the form of a shadow, a region which contains no reflections. That is, both completeness and redundancy are compromised. The problem is more noticeable for low symmetry crystals (triclinic, monoclinic, and orthorhombic), typical for organic materials. The situation can improve with a proper data collection strategy and the use of shorter X-ray wavelengths, which increases the effective size of the Ewald sphere, allowing to collect more reflections. The use of the DAC affects the final results of structure refinement, as statistics based on incomplete data are less reliable and a low data-to-parameter ratio limits the numbers of paramaters that can be refined. Moreover, systematic missing data results in the stretching of the electron density along the missing direction—perpendicular to the torus in reciprocal space—which can result for instance into a systematic error in the shape of the atomic displacement parameters.

In addition to the limited completeness, X-ray diffraction data collected at high pressures is affected by more absorption than for crystals collected in air (diamonds, gasket, baking plates, and PTM) and might show undesired diffraction from the diamonds and the gasket (and the baking plates if these are made out of beryllium).[44,45] Although modern software has proven successful in performing absorption correction

Figure 1.3: Cross section of a torus of reflections in the reciprocal space, characteristic of data collections of samples inside a DAC.
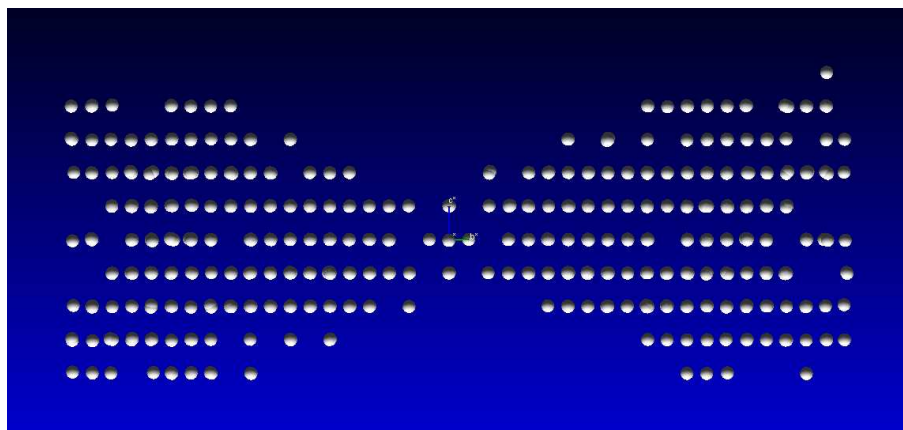
for high-pressure samples, it can be challenging considering the limited redundancy of the data.[43] Repetition of the data collection is usually a good approach to improve redundancy, although it is time consuming and might not be possible if radiation damage is present.

Considering the difficulties outlined before, the data-collection strategy for single crystals contained in a DAC is oriented towards increasing completeness and redundancy. Usually, as much data as possible is collected, independently of the symmetry of the crystals. For the purpose of this work, the guidelines from Dawson et al. were followed.[46] On a three-circle diffractometer, the maximum coverage of the reciprocal space, which primarily depends on the window opening of each DAC, is performed via ω scans, with additional ϕ scans added to increase redundancy. Each scan is actually double: first a scan is performed with the front of the DAC facing the incoming beam, and subsequently another one is performed with the back side facing the beam. In the absence of a fourth circle, the DAC is manually rotated on the goniometer head around the axis of the beam and the previous strategy is repeated for two or three different orientations. This significantly improves the redundancy of the data, but might introduce centring problems which need to be managed during the data reduction. In the case of diffractometers with only ϕ movement available, ϕ scans are performed, as wide as the window opening, on both sides of the DAC, combined with the aforementioned manual rotation. Shorter wavelengths and larger detectors are highly desirable in this configuration.

Data processing is performed in a similar way than it would for crystals measured at ambient-pressure conditions, although some differences need to be addressed.[46] As indicated previously, statistics tend to be less reliable when data completeness and redundancy is low. It is advisable in these cases to perform the complete integration, all needed corrections, and merging of different scans following different strategies, and use the end result of the refinement to asses which one of the strategies is the best approach.

Firstly, it should be considered whether different scans should be integrated together or separately. From the experience of this work, the latter tends to yield better results, due to the inherent difficulties of centring the DAC.[47] SAINT[48] has been the integration software of choice for most samples, although CRYSALIS PRO[49] has provided better results for samples collected using a Perkin-Elmer XRD 1621 amorphous-silicon flat panel detector, installed on the P02.2 beamline at the PETRAIII synchrotron (Ch. 3.2.2). Both programs include dedicated features to process high-pressure data, namely the ability to integrate only in the non-shadowed areas of the frames. The shadowed areas can be completely removed from the integration by specifying in the software the DAC opening angle. Specifying an angle smaller than the actual one eliminates reflections which suffer from partial shadowing, but this might negatively affect the total amount of integrated data. If the *true* opening angle is indicated in the integrations program, it is possible to use SHADE[50] after integration to filter problematic reflections. This program can also be used to remove partially-shadowed reflections or to remove reflections originating from the diamonds, which are sometimes integrated if overlapping with sample reflections. As part of this work a new version of this software, called SOMBRA, has been written to apply similar corrections to data collected using DACs with asymmetric windows (Appx. I.3).

Secondly, absorption correction is critical to reduce errors on the measured intensities. In this work, absorption corrections have been performed with SADABS[51] (or TWINABS[52] for twinned samples). When each scan is integrated separately, the point at which all data are analysed

together can affect the final result, as there is a different amount of data available to perform the absorption correction. Merging can be performed at the same time than absorption correction in SADABS, or afterwards, using a dedicated program, like SORTAV[53] or XPREP[54], or the refinement software, SHELXL[55] in this case. All options have been routinely tried for all samples, selecting the procedure that yielded the best results in each case.

1.2 Molecular dynamics simulations

§1.2.1, §1.2.2, and §1.2.4 constitute a brief review about molecular dynamics, in the same way the previous section was a review about high-pressure crystallisation and crystallography. §1.2.3 is aimed to demonstrate how energies can be computed from a simulated trajectory. Here I have used the teaching approach by P. G. Lafond, to whom I am grateful, which starting from the definition of energy works its way up demonstrating the partition functions of the NVE, NVT, and NPT ensembles. Subsequently, the relationship between the last two and the characteristic free energies of each ensemble—Helmholtz in NVT and Gibbs in NPT—are demonstrated. Finally, the variation of Gibbs free energy between two states in the NPT ensemble, based on their relative probabilities, is determined. Most of the equations in this section can be found in D. Frenkel and B. Smit, Understanding Molecular Simulation: From Algorithms to Applications, *Academic Press, 2001, although the information has been rearranged according to the aforementioned scheme, and additional steps have been included and certain concepts have been simplified in the demonstrations for clarity. The* raison d'êtrè *of this section is to provide a more concise approach to energy calculation than that found in text books, with additional explanations to make it easier to read for those with limited knowledge in statistical mechanics, and more importantly, is a personal exercise to test my understanding of the field.*

1.2.1 Equilibration fundamentals

MD is a computer simulation technique which, based on an appropriate force field describing the system under study, predicts the behaviour over time of such system by numerically solving the equations of motion.[56] Differently from other simulation techniques, MD focusses on the time

evolution of the system, a paramount factor when studying conformational changes in molecular systems. In MD, processes taking place at the molecular level are described by means of classical mechanics. The interaction between any two atoms in the system, separated a distance r, is described by a Lennard-Jones potential (U_{LJ}):

$$U_{\text{LJ}}(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right] \tag{1.1}$$

Where ϵ is the depth of the potential well and σ is the distance at which the potential is zero. The atoms experience attraction when they are within range of each other due to Van der Waals or covalent interactions, but repulsion becomes dominant below a certain distance due to the overlap of their atomic orbitals. When the two atoms are separated further apart than the range of the Lennard-Jones potential, a Coloumb potential (U_{C}) is better suited to describe their interaction:

$$U_{\text{C}}(r) = \frac{1}{4}\epsilon_0 \pi \left(\frac{q_i q_j}{r} \right) \tag{1.2}$$

Where ϵ_0 is the electrical permittivity of the medium and q_i and q_j are the charges of each atom. In a given system, each atom exerts a force $F(r)$ on any other atom according to the aforementioned potentials:

$$F(r) = -\frac{\mathrm{d}U(r)}{\mathrm{d}r} = -\frac{\mathrm{d}}{\mathrm{d}r} \left[U_{\text{LJ}}(r) + U_{\text{C}}(r) \right] \tag{1.3}$$

Once the forces are known, changes in the acceleration experienced by any atom due to interaction with other atoms can be determined from Newton's second law:

$$
\begin{aligned}
F(r) = m\frac{\mathrm{d}v}{\mathrm{d}t} = m\frac{\mathrm{d}^2 r}{\mathrm{d}t^2} \quad &\Rightarrow \\
\frac{\mathrm{d}v}{\mathrm{d}t} = \frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = \frac{F(r)}{m}
\end{aligned}
\tag{1.4}
$$

Where m is the mass of the atom, r its position, v its velocity, and t is the time. Then, the trajectory of any atom, composed of its position and velocity over time, can be calculated.

The large number of atoms which usually conform molecular sys-

tems makes it impossible to integrate the equations of motion exactly. Instead, numerical integration is used. Although there are different integration schemes, MD motors used in this work (NAMD[57] and GROMOS[58]) employ variations of the Verlet scheme[59] for its efficiency and numerical stability—that is, avoiding amplification of errors during the simulation.

Newton's laws of motion are defined for systems at constant energy, hence MD simulation are performed in the NVE ensemble (constant number of particles, volume, and energy) by default. Structural changes in molecules rarely occur at constant energy. In order to perform simulations in the NVT ensemble (at constant temperature), which is more useful from a chemical perspective, certain adjustments are needed.[59,60]

According to the equipartition theorem, the kinetic energy (K) associated to each degree of freedom of a system is:

$$K = \frac{1}{2}k_{\mathrm{B}}T \tag{1.5}$$

Where k_{B} is the Boltzmann constant and T is the temperature. For an atom, which can move in three directions:

$$K = \frac{3}{2}k_{\mathrm{B}}T \tag{1.6}$$

In a molecular system composed of N atoms:

$$K = \frac{3N}{2}k_{\mathrm{B}}T = \sum_{i=1}^{N}\frac{1}{2}m_i v_i^2 \quad \Rightarrow$$

$$T = \frac{1}{(3N-3)k_{\mathrm{B}}}\sum_{i=1}^{N}m_i v_i^2 \tag{1.7}$$

Three degrees of freedom have been removed, as they correspond to the movement of all the atoms as a single entity, which is not considered in MD. In addition to Eq. 1.7, the velocities of the atoms in the system must behave according to a Maxwell-Boltzmann distribution:

$$P(v) \propto \mathrm{e}^{-\beta K} \tag{1.8}$$

Where β is $1/(k_{\mathrm{B}}T)$.

In order to simulate a system in isothermal conditions, the velocities must be rescaled so that the temperature is in agreement with both Eqs. 1.7 and 1.8. This can be achieved in different ways. The simplest method consists in calculating the temperature after each step of the simulation, computing a scaling factor for the velocities in such a way that the temperature agrees with the aforementioned equations, and finally rescaling the velocities according to the scaling factor. While this method provides the right temperature, the velocity distribution is prone to errors. A better approach to reproduce the correct velocity distribution is the Andersen thermostat, which resets the velocities after a certain number of steps, taking new values directly from the Maxwell-Boltzmann distribution. This method has the disadvantage that non-realistic values for the velocities can be generated, which in turn generates non-physical trajectories. More realistic methods are based on the idea of coupling the simulated system with a bath. The Berendsen thermostat adds a heat bath which is at the target temperature and absorbs the temperature fluctuations of the system until both temperatures—of the system and the bath—are equilibrated. In reality this is not an isothermal process, and errors may arise for small systems, although the approximation is usually acceptable for larger systems. The method of choice in modern software is the Nosé-Hoover thermostat, which reformulates the equation of motion (Eq. 1.4):

$$\frac{\mathrm{d}v}{\mathrm{d}t} = \frac{F(r)}{m} - \zeta v \tag{1.9}$$

Being:

$$\frac{\mathrm{d}\zeta}{\mathrm{d}t} = \frac{1}{Q} \sum_i \left(m_i v_i^2 - 3N k_\mathrm{B} T \right) \tag{1.10}$$

Where ζ is the *friction* between the bath and the system and Q is the mass of the bath. The thermostat works by sinking part of the energy of the system into the bath, until the temperatures of both equalise. This method provides the right temperature and distribution of velocities.

The simulations reported in this work have been performed in the NPT ensemble (constant pressure). Isobaric conditions are achieved by approaches similar to those explained above for isothermal processes.[59,60] In the same way that the velocities—and hence the kinetic energy—can

be rescaled to generate the correct temperature, the volume of the simulation box can be rescaled over time to generate the target pressure. Similarly to the temperature case, simply rescaling the simulation box does not necessarily generate the right pressure distribution. The preferred approached is the Parrinello-Rahman barostat, which is formally equivalent to the Nosé-Hoover thermostat:

$$\frac{\mathrm{d}v}{\mathrm{d}t} = \frac{F(r)}{m} - \boldsymbol{M}v \tag{1.11}$$

Where $\boldsymbol{M}$ is a coupling matrix. The main advantage of this barostat is that pressure is treated as a stress tensor, allowing each dimension of the simulation box to change independently.

Another parameter to be addressed in MD simulations is the length of each time step. Numerical integration of the equations of motion works by determining the next point of the trajectory based on the immediately previous one, which generates accurate trajectories only when very short steps are used (0.1 or 0.2 fs). Such small time steps result in very long computation times, which are not practical for simulations of systems over thousands of atoms over hundreds of nanoseconds, as used here.

There are two well-known approaches to either allow for larger time steps or reduce the number of calculations in each step.[60] In rigid dynamics, chemical bonds are considered to be stiff. Under these conditions, the movement of molecules as a single unit is still successfully described by the equations of motion, although this does not apply to individual atoms, because their natural vibrations are constrained to be zero. Such constraints are introduced in the equations of motion in the form of extra energy terms. The major drawback of this approach is that extracting further information from the trajectory, like temperature, pressure, or energy, also requires to account for the constraints. Another approach is to coarse grain the potentials,[60] which removes from the simulation those atoms that have no effect, or more precisely, combines them with their parent atoms as a rigid body. For instance, neutral H atoms can be considered as part of the carbon atom they are attached to. By doing this, the system is greatly simplified without the need to modify the equations of motion. The only condition is to include

properly parametrised rigid bodies in the force field.

As introduced at the beginning of this section, atomic interactions in MD are treated differently depending on the distance between the atoms. In theory, an atom exerts a force on any other atom of the system. If said atoms are far apart, the force and the energy associated with their interaction might be small, but given the large number of long-range interactions that are present in a system, this term is not negligible and must be computed in order to derive proper energy values. This implies a large number of calculations and several strategies have been developed to decrease the computational cost.[59] Lennard-Jones interactions (Eq. 1.1) are simply ignored over a certain distance, under the assumption that the neglected contributions are likely to cancel out when added together. The energy of these interactions varies with r^{-6}, hence it is easy to stablish a cutoff distance over which the energy is statistically negligible. Coulombic interactions (Eq. 1.2) on the other hand, vary with r^{-1}, which implies a much slower decrease of energy with distance, making it more complex to establish a clear cutoff. These interactions are computed by means of Ewald summation. Ewald summation divides the electrostatic interactions in two terms: short- and long-range. The first term is computed using the Coulomb equation (Eq. 1.2) for a small number of atom pairs which are closer than a certain cutoff. The number of atom pairs over the cutoff is much larger and in the Ewald scheme their interactions are computed in reciprocal space, as convergence is much faster, making use of the fact that the system is considered periodic during the simulation. The implementation of the Ewald summation most frequently found in MD motors is particle mesh Ewald (PME). The main characteristic of this technique is that the computation is speeded up by using fast Fourier transform, which does not compute each interaction independently, but rather distributes them over the nodes of an imaginary grid, performing the calculation only at the nodes of the grid. Ewald summation is the most time-consuming part of any MD simulation.

The use of Ewald summation requires the simulated system to be in periodic boundary conditions (PBC). These conditions assume that the content of the simulation box is repeated infinitely in all three directions

of space, and that when a particle leaves the box from one side, it simple returns from the opposite side, maintaining its energy and trajectory. PBC not only reduce the computational cost of the simulation, but are also convenient to avoid undesired effects at the surface of the simulation box. Nevertheless, the use of PBC can also cause errors. For instance, in a simulation box containing a molecule or a cluster of molecules surrounded by solvent, the box must be large enough to avoid the molecules interacting with their images in the neighbouring box, which is physically unsound. In the absence of solvent, the simulation box must be large enough to accurately represent disorder.

1.2.2 Trajectory analysis

MD simulations of molecular systems produce trajectories in terms of position and velocity of atoms. For the simulation of crystals, as is the case in this work, it is usually more interesting to look at where the atoms move rather than the velocity with which they move. A common first step is a visual analysis of the trajectories, which can be watched as a video. When events of interest are found, the geometrical information can be extracted and processed to quantify the event. Because of the very specific nature of each event studied in this work, custom analysis scripts have been written tailored to each case. A detailed description of each analysis script is given in the *Experimental* section of each appropriate chapter.

1.2.3 Energy calculations

According to statistical mechanics, the energy of an NVE ensemble is:

$$
\begin{aligned}
\mathrm{d}E = T\mathrm{d}S - P\mathrm{d}V + \mu\mathrm{d}N \quad \Rightarrow \\
(\mathrm{d}E)_{NV} = T(\mathrm{d}S)_{NV}
\end{aligned}
\tag{1.12}
$$

From here, the temperature is:

$$
T = \left(\frac{\partial E}{\partial S}\right)_{NV}
\tag{1.13}
$$

Entropy is defined as:

$$S = k_{\mathrm{B}} \ln \Omega \tag{1.14}$$

Where Ω is the NVE partition function. Combining with Eq. 1.13:

$$\frac{1}{T} = \left(\frac{\partial S}{\partial E}\right)_{NV} = \left(\frac{\partial}{\partial E}(k_{\mathrm{B}} \ln \Omega)\right)_{NV} = k_{\mathrm{B}} \left(\frac{\partial(\ln \Omega)}{\partial E}\right)_{NV} \tag{1.15}$$

From which the thermodynamic β can be defined:

$$\beta = \frac{1}{Tk_{\mathrm{B}}} = \left(\frac{\partial(\ln \Omega)}{\partial E}\right)_{NV} \tag{1.16}$$

Given an NVT system consisting of a small simulation box, which can be in ν states of energy E_ν, immersed in a bath large enough so that its temperature can be considered constant in comparison to that of the box, the entropy of the bath is:

$$S_{\mathrm{bath}} = k_{\mathrm{B}} \ln \Omega_{\mathrm{bath}} \tag{1.17}$$

And its energy is:

$$\begin{aligned} E_{\mathrm{system}} = E_{\mathrm{bath}} + E_{\mathrm{box}} = E_{\mathrm{bath}} + E_\nu \quad \Rightarrow \\ E_{\mathrm{bath}} = E_{\mathrm{system}} - E_\nu \end{aligned} \tag{1.18}$$

The value of the entropy at E_{bath} can be approximated by two terms of a Taylor expansion:

$$\begin{aligned} S_{\mathrm{bath}}(E_{\mathrm{bath}}) = & S_{\mathrm{bath}}(E_{\mathrm{system}} - E_\nu) \approx \\ & S_{\mathrm{bath}}(E_{\mathrm{system}}) - E_\nu \frac{\partial S_{\mathrm{bath}}}{\partial E_{\mathrm{system}}} \end{aligned} \tag{1.19}$$

From Eqs. 1.13 and 1.16:

$$\begin{aligned} S_{\mathrm{bath}}(E_{\mathrm{bath}}) \approx & S_{\mathrm{bath}}(E_{\mathrm{system}}) - \frac{E_\nu}{T} = \\ & S_{\mathrm{bath}}(E_{\mathrm{system}}) - \beta k_{\mathrm{B}} E_\nu \end{aligned} \tag{1.20}$$

Applying to Eq. 1.17:

$$\ln \Omega_{\text{bath}} = \frac{1}{k_{\text{B}}} S_{\text{bath}} \approx \frac{1}{k_{\text{B}}} S_{\text{bath}}(E_{\text{system}}) - \beta E_{\nu} \quad \Rightarrow$$
$$\Omega_{\text{bath}} \approx \exp\left(\frac{1}{k_{\text{B}}} S_{\text{bath}}(E_{\text{system}}) - \beta E_{\nu}\right) \propto \mathrm{e}^{-\beta E_{\nu}} \tag{1.21}$$

Where $e^{-\beta E_{\nu}}$ is the so called Boltzmann factor. The total number of states of the system is given by:

$$\Omega_{\text{system}} = \Omega_{\text{bath}} \Omega_{\text{box}} \quad \Rightarrow$$
$$\Omega_{\text{box}} = \frac{\Omega_{\text{system}}}{\Omega_{\text{bath}}} \tag{1.22}$$

The probability of the simulation box being in a given state ν, applying Eq. 1.21, is:

$$P_{\nu} = \frac{1}{\Omega_{\text{box}}} = \frac{\Omega_{\text{bath}}}{\Omega_{\text{system}}} = \frac{\mathrm{e}^{-\beta E_{\nu}}}{\sum_{\nu} \mathrm{e}^{-\beta E_{\nu}}} = \frac{1}{Q} \mathrm{e}^{-\beta E_{\nu}} \tag{1.23}$$

Where Q, the partition function of an NVT ensemble, is:

$$Q = \Omega_{\text{system}} = \sum_{\nu} \mathrm{e}^{-\beta E_{\nu}} \tag{1.24}$$

A partition function allows to compute average properties of a system. For an NVT ensemble:

$$\langle X \rangle = \frac{1}{Q} \sum_{\nu} X_{\nu} \mathrm{e}^{-\beta E_{\nu}} \tag{1.25}$$

Then, the average energy of the system is:

$$\langle E \rangle = \frac{1}{Q} \sum_{\nu} E_{\nu} \mathrm{e}^{-\beta E_{\nu}} \tag{1.26}$$

The derivative of the Boltzmann factor with respect to β is:

$$\left(\frac{\partial \left(\mathrm{e}^{-\beta E_{\nu}}\right)}{\partial \beta}\right)_{NV} = -E_{\nu} \mathrm{e}^{-\beta E_{\nu}} \tag{1.27}$$

Applying to Eq. 1.26:

$$\langle E \rangle = \frac{1}{Q} \sum_\nu - \left(\frac{\partial \left(\mathrm{e}^{-\beta E_\nu} \right)}{\partial \beta} \right)_{NV} =$$
$$- \frac{1}{Q} \left(\frac{\partial}{\partial \beta} \sum_\nu \mathrm{e}^{-\beta E_\nu} \right)_{NV} \tag{1.28}$$

Applying Eq. 1.24:

$$\langle E \rangle = -\frac{1}{Q} \left(\frac{\partial Q}{\partial \beta} \right)_{NV} = - \left(\frac{\partial (\ln Q)}{\partial \beta} \right)_{NV} \tag{1.29}$$

It can be demonstrated that Q is related to the Helmholtz free energy:

$$A = E - TS \quad \Rightarrow$$
$$\beta A = \beta E - \beta T S = \beta E - \frac{1}{k_\mathrm{B}} S \tag{1.30}$$

Taking derivatives with respect to β:

$$\left(\frac{\partial (\beta A)}{\partial \beta} \right)_{NV} = \left(\frac{\partial (\beta E)}{\partial \beta} \right)_{NV} - \left[\frac{\partial}{\partial \beta} \left(\frac{1}{k_\mathrm{B}} S \right) \right]_{NV} =$$
$$E + \beta \left(\frac{\partial E}{\partial \beta} \right)_{NV} - \frac{1}{k_\mathrm{B}} \left(\frac{\partial S}{\partial \beta} \right)_{NV} \tag{1.31}$$

From Eq. 1.16, the differential of β can be expressed in terms of direct temperature:

$$\beta = \frac{1}{T k_\mathrm{B}} \quad \Rightarrow$$
$$\frac{\partial \beta}{\partial T} = \frac{\partial}{\partial T} \left(\frac{1}{T k_\mathrm{B}} \right) = - \frac{1}{T^2 k_\mathrm{B}} \quad \Rightarrow \tag{1.32}$$
$$\partial \beta = \frac{-\partial T}{T^2 k_\mathrm{B}}$$

Applying to Eq. 1.31:

$$\left(\frac{\partial (\beta A)}{\partial \beta} \right)_{NV} = E + \beta \left(\frac{\partial E}{-(\partial T)/T^2 k_\mathrm{B}} \right)_{NV} -$$
$$\frac{1}{k_\mathrm{B}} \left(\frac{\partial S}{-(\partial T)/T^2 k_\mathrm{B}} \right)_{NV} = \tag{1.33}$$
$$E - T \left(\frac{\partial E}{\partial T} \right)_{NV} + T^2 \left(\frac{\partial S}{\partial T} \right)_{NV}$$

By definition:

$$C_V = \left(\frac{\partial E}{\partial T}\right)_V \tag{1.34}$$

And then, recalling Eq. 1.12:

$$(\mathrm{d}E)_{NV} = T(\mathrm{d}S)_{NV} \quad \Rightarrow$$
$$\left(\frac{\partial E}{\partial T}\right)_{NV} = T\left(\frac{\partial S}{\partial T}\right)_{NV} \quad \Rightarrow \tag{1.35}$$
$$\left(\frac{\partial S}{\partial T}\right)_{NV} = \frac{1}{T}\left(\frac{\partial E}{\partial T}\right)_{NV} = \frac{C_V}{T}$$

Applying Eqs. 1.34 and 1.35 to Eq. 1.33 and equating the macroscopic ensemble to the microscopic level:

$$\left(\frac{\partial(\beta A)}{\partial \beta}\right)_{NV} = E - TC_V + TC_V = E = \langle E\rangle \tag{1.36}$$

Bringing together Eqs. 1.29 and 1.36:

$$\langle E\rangle = -\left(\frac{\partial(\ln Q)}{\partial \beta}\right)_{NV} = \left(\frac{\partial(\beta A)}{\partial \beta}\right)_{NV} \quad \Rightarrow$$
$$\ln Q = -\beta A \quad \Rightarrow \tag{1.37}$$
$$Q = \mathrm{e}^{-\beta A}$$

In a continuous system, Q can be integrated over all positions of all particles of the system:

$$Q = C \int \mathrm{e}^{-\beta E_\nu} \mathrm{d}x_1 \mathrm{d}y_1 \mathrm{d}z_1 \ldots \mathrm{d}x_N \mathrm{d}y_N \mathrm{d}z_N \tag{1.38}$$

Where C is a constant including the kinetic energy term and the number of particles. Defining the reduced position s as:

$$s_n = \frac{x_n}{L_x} \frac{y_n}{L_y} \frac{z_n}{L_z} \quad \Rightarrow$$
$$x_n y_n z_n = L_x L_y L_z s_n = V s_n \quad \Rightarrow \tag{1.39}$$
$$\mathrm{d}x_n \mathrm{d}y_n \mathrm{d}z_n = V \mathrm{d}s_n$$

Where L_x, L_y, and L_z are the dimensions of the system. Applying to

Eq. 1.38:

$$Q = C \int e^{-\beta E_\nu} V^N \mathrm{d}^{3N} s = C V^N \int e^{-\beta U_\nu} \mathrm{d}^{3N} s \qquad (1.40)$$

Given an NPT ensemble, consisting of a small box of variable volume, containing N particles, immersed in a large bath also of variable volume, containing M ideal particles, large enough so that its temperature can be considered constant in comparison to that of the box:

$$
\begin{aligned}
V_{\text{system}} &= V_{\text{bath}} + V_{\text{box}} \quad \Rightarrow \\
V_{\text{bath}} &= V_{\text{system}} - V_{\text{box}} = V_0 - V
\end{aligned}
\qquad (1.41)
$$

In terms of partition functions, applying Eq. 1.40:

$$
\begin{aligned}
Q_{\text{system}} = & Q_{\text{bath}} Q_{\text{box}} = \\
& C V_{\text{bath}}^M \int e^{-\beta U_\nu} \mathrm{d}^{3M} s \quad C V_{\text{box}}^N \int e^{-\beta U_\nu} \mathrm{d}^{3N} s = \\
& C(V_0 - V)^M \int e^{-\beta U_\nu} \mathrm{d}^{3M} s \quad C V^N \int e^{-\beta U_\nu} \mathrm{d}^{3N} s
\end{aligned}
\qquad (1.42)
$$

Because the particles in the bath are ideal, their potential energy is zero:

$$Q_{\text{system}} = C(V_0 - V)^M V^N \int e^{-\beta U_\nu} \mathrm{d}^{3N} s \qquad (1.43)$$

Assuming that $V_0 \gg V$ and $M \to \infty$:

$$
\begin{aligned}
(V_0 - V)^M = & V_0^M \left(\frac{V_0}{V_0} - \frac{V}{V_0} \right)^M = V_0^M \left(1 - \frac{V}{V_0} \right)^M \\
& \approx V_0^M \exp \left(\frac{-MV}{V_0} \right)
\end{aligned}
\qquad (1.44)
$$

From the law of ideal gases:

$$
\begin{aligned}
PV_0 &= M k_{\text{B}} T \quad \Rightarrow \\
\frac{M}{V_0} &= \frac{P}{T k_{\text{B}}} = \beta P
\end{aligned}
\qquad (1.45)
$$

Applying to Eq. 1.44:

$$(V_0 - V)^M \approx V_0^M e^{-\beta P V} \qquad (1.46)$$

And applying to Eq. 1.43:

$$Q_{\text{system}} = CV_0^M e^{-\beta PV} V^N \int e^{-\beta U_\nu} d^{3N} s =$$
$$CV_0^M V^N \int e^{-\beta U_\nu} e^{-\beta PV} d^{3N} s = \qquad (1.47)$$
$$CV_0^M V^N \int e^{-\beta U_\nu - \beta PV} d^{3N} s$$

By analogy with Q, for an NPT ensemble, the partition function Z is defined as:

$$Z = \sum_\nu e^{-\beta U_\nu - \beta PV} \qquad (1.48)$$

The average volume of an NPT ensemble is:

$$\langle V \rangle = \frac{1}{Z} \sum_\nu V e^{-\beta U_\nu - \beta PV} \qquad (1.49)$$

The derivative of the Boltzmann factor with respect to βP is:

$$\left(\frac{\partial \left(e^{-\beta U_\nu - \beta PV} \right)}{\partial(\beta P)} \right)_{NT} = -V e^{-\beta U_\nu - \beta PV} \qquad (1.50)$$

Applying to Eq. 1.49:

$$\langle V \rangle = -\frac{1}{Z} \sum_\nu \left(-\frac{\partial \left(e^{-\beta U_\nu - \beta PV} \right)}{\partial(\beta P)} \right)_{NT} =$$
$$-\frac{1}{Z} \left(\frac{\partial}{\partial(\beta P)} \sum_\nu e^{-\beta U_\nu - \beta PV} \right)_{NT} \qquad (1.51)$$

Applying Eq. 1.48:

$$\langle V \rangle = -\frac{1}{Z} \left(\frac{\partial Z}{\partial(\beta P)} \right)_{NT} = -\left(\frac{\partial(\ln Z)}{\partial(\beta P)} \right)_{NT} \qquad (1.52)$$

It can be demonstrated that Z is related to the Gibbs free energy:

$$G = H - TS = E + PV - TS \quad \Rightarrow$$
$$\beta G = \beta E + \beta PV - \beta TS = \beta E + \beta PV - \frac{1}{k_B} S \qquad (1.53)$$

Taking derivatives with respect to βP and equating the macroscopic

ensemble to the microscopic level:

$$\left(\frac{\partial(\beta G)}{\partial(\beta P)}\right)_{NT} = \left(\frac{\partial(\beta E)}{\partial(\beta P)}\right)_{NT} + \left(\frac{\partial(\beta PV)}{\partial(\beta P)}\right)_{NT} - \\ \left[\frac{\partial}{\partial(\beta P)}\left(\frac{1}{k_{\mathrm{B}}}S\right)\right]_{NT} = V = \langle V \rangle \tag{1.54}$$

Bringing together Eqs. 1.52 and 1.54:

$$\langle V \rangle = -\left(\frac{\partial(\ln Z)}{\partial(\beta P)}\right)_{NT} = \left(\frac{\partial(\beta G)}{\partial(\beta P)}\right)_{NT} \quad \Rightarrow \\ \ln Z = -\beta G \quad \Rightarrow \\ Z = \mathrm{e}^{-\beta G} \tag{1.55}$$

Differently from geometrical calculations, energy calculations are standardised to a great extent. The simplest approach is applied to frequent events. The idea is to perform a simulation, assuming that the event of interest will take place a reasonable number of times during the entire length of the simulation. From Eq. 1.55, an NPT system, as used in this work, evolving along a certain trajectory λ, is described by:

$$Z(\lambda) = \mathrm{e}^{-\beta G(\lambda)} \tag{1.56}$$

If the trajectory λ is not known, it can be described in terms of an order parameters χ, whose value can be computed from the trajectory a posteriori:

$$Z(\chi) = \mathrm{e}^{-\beta G(\chi)} \tag{1.57}$$

The total partition function is the sum over all possible values of χ:

$$Z = \sum_{\chi} Z(\chi) = \sum_{\chi} \mathrm{e}^{-\beta G(\chi)} \tag{1.58}$$

Once the partition functions are defined, it is possible to calculate the probability of any state χ_i as the ratio between its partition function and the total partition function:

$$P(\chi_i) = \frac{Z(\chi_i)}{Z} = \frac{\mathrm{e}^{-\beta G(\chi_i)}}{\sum_{\chi} \mathrm{e}^{-\beta G(\chi)}} \tag{1.59}$$

And the relative probability of two states χ_i and χ_j is:

$$\frac{P(\chi_i)}{P(\chi_j)} = \frac{\dfrac{e^{-\beta G(\chi_i)}}{\sum\limits_{\chi} e^{-\beta G(\chi)}}}{\dfrac{e^{-\beta G(\chi_j)}}{\sum\limits_{\chi} e^{-\beta G(\chi)}}} = \frac{e^{-\beta G(\chi_i)}}{e^{-\beta G(\chi_j)}} = \tag{1.60}$$

$$e^{-\beta[G(\chi_i)-G(\chi_j)]} = e^{-\beta \Delta G(\chi_{ij})}$$

And finally:

$$\Delta G(\chi_{ij}) = -\frac{1}{\beta} \ln \frac{P(\chi_i)}{P(\chi_j)} \tag{1.61}$$

Taking the most likely state as the reference value to determine Gibbs free energy differences:[61]

$$\Delta G(\chi_i) = -\frac{1}{\beta} \ln \frac{P(\chi_i)}{P(\chi_{\max})} \tag{1.62}$$

This equation allows to extract energy values directly from the likelihood of each state, which can be measured from the simulation. For a given event of interest, it is enough to determine how likely is it in the simulation, and then convert the probability to an energy term. In practice, the best approach is to obtain the probabilities of all related events (for instance, molecular reorientation, torsion angle rotation, formation of H bonds, etc.) and create a histogram, from which an energy curve can be directly computed.

The scheme for energy calculations outlined above requires a large sampling of all possible states, which in turn requires the events of interest to be as likely as any other event. Unfortunately, the events of interest can be very unlikely, so called *rare events*, and might not take place a significant number of times during the short time scale which is feasible in an MD simulation. For these situations, it is possible to use steered MD simulations. The idea behind a steered simulation is to input certain energy constraints to the system to force rare events to occur. Knowing the energy of the system during the trajectory and the extra energy input as a constraint, it is possible to determine the energy of the event.

Two very common schemes for steered MD are umbrella sampling and metadynamics.

Umbrella sampling[59,62] works under the assumption that the trajectory of the event is known, at least approximately. The event is defined in terms of an order parameter, and potential wells are applied to keep this order parameter evolving along the trajectory, effectively forcing the system to reproduce the event. By discounting the energy of the constraints from the total energy of the trajectory it is possible to know the energy landscape of the event. In practice, this is done by weighted histogram analysis method (WHAM).[63] Umbrella sampling provides a fast way to measure energy curves, but its major drawback is that it is necessary to know in advance the trajectory of the event.

For those events whose trajectory is unknown, metadynamics[64] offers a good alternative. In this scheme, an order parameter related to the event of interest is chosen an forced to map the entire energy landscape. For this, Gaussian hills are applied to the order parameter, forcing it to change and explore all possible states. Analogously to umbrella sampling, knowing the amount of energy input as Gaussian hills and the energy of the simulated trajectory, it is possible to retrieve the energy associated only to the order parameter. The main disadvantage of metadynamics is that it requires long simulation times.

1.2.4 Addressing the validity of molecular dynamics simulations

MD simulations are based on theoretical models, and as such, the validity of predicted trajectories should be compared with experimental results when feasible. Due to the very different nature of computational and experimental observations, comparison of both is far from being trivial,[65,66] especially considering that computational techniques are often used to access information which is not available experimentally. Experience has shown that upon a careful parametrisation, MD simulations can accurately reproduce crystal structures to the level of anisotropic displacement parameters.[67,68] Nevertheless, unless all relevant parameters are fine tuned, simulated results are better used in a qualitative man-

ner,[67] as most force fields had proven to reproduce the overall behaviour of molecular systems reasonably well.[69–71] Free energy calculations are particularly sensitive to the choice of force field and parameters, and results are difficult to validate experimentally.[72] In this work, Gibbs free energy differences determined from MD simulations are reported, mainly as a guidance, and discussion of results relies principally on overall shape of the free energy surfaces (FESs).

A common quick validation procedure, after ensuring that bond lengths and angles are properly reproduced, consists in checking how simulated unit-cell parameters compare with experimental ones. Unfortunately, most force fields oriented towards simulation of organic molecules have not been consistently tested at the pressures used in this work—up to 1.2 GPa. Instabilities arising from the use of extreme external conditions renders the comparison of unit-cell dimensions unreliable, a topic commonly addressed in the discussion forums of the different MD motors. For instance, at high pressure local densities might increase more than expected from experimental observations, creating *vacuum* regions in the model. Subsequently, the simulation box will shrink suddenly to eliminate such regions. This problem has been dealt with in the present work by minimising the energy of all molecular models under simulation conditions prior to the production runs, not to reach a real energy minimum, but to overcome the aforementioned instabilities.[67] As a result, at the start of the production runs unit-cell parameters of the model might be significantly different from the ones measured by X-ray diffraction. Nevertheless, molecular conformations and intermolecular distances at this stage are reproduced reasonably well.

References

[1] C. E. Weir, E. R. Lippincott, A. van Valkenburg and E. N. Bunting, *J. Res. Nat. Bur. Stand.*, 1959, **63A**, 55–62.

[2] P. W. Bridgman and I. Simon, *J. Appl. Phys.*, 1953, **24**, 405–413.

[3] G. J. Piermarini, *J. Res. Natl. Inst. Stand. Technol.*, 2001, **106**, 889–920.

[4] L. Merrill and W. A. Bassett, *Rev. Sci. Instrum.*, 1974, **45**, 290–294.

[5] R. M. Hazen and L. W. Finger, *Carnegie Inst. Washington Yearb.*, 1977, **76**, 655–656.

[6] W. A. Bassett, *High Pressure Res.*, 2009, **29**, 163–186.

[7] R. Boehler and K. de Hantsetters, *High Pressure Res.*, 2004, **24**, 391–396.

[8] A. Katrusiak, *Acta Cryst.*, 2008, **A64**, 135–148.

[9] H. Ahsbahs, *Z. Kristallogr.*, 2004, **219**, 305–308.

[10] D. R. Allan, R. Miletich and R. J. Angel, *Rev. Sci. Instrum.*, 1996, **67**, 840–842.

[11] R. Boehler, *Rev. Sci. Instrum.*, 2006, **77**, 2006–2008.

[12] T. Kenichi and N. Satoshi, *Rev. Sci. Instrum.*, 2003, **74**, 3017–3020.

[13] D. J. Dunstan, *Rev. Sci. Instrum.*, 1989, **60**, 3789–3795.

[14] G. J. Piermarini, S. Block, J. D. Barnett and R. A. Forman, *J. Appl. Phys.*, 1975, **46**, 2774–2780.

[15] H. K. Mao, J. Xu and P. M. Bell, *J. Geophys. Res.*, 1986, **91**, 4673–4676.

[16] R. Miletich, D. R. Allan and W. F. Kuhs, *Reviews Miner. Geochem.*, 2000, **41**, 445–519.

[17] P. M. Oger, I. Daniel and A. Picard, *BBA Proteins Proteom.*, 2006, **1764**, 434–442.

[18] D. E. Graf, R. L. Stillwell, K. M. Purcell and S. W. Tozer, *High Pressure Res.*, 2011, **31**, 533–543.

[19] C. A. Cameron, *PhD thesis*, University of Edinburgh, 2014.

[20] J. M. Besson, R. J. Nelmes, G. Hamel, J. S. Loveday, G. Weill and S. Hull, *Physica B*, 1992, **180-181**, 907–910.

[21] J. Binns, K. V. Kamenev, G. J. McIntyre, S. A. Moggach and S. Parsons, *IUCrJ*, 2016, **3**, 1–12.

[22] E. V. Boldyreva, *Acta Cryst.*, 2008, **A64**, 218–231.

[23] M. A. Neumann, J. van de Streek, F. P. A. Fabbiani, P. Hidber and O. Grassmann, *Nat. Commun.*, 2015, **6**, 7793.

[24] S. Saouane and F. P. A. Fabbiani, *Cryst. Growth Des.*, 2015, **15**, 3875–3884.

[25] F. P. A. Fabbiani, G. Buth, D. C. Levendis and A. J. Cruz-Cabeza, *ChemComm*, 2014, **50**, 1817–1819.

[26] F. P. A. Fabbiani, G. Buth, B. Dittrich and H. Sowa, *CrystEngComm*, 2010, **12**, 2541–2550.

[27] F. P. A. Fabbiani, D. C. Levendis, G. Buth, W. F. Kuhs, N. Shanklandd and H. Sowa, *CrystEngComm*, 2010, **12**, 2354–2360.

[28] F. P. A. Fabbiani and C. R. Pulham, *Chem. Soc. Rev.*, 2006, **35**, 932–942.

[29] H. Tomkowiak, A. Olejniczak and A. Katrusiak, *Cryst. Growth Des.*, 2013, **13**, 121–125.

[30] M. Andrzejewski, A. Olejniczak and A. Katrusiak, *Cryst. Growth Des.*, 2011, **11**, 4892–4899.

[31] A. Olejniczak and A. Katrusiak, *Cryst. Growth Des.*, 2011, **11**, 2250–2256.

[32] B. F. Johnston, W. G. Marshall, S. Parsons, A. J. Urquhart and I. D. H. Oswald, *J. Phys. Chem. B*, 2014, **118**, 4044–4051.

[33] A. J. Graham, D. R. Allan, A. Muszkiewicz, C. A. Morrison and S. A. Moggach, *Angew. Chem. Int. Ed.*, 2011, **50**, 11138–11141.

[34] F. P. A. Fabbiani, D. R. Allan, W. I. F. David, A. J. Davidson, A. R. Lennie, S. Parsons, C. R. Pulham, J. E. Warren, R. David, I. F. William, J. Alistair, A. R. Lennie, S. Parsons, C. R. Pulham and J. E. Warren, *Cryst. Growth Des.*, 2007, **7**, 1115–1124.

[35] S. A. Moggach, D. R. Allan, S. J. Clark, M. J. Gutmann, S. Parsons, C. R. Pulham and L. Sawyer, *Acta Cryst.*, 2006, **B62**, 296–309.

[36] S. A. Moggach, W. G. Marshall and S. Parsons, *Acta Cryst.*, 2006, **62**, 815–825.

[37] F. P. A. Fabbiani, D. R. Allan, S. Parsons and C. R. Pulham, *Acta Cryst.*, 2006, **B62**, 826–842.

[38] V. S. Minkov, S. V. Goryainov, E. V. Boldyreva and C. H. Görbitz, *J. Raman Spectrosc.*, 2010, **41**, 1748–1758.

[39] E. V. Boldyreva, *J. Mol. Struct.*, 2003, **647**, 159–179.

[40] E. V. Boldyreva, T. P. Shakhtshneider, H. Ahsbahs, H. Sowa and H. Uchtmann, *J. Therm. Anal. Calorim.*, 2002, **68**, 437–452.

[41] D. Seoung, Y. Lee, C.-C. Kao and T. Vogt, *Chem. Eur. J.*, 2013, **19**, 10876–10883.

[42] Y. Lee, T. Vogt, J. A. Hriljac, J. B. Parise, J. C. Hanson and S. J. Kim, *Nature*, 2002, **420**, 485–489.

[43] R. J. Angel, N. L. Ross, I. G. Wood and P. A. Woods, *Phase Transit.*, 1992, **39**, 13–32.

[44] A. Katrusiak, *Acta Cryst.*, 2004, **A60**, 409–417.

[45] R. J. Angel, *J. Appl. Cryst.*, 2004, **37**, 486–492.

[46] A. Dawson, D. R. Allan, S. Parsons and M. Ruf, *J. Appl. Cryst.*, 2004, **37**, 410–416.

[47] P. Dera and A. Katrusiak, *J. Appl. Cryst.*, 1999, **32**, 510–515.

[48] Bruker-AXS, *SAINT+ Integrating Engine v8.18C*, Bruker-AXS, Madison, Wisconsin, USA, 2011.

[49] *CRYSALIS PRO v171.36.28*, Rigaku Oxford Diffraction, Tokyo, Japan, 2012.

[50] S. Parsons, *SHADE*, The University of Edinburgh, Edinburgh, United Kingdom, 2005.

[51] G. M. Sheldrick, *SADABS v2008-1*, Bruker-AXS, Madison, Wisconsin, USA, 2008.

[52] G. M. Sheldrick, *TWINABS v2008-1*, Bruker-AXS, Madison, Wisconsin, USA, 2008.

[53] R. H. Blessing, *Cryst. Rev.*, 1987, **1**, 3–58.

[54] G. M. Sheldrick, *XPREP v2012-1*, Bruker-AXS, Madison, Wisconsin, USA, 2012.

[55] G. M. Sheldrick, *Acta Cryst.*, 2015, **C71**, 3–8.

[56] B. J. Alder and T. E. Wainwright, *J. Chem. Phys.*, 1959, **31**, 459–466.

[57] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781–1802.

[58] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindah, *SoftwareX*, 2015, **1-2**, 19–25.

[59] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Academic Press, 2001.

[60] M. P. Allen, *Computational Soft Matter: From Synthetic Polymers to Proteins, Lecture Notes*, John von Neumann Institute for Computing, Jülich, 2004, pp. 1–28.

[61] J. G. Kirkwood, *J. Chem. Phys.*, 1935, **3**, 300–313.

[62] G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187–199.

[63] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 1011–1021.

[64] A. Laio and M. Parrinello, *Proc. Nat. Acad. Sci. USA*, 2002, **99**, 12562–12566.

[65] W. F. van Gunsteren and A. E. Mark, *J. Chem. Phys.*, 1998, **108**, 6109–6116.

[66] W. F. van Gunsteren, J. Dolenc and A. E. Mark, *Curr. Opin. Struct. Biol.*, 2008, **18**, 149–153.

[67] A. Nemkevich, H. B. Bürgi, M. A. Spackman and B. Corry, *Phys. Chem. Chem. Phys*, 2010, **12**, 14916–14929.

[68] G. M. Lombardo, G. Portalone, M. Colapietro, A. Rescifina and F. Punzo, *J. Mol. Struct.*, 2011, **994**, 87–96.

[69] O. Guvench, E. Hatcher, R. M. Venable, R. W. Pastor and A. D. MacKerell, *J. Chem. Theory Comput.*, 2009, **5**, 2353–2370.

[70] N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark and W. F. Van Gunsteren, *Eur. Biophys. J.*, 2011, **40**, 843–856.

[71] W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.

[72] N. Hansen and W. F. Van Gunsteren, *J. Chem. Theory Comput.*, 2014, **10**, 2632–2647.

Formation of hydrates at high pressure

2.1 Introduction

CSP is the prediction of the crystal structure of a material based on the chemical diagram of its constituents.[1] Several computer programs have been developed for this task.[2] CSP is the subject of extensive research nowadays, as it provides another approach, complimentary to experimental research, to problems which are of great financial interest for certain industries. For instance, polymorph selection, a common problem in the pharmaceutical industry,[3–5] can be optimised if both in silico and experimental polymorph screening results are available. Similarly, cocrystal and solvate screening will also benefit from accurate CSP. It can also be used to calculate physicochemical properties of materials without the need of crystallisation, with the consequential economic saving. Despite all its advantages, CSP has proven so challenging as to be described as *hopelessly difficult.*[6]

In the field of molecular materials, CSP has proven successful on predicting the structure of crystals containing molecules which are either small or exhibit limited flexibility.[2] Recent results show that certain approaches are also able to tackle the prediction of larger molecules[7] and multicomponent systems, such as solvates and cocrystals.[8] The rate of success on predicting the structure of hydrates is acceptable,[2,9–12] although the problem remains challenging, because the number of candidate structures that need to be evaluated grows exponentially with the number of chemical species in the crystal.

Data mining strategies have been suggested as a way to reduce the number of candidates and optimise the search.[2] Data mining is based on the definition of structural predictors, which are evaluated over large datasets—the Cambridge Structural Database (CSD) for instance—by means of statistics, artificial intelligence and machine learning techniques. The predictors are then used to limit the search space to the most probable structures. Descriptors can be based on statistical distributions of structures[13] or physical properties, like intermolecular distances and geometries,[10,14] packing densities,[10] or space groups.[15] Actual accuracy of CSP based on data mining for solvate formation is around 75%.[13] CSP will benefit from better structural descriptors, which in turn rely on the continuous expansion and study of the CSD.

The biggest challenge of CSP is that it is becoming increasingly successful at computing accurate crystal energy landscapes, without being able to asses whether the lowest energy form will crystallise at all. Phase stability can be assessed by lattice and free energy calculations,[16] which are mainly limited by computational power, while modelling crystallisation is a more complex process which largely relies on kinetics, for which few models are available.[17] With regard to hydrates of organics molecules, substantial exploration of the CSD has been performed, in order to better understand how they form, and expecting that the gained knowledge will help to predict crystal structures along with the likelihood of their crystallisation.

The first surveys of the CSD showed that water tends to be a stronger H-bond acceptor than donor,[18] that H-bonded networks tend to use

all available donors and acceptors,[19] and that formation of H bonds is related to the ratio of donors and acceptors in a given molecule and the weight of the molecule.[20] More recently, a systematic analysis of the geometry of H-bonded water molecules in organic crystals found a mild correlation between their geometry and their chemical environment.[21,22] Although such an observation is not useful for prediction of hydrate formation, it led to a deeper analysis of chemical environments in hydrate structures, revealing that certain functional groups promoted hydration.[23,24] The last work of Infantes et al. on this topic showed that hydrate formation depends on the number of potential contacts for both donors and acceptors—its sum and its difference—in a molecule, rather than a simple donor-acceptor ratio; and on the polarity of the surface of the molecule, rather than its weight.[25] The validity of these hypotheses has been tested with the purpose of prediction of hydrate formation, although it has been found to be less convenient than comparison of excess free energies of mixture between water and the organic molecule.[26]

Van de Streek and Motherwell suggested a different approach, based on the comparison of the molecular volume of a molecule in a hydrate and an anhydrate structure.[27] The study found that, generally, larger organic molecules tend to use water as a space filler, based on the fact that the molecular volume for such molecules is larger in the anhydrate than in the hydrate. The same observation had already been made on the basis of molecular shape and hydration number.[28] Molecular volumes for smaller molecules tend to be larger in the hydrate, meaning that water is incorporated in the structure for its H-bonding capabilities, causing a loss of close packing. Other factors, such as molecular flexibility or number of H-bond donors and acceptors, were also explored, but the authors recognised that none of the measured parameters have robust predictive capabilities.

In this chapter, volumetric differences between hydrates and anhydrates are analysed for hydrates which have been studied at high pressure. Examples of such analysis already exist in the literature,[29–32] but here a new rationale to determine pressure-dependent volumes for water molecules is explored and the study is extended to all high-pressure

hydrates present in the CSD. This study stems from the observation that hydrate formation at high pressure is somewhat counterintuitive: on the one hand high pressure promotes efficient packing arrangements, which minimise molecular volumes, while on the other small molecules in hydrates generally exhibit larger molecular volumes than in the corresponding anhydrates. Analysing these differences can expose trends that can lead to a better understanding of hydrate crystal structures and hence be useful for subsequent CSP studies.

2.2 Experimental

2.2.1 List of structures determined at high pressure

The following terms are widely used in this text and require a clear definition:

- **Determination**: any ambient- or high-pressure (AP or HP) entry in the CSD with no restrictions in temperature.

- **Structural set**: group of entries with the same chemical composition. A structural set contains all determinations of all polymorphs of a substance, regardless of the pressure.

- **Redetermination**: any entry of a given structure having the same chemical composition, irrespective of the temperature and pressure at which it was measured.

- **Main component**: all the chemical entities in the chemical formula of a structure, except for water molecules.

- **Anhydrate-hydrate (A-H) set**: super-group of structures with a common main component and any variable number of water molecules.

A search of the CSD (v5.36 with updates to May 2015) was performed to find all the A-H sets of organic molecular substances which include at least one HP determination.

First, all HP entries were extracted from the CSD. The program CONQUEST (v1.17)[33] was used to find all entries containing information in the pressure field, limiting the search to organic substances and

entries free of errors. This search returns not only HP entries, but also AP ones for which a value of 0.0001 GPa has been given in the pressure field. Those entries were removed. The query field *organic* limits the output to entries which do not contain transition metals, although it includes structures containing alkaline and alkaline earth metals. Additional search criteria were added to remove such structures from the output, because they tend to form aquo-complexes rather than hydrates.

Second, each determination was exchanged by the structural set in which it is contained. The structural formula of the main component of each determination was used as a search query. The resulting list contained, for each substance found in the first step, all the AP and HP determinations of its anhydrate and hydrate forms. Such search also yielded cocrystals, salts, and solvates of the main component, which were removed, unless an A-H set existed for the cocrystal, salt, or solvate. Stereoisomerism plays an important role in crystal packing, hence isomers of the same main component were always considered to form independent structural sets.

Determinations fulfilling any of the following criteria were removed from the list, as they were not suitable for this study:

- Hydration number not reported.

- Deuterated structures. Accounting for isotope effects, which are known to cause volume changes,[34] was out of the scope of this study.

- No pressure value stated, instead simply *determined at high pressure* stated in the pressure field.

- Wrong or uncertain chemical formula, unit-cell volume, number of molecules per unit cell, space group, temperature, or pressure. The correctness of uncertain structures was checked in a case-by-case basis by referring to the original publication.

For the purpose of this study, the presence of deposited 3D coordinates was not a requirement, and a determination was deemed to be acceptable provided its unit-cell volume and number of chemical formulas per unit cell were reported. Nevertheless, structural sets for which no determina-

tion has deposited 3D coordinates were removed, as it was not possible to check for potential errors. Additionally, an A-H set was removed entirely if, after applying the previous criteria, all the determinations of the hydrate forms or all the HP determinations were removed.

The final list contained 1602 entries divided among 194 structural sets. Of those sets, 55 corresponded to the 23 substances which form hydrates (Appx. A). The list of A-H sets can be further divided into three sub-lists (Appx. B):

- **In situ-formed hydrates:** sets in which at least one of the hydrates has been reported to form at high pressure. This list contains 7 sets.

- **Compressed hydrates:** sets in which hydrates existing at ambient pressure have been subject of compression studies. This list contains 6 sets.

- **AP hydrates:** sets for which the anhydrate, but not the hydrates, has been studied under pressure. This list contains 10 sets.

Three other structures needed to be added manually to the list of in situ-formed hydrates: ^{t}BA (reported in Ch. 6), parabanic acid (the pressure of the determinations was not recorded in the CSD, but on the corresponding publication[35]), and 5,6-dimethylbenzimidazole (dMBzIm, published[32] although not available in the version of the CSD used).

2.2.2   Compressibility of lattice water

A new list was compiled from the hydrates, at either AP or HP, contained in the lists of in situ-formed hydrates and compressed hydrates. Molecular volumes were calculated as the volume within the Hirshfeld surface of a molecule (V_H). A Hirshfeld surface defines the region around a molecule inside which the electron density of the molecule exceeds that of any neighbouring molecule.[36] For each determination, V_H of each entity of the chemical formula was computed using CRYSTALEXPLORER.[37] All V_H were normalised to 293 K—the most frequent temperature recorded

for HP determinations—using Hofmann's model:[38]

$$V_H(293) = V_H(T)[1 - \bar{\alpha}(T - 293)] \tag{2.1}$$

Where $V_H(293)$ is the estimated molecular volume at 293 K; $V_H(T)$ is the molecular volume at the temperature T of the determination; and $\bar{\alpha}$ is the average thermal expansion coefficient determined by Hofmann, $0.95 \pm 0.03 \cdot 10^{-4}$ K^{-1}.

V_H were employed as a fast and reliable way to partition the unit-cell volume among all the chemical constituents of the crystal, albeit due to the nature of the Hirshfeld surface partitioning scheme, a small portion of the unit-cell volume remains unattributed (see §2.3.2 for further discussion). The unattributed volume of each determination ($V_{H,ntt}$) was calculated as:

$$V_{H,ntt} = V_Z - \sum V_{H,nwat} - \sum V_{H,wat} \tag{2.2}$$

Where V_Z is the volume of a chemical formula, obtained by dividing the unit-cell volume by Z; $V_{H,nwat}$ is the molecular volume of a molecule other than water; and $V_{H,wat}$ is the molecular volume of a lattice water molecule. $V_{H,ntt}$ was used to estimate the significance level during the analysis of the results (§2.3.2).

V_H of lattice water molecules were plotted in a PV diagram and a Murnaghan equation of state[39] was fitted to the data using EoSFit7:[40]

$$V_H = V_{H0} \left(1 + \frac{B_0'}{B_0}P\right)^{-1/B_0'} \tag{2.3}$$

Where V_H is the molecular volume of a lattice water molecule at a pressure P; V_{H0} is the corresponding volume at ambient pressure; B_0 is the bulk modulus at ambient pressure, in GPa; and B_0' is the pressure derivative of B. Only determinations below 5 GPa were used, as this equation does not properly reproduce the compression of lattice water at higher pressures (§2.3.1). The fitted equation was later used to estimate the volume of lattice water at any pressure.

2.2.3 Pressure-volume diagrams of anhydrate-hydrate sets

For each A-H set, all the possible phases were identified, and each determination assigned to its corresponding phase. For each determination, the molecular volume excluding water (V_{excl}) was calculated. For hydrates, it is calculated as:

$$V_{excl} = V_Z - N V_{H,wat} \tag{2.4}$$

Where N is the number of water molecules in a chemical formula and $V_{H,wat}$ is the molecular volume of a water molecule at the pressure of the determination, obtained from the fitted Murnaghan equation of state. For anhydrates, this volume corresponds simply to V_Z.

When several redeterminations of the same phase at the same pressure existed, a representative one was chosen according to the following criteria, in the stated order:[41]

- If there are four or more redeterminations, V_{excl} of the representative structure must not differ considerably from the median V_{excl} of all the redeterminations. This filters out redeterminations with poorly determined unit-cell parameters. If there are less than four redeterminations, it is not possible to ensure that V_{excl} is free from errors.

- Redeterminations with deposited 3D coordinates are preferred, as they are helpful when checking for errors in the structure. Whenever a phase has a single determination, it is accepted even if no 3D coordinates are available.

- Determinations at room temperature are preferred over those at low temperature, as most HP structures are determined at 293 K. This limits the errors introduced by the temperature normalisation.

- The redetermination with the lowest R factor is preferred.

2.3 Results and discussion

2.3.1 Compressibility of lattice water

Fig. 2.1 shows the compressibility of lattice water molecules in organic crystals, measured by means of Hirshfeld surfaces. Numerical data are provided in Appx. B. A Murnaghan equation (Eq. 2.3) was fitted to the experimental data from the CSD, obtaining the following parameters:

- $V_{H0} = 22.37$ Å^3

- $B_0 = 7.92$ GPa

- $B_0' = 3.91$

Figure 2.1: Compressibility of lattice water molecules in organic crystals, up to 5 GPa. The blue spots correspond to experimental molecular volumes within Hirshfeld surfaces. The red curve is the Murnaghan equation of state fitted to the data. The two green dotted curves correspond to the maximum and minimum uncertainties of the fit at the 2σ level.

The Murnaghan equation of state is used for its simplicity, although it is deemed appropriate only to model small compressions for which $V/V_{H0} > 0.9$.[42] For larger compressions the value of the fitted parameters (V_{H0}, B_0, and B_0') might not be representative of the physical properties of the system. For the compression of lattice water such condition is only fulfilled up to 2 GPa, with $V/V_{H0} \approx 0.75$ at 5 GPa. The aim of this study is to provide a simple, yet reasonably accurate, way of estimating the volume of a lattice water molecule at a certain pressure, for which an equation which provides a good fit to the experimental data is more important than the physical meaning of its parameters. It

has been considered that, from a mere mathematical perspective, the fit up to 5 GPa provides a better representation of the system. If the Murnaghan equation is fitted only to data below 2 GPa, a linear equation provides a considerably better fit. The average residual sums of squares (RSS) are 1.414 and 0.644, for the Murnaghan and linear fit respectively. The use of the Murnaghan equation is discouraged for such data, while the use of a linear fit misses the pressure dependence of the compressibility. A Murnaghan equation of state fitted to all data up to 5 GPa exhibits an average RSS of 0.769, acceptable if compared with that of the linear fit in the smaller pressure range. In addition, it allows to explore a larger range of pressures, a critical factor considering the limited number of available structures of hydrates at high pressure. Data over 5 GPa have not been considered, as there are only a few determinations available and their addition results in a considerable loss of quality in the fit.

In previous studies, the molecular volume of water in the liquid state or in ice VI (below and above 1.0 GPa, respectively) has been used as the volume of lattice water in organic structures.[29–32] Although a good approximation, it has been considered that the use of experimentally measured volumes, by means of Hirshfeld surfaces in this case, provides a more physically meaningful approach.

2.3.2 Assessment of errors

Several different sources of error can be identified in this study. Some of them originate from the structures in the CSD: accuracy of the refined structures, unit-cell errors, and uncertainties in the measurement of pressures. Other errors stem from the use of Hirshfeld surfaces to compute molecular volumes and the equation of state used to estimate the volume of water molecules at different pressures. A comprehensive error model has not been included, as the level of complexity required to address all the different sources combined is not justified given the limited amount of data employed in the study. Instead, each source of error is discussed in this section and a framework in which to analyse the results is set up.

The first source of error are the structures taken from the CSD. Structures are checked and errors are usually addressed upon deposition in the database, although there can be small errors, like misplaced hydrogen atoms or poorly refined disorder, leading to inaccurate geometries, which would affect the calculation of the molecular volumes. The refinement and molecular geometry of each structure used in this study was not checked on a case-by-case basis. Mogul[43] provides a tool that allows for quick comparison of the geometry of a structure against the rest of the CSD, but considerably user input is still required to analyse the results. A completely automated geometry check would be desirable to evaluate large datasets. After filtering potentially problematic structures according to the criteria established in §2.2.3, remaining errors are assumed to be small. Similarly, possible errors in the reported unit-cell parameters and pressure values are also assumed to be small, as there is no possibility to perform independent checks on these.

The second source of error originates from the use of Hirshfeld surfaces. Hirshfeld's model attempts to partition the electron density of a molecule among all its atoms[44] and has later been adapted to partition the electron density of a unit cell among all its molecules.[45] For a given molecule, a Hirshfeld surface defines the region around it inside which the electron density of the molecule exceeds that of any neighbouring molecule. Such a surface guarantees maximum proximity to the neighbouring molecules without an overlap.[36] Because Hirshfeld surfaces are based on the partition of the spherically averaged electron density, rather than the unit-cell volume, they do not achieve a full partition of the volume. It is estimated that about 5% of the unit-cell volume is not attributed to any molecule.[46] As a result, only differences in V_{excl} between the anhydrate and the hydrate which are larger than the unattributed volume can be considered significantly different. For the structures included in this study, the average $V_{H,ntt}$ is 7.01±0.41 Å^3 (at the 95% level of confidence).

An alternative scheme for the partition of the unit-cell volume has been proposed by Hofmann,[38] although its application in this study was not satisfactory. Said model proposes the use of averaged atomic

volumes, determined by directly partitioning the unit-cell volume over all atoms in the structure, across the entire CSD. This approach was used by Van de Streek,[27] to directly compute the packing efficiency of hydrates when compared to anhydrates. However, convergence to the appropriate atomic volumes requires a very large experimental dataset—the volumes obtained by this method for infrequent elements are not accurate. The number of high-pressure structures available in the CSD, for instance, is too small to compute the pressure dependence of atomic volumes. Molecular volumes determined from the partition of the electron density are better suited for small datasets. Some models can produce a complete partition, without empty volume, as proposed in Quantum Theory of Atoms in Molecules,[47] but require charge-density grade refinements,[36,48] which usually are not readily available.

Table 2.1: Value, variance, covariance, and correlation coefficient of the parameters of the Murnaghan equation fitted to the data of the compression of lattice water.

V_0	22.37
B_0	7.92
B_0'	3.91
$Var(V_0)$	0.15
$Var(B_0)$	6.88
$Var(B_0')$	3.20
$Cov(V_0, B_0)$	-0.71
$Cov(V_0, B_0')$	0.31
$Cov(B_0, B_0')$	-4.34
$Corr(V_0, B_0)$	-0.71
$Corr(V_0, B_0')$	0.46
$Corr(B_0, B_0')$	-0.92

The third source of error corresponds to the estimation of the volume of the water molecules. The uncertainty on the estimated volumes, $Var(V)$, is determined by error propagation on the Murnaghan equation (Eq. 2.3):

$$Var(V_H) = \left(\frac{\partial V_H}{\partial V_{H0}}\right)^2 Var(V_{H0}) + \left(\frac{\partial V_H}{\partial B_0}\right)^2 Var(B_0) +$$
$$\left(\frac{\partial V_H}{\partial B_0'}\right)^2 Var(B_0') + \frac{\partial V_H}{\partial V_{H0}}\frac{\partial V_H}{\partial B_0}Cov(V_{H0}, B_0) + \quad (2.5)$$
$$\frac{\partial V_H}{\partial V_{H0}}\frac{\partial V_H}{\partial B_0'}Cov(V_{H0}, B_0') + \frac{\partial V_H}{\partial B_0}\frac{\partial V_H}{\partial B_0'}Cov(B_0, B_0')$$

Where $Var(V_{H0})$, $Var(B_0)$, $Var(B_0')$, $Cov(V_{H0})$, $Cov(B_0)$, and $Cov(B_0')$ are the variances and covariances of the three parameters fitted by the equation (Table 2.1). The correlations between these parameters where measured as Pearson product-moment correlation coefficients (Eq. 2.6). Said correlations are -0.71 for the pair V_{H0}, B_0; 0.46 for V_{H0}, B_0'; and -0.92 for B_0, B_0'. With such large correlations, the use of covariances in the error propagation cannot be neglected. The partial derivatives needed for the propagation of uncertainties are as follows:

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} \quad (2.6)$$

$$\frac{\partial V_H}{\partial V_{H0}} = \left(1 + \frac{B_0'}{B_0}P\right)^{-1/B_0'} \quad (2.7)$$

$$\frac{\partial V_H}{\partial B_0} = \frac{V_{H0}P}{B_0^2} \left(1 + \frac{B_0'}{B_0}P\right)^{(-1/B_0')-1} \tag{2.8}$$

$$\frac{\partial V_H}{\partial B_0'} = -\frac{V_{H0}P}{B_0 B_0'} \left(1 + \frac{B_0'}{B_0}P\right)^{(-1/B_0')-1} +$$
$$\frac{V_{H0}}{B_0'^2} \left(1 + \frac{B_0'}{B_0}P\right)^{-1/B_0'} \ln\left(1 + \frac{B_0'}{B_0}P\right) \tag{2.9}$$

Combining Eqs. 2.7, 2.8, and 2.9 in Eq. 2.5 the uncertainties can be computed in all the pressure range. Said uncertainties, at the 2σ level (95% of confidence), are plotted in Fig. 2.1 as dotted green lines. The maximum uncertainty in all the pressure range is less than 1 Å^3, hence the error originating from the volume estimation can be neglected if compared with that coming from the use of Hirshfeld surfaces.

The estimation of the volume of water molecules could be considered unnecessary, as such volumes can be measured directly for each crystal structure, avoiding the uncertainty associated with any fitting. However, there are three main reasons for proceeding with an estimation:

- The use of estimated volumes from a fit to 75 experimental observations is likely to compensate for isolated errors in certain structures, which may arise if a single experimental volume is used.

- V_H calculated with CRYSTALEXPLORER have no uncertainty associated with them. That is compensated by the use of $V_{H,ntt}$ and the uncertainty of the estimation from the fit to the Murnaghan equation.

- An equation of state for the compressibility of water in molecular crystals is interesting for its predictive power. This may find applications in CSP, for instance in the calibration of the intra- and intermolecular potentials. This study also serves as a test of the validity of such an equation of state.

2.3.3 Pressure-volume diagrams of anhydrate-hydrate sets

Fig. 2.2 contains the PV diagrams for all in situ-formed and compressed hydrates. Numerical data corresponding to these diagrams are provided

Figure 2.2: Comparison between the molecular volumes, measured as V_H, of the anhydrate forms (in orange, full icons, different polymorphs labelled in Roman numerals or Greek letters, as defined in the original literature) and the hydrate forms (in blue, empty icons, labelled by their hydration numbers) of different materials. For structures belonging to a compression study, volumes at different pressures are linked by a solid line. For hydrates, V_Z and V_{excl} are plotted (above and below, respectively), linked by vertical dotted lines.
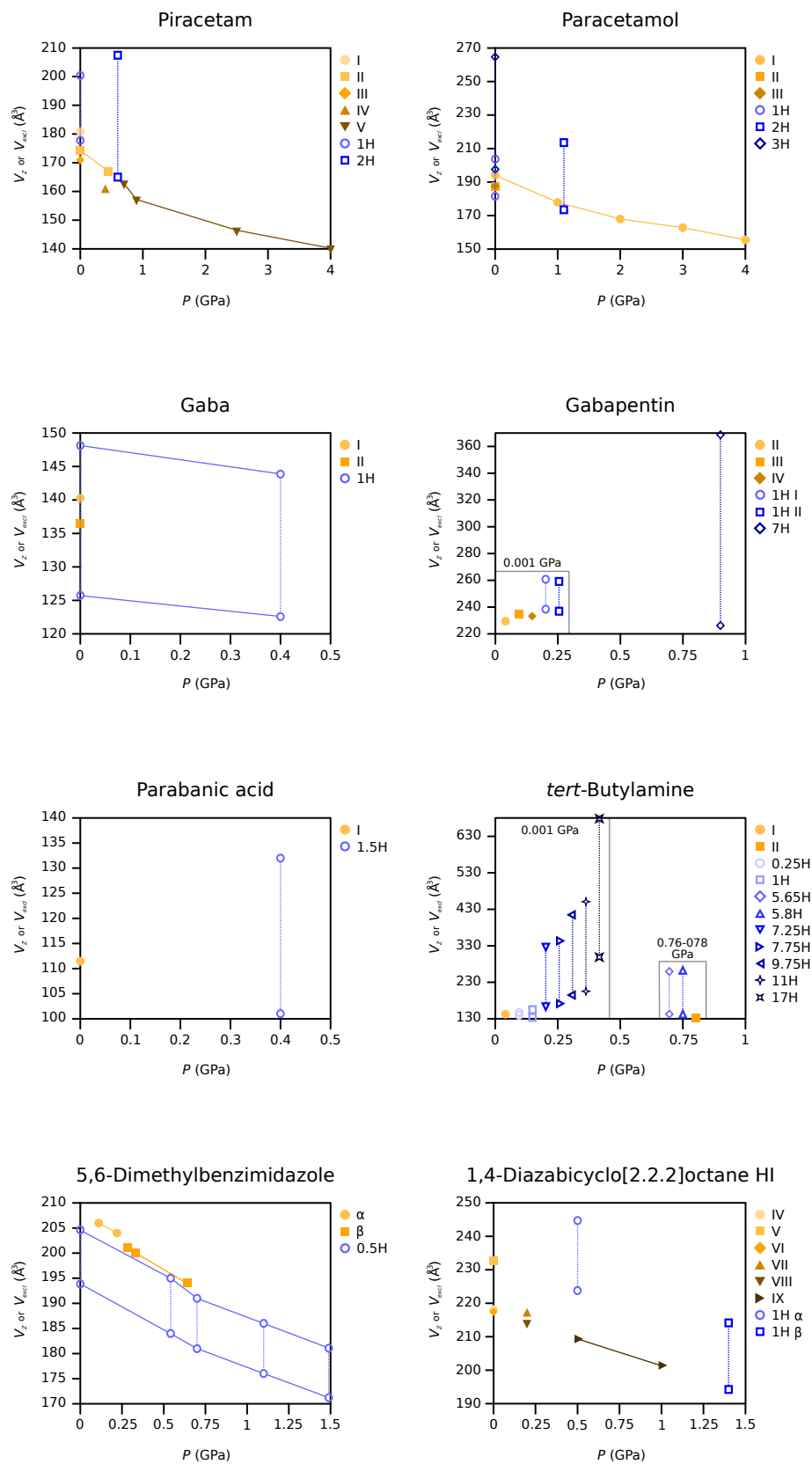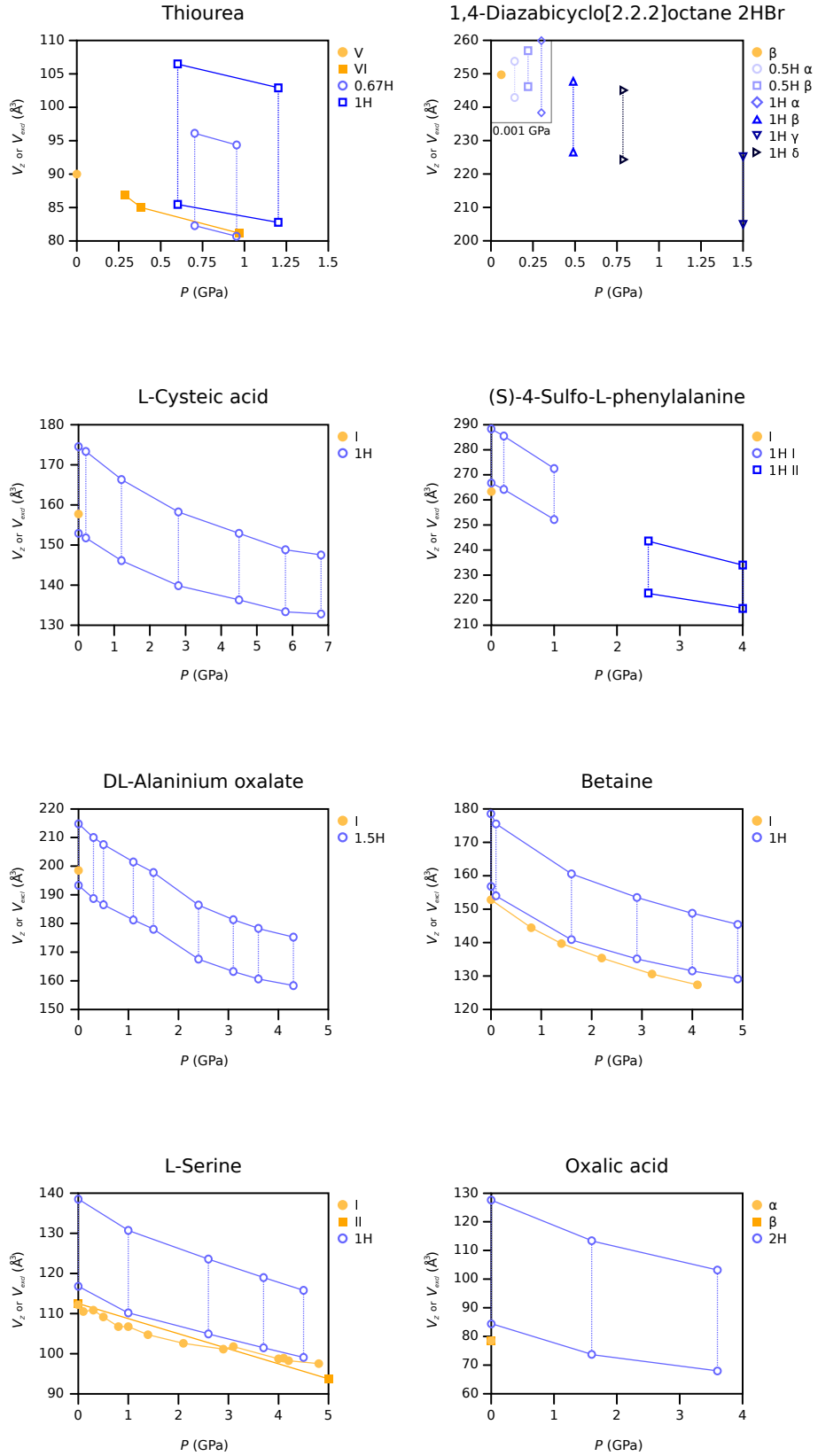
in Appx. C, were data from the group of AP hydrates are also provided. For hydrates, V_Z and V_{excl} are plotted, although only V_{excl} should be compared with V_Z of the anhydrate. V_Z of the hydrates are given to provide a reference for the total volume of each substance.

The analysis of the plots yields two general remarks:

- V_Z is generally larger for hydrates than for anhydrates. This was an expected result, as V_Z of a hydrate includes the volume of the water molecules.

- For the majority of structures, both in situ-formed and compressed, V_{excl} of a hydrate tends to be larger than V_Z of an anhydrate, although differences are usually smaller than the significance level established in the previous section, 7.01 ± 0.41 Å^3. The introduction of additional H bonds involving water molecules in the hydrate promotes a loss of close packing, which increases V_{excl},[27] albeit this loss cannot be too large, otherwise the structure would probably not crystallise at high pressure in the first place.

The only case in this study in which both V_Z and V_{excl} of the hydrate are smaller than V_Z of the anhydrate is dMBzIm.[32] The hydrate of this compounds can be crystallised only at high pressure (although it can be recovered to ambient pressure and kept in such conditions for long periods of time). Zielinski et al. argue that this is not a case of special behaviour of the hydrate, but rather of the anhydrate, which exhibits a loosely packed structure. In this case, water can be incorporated into the structure in already-available space, resulting in a more efficient packing.

The opposite effect—V_{excl} significantly larger for the hydrate at high pressure—is observed for 1,4-diazabicyclo[2.2.2]octane (dabcoHI) and ^{t}BA. In the case of dabcoHI,[30] V_{excl} for the α monohydrate is 14.4 Å^3 larger than that of phase IX at the same pressure. With such difference, the anhydrate may be expected to be the more favoured structure at high pressure. Two alternatives exist to explain the formation of the hydrate. First, the volume reduction still takes place during the experiment, although in the PTM (aqueous solution in this case) rather than in the crystal itself. This consideration may be generally relevant for materials which can form very dense solutions. Densities inside the DAC were

not measured in the study. Second, the hydrate, unfavoured in terms of volume, is considerably more stable from an energy point of view, e.g. through the formation of a more efficient H-bonded network promoted at high pressure. According to the authors the latter is the most likely explanation, although no evidence was provided. Furthermore, kinetic effects can also play a role in the formation of the hydrate, but were not investigated

^{t}BA (Ch. 6) is the only clathrate among the structures considered in this study. The 17-hydrate, obtained at ambient pressure and low temperature, exhibits a V_{excl} of 299.1 Å^3, twice larger than that of the anhydrate phase I. This is unsurprising, as all the small cages in the clathrate are empty, and the resulting space available is distributed over the guest molecules according to the partition scheme used herein (§2.3.2). The rest of the ambient-pressure ^{t}BA clathrates (7.25-, 7.75-, 9.75-, and 11-hydrate) behave in a similar manner. In contrast, V_{excl} for the two high-pressure clathrates (5.65- and 5.8-hydrate) are only 12.2 and 9.6 Å^3 larger, respectively, than the volume of the anhydrate phase II, crystallised at a similar pressure. Although the difference in volume is larger than that found in most other examples, it is very small compared to the AP clathrates, exemplifying how efficient packing is enforced at high pressures.

For the remainder of the structures analysed, V_{excl} of the hydrate is either smaller or not significantly larger than V_Z of the anhydrate at similar pressures. Other interesting features might be identifiable; however, a larger number of structures would be necessary to attempt further analysis.

2.4 Conclusions and outlook

Crystallisation is a complex process that is influenced by many different factors. Packing efficiency and energy stabilisation are two of such factors which are commonly used to systematise crystallisation outcome. Except for materials with a substantial amount of void space in their anhydrous crystal structure, hydration tends to promote looser packings,

which can be compensated with the formation of stabilising H bonds between organic molecules and lattice water. This study is focussed on the formation of hydrates at high pressure, conditions under which packing efficiency gains more importance, and proposes an analysis based on molecular volumes.

An equation of state for the compressibility of lattice water has been determined. Previous comparisons of the volumes of organic molecules in different hydration states have resorted to liquid water or ice VI to estimate the molecular volume of lattice water. In spite of the uncertainty associated to the fit, the approach proposed here is advantageous, as it is more likely to reflect the behaviour of water in crystals of organic structures.

The number of structures of hydrates at high pressure is still very limited, but most of the examples show that the volume of a formula unit in a crystal—after removing the volume occupied by water molecules—is smaller or not significantly larger than the same volume in an anhydrate phase at the same pressure. This may indicate that energy contributions can easily compensate any loss in packing efficiency caused by addition of water in the structure. Only three exceptions to this behaviour have been found. Both ^{t}BA and dabcoHI exhibit molecular volumes in the hydrate phases larger than in the anhydrate one. This effect can be easily explained in the case of ^{t}BA based on the formation of hydrate clathrates. The behaviour of dabcoHI poses a bigger challenge, and as for now there is not a satisfactory explanation for the formation of the HP hydrate. The opposite effect is observed in dMBzIm, where the hydrate is smaller than the anhydrate, even before removing the volume of water molecules. This is explained by a rather loose packing of the anhydrate.

Based on the results of this study, it would be desirable that all reports of high-pressure crystal structures include compression studies and molecular volumes measured by means of Hirshfeld surfaces, as to facilitate comparative studies. More experimental data will allow for a more comprehensive study, which in turn is likely to reveal more trends of hydrate formation at high pressure, from which structural chemistry and CSP can greatly benefit. Notwithstanding, volume considerations,

although very important at high pressure, are only one of many contributions to hydrate formation. Further studies, considering lattice energies and kinetic effects, will also be needed to deepen our understanding of hydrate formation under high-pressure conditions.

References

[1] S. L. Price, *Chem. Soc. Rev.*, 2014, **43**, 2098–2111.

[2] D. A. Bardwell, C. S. Adjiman, Y. A. Arnautova, E. Bartashevich, S. X. M. Boerrigter, D. E. Braun, A. J. Cruz-Cabeza, G. M. Day, R. G. Della Valle, G. R. Desiraju, B. P. Van Eijck, J. C. Facelli, M. B. Ferraro, D. Grillo, M. Habgood, D. W. M. Hofmann, F. Hofmann, K. V. J. Jose, P. G. Karamertzanis, A. V. Kazantsev, J. Kendrick, L. N. Kuleshova, F. J. J. Leusen, A. V. Maleev, A. J. Misquitta, S. Mohamed, R. J. Needs, M. A. Neumann, D. Nikylov, A. M. Orendt, R. Pal, C. C. Pantelides, C. J. Pickard, L. S. Price, S. L. Price, H. A. Scheraga, J. van de Streek, T. S. Thakur, S. Tiwari, E. Venuti and I. K. Zhitkov, *Acta Cryst.*, 2011, **B67**, 535–551.

[3] J. Bernstein, *Polymorphism in Molecular Crystals*, Oxford Univ. Press, 2007.

[4] A. Y. Lee, D. Erdemir and A. S. Myerson, *Annu. Rev. Chem. Biomol. Eng.*, 2011, **2**, 259–280.

[5] A. J. Cruz-Cabeza, S. M. Reutzel-Edens and J. Bernstein, *Chem. Soc. Rev.*, 2015, **44**, 8619–8635.

[6] T. S. Thakur, R. Dubey and G. R. Desiraju, *Annu. Rev. Phys. Chem.*, 2015, **66**, 21–42.

[7] M. A. Neumann, J. van de Streek, F. P. A. Fabbiani, P. Hidber and O. Grassmann, *Nat. Commun.*, 2015, **6**, 7793.

[8] E. Gibney, *Nature*, 2015, **527**, 20–21.

[9] B. P. van Eijck and J. Kroon, *Acta Cryst.*, 2000, **B56**, 745.

[10] A. T. Hulme and S. L. Price, *J. Chem. Theory Comput.*, 2007, **3**, 1597–1608.

[11] K. M. Anderson, G. M. Day, M. J. Paterson, P. Byrne, N. Clarke and J. W. Steed, *Angew. Chem. Int. Ed.*, 2008, **47**, 1058–1062.

[12] D. E. Braun, P. G. Karamertzanis and S. L. Price, *ChemComm*, 2011, **47**, 5443–5445.

[13] K. Takieddin, Y. Z. Khimyak and L. Fábián, *Cryst. Growth Des.*, 2015, **16**, 70–81.

[14] D. W. M. Hofmann and J. Apostolakis, *J. Mol. Struct.*, 2003, **647**, 17–39.

[15] A. J. Cruz Cabeza, E. Pidcock, G. M. Day, W. D. S. Motherwell and W. Jones, *CrystEngComm*, 2007, **9**, 556–560.

[16] J. Nyman and G. M. Day, *CrystEngComm*, 2015, **17**, 5154–5165.

[17] J. D. Dunitz, *ChemComm*, 2003, 545–548.

[18] G. A. Jeffrey and H. Maluszynska, *Acta Cryst.*, 1990, **B46**, 546–549.

[19] M. C. Etter, *Acc. Chem. Res.*, 1990, **23**, 120–126.

[20] G. R. Desiraju, *J. Chem. Soc., Chem. Commun.*, 1991, **6**, 426–428.

[21] L. Infantes and S. Motherwell, *CrystEngComm*, 2002, **4**, 454–461.

[22] A. L. Gillon, N. Feeder, R. J. Davey and R. Storey, *Cryst. Growth Des.*, 2003, **3**, 663–673.

[23] L. Infantes, J. Chisholm and S. Motherwell, *CrystEngComm*, 2003, **5**, 480–486.

[24] L. Infantes and W. D. S. Motherwell, *ChemComm*, 2004, 1166–1167.

[25] L. Infantes, L. Fábián and W. D. S. Motherwell, *CrystEngComm*, 2007, **9**, 65–71.

[26] Y. A. Abramov, *CrystEngComm*, 2015, **17**, 5216–5224.

[27] J. van de Streek and S. Motherwell, *CrystEngComm*, 2007, **9**, 55–64.

[28] C. H. Görbitz and H.-P. Hersleth, *Acta Cryst.*, 2000, **B56**, 526–534.

[29] M. Andrzejewski, A. Olejniczak and A. Katrusiak, *Cryst. Growth Des.*, 2011, **11**, 4892–4899.

[30] A. Olejniczak and A. Katrusiak, *Cryst. Growth Des.*, 2011, **11**, 2250–2256.

[31] H. Tomkowiak, A. Olejniczak and A. Katrusiak, *Cryst. Growth Des.*, 2013, **13**, 121–125.

[32] W. Zieliński and A. Katrusiak, *CrystEngComm*, 2015, **17**, 5468–5473.

[33] I. J. Bruno, J. C. Cole, P. R. Edgington, M. Kessler, C. F. Macrae, P. McCabe, J. Pearson and R. Taylor, *Acta Cryst.*, 2002, **B58**, 389–397.

[34] V. S. Kogan, *Sov. Phys. Usp.*, 1963, **5**, 951–975.

[35] F. P. A. Fabbiani, D. R. Allan, W. G. Marshall, S. Parsons, C. R. Pulham and R. I. Smith, *J. Cryst. Growth*, 2005, **275**, 185–192.

[36] M. A. Spackman and D. Jayatilaka, *CrystEngComm*, 2009, **11**, 19–32.

[37] S. Wolff, D. Grimwood, J. McKinnon, M. Turner, D. Jayatilaka and M. Spackman, *CRYSTAL EXPLORER v3.1*, University of Western Australia, 2012.

[38] D. W. M. Hofmann, *Acta Cryst.*, 2002, **B57**, 489–493.

[39] F. Murnaghan, *Am. J. Math.*, 1937, **49**, 235–260.

[40] R. J. Angel, J. Gonzalez-Platas and M. Alvaro, *Z. Kristallogr.*, 2014, **229**, 405–419.

[41] J. van de Streek, *Acta Cryst.*, 2006, **B62**, 567–579.

[42] R. J. Angel, *High-Pressure Crystallography*, Springer Science & Bussiness Media, 2004, ch. 1.2, pp. 21–36.

[43] I. J. Bruno, J. C. Cole, M. Kessler, J. Luo, W. D. S. Motherwell, L. H. Purkis, B. R. Smith, R. Taylor, R. I. Cooper, S. E. Harris and A. G. Orpen, *J. Chem. Inf. Comput. Sci.*, 2004, **44**, 2133–2144.

[44] F. L. Hirshfeld, *Theor. Chim. Acta*, 1977, **44**, 129–138.

[45] M. A. Spackman and P. G. Byrom, *Chem. Phys. Lett.*, 1997, **267**, 215–220.

[46] J. J. McKinnon, A. S. Mitchell and M. A. Spackman, *Chem. Eur. J.*, 1998, **4**, 2136–2141.

[47] R. Bader, *Atoms in Molecules – A Quantum Theory*, Oxford Univ. Press, 1990.

[48] C. Gatti, V. R. Saunders and C. Roetti, *J. Chem. Phys.*, 1994, **101**, 10686–10696.

Hydrates of β-cyclodextrin at high pressure

3.1 Introduction

At ambient pressure, β-CD (Fig. 3.1) hydrates incorporate a variable and non-stoichiometric number of water molecules, ranging from 9.4 to 12.3. Despite their variable water content, all hydrates belong to the same polymorph, named form I in this work. The lattice parameters were first reported by Hamilton et al. in 1968[1] to belong to space group $P2_1$, $a = 10.31(2)$ Å, $b = 20.86(1)$ Å, $c = 15.2(09)$ Å, and $\beta = 109.29(15)°$. The hydration number was estimated at twelve water molecules by density measurements. These observations were subsequently confirmed by X-ray single-crystal diffraction;[2,3] the refined structure exhibited a large degree of disorder in the water molecules, and clearly indicated that the β-CD molecule adopts a *round* shape. Such observations evidenced that different CDs, namely α- and β-CD, behaved rather differently in the solid state. While α-CD was known to adopt a round shape when forming

inclusion complexes and a slightly distorted one in the hydrated state;[4] β-CD was consistently characterised as round in crystals of either the hydrate or any inclusion compound.[5,6] This difference suggested that complexation could take place via different mechanisms depending on the size of the CD.[3] Water molecules in β-CD form I, particularly those located inside the cavity, were not only disordered, but also deemed to form an inefficient H-bonded network. A similar situation is encountered in hydrates of α-CD, albeit it is partially compensated by the conformational change of the macrocycle. This is one of the reasons why CDs are such efficient hosts in the formation of inclusion complexes; they have a natural tendency to displace water from the interior of the cavity (in favour of another guest) to the intermolecular spaces, where it can contribute towards more efficient H bonds.[2]

Figure 3.1: Chemical diagram of β-CD (right) and an isolated glucose unit (left). Each atom is given a label m_n, indicating that it is the mth atom of the nth glucose unit.
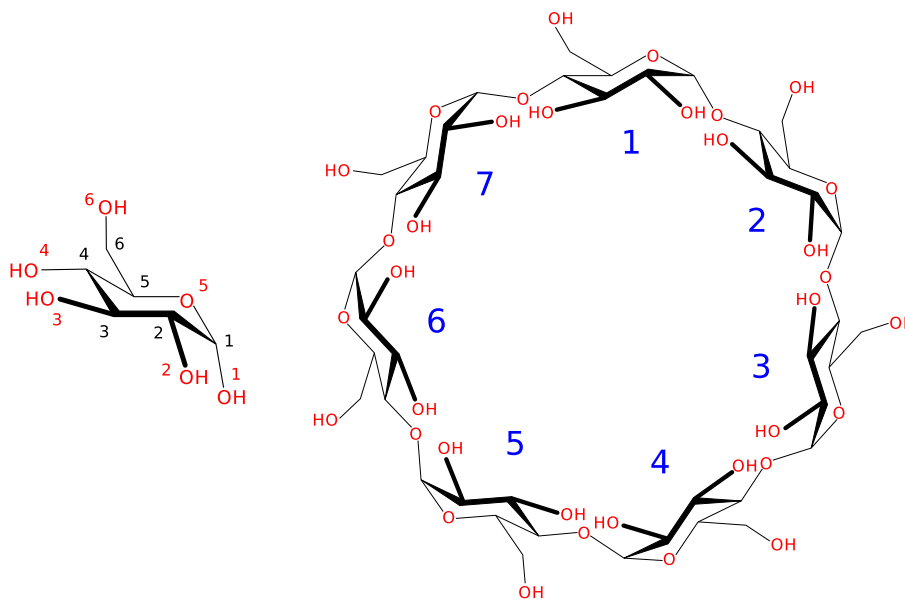
To further explore the behaviour of β-CD form I, a hydration study was carried out by Steiner and Koellner.[7] It was found that β-CD undergoes a reversible dehydration process—from 12.3 to 9.4 water molecules, without a significant change in lattice constants—in response to changes in the humidity of the environment surrounding the crystal. The structure of the hydrate is based on a herringbone-type packing of β-CD molecules (Fig. 3.2), which lacks of channels that could allow water to circulate through the crystal. Based on packing consideration and

neutron-diffraction observations,[8] Steiner and coworkers postulated that the hydration/dehydration process could take place through an H bond flip-flop mechanism. In such mechanism, H-bonds flip between two configurations, causing a concerted chain reorganisation across the crystal able to alternatively open and close certain regions of the structure in a fast manner, moving water molecules along.
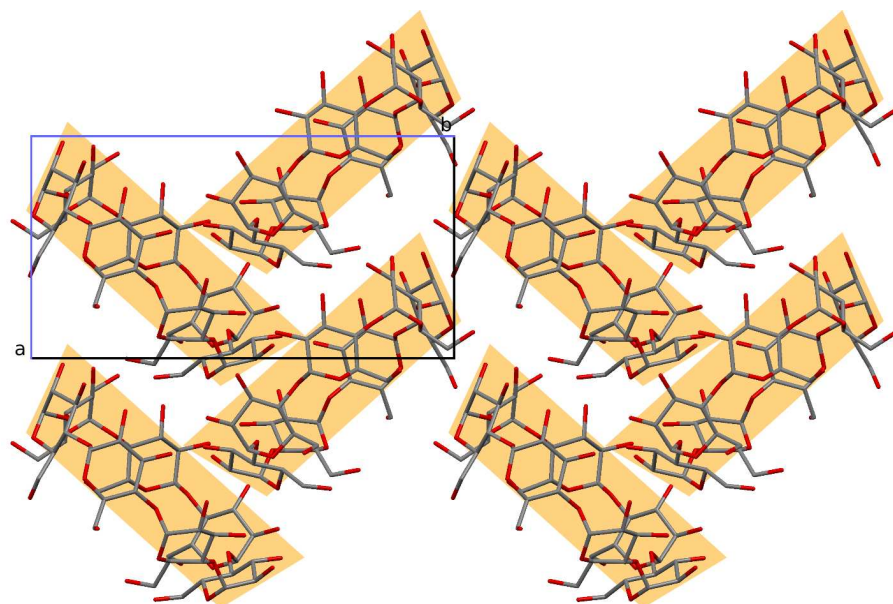
Figure 3.2: Herringbone packing of β-CD form I. Projection perpendicular to the c axis. Trapeziums are used as a simplified way of representing the outline of projected β-CD molecules and to show the packing arrangement more clearly. H atoms, water molecules, and disorder have been omitted for clarity.

In this chapter, two new hydrates of β-CD, named forms II and III, crystallised in situ at high pressure, are reported. Both forms exhibit structural parameters and packing types that have already been described in the literature, although they are substantially different from the classic herringbone motif associated with hydrates and complexes incorporating small guests.[9] The change in packing type can be linked to the super-hydration exhibited by both forms, a feature that seems to be characteristic of β-CD at high pressure.[10] Super-hydration induced by pressure is known to occur in certain zeolites;[11,12] it is also known that MOFs can accommodate large quantities of solvent under pressure.[13] While hydrate and solvent formation of organic molecules can be enhanced under high pressure (§1.1.2), β-CD is, to the best of our knowledge, the first example of a purely organic molecule that exhibits super-hydration. The concept of super-hydration is not unambiguously defined in the literature; hence for the scope of this work a *super-hydrate*

is defined as a crystal structure that incorporates a substantially larger number of water molecules per formula unit, compared to other hydrates of the same molecule, and which cannot be described as a water clathrate or semi-clathrate.

Because of the high degree of disorder of the water network in forms II and III and the limitations imposed by the use of a DAC, the hydration numbers cannot be accurately determined. The present chapter elaborates on the determination of the water content and its implications on the crystallisation of forms II and III.

3.2 Experimental

3.2.1 Crystal growth

Hydrates of β-CD were produced by following the in situ high-pressure crystallisation procedure outlined in §1.1.2. Several crystallisation experiments were performed and three crystals (named **A**, **B** and **C**) were ultimately measured by X-ray diffraction. Crystals **A** and **B** corresponded to form II at 0.93 and 0.92 GPa, respectively, while crystal **C** corresponded to form III at 0.72 GPa.

Single crystals of form I were produced from an aqueous saturated solution of β-CD, prepared by dissolving ∼40 mg of the compound (Sigma-Aldrich, used as received; solubility at 298 K is 18.5 mg/mL) in ∼1 mL of hot (∼343 K) demineralised water and then cooling down to room temperature. Subsequent to crystallisation, both crystals and the solution were employed for the high-pressure experiments.

Crystal **A** was grown in a square DAC (Appx. I.1) equipped with an Inconel 718 gasket 250 μm thick pre-indented to a thickness of 170 μm and with a hole 300 μm in diameter. Given the small size of the sample chamber in the DAC and the relatively low solubility of β-CD (if compared with other CDs), the amount of β-CD present in the chamber when loading the DAC with the solution only was not sufficient to grow a crystal of a suitable size for X-ray diffraction. This problem was overcome thanks to the inverse solubility of form I with pressure; that is, crystals dissolve, rather than grow, upon increasing pressure. A cer-

tain amount of crystals of form I were loaded in the cell along with the saturated aqueous solution (Fig. 3.3a). Pressure increase resulted in the dissolution of the crystals, in turn increasing the amount of material in solution (Fig. 3.3b-e). Further increase in pressure produced a new crystalline phase—form II—above 0.90 GPa (pressure value depending on amount of β-CD inside the DAC).

Experimental observations showed that form II tends to grow poly-crystalline (Fig. 3.3f) and that it is very insoluble at the pressure of crystallisation. Because of this, obtaining a single crystal of form II from its polycrystalline precipitate by temperature cycling is impractical, as it requires very high temperatures. In addition, it was observed that the dissolution of the polycrystalline form II always left residual particles in the sample chamber (Fig. 3.3g-i), which acted as nucleation points during recrystallisation on cooling, leading inevitably to the growth of several crystals. The origin of the residue is unknown, although it is suspected that, due to its excellent complex-formation capabilities, commercial β-CD might contain impurities which precipitate at high pressure. The amount of residue was always very small and difficult to recover once the DAC was opened, hence identification of its chemical composition was not possible. Based on these observations, it was concluded that polycrystalline precipitation of form II should be avoided.

The procedure finally used to grow crystal **A** consisted in increasing pressure gradually while monitoring the dissolution of the crystals loaded at the beginning of the experiment. Above 0.60 GPa, pressure was further increased in less than 0.05 GPa steps and the DAC was left to equilibrate for several hours in between each step. At 0.93 GPa, the crystals of form I dissolved completely and form II grew as a single crystal, which was measured by X-ray diffraction (Fig. 3.3j). A precise relationship of pressure steps and waiting times cannot be given, because it depends on the total amount of β-CD present in the sample chamber, which was not quantified, and the temperature at which the experiment is conducted (293-298 K in this case).

The other two crystals were grown following a similar procedure. Crystal **B** was grown in a triangular DAC (Appx. I.2) equipped with

Figure 3.3: Crystallisation of β-CD form II (Crystal **A**). **a)** DAC loaded with crystals of form I and saturated solution. **b-e)** Dissolution as pressure increases. **f)** Crystallisation of form II in polycrystalline form. **g,h)** Isobaric dissolution of the crystals of form II upon heating. **i)** Residue left after the dissolution of the polycrystalline form II. **j)** Single crystal of form II at 0.93 GPa.
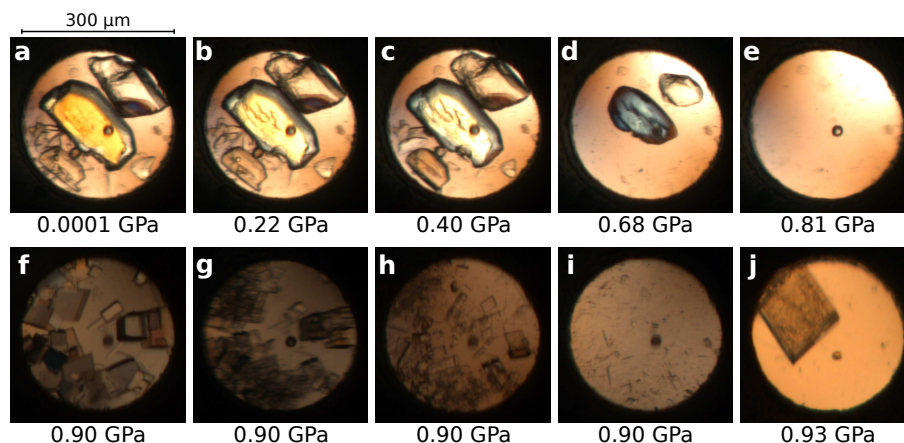
a 200 μm thick gasket pre-indented to 160 μm with a hole 320 μm in diameter. This crystal was grown and measured at 0.92 GPa (Fig. 3.4a). Crystal **C** was prepared in a 90/100° DAC (see Appx. I.3 for a technical description of the cell and further information on its use) with gasket of thickness 200 μm pre-indented to 120 μm and with a hole 270 μm in diameter. This crystal (Fig. 3.4b), obtained and measured at 0.72 GPa, turned out to be a different HP phase—form III—despite not showing any optical difference (colour, morphology, etc.) with respect to form II.

Figure 3.4: a) Single crystal of β-CD form II at 0.92 GPa. **b)** Single crystal of β-CD form III at 0.72 GPa.
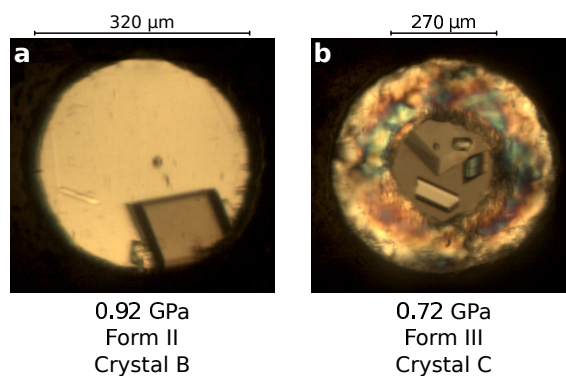
3.2.2 X-ray data collection and processing

Data on crystal **A** were collected at beamline F1 at DORIS (DESY) on a Huber four-circle diffractometer, with 1° ϕ scans at two different values of χ (0 and 90°). Exposure time was 80 s. Radiations of 0.5 and 0.65 Å with a beam size of 1×1 mm was used. Sample-detector distances were 200 and 130 cm for diffraction with 0.5 and 0.65 Å radia-

tion, respectively. The detector was a MAR165 CCD. Before integration, information in frame headers was updated with CMH and frames were converted from MAR to Bruker-Nonius Smart format using M2B.[14] The frames were finally converted into SFRM format using the tools included in APEX2.[15]

Crystals **B** and **C** were measured at the P02.2 beamline at PETRAIII (DESY) on a one-circle diffractometer. A $5{\times}2~\mu$m beam of 0.28962 Å wavelength was used. The collection strategy consisted of 1° ϕ scans with the DACs mounted on the goniometer head in three different orientations (60° rotations around the direction of the beam). Exposure time was 1 s. Two detectors were used for these experiments: a MAR345 image plate and a Perkin-Elmer XRD 1621 amorphous-silicon flat panel. MAR frames were converted to SFRM format directly in APEX2. ESPERANTO frames from the Perkin-Elmer detector were converted into MAR format first using CRYSALIS PRO,[16] and then to SFRM format in APEX2.

All data were integrated following the strategy outlined in §1.1.3.

3.2.3 Structure refinement

The structure of form II, determined from crystal **A**, was at first considered a pseudo-merohedral monoclinic twin on the basis of unexpectedly high R factors. The data could be integrated in two possible monoclinic lattices, P and C, both resembling an orthorhombic C lattice, due to $a \approx c$ in the first case and $\beta \approx 90°$ in the second. Due to the limited data available, it was not possible to elucidate the correct symmetry and the structure could be refined in space groups $P2_1$, $C2$, or $C222_1$. Maximum resolution of the data was only 1.4 Å with a completeness of 44% and a redundancy of 1.2 in space group $P2_1$ and 2 in $C2$. Crystal **B** provided data with a resolution up to 1.1 Å, which was sufficient to compare the refinement in the three possible space groups and determine that the orthorhombic one provides the best refinement. Data from crystals **A** and **B** were not merged. See discussion in §3.3.1 for more information.

Both forms II and III, from crystals **B** and **C** respectively, were refined following the same strategy. Models of β-CD molecules, taken from structures with the same lattice constants and space groups, were used

as a starting point. The model for form II was provided by Saouane,[10] while the model for form III corresponds to the structure deposited as WISREV[17] in the CSD. Full-matrix least-squares refinement against F^2 was performed in SHELXL v2014/7,[18] using the GUI SHELXLE.[19] All ordered non-H atoms of the CD molecules were refined anisotropically, with their atomic displacement parameters (ADPs) subject to similarity restraints. Disordered atoms of the CD molecules—the primary hydroxy group (O6) in some of the glucose units—and all the water molecules were refined isotropically. Restraints were applied to all bonds, angles and torsion angles, with the exception of the torsion angles involving the atom O6, which is known to be conformationally flexible. Restraints were automatically generated on the GRADE WEB SERVER.[20] In addition, the geometry of all the glucose rings was restrained to be similar. The large number of restraints helps to compensate for the low data-to-parameter ratio of the high-pressure data and is justified on the basis that β-CD molecules exhibit the same conformation in all their structures—apart from the orientation of the O6 groups—hence the restraints do not bias the model.[5,6] H atoms of the β-CD molecule were placed geometrically and constrained using the riding model. No H atoms have been refined for water molecules. For further crystallographic data see Table 3.1.

An important consideration during the refinement of β-CD hydrates is the placement of water molecules. When considering the large solvent-accessible volume in their structures and the unexpectedly low density for materials at 0.92 and 0.72 GPa, both forms II and III appear to contain a large amount of disordered water molecules, considerably larger than form I, which cannot be completely refined due to their extensive disorder and the limitations of high-pressure data. Refinement of disordered solvent is likely to cause a decrease in the R factor, due to the increase in the number of refined parameters,[22] even when the atomic positions do not match the electron density. To avoid over-fitting of the refined model, macromolecular crystallography (which faces problems similar to those of the high-pressure samples in this study: low-resolution and large solvent volume) makes use of the R_{free} factor.[23]

	Form II	Form III
Crystal data		
Chemical formula	$C_{42}H_{70}O_{35} \cdot 25.5H_2O$	$C_{42}H_{70}O_{35} \cdot 16.25H_2O$
Molecular mass (g mol^{-1})	1594.38	1427.73
Temperature (K)	298(2)	298(2)
Pressure (GPa)	0.92(5)	0.72(5)
Crystal system	orthorhombic	triclinic
Space group	$C222_1$	$P1$
a (Å)	19.7458(6)	17.8952(19)
b (Å)	24.0901(6)	15.4889(12)
c (Å)	32.980(5)	15.5473(12)
α (°)		101.061(7)
β (°)		113.078(9)
γ (°)		101.440(8)
V (Å^3)	15688(2)	3708.0(6)
Z	8	8
Crystal size (mm)	∼0.11×0.10×0.06	∼0.09×0.07×0.05
Data collection		
Radiation wavelength (Å)	0.28962	0.28962
Absorption correction	Semi-empirical (using I meas.), SADABS[21]	
T_{min}, T_{max}	0.8811, 1.0000	0.9586, 1.0000
Number of $[I > 2\sigma(I)]$ reflections:		
- measured	29976	12992
- independent	3977	7684
- observed	3781	5113
Completeness (%)	64.7	73.2
$(sin\theta/\lambda)_{max}$ (Å^{-1})	0.45	0.45
R_{int} (%)	4.35	8.26
Refinement		
$R[F^2 > 2\sigma(F^2)]$ (%)	9.46	12.82
$wR(F^2)$ (%)	24.85	32.06
S	1.203	1.112
Number of reflections	3977	7684
Number of parameters	876	1583
Number of restraints	2194	5313
$\Delta\rho_{min}$, $\Delta\rho_{max}$ (e Å^{-3})	−0.390, 0.492	−0.443, 0.510

Table 3.1: Crystallographic data and refinement parameters of β-CD forms II and III.

The R factor is defined as:

$$R = \frac{\sum_{h,k,l} \|F_{obs}(h,k,l) - F_{calc}(h,k,l)\|}{\sum_{h,k,l} |F_{obs}(h,k,l)|} \tag{3.1}$$

Where $F_{obs}(h,k,l)$ and $F_{calc}(h,k,l)$ are the observed and calculated structure factor amplitudes, respectively. R_{free} is defined in the exact same way, but is computed using a set of reflections which are not used in the refinement. The rest of the reflections, against which the model is refined, allow to compute R_{work}, which is equivalent to R.

For the purpose of this study, every tenth reflection has been moved to the R_{free} set using XPREP.[24] Refining in this way, only water molecules

which provide a decrease in both R_{work} and R_{free} have been included in the model. An increase in R_{free} after the introduction of certain water molecules is an indication that the electron density in the corresponding region is too smeared to model the water content. None of the R factors changed significantly with the occupancies of the water molecules, which were set to either 25, 50, 75, or 100% on the basis of the relative values of their isotropic ADPs.

This strategy was used until all possible water molecules were located. Then, several cycles of refinement against the full dataset were performed until convergence. This type of refinement evidences that the exact water content of forms II and III of β-CD cannot be determined by means of single-crystal X-ray diffraction. Several approaches to estimate the water content in this structures are discussed in the following section.

3.3 Results and discussion

3.3.1 Twinning in β-cyclodextrin hydrate form II
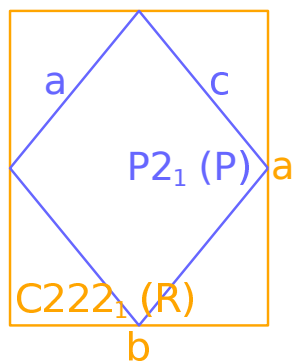
Figure 3.5: Relationship between the monoclinic $P2_1$ (**P** setting, blue) and the orthorhombic $C222_1$ (**R** setting, orange) unit cells of β-CD form II.

To address the possible existence of pseudo-merohedral twinning in the structure of β-CD form II, diffraction data were first indexed and integrated in the three possible settings (Table 3.2). The three settings are mutually related, as both space groups $P2_1$ (**P**) and $C2$ (**Q**) are subgroups of $C222_1$ (**R**). Data statistics after integration, determined using XPREP, are summarised in Table 3.3. XPREP was also used to determine whether any of the monoclinic settings should be transformed into the orthorhombic unit cell (Table 3.4). The following matrices can be used to transform between the different settings (Fig. 3.5):

$$(\mathbf{P} \to \mathbf{R}) \qquad P2_1 \to C222_1 \qquad \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \qquad (3.2)$$

$$(\mathbf{Q} \to \mathbf{R}) \qquad C2 \to C222_1 \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad (3.3)$$

Setting	Space group	a (Å)	b (Å)	c (Å)	β (°)	V (Å³)
P	$P2_1$	15.5933(4)	32.995(4)	15.5927(4)	101.313(3)	7866.6(10)
Q	$C2$	19.7575(5)	24.1014(7)	32.983(5)	90.013(5)	15706(3)
R	$C222_1$	19.7458(6)	24.0901(6)	32.980(5)	90	15688(2)

Table 3.2: Possible settings for β-CD form II.

Setting	Max. res. (Å)	Compl. (%)	Red.	Mean I	Mean I/σ	R_{int} (%)	R_σ (%)
P	1.10	66.8	5.04	36.2	29.21	4.63	2.62
Q	1.10	65.2	5.02	35.2	25.74	5.05	2.93
R	1.10	66.9	9.38	37.0	36.46	4.95	2.13

Table 3.3: Comparison of the data indicators after integrtion of β-CD form II in the three possible settings.

Setting	Mean $|E^2 - 1|$	FOM	R_{sym} (%)
P	0.772	0.002	3.7
Q	0.769	0.013	5.5
R	0.761	0.000	2.8

Table 3.4: XPREP statistics for unit-cell transformations to setting **R** of the integrated data of β-CD form II.

Eq. 3.3 is a unity matrix, as the unit cell does not need to be transformed, only the β angle should be constrained to 90°.

The three settings yield very similar statistics, however, **R** seems a better choice, as it provides low R_{int} and R_σ—even considering the higher symmetry of the space group $C222_1$—but exhibits slightly larger mean intensity and intensity-to-noise ratio. Information in Table 3.4 also indicates that setting **R** should be the preferred choice, as both **P** and **Q** transformed to **R** exhibit larger R_{sym} than **R** itself. Furthermore, the mean $|E^2 - 1|$ is close to the expected value for a non-centrosymmetric space group (ideally 0.736), while this value tends to differ in cases of twinning. Finally, given a choice of space groups belonging to different crystal systems and in the absence of twinning, it is good crystallographic practice to refine in the highest possible symmetry space group.[25]

A final test was performed by refining the structure in the **P** setting, with and without a twin law. This test was not performed with the **Q** setting, as it is the unlikeliest of the three possibilities. The R factor of the refinement went from 11.21 to 16.06% upon the introduction of the twin law. Moreover, the final refinement in the **R** setting, space group $C222_1$, provided the best refinement, and hence the possibility of twinning in the structure of β-CD form II was discarded.

3.3.2 Description of the structure of β-cyclodextrin form II

β-CD form II, a 25.5 hydrate, crystallises in an orthorhombic unit cell with $a = 19.7485(6)$ Å, $b = 24.0901(6)$ Å, $c = 32.980(5)$ Å, and space group $C222_1$. Eight symmetry-related β-CD molecules are contained in a unit cell. The asymmetric unit consists of a single β-CD molecule and a disordered network of 25.5 water molecules. All glucose units within the β-CD molecule exhibit the same chair conformation. Five out of seven O6 groups are in (−)-gauche conformation, that is, pointing outwards the cavity, which is the preferred conformation. The remaining two O6 groups exhibit disorder between the (−)- and (+)-gauche conformation, the latter being intended to stabilize water molecules inside the cavity by H-bond formation (Fig. 3.6).[26] O6_5 has relative occupancies of 60 and 40% for the (−) and (+) conformations, respectively, while for O6_6 both conformations are equally occupied at 50%. As encountered in most CD molecules, the seven O-glycosidic atoms (O4) are fairly coplanar. The root mean square (RMS) between the O4 atoms and their mean plane (O4 plane) is 0.020 Å.

Figure 3.6: Graphic definition of the (+)- and (−)-gauche conformations of the O6 groups in CD molecules as Newman projections of the C5–C6 bond. The O6 group points towards the cavity (positive O5–C5–C6–O6 torsion angle) in (+)-gauche conformation. In (−)-gauche conformation, which is the preferred one, the O6 group points outwards the cavity (negative O5–C5–C6–O6 torsion angle).
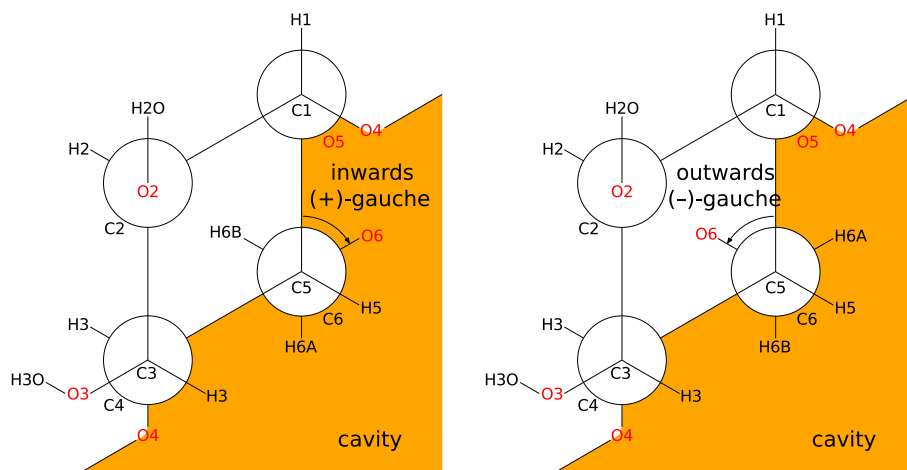
The main packing motif is a dimer of β-CD molecules (Fig. 3.7) opposed by their wider openings (O2-O3 side, usually termed as *head-to-head* dimer). The two molecules in the dimer are related by a two-fold symmetry axis parallel to the *b* axis of the unit cell and are linked by H bonds between the secondary hydroxy groups (O2 and O3). Both

molecules are close enough (the mean plane of the atoms O2 and O3 of each molecule are separated ∼2.6 Å) to form a close barrel-like cavity. These dimers arrange themselves in screw channels along the *c* axis (Fig. 3.8), thanks to the formation of H bonds between the O6 groups of two adjacent dimers, involving water molecules situated directly on top and below the dimer. The space in between dimers can be described as a wedge with an opening angle of ∼22° (defined by the mean plane of the C6 atoms of each molecule). The dimers are separated ∼2.5 Å at the closest point and ∼6.2 Å at the furthest. From the perspective of the water molecules, the structure is better described by the combination of two different channels, one defined by the screw channel running trough the cavities of the *β*-CD molecules, and another defined by the space in between four adjacent screw channels (Fig. 3.9). Water molecules are likely to move along the channels, and thanks to the existence of the wedges in between *β*-CD dimers, they can also move from one channel to another.
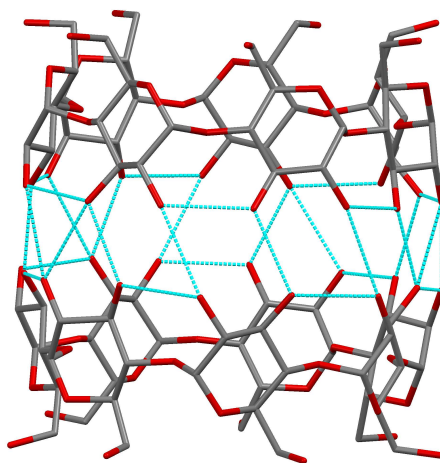
Figure 3.7: *Head-to-head* dimer of *β*-CD molecules in form II. H atoms, water molecules, and disorder have been omitted for clarity. H bonds between O2 and O3 groups are represented as dashed blue lines.

The packing arrangement in structures of *β*-CD host-guest complexes crystallising in the same space group than form II and with similar unit-cell parameters is usually referred to as a *chessboard*;[6] this particular pattern becomes apparent when the structure of the hydrate form II is projected perpendicularly to the *a* axis (Fig. 3.10). The screw-channel packing, which is characteristic of certain *β*-CD structures crystallising in space group $P2_1$, can also be observed in the *chessboard* $C222_1$

Figure 3.8: Screw-channel packing arrangement characteristic of structures of β-CD crystallising in space group $C222_1$, exemplified here for hydrate form II. Projection perpendicular to the diagonal of the *ab* face. A channel inside β-CD cavities is represented in blue. Wedges connecting the channels inside and outside β-CD molecules are depicted in red. H atoms, water molecules, and disorder have been omitted for clarity.
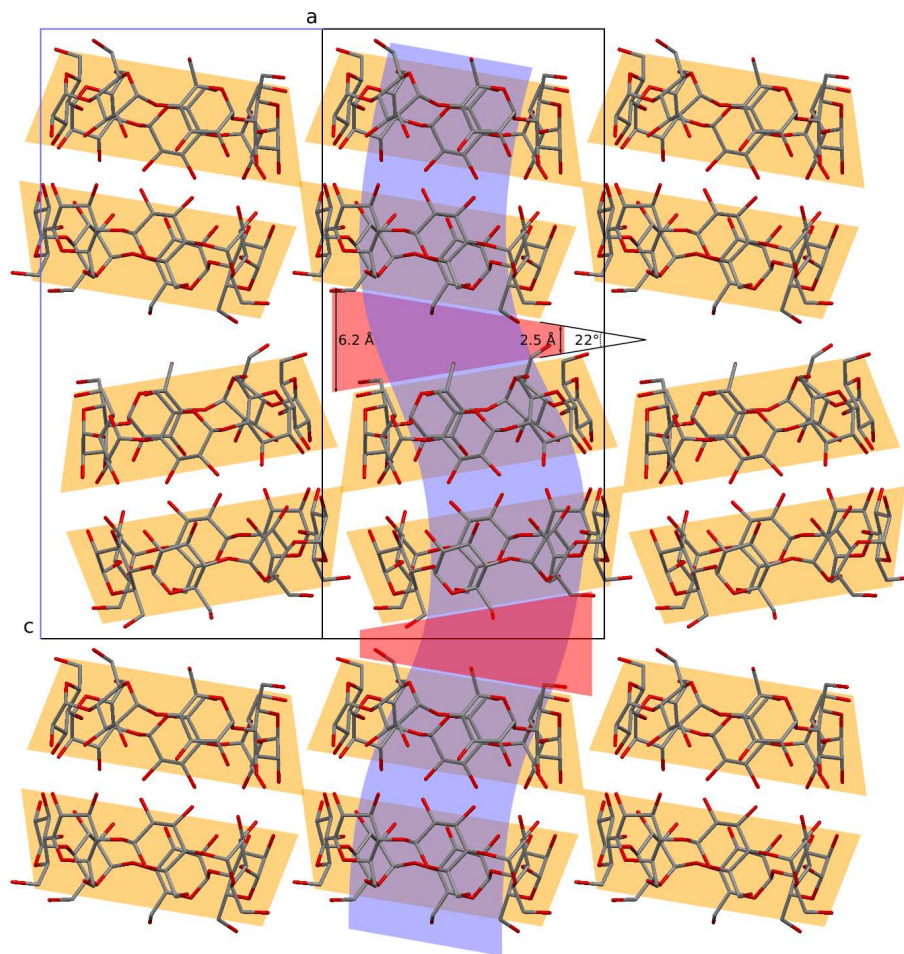
structures, in a projection perpendicular to de diagonal of the *ab* face. These and other considerations can be found in a thorough discussion of the packing types in β-CD inclusion complexes by Saouane.[10] Such discussion stemmed from the structural analysis of form V of β-CD·paracetamol complex, also crystallised in situ at high pressure, which is isostructural with β-CD hydrate form II reported herein. The main conclusion of the study is that the structure of β-CD·paracetamol form V, and by extension that of β-CD hydrate form II, is slightly different than that of the rest of the $C222_1$ complexes, albeit having the same unit-cell constants. In particular, some details of the packing arrangement are closer to those of the structures crystallised in space group $P2_1$. In fact, the screw channels in form II hydrate are as narrow as those in the $P2_1$ structures, but they are oriented like those in the $C222_1$ structures. The relationship between the packing in both space groups is so
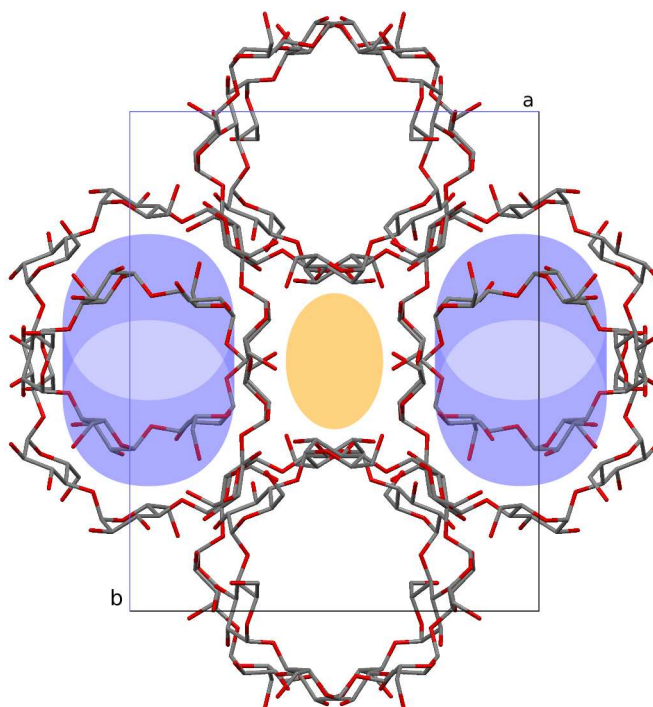
Figure 3.9: View of the channels in the structure of β-CD hydrate form II, along the axes of the channels. Cavity channels are depicted in blue, while intermolecular channels are depicted in orange.

close that it is possible to refine the structure in both $P2_1$ and $C222_1$, as discussed in §3.3.1.

A possible reason for the particular packing arrangement observed in the hydrate form II and the paracetamol complex form V is that it allows to maximise the *void* volume in the structure, hence allowing for a greater water content. In order to test this hypothesis, solvent accessible volumes have been calculated using PLATON[27] in all $P2_1$ and $C222_1$ structures, after removing water and guest molecules (Table 3.5). The calculation shows that voids per β-CD molecule in $P2_1$ structures are, on average, 58.4 Å^3 smaller than in $C222_1$ structures, and these are in turn 70.6 Å^3 smaller than those in the two HP structures. Such differences in void volume can account for a minimum of 2.6 and 3.2 additional water molecules, respectively. Although this is a significant difference, the change in packing alone cannot explain the super-hydration observed in the hydrate form II.

Another interesting observation is that β-CD hydrate form II and β-CD·paracetamol form V are isostructural and that they both show the same super-hydration effect. The *Crystal Packing Similarity* module

Figure 3.10: *Chessboard* packing type characteristic of structures of β-CD in space group $C222_1$, here exemplified with hydrate form II. Projection perpendicular to the *a* axis. H atoms, water molecules, and disorder have been omitted for clarity.
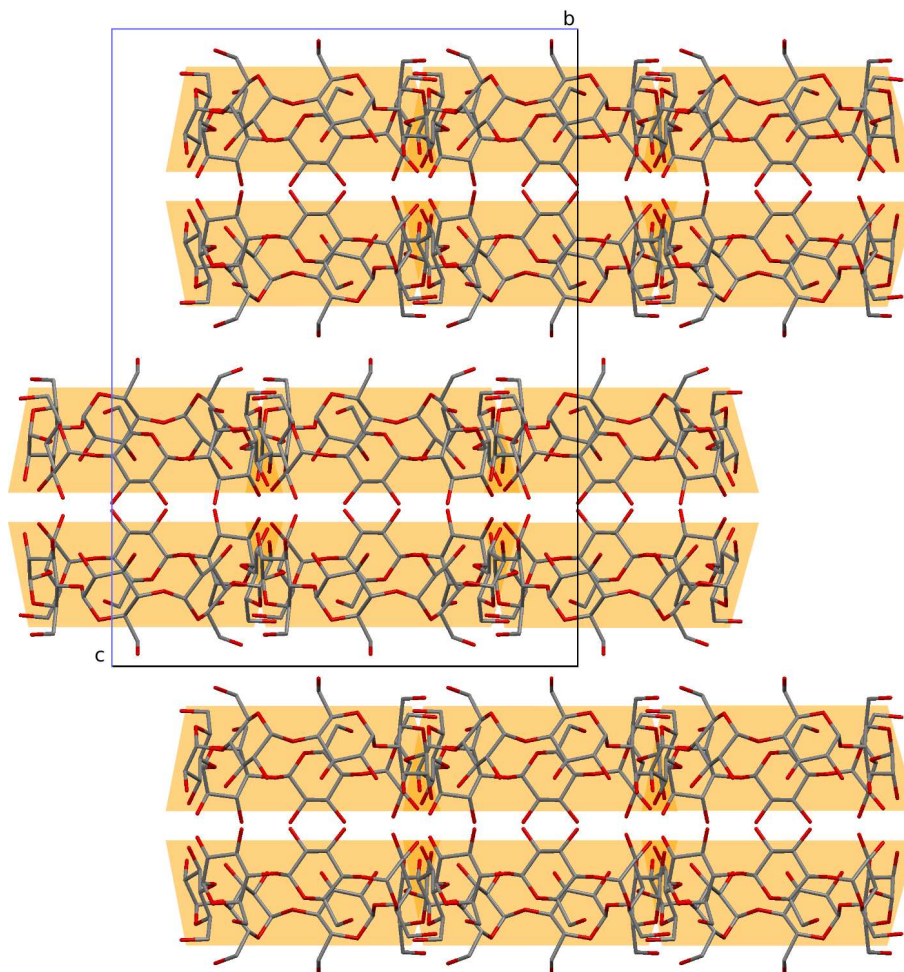
implemented in MERCURY yields an RMS difference of 0.078 Å between the cores of the two structures (the core is defined as the β-CD molecule after O2, O3, C5, and O6 atoms are removed, along with water and guest molecules). In this case, the presence of the guest does not seem to alter the structure of the hydrate, which can be explained by the low occupancy of the guest, 30%.

3.3.3 Description of the structure of β-cyclodextrin form III

Form III of β-CD, which is a 16.25 hydrate, crystallises in a triclinic unit cell with lattice parameters $a = 17.8952(19)$ Å, $b = 15.4889(12)$ Å, $c = 15.5473(12)$ Å, $\alpha = 101.061(7)°$, $\beta = 113.078(9)°$, $\gamma = 101.440(8)°$, space group $P1$. The are two independent β-CD molecules in the unit

Space group	Ref. code	V_Z (Å^3)	$V_{Z,void}$ (Å^3)	$V_{Z,void}$ (%)
HP $C222_1$	Hydrate form II	1958.6	816.1	41.7
	Paracetamol form V	1961.0	828.5	42.3
	$\mu\pm\sigma$		822.3 ± 8.8	42.0 ± 0.4
$C222_1$	BOSZUF[28]	1867.1	732.1	39.2
	BOTBAO[28]	1889.9	748.9	39.6
	DUYVID[29]	1879.1	741.0	39.4
	FALDUS[30]	1928.5	785.1	40.7
	LACTAL[31]	1925.3	784.1	40.7
	MEGQUK[32]	1872.4	732.5	39.1
	WEZSIE[33]	1869.9	738.1	39.5
	$\mu\pm\sigma$		751.7 ± 23.2	39.8 ± 0.7
$P2_1$	BEGWEQ[34]	1819.7	657.6	36.1
	BOTBES[28]	1819.7	670.5	36.8
	FODBIK[35]	1880.8	725.9	38.6
	XEYDIQ[36]	1871.6	722.8	38.6
	YOJHUC[37]	1842.4	689.6	37.4
	$\mu\pm\sigma$		693.3 ± 30.6	37.5 ± 1.1

Table 3.5: Solvent accessible volumes per β-CD molecule in structures of β-CD which share lattice parameters with form II, calculated with PLATON[27].

The list of structures has been taken from Saouane,[10] removing structures for which not all H atoms have been placed. Bond distances involving H atoms have been elongated to neutron values using MERCURY[38] prior to the calculation of V_{void}.

cell, whose conformation follows the same pattern as described before, with most of the O6 groups in $(-)$-gauche conformation. Only three O6 groups exhibit disorder. O6_13 and O6_26 both exhibit an occupancy of 70% for the $(-)$ conformation and 30% for the $(+)$. O6_22 shows relative occupancies of 60 and 40% for the $(-)$ and $(+)$ conformations, respectively. Both β-CD molecules in the unit cell are planar, with an RMS of the O4 atoms with respect to the O4 plane of 0.028 and 0.022 Å for each molecule.

The same *head-to-head* dimer observed in other β-CD structures is repeated in this case. Even in the absence of symmetry elements relating the two β-CD molecules in the $P1$ structure, the two components of the dimer are parallel, as evidenced by the 0.41° dihedral angle between the the two O4 planes. The core dimers of both forms II and III overlap with an RMS of 0.081 Å. The packing arrangement in form III consists of layers of dimers perpendicular to the a axis separated by layers of disordered water and displaced laterally with respect to the layers above and below (Fig. 3.11). This packing type is usually called *intermediate*.[6]

To check whether high pressure has any observable effect on the structure, a comparison between the packing of form III and 45 other structures of β-CD complexes crystallising in space group $P1$ with sim-

Figure 3.11: *Intermediate* packing arrangement of β-CD hydrate form III. Projection perpendicular to the *b* axis. H atoms, water molecules, and disorder have been omitted for clarity.
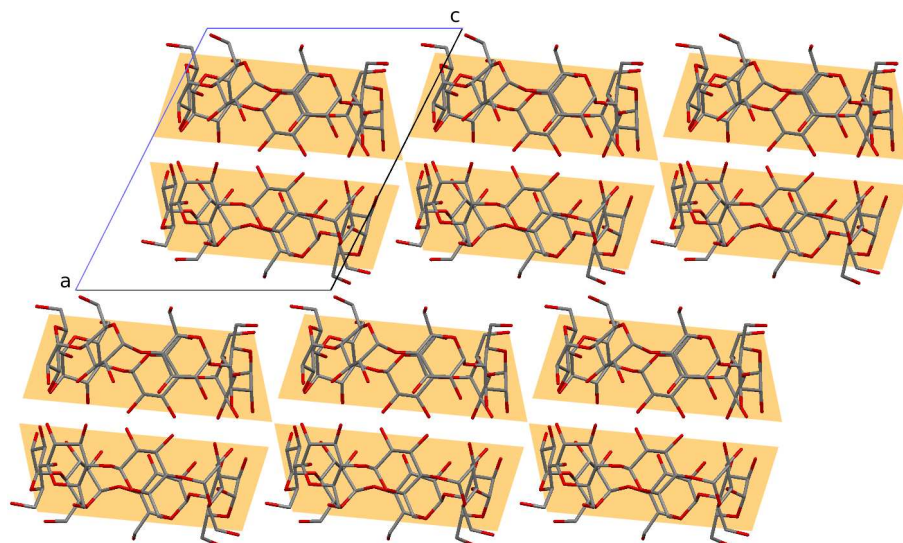
ilar unit-cell parameters has been performed with the *Crystal Packing Similarity* module implemented in MERCURY, using only the core β-CD structure. Based on the RMS values obtained from the comparison (Table 3.6), form III can be considered identical to most other related β-CD structures. Only two exceptions have been found: GUMXIX,[39] which exhibits the shortest of all the reduced *a* axes (14.9878 Å against an average of 15.35±0.09 Å) and also the smallest of the reduced α angles (109.51° against an average of 113.1±5°); and OFOWAI, which exhibits the smallest of the reduced β angles (96.882° against an average of 99.2±5°) and a lack of coplanarity of the O4 atoms (although this was not reported as a special feature of the structure).

Similarly to what was observed in form II, the fact that this form is grown at high pressure does not seem to alter its internal structure with respect to similar structures, but still causes a dramatic rise in the hydration number.

3.3.4 Hydration numbers of forms II and III of β-cyclodextrin

To better understand the hydration of β-CD under high pressure, the criteria outlined in Ch. 2 have been applied. V_{excl} for all three forms of β-CD are calculated in Table 3.7. There are 20 different determinations of the hydrate form I of β-CD recorded in the CSD (Appx. D). It has a

Ref. code	Match[a]	RMS	PXRD[b]	Ref. code	Match	RMS	PXRD
AGAZIR[40]	15	0.403	0.995	FODBEG[35]	15	0.420	0.996
AGAZOX[40]	15	0.372	0.998	GOTMOT[41]	15	0.376	0.997
AGAZUD[40]	15	0.505	0.994	GUMXIX[39]	12	1.511	0.979
AJUVAC[42]	15	0.456	0.995	HAHVES[43]	15	0.378	0.997
ANAXAP[44]	15	0.475	0.996	HAKPUG[45]	15	0.414	0.998
BIDMOQ[46]	15	0.415	0.995	HEGXUM[47]	15	0.393	0.993
BOGCAB[48]	15	0.396	0.998	ICIYAW[49]	15	0.371	0.998
BOSZEP[28]	15	0.451	0.996	LONGIE[50]	15	0.464	0.993
BOSZIT[28]	15	0.427	0.997	MASBIR[51]	15	0.445	0.996
BOSZOZ[28]	15	0.382	0.998	MASBOX[51]	15	0.453	0.995
BURXOC[52]	15	0.356	0.994	NILQUV[53]	15	0.479	0.996
CACPOM[54]	15	0.378	0.994	OFOWAI[55]	9	1.000	0.993
CUPYOC[56]	15	0.440	0.997	OFOWIQ[57]	15	0.378	0.997
DIFHOP[58]	15	0.440	0.990	TAFZIK[59]	15	0.478	0.996
DOCVUM04[60]	15	0.448	0.993	TEJHAR[61]	15	0.390	0.995
DOCVUM06[62]	15	0.463	0.994	UJEFEV[63]	15	0.466	0.997
DOCVUM07[62]	15	0.450	0.995	UKAGOC[64]	15	0.406	0.995
DOCVUM08[62]	15	0.386	0.997	VOFJIK[65]	15	0.421	0.997
DOCVUM09[62]	15	0.378	0.998	VOQDOU[60]	15	0.395	0.994
DOCVUM10[62]	15	0.384	0.998	VOQDUA[66]	15	0.457	0.990
DOCVUM11[62]	15	0.399	0.998	WISREV[17]	15	0.445	0.994
EYAQAX[67]	15	0.486	0.996	WISRIZ[17]	15	0.391	0.994
EYAQEB[67]	15	0.436	0.997				

[a] Out of 15 molecules.
[b] Degree of similarity between the PXRD patterns of both forms compared.

Table 3.6: Comparison of the packing of the core structure of β-CD hydrate form III against 45 other structure of β-CD in space group $P1$ with similar unit-cell parameters using the *Crystal Packing Similarity* module implemented in MERCURY. A cluster of 15 molecules was used in each comparison.

hydration number of 12, which is representative of the amount of water that the structure contains if no dehydration occurs.[7]

Fig. 3.12 shows the *PV* diagram for β-CD hydrates (see §2.3.3 for a general description of the diagram). V_{excl} for the two super-hydrates obtained at high pressure, forms II and III, are considerably larger than that of form I. This is an unexpected behaviour, as molecular volumes in structures obtained in situ at high pressure tend to be either similar or smaller than their AP counterparts (Ch. 2). Furthermore, the super-hydration could be explained as a way to increase the density of the crystal, but this does not seem to be the case here.

The hydration numbers of forms II and III are underestimated from the refinement, because part of the electron density is too smeared to be modelled, which leads to larger V_{excl}. Volume considerations have been

Form	I	II	III
P (GPa)	0.0001	0.9	0.72
V_{cell} (Å^3)	3072.509	15688	3708
Z	2	8	2
V_Z	1536.254	1961	1854
N	12	25.5	16.25
$V_{H,wat}$ (Å^3)	268.44	519.2	336.3
V_{excl} (Å^3)	1267.82	1442	1518

Table 3.7: Molecular volume of β-CD excluding water in forms I, II and III.

Figure 3.12: *PV dia-gram of β-CD hydrates. Empty icons represent V_Z (higher) and V_{excl} (lower). Solid icons correspond to V_{excl} calculated with the minimum estimation of N. Solids icons with orange edge correspond to V_{excl} calculated with the max-imum estimation of N. All the points correspond-ing to the same form are linked by a vertical dotted line, thinner for the estimated volumes. A red dashed line indicating V_{excl} of form I at ambient pressure has been included as a reference.*
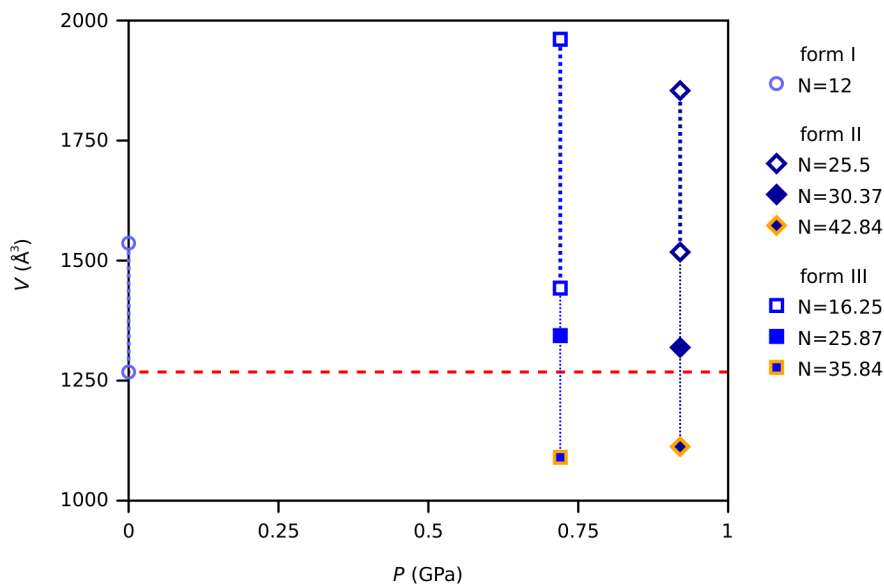
used to provide a better estimation of the hydration numbers. After re-moving all the water content from both structures and normalising the bonds involving H atoms to neutron distances, the solvent accessible volumes (V_{void}) of the structures have been calculated with PLATON.[27] V_{void} is defined as the space inside a surface drawn 1.2 Å away from the van der Waals surface of the CD molecules. The hydration number can be determined from the molecular V_{void} ($V_{void,Z}$) applying the follow-ing equation, assuming that all the available space is filled with water molecules:

$$N = \frac{V_{void,Z}}{V_{wat}} \tag{3.4}$$

Where V_{wat} is the volume of a single water molecule. There are indi-cations that the behaviour of the water molecules inside the cavity of the β-CD molecule is different than those of the molecules in the inter-molecular space.[7] Together with the observation that β-CD molecules are rigid, it is sensible to consider that water molecules inside the cav-ity might not compress as much as those outside. Hence, minimum and maximum possible V_{wat} have been considered to obtain a range of possi-ble hydration numbers. The maximum V_{wat}, leading to the minimum N, is 28.67 Å^3, which is the average volume of a water molecule in the struc-ture of form I ($V_{void,Z} = 344.05$ Å^3, occupied by 12 water molecules), assuming that there is no compression of the lattice water. This volume

Form	II		III	
P (GPa)	0.92		0.72	
$V_{void,Z}$ (Å^3)	870.85		741.65	
V_{wat} (Å^3)	N	V_{excl} (Å^3)	N	V_{excl} (Å^3)
28.67^a	30.37	1344	25.87	1319
20.33 and 20.70^a	42.84	1090	35.84	1112

^{a} Average volume of a water molecule in the structure of form I.
^{b} Determined from the compressibility curve of lattice water at 0.92 and 0.72 GPa, respectively.

Table 3.8: Hydration numbers (N) of β-CD hydrates estimated from volume considerations.

is considerably larger than the average 22.37 Å^3 estimated in §2.3.1, but is deemed more appropriate considering the particularities of β-CD. The minimum V_{wat}, leading to the maximum N and determined from the compressibility curve of lattice water, is 20.33 Å^3 at 0.92 GPa for form II and 20.70 Å^3 at 0.72 GPa for form III. Estimated hydration numbers, and the values of V_{excl} they lead to, can be found in Table 3.8.

The estimated V_{excl} are also included in Fig. 3.12. The results indicate that the hydration numbers refined from X-ray diffraction data are underestimated. This can also be checked by dividing $V_{void,Z}$ by N as determined during the refinement (Table 3.7), which gives V_{wat} of 34.15 and 45.64 Å^3 for forms II and III, respectively, considerably larger than the maximum V_{wat} determined above. Nevertheless, considering the lower end of the range of estimated values for the hydration number yields V_{excl} larger than that of form I. For the β-CD molecule to exhibit lower V_{excl} in forms II and III than in form I it will be necessary to have hydration numbers larger than 32.30 and 26.60, respectively. Those numbers should still be larger to account for the fact that V_{excl} of form I will be smaller at higher pressures. This last contribution cannot be estimated, as the compressibility curve of β-CD form I has not been measured. Based in similar experiences (§5.2.1), such an experiment is not straightforward, given that when using solvents such as water, methanol, or ethanol, form I dissolves with increasing pressure. The experiment would require finding a non penetrating, non interacting PTM and ensuring that the hydration number remains constant during compression.

With this information, two different scenarios can be suggested. The first option is that β-CD forms II and III exhibit in fact smaller V_{excl} than

form I. If this were the case, the hydration numbers would be severely underestimated in this study. It would be necessary to collect better data or provide better estimations for the hydration numbers and compare the resulting V_{excl} with those in the compressibility curve of form I. This scenario seems unlikely given the observations made in §3.3.2. In the first place, form II shows the same unit-cell parameters than other structures of β-CD complexes at ambient pressure. This could indicate that the crystal does not compress, but rather acts as a storage of water, for the solution to increase its density as much as possible. The solution surrounding the crystal at 0.9 GPa can increase its density by having most of its water H-bonded to β-CD molecules, hence decreasing their mobility. The excess water is then passed on to the crystal. Second, the analysis of void volumes suggests that either all structures of β-CD reported up to date have been refined with underestimated hydration numbers, or else the volume occupied by water molecules in such structures is larger than the one observed in crystals of hydrates of other organic substances (Table 3.9). The latter, more likely, might indicate a large availability of space which can be filled with extra water molecules to achieve the super-hydrated state without the need of compressing the structure of the crystal; or, if water molecules are less compressible than expected in β-CD crystals, that the hydration number, at least in form II, tends towards the lower end of the range given above.

Table 3.9: Average volume of lattice water molecules in β-CD structures, calculated from V_{void} and N determined from diffraction data (Table 3.5^a)

Space group	Ref. code	$V_{Z,void}$ (Å^3)	N	V_{wat} (Å^3)
$C222_1$	BOSZUF[28]	426.9	13.5	31.6
	BOTBAO[28]	324.6	10.5	30.9
	DUYVID[29]	292.0	10.1	28.9
	WEZSIE[33]	364.5	10.0	36.4
HP $P2_1$	BEGWEQ[34]	347.0	14.0	24.8
	BOTBES[28]	289.4	11.3	25.7
	FODBIK[35]	347.2	10.8	32.3
	YOJHUC[37]	281.5	11.8	24.0

^{a} Differently from Table 3.5, $V_{Z,void}$ are calculated here removing only water molecules, but leaving the guests. Structures FALDUS and MEGQUK have been removed because the content of their cavities has not been determined experimentally. Structures LACTAL and XEYDIQ have been removed because both yielded very large volumes for water molecules, probably due to a small guest which does not fill completely the cavity. In such situation the remaining void volume of the cavity is distributed over all the water molecules.

In the second option, V_{excl} for forms II and III are larger than that of form I. If this were the case, the density and the concentration of the solution surrounding crystals of forms II or III inside the sample chamber should be considerably larger than that of a saturated aqueous solution of β-CD at ambient pressure. This is because the application of pressure leads invariably to a decrease in the volume of the system. If the HP crystals are not denser than their AP counterparts, the solution around them must be denser. An experiment to test this scenario has been designed but not carried out yet. The basic idea is to determine the hydration number of form II by measuring the mass of β-CD and water in the crystal. These parameters cannot be measured directly, but can be derived from the size of the crystals and the amount of solution and its density, which can be determined spectroscopically. The main problem of this method is the determination of the volume of the sample chamber, as its depth cannot be measured accurately. This can be overcome if the total mass of β-CD and water inside the sample chamber are known.

3.4 Conclusions and outlook

Two new hydrates of β-CD, with 25.5 and 16.25 water molecules per CD molecule, have been crystallised in situ at high pressure and their crystal structures determined and described. The results provide further evidence of the ability of high-pressure techniques to allow access to crystalline forms which cannot be observed by other means. The two structures provide for the first time evidence of super-hydration in a purely organic compound. Furthermore, the use of pressure has exposed the complexity of an otherwise well-characterised system.

Crystals of β-CD form I, a dodecahydrate, are known to undergo a reversible dehydration process, controlled by the humidity of their environment. The movement of water molecules through the crystal structure, hindered by the tightness of the herringbone packing, is made possible by a concerted conformational change of flip-flop H bonds across the crystal. Forms II and III, 25.5 and 16.25 hydrates, respectively, can be crystallised at 0.92 and 0.72 GPa from β-CD aqueous saturated solu-

tion. Both forms were obtained following an established crystallisation protocol, albeit form III crystallised serendipitously and more experiments are needed to make the crystallisation of this form reproducible, in contrast with form II, which can be obtained routinely.

Both forms can be considered super-hydrates and exhibit a rather different behaviour than form I. Unfortunately, single-crystal X-ray diffraction studies have not allowed to accurately determine the hydration numbers of these new forms. Notwithstanding this, the hydration numbers inferred from structure refinement allow to at least make a number of valid speculations. On the one hand, if the hydration numbers were indeed to be correct, both forms would be considerably less dense than the AP form I, which contravenes the usual behaviour observed for crystals obtained under high pressure, and would force the solution surrounding the crystals at high pressure to be extremely dense. On the other hand, if the crystals were to exhibit the expected density at their formation pressures, the hydration numbers should be considerably higher, emphasizing the super-hydration features of β-CD molecules. The collection of diffraction data on more specimens at higher resolutions, possibly combined with the use of low temperature (e.g. with a close-cycle cryostat), which should enable better modelling of the disorder, or the measurement of the densities of the solution inside the DAC along the crystallisation process, by spectroscopic techniques, might allow to determine more accurate hydration numbers and shed light on the formation and behaviour of both forms II and III, which in turn will provide a better understanding on how the hydration process occurs.

References

[1] J. A. Hamilton, L. K. Steinrauf and R. L. Van Etten, *Acta Cryst.*, 1968, **B24**, 1560–1562.

[2] K. Lindner and W. Saenger, *Angew. Chem. Int. Ed.*, 1978, **17**, 694–695.

[3] K. Lindner and W. Saenger, *Carbohydr. Res.*, 1982, **99**, 103–115.

[4] W. Saenger, M. Noltemeyer, P. C. Manor, B. Hingerty and B. Klar, *Bioorg. Chem.*, 1976, **5**, 187–195.

[5] W. Saenger, *Angew. Chem. Int. Ed.*, 1980, **19**, 344–362.

[6] D. Mentzafos, I. M. Mavridis, G. Le Bas and G. Tsoucaris, *Acta Cryst.*, 1991, **B47**, 746–757.

[7] T. Steiner and G. Koellner, *J. Am. Chem. Soc.*, 1994, **116**, 5122–5128.

[8] C. Betzel, W. Saenger, B. E. Hingerty and G. M. Brown, *J. Am. Chem. Soc.*, 1984, **106**, 7545–7557.

[9] W. Saenger and T. Steiner, *Acta Cryst.*, 1998, **A54**, 798–805.

[10] S. Saouane, *PhD thesis*, Georg-August-Universität Göttingen, 2015.

[11] Y. Lee, J. A. Hriljac, T. Vogt, J. B. Parise and G. Artioli, *J. Am. Chem. Soc.*, 2001, **123**, 12732–12733.

[12] D. Seoung, Y. Lee, C.-C. Kao and T. Vogt, *Chem. Eur. J.*, 2013, **19**, 10876–10883.

[13] A. J. Graham, D. R. Allan, A. Muszkiewicz, C. A. Morrison and S. A. Moggach, *Angew. Chem. Int. Ed.*, 2011, **50**, 11138–11141.

[14] C. Paulmann, *MARTOOLSCP*, 2007.

[15] Bruker-AXS, *APEX2*, Bruker-AXS, Madison, Wisconsin, USA, 2007.

[16] *CRYSALIS PRO v171.36.28*, Rigaku Oxford Diffraction, Tokyo, Japan, 2012.

[17] S. Makedonopoulou and I. M. Mavridis, *Acta Cryst.*, 2000, **B56**, 322–331.

[18] G. M. Sheldrick, *Acta Cryst.*, 2015, **C71**, 3–8.

[19] C. B. Hübschle, G. M. Sheldrick and B. Dittrich, *J. Appl. Cryst.*, 2011, **44**, 1281–1284.

[20] *GRADE*, http://grade.globalphasing.org.

[21] G. M. Sheldrick, *SADABS v2008-1*, Bruker-AXS, Madison, Wisconsin, USA, 2008.

[22] W. C. Hamilton, *Acta Cryst.*, 1965, **18**, 502–510.

[23] A. T. Brünger, *Nature*, 1992, **355**, 472–475.

[24] G. M. Sheldrick, *XPREP v2012-1*, Bruker-AXS, Madison, Wisconsin, USA, 2012.

[25] D. Watkin, *Acta Cryst.*, 1994, **A50**, 411–437.

[26] K. K. Chacko and W. Saenger, *J. Am. Chem. Soc.*, 1981, **103**, 1708–1715.

[27] A. L. Spek, *Acta Cryst.*, 2009, **D65**, 148–155.

[28] Y. L. Zhao, D. Benítez, I. Yoon and J. F. Stoddart, *Chem. Asian J.*, 2009, **4**, 446–456.

[29] G. D. Enright, K. A. Udachin and J. A. Ripmeester, *CrystEngComm*, 2010, **12**, 1450–1453.

[30] L. Damodharan, V. Pattabhi and K. Nagarajan, *Mol. Cryst. Liq. Crys.*, 2004, **423**, 17–35.

[31] Y. Liu, Y. L. Zhao, H. Y. Zhang, X. Y. Li, P. Liang, X. Z. Zhang and J. J. Xu, *Macromolecules*, 2004, **37**, 6362–6369.

[32] M. M. Bobek, G. Giester, H. Kählig and U. H. Brinker, *Tetrahedron Lett.*, 2000, **41**, 5663–5667.

[33] P. M. Dean, *Aust. J. Chem.*, 2007, **60**, 133–138.

[34] M. R. Caira, E. De Vries, L. R. Nassimbeni and V. W. Jacewicz, *J. Incl. Phenom. Macrocycl. Chem.*, 2003, **46**, 37–42.

[35] Y. Liu, Y.-L. Zhao, Y. Chen and D.-S. Guo, *Org. Biomol. Chem.*, 2005, **3**, 584–591.

[36] M. Ceborska, M. Asztemborska and J. Lipkowski, *Chem. Phys. Lett.*, 2012, **553**, 64–67.

[37] T. Cappadona, L. Daniels and T. Siddiquee, *Appl. Sci.*, 2012, **2**, 773–779.

[38] C. F. Macrae, I. J. Bruno, J. A. Chisholm, P. R. Edgington, P. McCabe, E. Pidcock, L. Rodriguez-Monge, R. Taylor, J. van de Streek and P. A. Wood, *J. Appl. Cryst.*, 2008, **41**, 466–470.

[39] N. V. Sule, A. Ugrinov, S. Mallik and D. K. Srivastava, *Biochim. Biophys. Acta*, 2015, **1850**, 141–149.

[40] J. M. Alexander, J. L. Clark, T. J. Brett and J. J. Stezowski, *Proc. Nat. Acad. Sci. USA*, 2002, **99**, 5115–5120.

[41] J. A. Fernandes, A. I. Ramos, P. R. Claro, F. A. Almeida Paz and S. S. Braga, *CrystEngComm*, 2015, **17**, 937–946.

[42] M. R. Caira, E. J. C. de Vries and L. R. Nassimbeni, *Chem. Commun.*, 2003, 2058–2059.

[43] X.-J. Shen, H.-L. Chen, F. Yu, Y.-C. Zhang, X.-H. Yang and Y.-Z. Li, *Tetrahedron Lett.*, 2004, **45**, 6813–6817.

[44] M. L. Singleton, J. H. Reibenspies and M. Y. Darensbourg, *J. Am. Chem. Soc.*, 2010, **132**, 8870–8871.

[45] E. Wang, G. Chen and H. Liu, *Chin. J. Chem.*, 2009, **27**, 2097–2101.

[46] M. B. Hursthouse, C. Z. Smith, M. Thornton-Pett and J. H. P. Utley, *J. Chem. Soc., Chem. Commun.*, 1982, 881–882.

[47] A. Rontoyianni, I. M. Mavridis, E. Hadjoudis and A. J. M. Duisenberg, *Carbohydr. Res.*, 1994, **252**, 19–32.

[48] J. A. Hamilton and M. N. Sabesan, *Acta Cryst.*, 1982, **B38**, 3063–3069.

[49] S. M. Bruno, J. A. Fernandes, J. Marques, S. C. Neto, P. J. Ribeiro-Claro, M. Pillinger, F. A. Almeida Paz, M. P. M. Marques, S. S. Braga and I. S. Gonçalves, *Eur. J. Inorg. Chem.*, 2011, 4955–4963.

[50] S. Makedonopoulou, A. Tulinsky and I. M. Mavridis, *Supramol. Chem.*, 1999, **11**, 73–81.

[51] T. J. Brett, J. M. Alexander and J. J. Stezowski, *J. Chem. Soc., Perkin Trans. 2*, 2000, 1105–1111.

[52] A. Paulidou, D. Maffeo, K. Yannakopoulou and I. M. Mavridis, *CrystEngComm*, 2010, **12**, 517–525.

[53] Z. Fan, C.-H. Diao, M.-J. Guo, R.-J. Du, Y.-F. Song, Z.-L. Jing and M. Yu, *Carbohydr. Res.*, 2007, **342**, 2500–2503.

[54] S. Makedonopoulou and I. M. Mavridis, *Carbohydr. Res.*, 2001, **335**, 213–220.

[55] Z.-X. Yang, Y. Chen and Y. Liu, *Carbohydr. Res.*, 2008, **343**, 2439–2442.

[56] N. R. Lien and J. R. Telford, *Carbohydr. Res.*, 2009, **344**, 2606–2608.

[57] T. Aree and N. Chaichit, *Carbohydr. Res.*, 2008, **343**, 2285–2291.

[58] F. Nishioka, I. Nakanishi, T. Fujiwara and K.-I. Tomita, *J. Inclusion Phenom.*, 1984, **2**, 701–714.

[59] T. Bojinova, H. Gornitzka, N. Lauth-de Viguerie and I. Rico-Lattes, *Carbohydr. Res.*, 2003, **338**, 781–785.

[60] J. L. Clark and J. J. Stezowski, *J. Am. Chem. Soc.*, 2001, **123**, 9880–9888.

[61] A. Rontoyianni and I. M. Mavridis, *Acta Cryst.*, 1996, **C52**, 2277–2281.

[62] J. L. Clark, J. Peinado, J. J. Stezowski, R. L. Vold, Y. Huang and G. L. Hoatson, *J. Phys. Chem. B*, 2006, **110**, 26375–26387.

[63] E. J. Wang, G. Y. Chen and C. R. Han, *Acta Chim. Sinica*, 2009, **67**, 2527–2530.

[64] Y. Liu, Y.-L. Zhao, H.-Y. Zhang and H.-B. Song, *Angew. Chem. Int. Ed.*, 2003, **42**, 3260–3263.

[65] Y. Zhang, S. Yu and F. Bao, *Carbohydr. Res.*, 2008, **343**, 2504–2508.

[66] J. L. Clark, B. R. Booth and J. J. Stezowski, *J. Am. Chem. Soc.*, 2001, **123**, 9889–9895.

[67] E. Wang, G. Chen and C. Han, *Chin. J. Chem.*, 2011, **29**, 617–622.

β-Cyclodextrin dimethylformamide 12.5 hydrate inclusion complex

*The contents of this chapter are published in R. Granero-García amd F. P. A. Fabbiani, Acta. Cryst., 2014, B**70**, 586–594. Only my personal contribution to the results presented in this paper are included here.*

4.1   Introduction

Already in 1991 Mentzafos et al. showed that certain packing types of β-CD are associated with certain space groups and unit-cell parameters and presented a classification of the dimeric β-CD complexes according to this analysis[1]. In their seminal 1998 review article, Saenger and Steiner summarized the most likely packing types adopted by β-CD complexes depending on the size of the macrocycle and the nature of the guest.[2] When water or small guests are included, the preferred mode of packing of β-CD molecules follows a herringbone motif, whilst in the presence of bigger guests a channel arrangement is favoured. In the

particular case of space group $P2_1$, in which the here reported β-CD dimethylformamide (DMF) complex crystallizes, two types of packing are predominant: herringbone, associated with the unit-cell parameters $a{\sim}15$ Å, $b{\sim}10$ Å, $c{\sim}21$ Å, $\beta{\sim}110°$; and dimeric layers, associated with $a{\sim}15$ Å, $b{\sim}32$ Å, $c{\sim}15$ Å, $\beta{\sim}102°$. Exceptions exist where these packing types appear and the resulting unit-cell parameters are slightly different from the ones just outlined. The structure of β-CD–DMF belongs to one of these exceptions and shows the largest deviations observed thus far. This chapter presents a thorough analysis of the packing motifs of structurally related β-CD complexes using the structural similarity concept incorporated in the XPac method.[3]

4.2 Experimental

4.2.1 Crystal growth

Crystals of the β-CD–DMF complex were grown by slow evaporation from an aqueous saturated solution of β-CD (Sigma-Aldrich, used as received) to which a few drops of DMF (Sigma-Aldrich, used as received) were added. The crystals grew in the shape of long, thin, transparent laths.

4.2.2 X-ray diffraction analysis

X-ray diffraction data were collected on one of these crystals at 120 K on a Bruker ApexII CCD diffractometer equipped with a Mo microsource (λ = 0.71073 Å). The data collection strategy consisted of ω scans with scan steps of 0.3° and exposure times of 60 s per frame. Data were integrated following the strategy outlined in §1.1.3, without the high-pressure corrections, which were not needed in this case. The structure was solved using the integrated Patterson and direct methods procedure implemented in the program PATSEE[4] and full-matrix least-squares refinement against F^2 was subsequently performed using SHELXL v2013[5] through the GUI SHELXLE.[6]

All non-H atoms were refined anisotropically, with the exception of some of the disordered atoms, namely: O60_1 in the β-CD macrocycle,

the disordered atoms in the DMF molecule (C2, C2′, N3, N3′, C5, and C5′), and the O atoms of some of the water molecules (O9, O9′, O11, O11′, and O11″; Fig. 4.1). The geometry of all the glucose rings was restrained to be the same. β-CD H atoms were initially placed geometrically and constrained using the riding model; some positions were later improved by looking at difference Fourier maps. H atoms belonging to water molecules were located in Fourier difference maps when possible, their positions optimized according to the formation of possible H bonds and then refined subject to restraints. In the case of the water molecules labelled O10, O11 and O12 (with their respective disordered positions) H atoms could not be located. Site-occupancy factors were initially refined and subsequently fixed. An iterative refinement procedure of the highly correlated site-occupancy factors and displacement parameters using different starting models ensured refinement convergence and a reasonable model. A final R factor of 5.62% was obtained. For further crystallographic data see Table 4.1.

Figure 4.1: Chemical diagram of β-CD with numbering scheme. For simplicity only one of the glucose fragments is labelled.

4.2.3 XPAC analysis

A structure comparison among complexes of β-CD stored in the CSD (v5.34 with updates to May 2013) was carried out in an attempt to clarify the relationship between the unit-cell parameters, the crystal pack-

Table 4.1: Crystallographic data and refinement parameters of β-CD–DMF·12.5H$_2$O.

Crystal data	
Chemical formula	C$_{42}$H$_{70}$O$_{35}$–C$_3$H$_7$NO·25.5H$_2$O
Molecular mass (g mol^{-1})	1433.27
Temperature (K)	120(2)
Crystal system, space group	monoclinic, $P2_1$
a, b, c (Å)	15.500(3), 9.794(2), 21.689(5)
β (°)	90.138(3)
V (Å^3)	3292.5(12)
Z	2
Crystal size (mm)	∼0.2×0.1×0.05
Data collection	
Radiation wavelength (Å)	Mo Kα
Absorption correction	Semi-empirical (using I meas.), Sadabs[7]
T_{min}, T_{max}	0.688, 0.745
Number of [$I > 2\sigma(I)$] reflections reflections:	
- measured	38760
- independent	13449
- observed	11103
$(sin\theta/\lambda)_{max}$ (Å^{-1})	0.626
R_{int} (%)	5.29
Refinement	
$R[F^2 > 2\sigma(F^2)]$ (%)	5.60
$wR(F^2)$ (%)	13.00
S	1.12
Number of reflections	13449
Number of parameters	959
Number of restraints	564
$\Delta\rho_{min}$, $\Delta\rho_{max}$ (e Å^{-3})	−0.34, 0.46

ing and the chemical composition of this type of material. The XPac method[3] was used for this purpose. This method is based on the comparison of a cluster of molecules built around a central one, for which a set of atomic positions can be freely defined by the user. Clusters for every structure to be compared are analysed in terms of all possible angles and interplanar angles between molecules in the clusters, on the basis of user-selected predefined atoms. Values are compared across all clusters and mean differences are output as a series of parameters. This methodology allows a quantitative comparison across a large number of structures to be performed in a very efficient way. Additionally, the method yields a dissimilarity index[8] that can be used to quickly assess the degree of structural similarity. The method also identifies whether supramolecular constructs that are common to the compared structures exist, if any, and then determines their dimensionality (zero-, one-, two- or three-dimensional).

Several authors presenting structures of β-CD inclusion complexes report parameters based on the mean plane for the seven glycosidic O atoms (O4 plane) in order to characterize molecular shape.[9–15] Hence, these were the atoms also used to define molecules in the XPac study, along with the carbon atoms linked to them (C1 and C4 of every glucose ring). This choice of atoms also allows to take into account slight distortions of the macrocycle. A subset of 38 β-CD structures was obtained from the CSD. Only structures in the space group $P2_1$, exhibiting the herringbone packing motif, with similar unit-cell parameters, and with deposited three-dimensional coordinates were chosen. A list of all the structures included in the XPac calculation can be found in Table 4.2 (for complete information on unit-cell parameters see Appx. E). MUPNEQ01, whose structure was determined by powder diffraction,[11] was found to fulfil all the prerequisites of the CSD search but was not included in the XPac search because of the highly distorted geometry of the β-CD molecule. All XPac calculations were performed with the program's default values.

ARUXIU[14]	DEWCOY[16]	MIFHAK[17]	**UPULEX**[18]
ARUXOA[14]	DIRVOP[19]	**NUFVEQ**[20]	VACZIJ[12]
BCDEXD03[21]	GUXZOO[22]	**NUFVIU**[20]	VOJLIQ[23]
BCDEXD04[21]	HAQNUK[24]	**PIJGIY**[25]	*XEGTAF*[26]
BCDEXD05[21]	IZALEB[27]	**PIJGOE**[25]	**YIYSII**[28]
BCDEXD10[9]	KIJSEC[29]	**POBRON**[21]	**YIYSOO**[28]
BOBPEN[30]	KOBRIC[10]	POBRON01[31]	**ZIGZIY**[32]
BOBPIR[30]	**KUTKOZ**[33]	**POVSIC**[34]	β-CD–DMF
BUVSEQ01[21]	LADBAV[15]	**SUJDAD**[35]	
DEBGOG[36]	**LADQOX**[37]	TUSHUL[38]	

Table 4.2: CSD reference codes for the $P2_1$ structures compared with the XPac method.

Reference codes in bold correspond to structures in which the a axis is longer than the c axis. In all cases the b axis is the shortest one, with the exception of the reference code in italics (*XEGTAF*), which exhibits a unit cell $a = 10.601(4)$ Å, $b = 26.402(9)$ Å, $c = 15.586(6)$ Å, $\beta = 93.145(6)°$.

4.3 Results and discussion

4.3.1 Structure description

The β-CD macrocycle does not show any remarkable distortion when compared with other structures available in the CSD. One of the primary hydroxyl groups (O6 group) is disordered between two positions

with partial occupancies of 91 and 9%, respectively. This type of disorder appears in a third of all the reported β-CD complexes. All glucose rings within the β-CD macrocycle adopt a chair conformation according to the Cremer-Pople parameters.[39] The relative orientations of the glucose rings with respect to the macrocycle have been analysed in terms of angles between mean planes, in accordance with previous reports.[10–12,12,15] The macrocycle is defined by the O4 plane and each glucose is defined by the mean plane of the atoms O4_n, C4_n, C1_n and O4_$(n+1)$. The distances between any of the O4 atoms and the O4 plane is never larger than 0.168(3) Å. Only glucose 3 is found to be significantly more tilted after comparison of tilting angles across all β-CD structures. All O6 groups exhibit $(-)$-gauche conformation, that is, pointing towards the outside of the cavity.[2] A heptagon can be defined where the vertices coincide with the seven O4 atoms, with inner angles ranging from 120.82(9) to 136.8(10)°, in agreement with 128.57° for the regular polygon. See Table 4.3 for all geometric parameters.

Table 4.3: Geometric parameters (calculated with PLATON[40], in Å and °) of β-CD–DMF.

Glucose ring	1	2	3	4
Q^a	0.567(5)	0.584(5)	0.592(5)	0.550(5)
θ^b	7.4(5)	4.0(5)	3.1(5)	5.4(5)
O4 torsion angles^c	119.0(4)	99.4(4)	101.4(4)	115.2(4)
	111.2(4)	131.8(4)	108.4(4)	154.7(4)
Tilt angle^d	17.3(2)	5.6(2)	31.8(2)	9.6(2)
O4 deviation^e	−0.146(4)	0.153(4)	0.039(4)	−0.168(3)
O1 torsion angles^f	55.1(6), 168(3)	50.2(5)	53.5(6)	51.7(6)
	−65.3(6), 48(3)	−69.5(5)	−65.0(5)	−69.7(5)

Glucose ring	5	6	7
Q^a	0.575(5)	0.554(5)	0.566(5)
θ^b	3.9(5)	4.5(5)	4.1(5)
O4 torsion angles^c	102.7(4)	120.2(4)	104.0(4)
	123.4(4)	118.6(4)	142.3(4)
Tilt angle^d	9.8(2)	21.4(2)	6.9(2)
O4 deviation^e	0.082(4)	0.068(3)	−0.028(4)
O1 torsion angles^f	56.1(5)	62.0(6)	52.1(5)
	−62.4(5)	−59.3(5)	−68.4(4)

^{a} Cremer-Pople puckering amplitude.
^{b} Index to measure the deviation from the ideal chair conformation (ideal value: $\theta = 0$)
^{c} Torsion angles C4_n–O4_n–C1_$(n$-1$)$–O5_$(n$-1$)$ and C3_n–C4_n–O4_n–C1_$(n$-1$)$.
^{d} Defined as the angle formed by the O4 plane and the mean plane of the atoms O4_n, C4_n, C1_n and O4_$(n+1)$.
^{e} Distance from the atom O4_n to the O4 plane.
^{f} Torsion angles C4_n–C5_n–C6_n–O6_n and O5_n–C5_n–C6_n–O6_n. Values in italics correspond to the minor disordered component.

The DMF molecule is located inside the cavity of the β-CD molecule. An H bond (O2_2–H2O_2...O1_8) links a DMF molecule (H-bond acceptor) and a neighbouring β-CD related to the central host molecule by translation along the b axis. The centres of mass of the host and guest molecules almost coincide and differ only by 0.977 Å. The DMF molecule is almost equally disordered over two positions—refined relative occupancies are 56(2) and 46(2)%. Even allowing for some motion, as indicated by the ADPs, the carbonyl group remains in its position and acts as a pivot point.

The unit formed by β-CD and DMF is surrounded by 12.5 molecules of water. Of them, six are not disordered and show full occupancy; one (O7) has an occupancy of 50%; two (O8 and O9) are fully occupied, but disordered between two positions each (relative occupancies of 60 and 40% and 70 and 30%, respectively); the rest are unevenly distributed along two types of channels: one filled with O10 and the other filled with O11 and O12, each of them disordered between three positions (Fig. 4.1). All the water molecules are located in the space left between β-CD molecules and form a net of H bonds; there are no water molecules occupying the macrocycle cavity.

4.3.2 Structure comparison

Of the 38 structures analysed with XPac (Table 4.2), 15 show a unit cell with dimensions $a{\sim}15$ Å, $b{\sim}10$ Å, $c{\sim}21$ Å, and $\beta{\sim}110°$. Another 19 structures are described with different settings, in which the a and c axes are interchanged. In addition, two other structures—DIRVOP and POVSIC—show a considerably smaller monoclinic angle of ${\sim}102°$, which is still much larger than the $90.138(3)°$ found in the β-CD–DMF structure reported here. Another structure—XEGTAF—showing further differences in the unit-cell parameters was included in the study. XEGTAF shows a long unique axis of 26.402(9) Å and a small β angle of 93.145(6)°. Since XPac compares only the arrangements of molecules independently of the space group they crystallize in or their unit cells, consideration of these structures for the comparison is still possible and meaningful. The presence of redeterminations of the same compound accounts for

a different number of water molecules, which may influence the crystalline packing. Variable water content is characteristic of β-CD structures, where water molecules have a high mobility thanks to flip-flop H bonds[21].

Of all the structures considered when using a unified unit-cell setting, β-CD–DMF exhibits the longest *b* and *c* axes, while the *a* axis is the shortest one. This inclusion compound also exhibits the largest unit-cell volume. For this compound, unit-cell parameters have been checked for different crystals at different temperatures; five crystals from three crystallization batches were tested and no significant variation of unit-cell parameters as a function of temperature was recorded. Full data collection was not performed and hence no comments on the possible variation of water content can be made. For most of the compounds reported up to date, a description of the crystal structure is usually limited to the identification of the characteristic herringbone packing along the unique axis. In some cases[10,19,36] the position of a β-CD unit is given relative to the unit-cell axes. The absence of an in-depth structure comparison has motivated the XPAC analysis presented here.

Table 4.4 shows a summary of the results of the XPAC analysis (for complete information refer to the supplementary material in R. Granero-García amd F. P. A. Fabbiani, *Acta. Cryst.*, 2014, B**70**, 586–594). The following terminology is introduced to accompany the interpretation of the results (Fig. 4.2). A *strand* defines a linear sequence of single β-CD molecules. A strand can be *stacked* along the *b* axis or *side-to-side* along any of the other two axes, depending on the settings. A *herringbone* refers to two neighbouring stacked strands related by the screw operation along the *b* axis. A *plane* consists of several parallel herringbones. The *full structure* is formed when several planes are stacked one on top of each other in the third crystallographic dimension. A similarity with three dimensions implies that the two structures match at the level of full structures. Two-dimensionality means that only planes match between the two structures. One-dimensionality refers to a match of herringbones or strands only. Zero-dimensionality implies that the match is reduced to isolated molecules, although zero-dimensionality has also been used in

the literature to refer to a finite unit comprising two or more molecules.

All structures are characterized by the herringbone-type packing, as confirmed by inspection with any structure viewing program. The degree of similarity between structures can, however, be quantified further. On the basis of the XPac calculations, the following observations can be made:

i In most cases packing differences are translated into substantial differences in the unit-cell parameters. There are exceptions to this: for DEWCOY, KOBRIC, or UPULEX the unit cells do not show any differences and the packing is still distinct (zero-dimensional similarity).

ii DIRVOP and POVSIC, which exhibit the shortest among the longest axis—20.118(2) and 20.042(6) Å, respectively, against the mean value of 21.0(9) Å—and different β angles—102.14(1) and 102.30(2)°, respectively, against a mean of 108(5)°—exhibit three-dimensional similarity with each other (dissimilarity index of 1.9), two-dimensional similarity with β-CD–DMF, zero-dimensional similarity with DEWCOY, KOBRIC, UPULEX, and XEGTAF, and one-dimensional similarity with the remaining 31 structures.

	BCDEXD03	BOBPEN	DEWCOY	DIRVOP	KIJSEC	KOBRIC	POVSIC	UPULEX	XEGTAF
BOBPEN	3.8								
DEWCOY									
DIRVOP	7.0	7.8							
KIJSEC	2.3	3.3		7.3					
KOBRIC									
POVSIC	9.4	9.8		1.9	9.4				
UPULEX	18.0	19.0			18.6				
XEGTAF			8.8			6.6			
β-CD–DMF	7.9	10.4		8.9	10.0		9.4		

Table 4.4: Selected results from the comparisons performed with XPac.

The dimensionality of structural similarity is colour coded in white (three-dimensional), yellow (two-dimensional), orange (one-dimensional) or red (zero-dimensional). The dissimilarity index is indicated for each comparison, except for the zero-dimensional cases, where no dissimilarity index is ouput. Large dissimilarity indices (like those for UPULEX) point to significant geometrical differences in the arrangement of β-CD molecules that are nevertheless still consistent with the overall herringbone-packing arrangement. Only results described in detail in §4.3.2 of the text are shown.

iii The one-dimensional match for DIRVOP always corresponds to a strand, while for povsic it is extended to a herringbone for 18 of the 31 structures. A visual inspection of the compared structures indicates that the herringbone motif is always preserved, but the distance between stacked strands can be different, increasing the dissimilarity. This could be ascribed to the nature of the guest. The guests molecules in DIRVOP and POVSIC are rather small molecules (hexamethylenetetramine and trans-cyclohexane-1,4-diol, respectively), which are unusually located towards the opening of the *β*-CD near the secondary alcohol groups (usually the guest is displaced towards the other side). A similar displacement of the guest is found in other structures (e.g. NUFVIU, PIJGIY, and PIJGOE), yet these do not show any significant packing differences. These observations point out, in contrast to previous reports,[12] that a guest that is fully included in the cavity has no direct effect on the herringbone packing. It is rather the formation of the H-bonded network, and hence the location of hydration water molecules, which ultimately determines the separation between neighbouring herringbones. Previous studies proposed that the formation of H-bonded networks in *β*-CD complexes is highly influenced by the guest and that guests that are similar in size and shape can lead to unrelated H-bonded networks, unit cells and space groups.[24,41] This idea is supported by observations on XEGTAF (see vi).

Figure 4.2: Schematic explanation of terminology used in the text. The *strand* corresponds to the motif enclosed in the red area, the *herringbone* corresponds to the blue area and the *plane* to the green area.
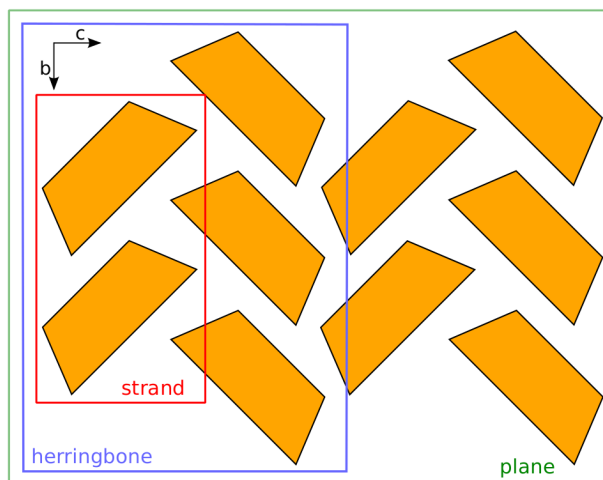
iv One-dimensional similarity is observed between UPULEX and most of the structures. However, the reasons for the one-dimensional similarity are of a different nature compared with DIRVOP and POVSIC. This is confirmed by the fact that pairs UPULEX-DIRVOP and UPULEX-POVSIC show zero-dimensional similarity. The reason is a different tilt of the β-CD molecule with respect to the unit-cell axes.

v Amongst structures with similar unit-cell parameters, DEWCOY and KOBRIC show zero-dimensional similarity with all other structures with the exception of XEGTAF (two-dimensional similarity), despite XEGTAF's long monoclinic axis. In the case of DEWCOY, the match is through a stacked strand along the shortest axis, while for KOBRIC it is a side-to-side strand. This exceptional difference might be produced by a different type of guest, which is KOH in KOBRIC. K^+ is not included in the cavity. All other structures, with the exception of BOBPEN, have neutral molecules as guests. In BOBPEN, the guest is HI, but I^- is located inside the cavity, emulating the organic inclusion complexes, and it does not exhibit differences in packing. This explanation does, however, not hold for DEWCOY, which shows a rather big guest partially outside the cavity. Other structures have this same feature and show no difference in packing.

vi XEGTAF, which has a different unit cell, does only match with DEWCOY and KOBRIC (one-dimensional similarity). In this case the guest is quite large so that approximately half of it is located outside the cavity, resulting in the elongation of the b axis. These observations were reported originally by Fan et al.[26] and a similar explanation was given for the elongation of the b axis in KOBRIC.[10]

vii The results of the packing similarity analysis of the β-CD–DMF structure can be partially interpreted using a combination of the observations i-vi above. The structure exhibits the smallest β angle, 90.138(3)°, the shortest among the short axes, 9.794(2) Å, and also the shortest among the middle-length axes, 15.500(3) Å. This is the only case where differences in terms of unit-cell parameters are found in axes that are not the longest one. The XPAC method determines that the packing similarity for the β-CD–DMF complex with most

of the structures is one-dimensional, in agreement with the observations for the other two cases with a different β angle (DIRVOP and POVSIC). It shows zero-dimensional similarity with the previously described outliers (DEWCOY, KOBRIC, UPULEX and XEGTAF), in line with the considerations for POVSIC and DIRVOP (see ii and iii). Indeed the match with both POVSIC and DIRVOP is two-dimensional. Two-dimensional similarity is also observed with BOBPEN and KIJSEC; however, we could not find an obvious explanation for this.

4.4 Conclusions and outlook

A new complex of β-CD with DMF has been crystallized and its structure determined by single-crystal X-ray diffraction. The structure exhibits the herringbone-type packing that appears frequently for these complexes, but has significant differences in both the unit-cell parameters and the relative distances and angles of the packed β-CD molecules.

A detailed comparison among 38 structurally related β-CD complexes has been carried out using the XPac method in an attempt to find the origin of the packing differences and similarities. It is now clear that the herringbone packing is maintained in all the structures and differences arise from the relative orientation of β-CD molecules in the herringbone motif, that is, along the longest axis, but never in the other two (with the exception of XEGTAF). However, it cannot be said that differences in packing are always reflected in different unit-cell parameters, whereas the opposite is always true. The conformation of the β-CD molecule is not significantly different in the crystal structures considered in this study; hence differences in packing are independent of β-CD geometry.

The results of the XPac analysis also confirm previous observations that the guest molecule does not directly influence the crystal packing, since differences in guest type and position within the cavity appear in structures with both zero- and three-dimensional similarity. It seems that the position of water molecules and the H-bonded network, which might also be influenced by the H-bonding capability of the guest, play

an important structure-directing role. Such an analysis has not been carried out in this study, but this theory could be tested in the future. In this context, Steiner et al., based on the observation that the position of hydration water molecules located outside the cavity of the β-CD molecule is invariant in six structures, reported that it is the packing that determines the position of the water molecules.[33] A similar observation was reported by Gessler et al.; however, in this case it was found that the guest molecule is likely to be located in such a way that possible H-bond donor or acceptor atoms are located in the same position than the water molecules displaced during the inclusion process.[25] Fan et al. have also reported that structural characteristics of β-CD inclusion complexes arise from a relation between guest/solvent type and position, and H-bonded network formation.[26] Rationalisation of the structure complexity of β-CD inclusion complexes continues to be an area of intense research activity.

References

[1] D. Mentzafos, I. M. Mavridis, G. Le Bas and G. Tsoucaris, *Acta Cryst.*, 1991, **B47**, 746–757.

[2] W. Saenger and T. Steiner, *Acta Cryst.*, 1998, **A54**, 798–805.

[3] T. Gelbrich and M. B. Hursthouse, *CrystEngComm*, 2005, **7**, 324–336.

[4] E. Egert and G. M. Sheldrick, *Acta Cryst.*, 1985, **A41**, 262–268.

[5] G. M. Sheldrick, *Acta Cryst.*, 2015, **C71**, 3–8.

[6] C. B. Hübschle, G. M. Sheldrick and B. Dittrich, *J. Appl. Cryst.*, 2011, **44**, 1281–1284.

[7] G. M. Sheldrick, *SADABS v2008-1*, Bruker-AXS, Madison, Wisconsin, USA, 2008.

[8] F. P. A. Fabbiani, B. Dittrich, A. J. Florence, T. Gelbrich, M. B. Hursthouse, W. F. Kuhs, N. Shankland and H. Sowa, *CrystEngComm*, 2009, **11**, 1396–1406.

[9] K. Lindner and W. Saenger, *Carbohydr. Res.*, 1982, **99**, 103–115.

[10] P. Charpin, I. Nicolis, F. Villain, C. de Rango and A. W. Coleman, *Acta Cryst.*, 1991, **C47**, 1829–1833.

[11] M. M. Pop, K. Goubitz, G. Borodi, M. Bogdan, D. J. A. De Ridder, R. Peschar and H. Schenk, *Acta Cryst.*, 2002, **B58**, 1036–1043.

[12] T. Aree and N. Chaichit, *Carbohydr. Res.*, 2002, **337**, 2487–2494.

[13] T. Aree and N. Chaichit, *Carbohydr. Res.*, 2008, **343**, 2285–2291.

[14] T. Aree, B. Schulz and G. Reck, *J. Incl. Phenom. Macro.*, 2003, **47**, 39–45.

[15] A. Paulidou, K. Yannakopoulou and I. M. Mavridis, *J. Incl. Phenom. Macro.*, 2010, **68**, 297–303.

[16] E. J. Wang, Z. X. Lian and J. Cai, *Carbohydr. Res*, 2007, **342**, 767–771.

[17] M. Crisma, R. Fornasier and F. Marcuzzi, *Carbohydr. Res.*, 2001, **333**, 145–151.

[18] P. Guo, Y. Su, Q. Cheng, Q. Pan and H. Li, *Carbohydr. Res.*, 2011, **346**, 986–990.

[19] K. Harata, *Bull. Chem. Soc. Jpn.*, 1984, **57**, 2596–2599.

[20] T. Aree and N. Chaichit, *Supramol. Chem.*, 2009, **21**, 384–393.

[21] T. Steiner and G. Koellner, *J. Am. Chem. Soc.*, 1994, **116**, 5122–5128.

[22] J. L. Wang, F. M. Miao, W. H. Zhou and S. K. Ma, *Chin. J. Chem.*, 2002, **20**, 358–361.

[23] Y. Zhang, S. Yu and F. Bao, *Carbohydr. Res.*, 2008, **343**, 2504–2508.

[24] Y. M. Zhang, Z. X. Yang, Y. Chen, F. Ding and Y. Liu, *Cryst. Growth Des.*, 2012, **12**, 1370–1377.

[25] K. Gessler, T. Steiner, G. Koellner and W. Saenger, *Carbohydr. Res.*, 1993, **249**, 327–344.

[26] Z. Fan, C. Diao, H. Song, Z. Jing, M. Yu, X. Chen and M. Guo, *J. Org. Chem.*, 2006, **71**, 1244–1246.

[27] M. K. Grachev, I. A. Senyushkina, G. I. Kurochkina, K. A. Lysenko, L. K. Vasyanina and E. E. Nifant'ev, *Russi. J. Org. Chem.*, 2010, **46**, 1506–1510.

[28] T. Steiner, G. Koellner, K. Gessler and W. Saenger, *J. Chem. Soc., Chem. Commun.*, 1995, 511–512.

[29] Y. V. Lisnyak, A. V. Martynov, V. N. Baumer, O. V. Shishkin and A. V. Gubskaya, *J. Incl. Phenom. Macro.*, 2007, **58**, 367–375.

[30] K. Lindner and W. Saenger, *Carbohydr. Res.*, 1982, **107**, 7–16.

[31] L. Damodharan, V. Pattabhi and K. Nagarajan, *Mol. Cryst. Liq. Crys.*, 2004, **423**, 17–35.

[32] T. Steiner and W. Saenger, *J. Chem. Soc., Chem. Commun.*, 1995, 2087–2088.

[33] T. Steiner, G. Koellner and W. Saenger, *Carbohydr. Res.*, 1992, **228**, 321–332.

[34] T. Steiner and W. Saenger, *J. Chem. Soc., Perkin Trans. 2*, 1998, **2**, 371–378.

[35] R. W. Seidel and B. B. Koleva, *Acta Cryst.*, 2009, **E65**, o3162–o3163.

[36] K. Harata, K. Uekama, M. Otagiri, F. Hirayama and Y. Ohtani, *Bull. Chem. Soc. Jpn.*, 1985, **58**, 1234–1238.

[37] G. Kurokawa, M. Sekii, T. Ishida and T. Nogami, *Supramol. Chem.*, 2004, **16**, 381–384.

[38] J. Y. Li, D. F. Sun, A. Y. Hao, S. H. Y. and S. J., *Carbohydr. Res.*, 2010, **345**, 685–688.

[39] D. Cremer and J. A. Pople, *J. Am. Chem. Soc.*, 1975, **97**, 1354–1358.

[40] A. L. Spek, *Acta Cryst.*, 2009, **D65**, 148–155.

[41] Y. L. Zhao, D. Benítez, I. Yoon and J. F. Stoddart, *Chem. Asian J.*, 2009, **4**, 446–456.

CHAPTER 5

Hydration of α-cyclodextrin at high pressure

5.1 Introduction

α-CD hydrates have been the object of extensive research; four different forms (named I to IV) have been characterised over the course of 25 years. All of them belong to the space group $P2_12_12_1$ and their unit cells, although distinct, hold a reasonable resemblance. In spite of this, the four forms exhibit three different crystal packings (forms I and III exhibit the same packing arrangement) and slightly different conformations of the macrocycle. This latter feature was revealed to be the key factor in determining the complex-formation capabilities of α-CD.

The structure of form I was originally reported by Manor and Saenger in 1974,[1] and subsequently redetermined by neutron diffraction by Klar et al.[2] This form is easily obtained by recrystallisation from a saturated aqueous solution and is characterised by the herringbone packing motif.[3] In form I, α-CD exists as an hexahydrate, with two water molecules

105

crystallising inside the macrocycle cavity and four in the intermolecular space.

What set this structure apart from any other CD structure known at the time was a different conformation of the α-CD molecule (Fig. 5.1). CD molecules in all other α- and β-CD complexes, including hydrates of the latter, exhibit a similar conformation, in which the macrocycle is *round* and all the glucose rings are equaly tilted with respect to the equatorial plane. One of the glucose rings of α-CD in form I is considerably more tilted than the rest, which causes the macrocycle to adopt a *tense* conformation.[1,4] This *tense* conformation was deemed as a high-energy state, resembling the conformation of α-CD in aqueous solution. In the solid state the conformation is stabilised by the formation of H bonds between two of the primary hydroxy groups (O6 groups) and the water molecules inside the cavity, which requires said groups to be facing towards the cavity—(+)-gauche conformation—a situation energetically less favourable than facing outwards—(−)-gauche conformation.[5] Additionally, the distance between secondary hydroxy groups (O2 and O3) is larger in α- than in β-CD, which weakens the extended H bonded network of the α-CD structure,[6] allowing for the observed distortion.

Figure 5.1: Molecule of α-CD in its *tense* conformation from the structure of form I at 0.20 GPa (§5.3.1). Water molecules outside the cavity and H atoms have been removed for clarity. The two O6 groups in the (+)-gauche conformation have been labelled; O6_1 adopts both (+) and (−) conformations with a 50% probability. Possible H bonds between O6 groups and the water molecules inside the cavity have been plotted as blue dashed lines. It can be seen how glucose ring 5 is tilted with respect to the rest of the glucose rings.
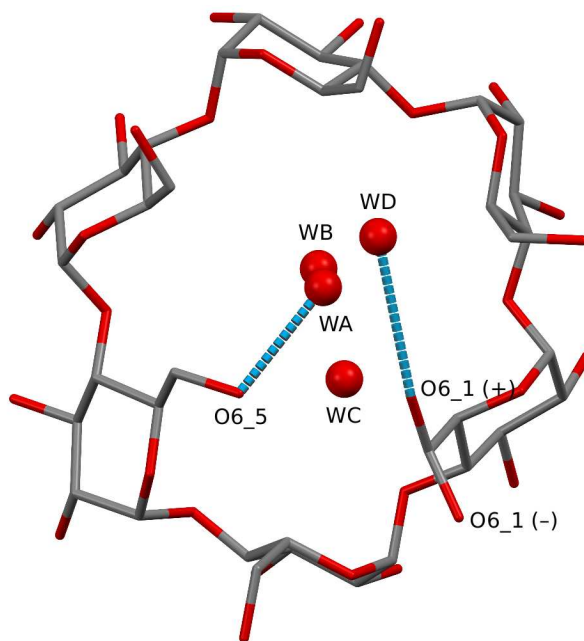
Form II[7] co-precipitates with form I, albeit it only represents a 2% of the crystallisation outcome and its formation is not well understood. It is also an hexahydrate, although the distribution of the water molecules is different, namely five molecules are located outside the cavity and one inside. In the absence of other guests the cavity is partially filled by an O6 group from a neighbouring molecule, which is made possible via a distortion of the herringbone packing. Despite the changes in the packing motif, the conformation of the α-CD molecules is nearly identical in both forms I and II.

Form III[5] was obtained while attempting to crystallise a complex of α-CD with barium cations, which instead led to a 7.25-hydrate. In this structure, there are five water molecules in the intermolecular spaces and 2.57 inside the cavity of the macrocycle. In this case, the herringbone packing is identical to the one observed in form I, but the conformation of the α-CD molecule differs, as it is *round* in form III, with all the O6 groups in $(-)$-gauche conformation.

The last form, IV,[8] was obtained in a similar manner than form III, while trying to crystallise complexes of α-CD with organic molecules containing alkylic chains. The stoichiometry corresponds to an 11-hydrate, with up to 7 molecules inside the cavity, a value which is very likely to be overestimated. The conformation of the α-CD molecules in form IV is also *round*, although the packing is of the brick type, unrelated to the herringbone type observed in the other three forms.[3]

The distinct conformation of α-CD in the hydrate forms I and II was used by Saenger et al. to explain inclusion-complex formation in α-CD by the *induced-fit* mechanism.[4] According to this mechanism, characteristic of enzymatically-catalysed reactions,[9] α-CD exists in solution in its *tense* form, with a low occupancy of the cavity, similar to the one observed in crystal structures. In the presence of a guest of suitable size, water, which is not fully coordinated by H bonds, easily leaves the cavity allowing the formation of an inclusion complex. The existence of both forms III and IV confirms the induced-fit mechanism, as the α-CD molecule in the *round* shape filled with more than two water molecules is a likely intermediate state during the complexation process.

The induced-fit mechanism explains how inclusion takes place, but not why; the driving forces of the inclusion process have been matter of a thourough debate.[10] Van der Waals and hydrophobic interactions seem to be the major driving forces, while electrostatic interactions and hydrogen bonding mainly affect the final conformation of the complex.

Prior to the commencement of the work reported here, we showed that by recrystallising α-CD form I at 0.65 GPa it was possible to increase the number of water molecules inside the cavity by 0.8, similarly to what is observed in form III, albeit while retaining the *tense* conformation of the macrocycle. Such structure had been previously predicted by computational means.[11] The inclusion of extra water was accompanied by a change in the conformation of an O6 group, from 100% (+) to 50:50% (+):(−), and in turn by a noticeable increase of the solvent-accessible volume of the α-CD cavity. This chapter presents a deeper crystallographic insight into the compression of α-CD form I, showing that the inclusion of additional water can occur in the solid state, without the need of recrystallisation, and attempts to provide a new mechanism for the inclusion of water by means of MD simulations.

5.2 Experimental

5.2.1 Crystal growth and compression study

Crystals of hydrate form I of α-CD were first grown at ambient conditions by slow evaporation of a saturated aqueous solution of α-CD (Sigma-Aldrich, used as received; solubility at 298 K is 145 mg/mL). A crystal grown under these conditions was loaded into a 90/100° DAC (see Appx. I.3 for a technical description of the cell and further information on its use) along with its saturated solution and directly compressed to 0.20 GPa. The DAC was equipped with a brass gasket of thickness 200 μm pre-indented to 120 μm and with a hole 300 μm in diameter. α-CD exhibits a similar behaviour than β-CD, dissolving upon increasing pressure (§3.2.1);[12] although dissolution occurs as early as 0.20 GPa, the final size of the crystal was acceptable for data collection (Fig. 5.2).

300 μm
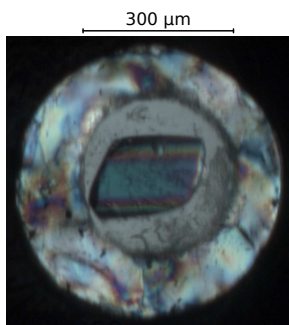
Figure 5.2: Single crystal of α-CD form I at 0.20 GPa.

The pressure range in which hydration takes places—from 0.0001 to 0.65 GPa—has been explored using three different crystals, measured at 0.0001, 0.20, and 0.65 GPa[12] (named **St1**, **St2**, and **St3**, respectively). To ensure that the differences observed between the lattice parameters of the three specimens are within the limits of reasonable experimental error, a compression study of α-CD was also performed. A triangular DAC with 800 μm diamond culets was used for this experiment (Appx. I.2). The cell was equipped with an Inconel 718 gasket pre-indented to a thickness of 160 μm and sample chamber 495 μm in diameter. Compression studies are usually performed by collecting X-ray diffraction data at different points on increasing pressure, but due to the dissolution of α-CD, the study was performed on decompression. The DAC was loaded with a crystal grown at ambient pressure, as described above, together with the saturated aqueous solution and pressure was increased to 0.60 GPa; complete dissolution of the crystal was observed at 0.40 GPa. Polycrystalline material was obtained on decompression and a single crystal suitable for diffraction was obtained at 0.54 GPa following the usual procedure of temperature cycling (§1.1.2). Data were collected on said specimen, and subsequently pressure was slowly released, collecting data at 0.40, 0.33, and 0.10 GPa.

5.2.2 X-ray data collection and processing

Data on the crystal at 0.20 GPa were collected at beamline F1 at DORIS (DESY) on a Huber four-circle diffractometer, with a combination of 0.5° ω and ϕ scans at two different values of χ (0 and 120°). Radiation of 0.5 Å with a beam 1×1 mm was used, exposing each frame for 40 s. The detector was a MAR165 CCD. Information in frame headers were updated with CMH and frames were converted from MAR to Bruker-Nonius Smart format using M2B.[13] A final conversion to SFRM format was performed with the tools included in APEX2.[14]

Data for the 0.0001 and 0.65 GPa pressure points come from two further crystal specimens which had been measured previously, as explained above, and the resulting structures published.[12]

Data for the compression study were collected using a MAR345 image plate and an INCOATEC microfocus (IμS) source of Ag radiation (0.56085 Å). Each collection consisted of 0.5° ϕ scans with 60 s exposure time. The setup of this diffractometer only allowed for ϕ scans, which was compensated by recollecting data with the DAC rotated 120° around the axis of the beam. MAR frames were directly converted to SFRM format using APEX2.

Data were integrated following the strategy outlined in §1.1.3.

5.2.3 Structure refinement of α-cyclodextrin hydrate at 0.2 GPa

At 0.2 GPa α-CD does not undergo any major structural transformation apart from an increase in hydration number, hence a low-temperature determination of form I (CSD reference code: CHXAMH04[12]) has been used as starting point for refinement. Full-matrix least-squares refinement against F^2 has been carried out using SHELXL v2014/7,[15] through the GUI SHELXLE.[16]

All non-H atoms were refined anisotropically, with the exception of three out of eight water molecules, for which no H atoms could be determined. Two of these three water molecules have an occupancy of 20%. Occupancy factors were initially refined and then fixed to the refined values. All bond distances and angles were restrained to target values automatically generated on the GRADE WEB SERVER.[17] Torsion angles were restrained in a similar manner, excluding those involving O6, which are known to be conformationally flexible. In addition, the geometry of all the glucose rings was restrained to be the same. α-CD H atoms were initially placed geometrically and constrained using the riding model; some positions were later improved by looking at difference Fourier maps. H atoms belonging to water molecules were located in Fourier difference maps when possible, their positions optimized according to the formation of possible H bonds and then refined subject to restraints. For further crystallographic data see Table 5.1.

Crystal data	
Chemical formula	$C_{36}H_{60}O_{30} \cdot 6.4H_2O$
Molecular mass (g mol^{-1})	1088.14
Temperature (K)	298(2)
Pressure (GPa)	0.20(5)
Crystal system, space group	orthorhombic, $P2_12_12_1$
a, b, c (Å)	14.8069(13), 33.970(3), 9.4140(8)
V (Å^3)	4735.2(7)
Z	4
Crystal size (mm)	$\sim$0.20$\times$0.15$\times$0.10
Data collection	
Radiation wavelength (Å)	0.5000
Absorption correction	Semi-empirical (using I meas.), SADABS[18]
T_{min}, T_{max}	0.5898, 0.7440
Number of $[I > 2\sigma(I)]$ reflections:	
- measured	35544
- independent	3157
- observed	2913
$(sin\theta/\lambda)_{max}$ (Å^{-1})	0.556
R_{int} (%)	4.65
Refinement	
$R[F^2 > 2\sigma(F^2)]$ (%)	3.57
$wR(F^2)$ (%)	9.08
S	1.03
Number of reflections	3157
Number of parameters	1797
Number of restraints	691
$\Delta\rho_{min}$, $\Delta\rho_{max}$ (e Å^{-3})	$-$0.205, 0.313

Table 5.1: Crystallographic data and refinement parameters of α-CD$\cdot$6.4H$_2$O.

5.2.4 Molecular dynamics

For a description of fundamental concepts in MD, refer to §1.2; the present section contains only technical details on how the simulations were implemented. All the simulations in this chapter use the CHARMM (Chemistry at HARvard Macromolecular Mechanics) force field.[19,20] Water molecules, either in the lattice or the bulk of the solvent, were parametrised as TIP3P.[21] The force field has been used without any modification of atomic charges or force constants. The same force field has been used to minimise the energy of all the molecular systems prior to simulation. All simulations in this chapter have been performed using NAMD (v2.9).[22]

5.2.4.1 Equilibrations of infinite crystals

These simulations are simple equilibrations of an infinite lattice of α-CD molecules (that is, without bulk solvent) let to evolve at constant pres-

sure values. Four different equilibrations have been carried out, by simulating two different setups at two different pressures each. The two setups are termed AP and HP and are based on the crystal structures at 0.0001 and 0.65 GPa, respectively.
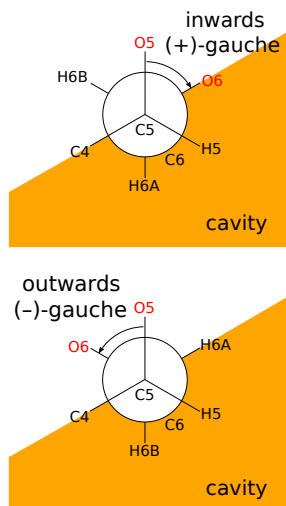
The nearly-cubic simulation box of the AP setup consists of six unit cells (two replicas on the *a* axis, one on *b*, and three on *c*) of the AP crystal structure of α-CD form I (CSD reference code: CHXAMH02[2]). According to the crystal structure, O6_1 can adopt both (+)- and (−)-gauche conformations (Fig. 5.3); the likelihood of the latter being only 10%. Such an small disorder was deemed to have no effect in the outcome of the simulation; hence it was not modelled in the setup. Lattice water in the crystal structure has been included in the setup; there are six water molecules per α-CD molecule, two (termed **WA** and **WB**) located inside the cavity of the macrocycle and four (termed **W1**, **W2**, **W3**, and **W4**) in intermolecular spaces.

α-CD form I at 0.65 GPa also exhibits disorder of the O6_1 group, with the (−) and (+) conformations equally populated. Furthermore, there are two additional lattice water molecules per α-CD molecule in the structure (termed **WC** and **WD**), located inside the α-CD cavity with an occupancy of only 40% each. The six water molecules **W1**-**4**, **WA**, and **WB** are also present in this structure at the same positions. The disorder is what distinguishes the HP structure from the AP one, hence it must be modelled in the setup. Because NAMD does not allow to model disorder by means of percentage populations in a unit cell, and rather requires fully occupied atoms which might or not be present in certain unit cells within a super cell, the simulation box of the HP setup was required to be larger than that of the AP setup (smallest setups were preferred whenever possible to reduce the computational cost of the calculations). The simulation box of the HP setup is composed of ten unit cells (two replicas on the *a* axis, one on *b*, and five on *c*) of the HP structure of α-CD form I determined at 0.65 GPa (CSD reference code: REGPAW[12]), which accounts for 40 α-CD molecules and 320 water positions, 240 of which are fully occupied (**W1**-**4**, **WA**, and **WB**) and 80 partially filled with only 32 water molecules (**WC** and **WD**). By

Figure 5.3: Definition of the (+)- and (−)-gauche conformations of the O6 groups in CD molecules.

random selection, the conformation of the 06_1 group was made (+) in half the α-CD molecules and (−) in the other half. From the latter, 16 positions **WC** and 16 **WD** (out of 20 each) were filled with water. This distribution of disordered features is in agreement with the hypothesis that the O6_1 needs to change from the (+) to the (−) conformation for the additional water to access the cavity of the macrocycle. For a complete description on how the disorder has been reproduced, see Appx. F.

Both setups were simulated in the NPT ensemble at constant pressures of 0.0001 and 0.65 GPa. The simulation of each setup as its respective pressure, AP at 0.0001 GPa and HP at 0.65 GPa, are intended to reproduce the behaviour of the crystal in real conditions. The other two simulations, with the pressures interchanged, allow to control for model bias (§5.3.2.1). Further details on how the simulations were performed can be found in Table 5.2.

Simulation length	100 ns
Time step	1 fs
Frame output	every 2000 steps (2 ps)
Number of frames	50000
Minimisation	100 steps (0.1 ps)
Force field	CHARMM
PBC	yes
Ensemble	NPT
Temperature control	Langevin
Temperature	298 K
Pressure control	Langevin (not flexible)
Pressure	0.0001 or 0.65 GPa
Treatment of interactions	
- Covalent	first and second neighbours
- Mixed	fourth neighbours
- Coulomb and Van der Waals	fifth and higher neighbours (PME)
Interactions cutoff	12 Å
Smoothing after cutoff	10 Å
Treatment of bonds with H atoms	rigid

Table 5.2: Details of the equilibrations carried out on infinite crystals.

The trajectories obtained from these simulations serve a double purpose. On the one hand, they are simple enough to test the performance of the force field at the simulation conditions. On the other hand, they provide and insight on the pressure dependence of the dynamics of the O6_1 group in the absence of bulk solvent.

A PYTHON script has been written to analyse the evolution of the disorder of the O6 groups during the trajectories. Although the group

of interest is 06_1, the analysis was applied to all O6 groups, to ensure that the entire structure was properly reproduced by the simulation. The script loops over all frames, identifying all O6 groups of all α-CD molecules in a frame and establishing their conformations, either (+)- or (−)-gauche by means of the torsion angle around the bond C5–C6. (Fig. 5.3).

Once the conformation of all O6 groups is known, a histogram of the torsion angles is created for each group in the simulation box to identify the equilibrium positions, which correspond to the maxima of the histograms. Apart from the expected (+) and (−) conformations, a third local maximum is seen in many histograms and was identified as an intermediate state in the conformational change between (+) and (−), based on shorter residence times (evaluated visually as area of the histogram). Each equilibrium position of each O6 was defined by the mean value of the torsion angle plus and minus 3σ calculated from a normal curve fitted to the corresponding maximum of the histogram. The script loops over the complete list of torsion angles, assigning them to the corresponding conformation of each O6 group. Torsion angles outside equilibrium positions are rejected as outliers—they correspond to transition conformations. The total residence time of each O6 group in each conformation can be calculated from the number of torsion angles assigned to it. Finally, residence times in each conformation of O6 groups corresponding to the same glucose ring of all α-CD molecules in the simulation box were combined and expressed as percentages, which is equivalent to the usual expression of disorder in crystal structures (space and time average).

The script can function in an entirely automatic way, but graphic output is provided to permite user input in cases where more than three maxima existed in the histogram. Nevertheless, such cases were rare and statistically not significant.

5.2.4.2 Equilibrations of molecular clusters in water

These equilibrations follow the same pattern described in the previous section: two setups, AP and HP, simulated both at 0.0001 and 0.65 GPa,

although the setups employed were different. The boxes in this set of simulations consisted of a finite cluster of α-CD molecules submerged in liquid water, in an attempt to replicate the behaviour of a crystal surrounded by bulk solvent. PBC were still in use to avoid surface effects on the borders of the box.[23] The molecular clusters need to be larger than the ones described in the previous section, to accurately describe the properties of the crystal, while still keeping the amount of molecules to a minimum. Both AP and HP clusters were built from the setups described in the previous section, by doubling the number of unit cells repeated along the a axis from two to four. Such clusters contain at least four layers of α-CD along each axis; in this way there is a surface of α-CD molecules surrounding a small core of two layers of molecules, which cannot be directly influenced by the solvent. Although much larger clusters would be desirable, these were not feasible due to computational limitations. The disordered features of the HP cluster were remodelled, following the same criteria outlined previously, to ensure a random distribution over the entire cluster (Appx. G).

Each cluster was completely surrounded by a 12 Å-thick layer of water, containing over 29000 water molecules. The simulations were performed as described in Table 5.2, although the minimisation time was increased to 20 ps (the system used in the previous section was already close to the energy minimum, while that is not the case here) and the simulation length was limited to 50 ns (to reduce consumption of computational resources), producing only 25000 frames.

The aim of these simulations was to study the effects of hydrostatic pressure applied on the crystal via bulk solvent. By analysing the disorder of the O6 groups and comparing the results with those obtained from the crystal without bulk solvent, it would be possible to determine whether the disorder is caused by pressure alone or if it requires the presence of solvent. Furthermore, inclusion events, that is, water molecules moving from the bulk solvent into the crystal, could in principle occur and be monitored, which would allow to quantify the hydration process undergone by the crystal.

The disorder of the O6 groups in the trajectories of these simulations was analysed in the same way explained in the previous section. In addition, the evolution of the water content of the α-CD cavities was also measured with a custom-made PYTHON script. In a given frame, the script first identifies all the α-CD molecules by means of their six O-glycosidic atoms (O4 atoms); it then finds the centre of mass of said atoms (which is a computationally-efficient approximation to the real centre of mass of each molecule) and finally creates a virtual sphere—representing the α-CD cavity—with centre at the centre of mass and radius equal to the average distance between the centre and the O4 atoms (calculated dynamically) minus the Van der Waals radius of an O atom (1.5 Å). In the next step, the script loops over all water molecules in the frame and creates a list of the water content of each sphere, by comparing the coordinates of the sphere and the water molecules. The script repeats this process for each frame of the trajectory to output a graph of the evolution of the water content of each α-CD during the simulation length. The evolution is calculated as a moving average of the instantaneous water content in each frame. Finally, results for each independent α-CD molecule are averaged, similarly to the space average observed in an unit cell obtained by X-ray diffraction.

5.2.4.3 Umbrella sampling

The simulations described in the two previous sections aimed at determining properties of the hydration process which α-CD undergoes under pressure from the point of view of geometric parameters—position of O6 groups and water molecules—but do not allow derivation of energy values, mainly due to insufficient sampling of intermediate states between the equilibrium positions. Umbrella sampling is a type of steered MD simulation which overcomes the sampling problem and allows for easy determination of Gibbs free energy curves along known paths (§1.2.3). In the context of this work, umbrella sampling was used to determine the energy curve of the conformational change of O6 groups.

Three different energy curves of a single O6_1 group were derived from umbrella sampling calculations:

- **US1:** AP setup at 0.0001 GPa

- **US2:** HP setup at 0.65 GPa

- **US3:** AP setup at 0.65 GPa

All setups in this section correspond to those described in §5.2.4.1; the energy curves correspond to an O6 group deep within the crystal which feels no effects from the bulk solvent. The calculations follow the same principle stated previously: to replicate the behaviour of the corresponding stable form at ambient- and high-pressure conditions (**US1** and **US2**), plus an additional reference calculation to control for model bias (**US3**).

Each calculation is composed of 120 independent simulations, 1.5 ns in length each. The rest of the simulation parameters were identical to those in Table 5.2 (frame output was increased to every 5000 steps, producing only 300 frames for visual inspection; generation of energy curves does not rely on frame output). In each simulation the torsion angle C4–C5–C6–O6 of the α-CD unit 1 was set to a different value and restrained by a harmonic potential with force constant 2 kJ/mol, effectively restraining the angle to rotate within $\pm 4°$ of its initial position. The angle was set to 180° in the first simulation, and subsequently moved by 3° to complete a full 360° rotation in 120 simulations. Because the degree of rotation allowed by the harmonic potential is $\pm 4°$, but the central position is changed only 3° in every simulation, there is a certain degree of overlap of the sampled space between contiguous simulations, which is necessary to compute reliable energy curves. Stronger potentials are usually not desirable, as they would excessively restrict the flexibility of the torsion angle, limiting the overlap between simulations and forcing to perform more of them, with the consequent increase in computational cost. Weaker potentials tend to be a better choice, as they allow for wider sampling in each simulation, which in turn allows to space out the potentials reducing the number of simulations needed; albeit for this particular system it was found that the torsion

angle skewed towards certain values if weaker potentials were used. The weighted histogram analysis method, as implemented in WHAM,[24] was used to derive the energy curves from the simulations.

5.2.4.4 Metadynamics

In the same way that umbrella sampling has been used to estimate the energy change associated with the conformational change of O6_1, metadynamics calculations were implemented to determine the energy cost of incorporating a water molecule into the cavity of an α-CD molecule. The main difference with respect to the previous sections is that, while the path followed by the torsion angle on its rotation can be easily described, the path followed by a water molecule during an inclusion event remains unknown. Metadynamics, although very time consuming, provides a way to measure the Gibbs free energy landscape of a complete unit cell, from which the minimal energy path that a molecule can follow can be extracted (§1.2.3).

A single 400 ns simulation was performed at 0.65 GPa, based on the HP setup described in §5.2.4.1. The rest of the simulation parameters were identical to those in Table 5.2. 3D Gaussian potentials, simply referred as hills—0.5 kJ/mol in height and 0.4 Å in width in all directions—were applied to the position of the water molecule **WC** of α-CD unit 10 (to define the unit cell to be explored in the centre of the simulation box) every 1 ps, forcing the molecule to move continuously across the unit cell. The energy calculation was limited to a single unit cell inside the simulation box by creating soft potential walls (force constant of 200 kJ/mol) 1 Å away of the unit cell faces, which do not allow the target water molecule to cross through them. 1 Å-thick soft walls were used to ensure that the target molecule gently bounces back inside the limited region rather suffering an abrupt change in trajectory, which might cause errors when computing energies near the walls. Well-tempered metadynamics was used, that is, the Gaussian hills gradually decreased in size as the simulation progressed. This generates a complete, albeit low resolution, energy landscape in a relatively short time, while the rest of the simulation length is devoted to increase the level of detail to produce more

accurate energies, as opposed to mapping each area at high resolution before moving to the next one. Well-tempered dynamics offers two main advantages: it is possible to stop the simulation at any time and still obtain results, and it provides a simple way to monitor the convergence of the calculation, by looking at the moving average of the height of the Gaussian hills, which tends to zero (Fig. 5.4).
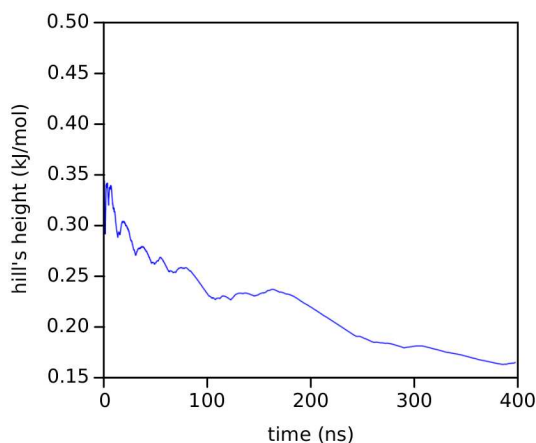
Figure 5.4: Convergence of the metadynamics calculation. Hill's height decreases rapidly at the beginning, halving in 100 ns. Decrease over 300 ns occurs very slowly. Convergence has not been reached after 400 ns; however, qualitative results can be considered accurate at this point.

Free energies are derived from the value of the potential applied in each step and the position in which it is applied, measured as the position of the target molecule relative to its initial position. Therefore, it is exceedingly important to avoid any rotation or drift of the system inside the simulation box. For an infinite crystal, as it is the case here, this can be achieved by fixing the position of a single atom. Fixing atoms can bias the results, and for this reason the fixed atom (C1_1 of α-CD unit 1) has been chosen outside the unit cell in which the calculation is performed.

NAMD has no implementation to use atomic coordinates as targets of energy calculations. PLUMED (v2.1.1),[25] a plugin which implements such functionality, was used in combination with NAMD for this simulation. SUM_HILLS, a tool integrated in PLUMED, was used to process the trajectory into a free energy landscape. To represent the 3D energy landscape the sampled space has been divided in to voxels (cubes 0.14 Å in linear length), each of them being characterised by its position within a 3D grid and energy values. A PYTHON script implementing MAYAVI[26] has been written to create a coloured 3D representation of the energy

119

landscape from the SUM_HILLS output. The 3D landscape can only be accessed electronically, and for the purpose of 2D representation the script allows to produce slices of the landscape.

5.3 Results and discussion

5.3.1 Crystallographic evidence for hydration

Crystallographic results are based on three structures of α-CD at three different pressures:

- **St1:** at 0.0001 GPa (CSD reference code: CHXAMH02 [2])

- **St2:** obtained by direct compression to 0.20 GPa of a crystal grown under ambient-pressure conditions, as described in §5.2.1

- **St3:** obtained by in-situ crystallisation at 0.65 GPa (CSD reference code: REGPAW [12])

The three structures were determined from different crystals, hence a compression study was performed to check whether the lattice parameters were consistent across all structures. Lattice parameters of all structures can be found in Table 5.3; Fig. 5.5 shows the results in graphic form. Axes *a* and *c* compress under the application of pressure, while axis *b* elongates; overall, the unit-cell volume decreases. The variation of lattice parameters follows a similar trend for all experiments, with the exception of the *b* axis, which is longer at 0.55 GPa (from the compression study) than at 0.65 (from the in situ crystallisation experiment). This could be due to the fact that these pressure points were measured with different crystals, or could also be due to an experimental error because of the very low completeness of the 0.55 GPa dataset (31%). Nevertheless, given that the rest of the parameters follow similar trends, even when coming from different crystals, it is considered that the results coming from different crystals are consistent.

A comparison of the packing of **St1**, **St2**, and **St3**, using the *Crystal Packing Similarity* module in MERCURY [27] (using only the core α-CD structure, without O2, O3, C6, and O6 atoms) shows that the three structures are very similar. The RMS between **St1** and **St2** is 0.097,
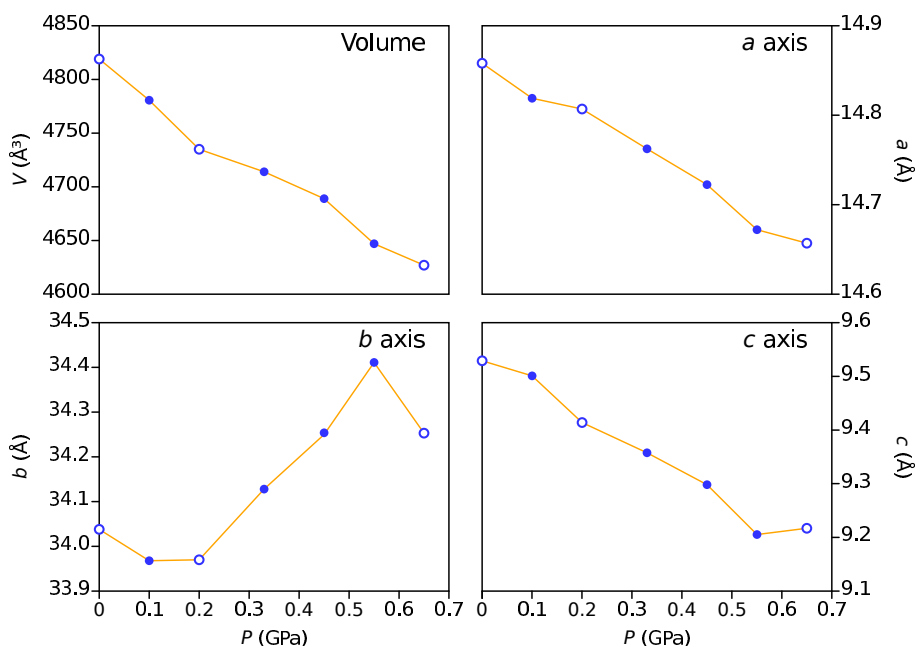
Figure 5.5: Variation of the lattice parameters and unit-cell volume of α-CD form I with pressure. Parameters corresponding to structures **St1**, **St2**, and **St3** are indicated with an empty icon. Parameters corresponding to structures from the compression study, with a full icon. A line linking all icons has been included for visual reference only.

between **St1** and **St3** is 0.289, and between **St2** and **St3** is 0.196; the small difference can be attributed to the effects of pressure. In any case, the 15 molecules constituting the cluster used in the comparison match across the three structures.

The hydration process undergone by α-CD under pressure is characterised by the following (see also Table 5.4 and Fig. 5.6):

- The water content in the cavity of the α-CD molecules remains constant at 2 molecules (**WA** and **WB**) up to 0.10 GPa. Above this pressure, the water content gradually increases up to 2.8 molecules over 0.55 GPa. Inclusion of water molecules takes place via two partially-occupied positions (**WC** and **WD**), with similar fractional coordinates observed across all structures.

- The water content outside the cavity, 4 water molecules per α-CD

Structure	P (GPa)	a (Å)	b (Å)	c (Å)	V (Å³)
St1[2]	0.0001	14.858(4)	34.038(10)	9.529(2)	4819(2)
St2	0.20	14.8069(13)	33.970(3)	9.4140(8)	4735(1)
St3[12]	0.65	14.657(5)	34.253(14)	9.217(4)	4627(3)
Compression	0.10	14.819(2)	33.968(3)	9.501(1)	4782(2)
	0.33	14.762(2)	34.128(3)	9.3576(11)	4714(2)
	0.45	14.722(2)	34.254(4)	9.2982(11)	4689(2)
	0.55	14.672(3)	34.411(5)	9.2053(16)	4647(2)

Table 5.3: Lattice parameters of α-CD form I at different pressures.

molecule, remains invariant during compression.

- The group O6₁ can adopt two conformations: (+) and (−), pointing inwards and outwards the cavity, respectively. As the inclusion of water progresses, the (−) conformation becomes increasingly populated, from 10% at 0.10 GPa to 50% at 0.55 GPa.

- Changes in structure and hydration can occur in the solid state and do not require recrystallisation.

Table 5.4: Summary of the differences observed in α-CD form I at different pressures.

| Structure | P (GPa) | Water molecules[a] | | O6_1 conformation (%) | |
		In	Out	(+)	(−)
St1[2]	0.0001	2	4	90	10
St2	0.20	2.4	4	80	20
St3[12]	0.65	2.8	4	50	50
	0.10	2	4	90	10
Compression	0.33	2.6	4	75	25
	0.45	2.7	4	65	35
	0.55	2.8	4	50	50

[a] Number of water molecules located inside and outside of the α-CD cavity.

Figure 5.6: Increase in the occupancy of **WC** and **WD** (blue) and the (−)-gauche conformation of O6₁ (orange) with pressure. Parameters corresponding to structures **St1**, **St2**, and **St3** are indicated with an empty icon. Parameters corresponding to structures from the compression study, with a full icon. Straight lines are given for visual reference only.
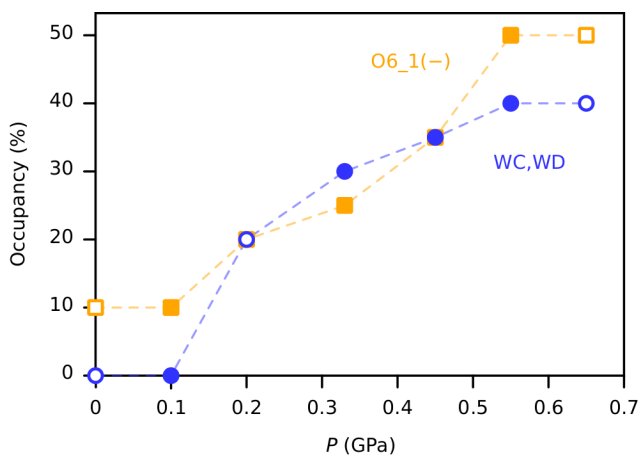
These observations evidence that the inclusion of water in the α-CD cavities and the increased disorder of the 06₁ group must be related to the increase of pressure. Previous observations showed that the displacement of the O6₁ group from the (+) to the (−) conformation increases the solvent-accessible volume of the cavity, however, that is not sufficient to explain how the inclusion happens, because there are no channels in the structure which could allow for the movement of water to reach the cavities.[12] Three different mechanisms can be proposed to explain the hydration in the solid state:

- **Sequential, disorder-driven:** the conformational change of O6_1 is caused by pressure, opening the cavity and allowing water to flow in. The opening of the cavities should happen in such a way that channels across the crystal are created.

- **Sequential, inclusion-driven:** pressure forces the water inside the crystalline structure and the molecules are accommodated in the α-CD cavities; the conformation of O6_1 must change in order to allow the flow of water.

- **Concerted:** pressure causes both effects—water flow inside the crystal and O6_1 conformational change—at the same time, without said effects being related directly.

The likelihood of each mechanism is discussed in detail in the next section based on the results from the MD calculations.

5.3.2 Hydration mechanism by molecular dynamics

The validity of the results obtained by MD depends on the performance of the force field at the conditions under which each simulation is carried out, as introduced in §1.2.4. The results outlined in this section are in qualitative agreement with experimental values measured by single-crystal X-ray diffraction; however, a quantitative interpretation would not be appropriate, as numerical results from the simulations tend to be shifted with respect to the experimental observations. For instance, simulated unit-cell volumes can differ up to ± 0.5 Å^3 from the experimental ones and Gibbs free energy minima determined from umbrella sampling are systematically shifted (§5.3.2.3). These quantitative differences are likely to stem from the fact that the CHARMM force field has been used here without any modification, and the force constants probably need to be optimised for each specific pressure. Notwithstanding, these results are consistent in all simulations, as explained in the following sections. Such consistency is an argument in favour of the simulations here presented; although all conclusions are drawn from comparison of simulations performed in a similar fashion and the intention here is not to give absolute results. Differences between simulated and experimental results are discussed when required.

5.3.2.1 Equilibrations of infinite crystals

The simulations of a crystal of α-CD in the absence of bulk solvent allow to determine whether the conformational change undergone by 06_1 is triggered by pressure alone or if it requires the inclusion of additional water. Table 5.5 contains information on the relative occupancy of the (−)-gauche conformation of the O6 groups (pointing outwards the α-CD cavity) determined from the four simulations, that is, two different initial setups at two pressures each, as described in §5.2.4.1.

Table 5.5: Relative occupancy of the (−) conformation of the O6 groups in the simulation of α-CD infinite crystals and in the experimental crystal structures. Values approximated to the nearest integer (Appx. H for complete results).

Setup	AP	HP	AP	HP	Crystal structures	
P (GPa)	0.0001	0.65	0.65	0.0001	0.0001 GPa	0.65 GPa
06_1	0	48	11	45	0	50
06_2	99	99	95	99	100	100
06_3	0	2	0	4	0	0
06_4	99	99	99	99	100	100
06_5	95	97	99	98	100	100
06_6	99	99	97	99	100	100

The largest change observed with respect to the initial state is an 11% increase in the disorder of O6_1 of the AP setup simulated at 0.65 GPa. This is considered not significant, as it is known that O6_1 can be disordered between the two gauche conformations up to 10% even at ambient pressure.[2] The slight differences observed between the simulations of the same setup at different pressures evidence the absence of model bias. It is clear that the pressure at which the simulations are carried out does not have a direct effect on the disorder of the O6 groups, as these remain in a similar state as in the initial setup. This observation points to the *sequential, inclusion-driven* mechanism of hydration.

5.3.2.2 Equilibrations of molecular clusters in water

These simulations encountered two major problems at both ambient and high pressure: dissolution and movement of the cluster. Dissolution should be avoided, as it reduces the number of α-CD molecules in the cluster which can be analysed at the end of the simulation. Furthermore, because dissolution is not reproducible in different simulations, it makes comparison of the results more difficult. Movements of the crystal, mainly rotation and drift, introduce unnecessary complications in

the analysis of the trajectories (e.g.: redefinition of the reference frame of the crystal in every frame) and can cause instabilities if two clusters from adjacent simulations boxes get in touch. The latter can be avoided by using a simulation box sufficiently large to contain the full rotation of the cluster, at the cost of increasing the amount of bulk solvent and in turn the computational cost of the simulation. The observation of these two effects corresponds to a physically meaningful processes. It can only be controlled by constraining certain simulation parameters, albeit the number of constrains should be kept to a minimum, otherwise the simulated system might not be representative of a crystal.

In this case it was found appropriate to fix a single atom (C1_1) of the eight α-CD molecules at the corners of the cluster. This is enough to completely avoid movements of the cluster, and also greatly reduces dissolution, as this tends to happen in the corners, where CD molecules are more exposed to the solvent. As indicated in §5.2.4.4, by fixing a single atom per α-CD molecule the flexibility of the macrocycle is not compromised. The use of this type of constraints poses an additional problem: because the volume of the cluster is virtually fixed, applying pressure on the system does not lead to the compression of the cluster; it is essentially treated like a very incompressible material. In any case, because of its very small size, compression of the cluster would be negligible when compared to that of the bulk solvent, for which no constraints were applied during the simulation.

Even after the introduction of the constraints, this set of simulations failed to reproduce experimental results. With respect to the O6 groups, it was observed that all of them underwent conformational change to a certain extent, independently of the pressure. Because the final results are an average of all the groups in the cluster, the analysis was repeated separately for the α-CD molecules on the outer layer and the core of the cluster, showing that the conformational change only took place in the outer layer, while molecules of the central core remained unaltered.

When computing the number of water molecules inside the α-CD cavities, it was found to be 2 at 0.0001 GPa and 2.3 at 0.65 GPa. Expecting similar surface effects, the calculation was also performed sepa-

rately on the outer and core layers of the cluster; here it was found that, while inclusion of water molecules does take place only at high pressure, it is limited to the surface of the crystal. If these simulations were to be scaled to the size of the crystals measured experimentally, a simple surface effect cannot account for an increase of the hydration number from 6 to 6.8. This set of simulations was discarded, on the basis that it could not reliably reproduce experimental results.

5.3.2.3 Umbrella sampling

Fig. 5.7 shows the three Gibbs free energy curves of O6_1 at different conditions calculated by umbrella sampling. All the curves exhibit three minima, with different relative stabilities. The minimum at 187° corresponds to the (+)-gauche conformation, for which an experimental torsion angle of 170.46° is measured at 0.0001 GPa and 166.22° at 0.65 GPa. The minimum at 53° corresponds to the (−) conformation, which exists at 0.65 GPa with a torsion angle of 37.62°. There is an additional minimum at 282°, which due to its higher relative free energy can be considered as an intermediate state. The existence of this state has already been addressed in §5.3.2.1. There is a clear systematic difference of ∼15° between the values determined from the simulation and those measured experimentally. The origin of this shift is currently unknown, however, the curves can be considered qualitatively correct, as discussed at the beginning of this section.

At 0.0001 GPa, from **US1** (see §5.2.4.3 for definitions), the preferred conformation of O6_1, with the lowest Gibbs free energy, is (+)-gauche, that is, pointing towards the cavity. Although this is usually not the preferred orientation, it is favoured for this particular glucose ring to stabilise the water inside the α-CD cavity by H bond formation. The (−) conformation is slightly higher in energy, 4.5 kJ/mol, and can also exist at ambient pressure, as known from the fact that a small degree of disorder in between the two conformations is observed in the crystal structure at 0.0001 GPa. Notwithstanding this, the energy barriers for the rotation are high enough to hinder the conformational change, that is, the disorder is static rather than dynamic. At 0.65 GPa, from **US2**,
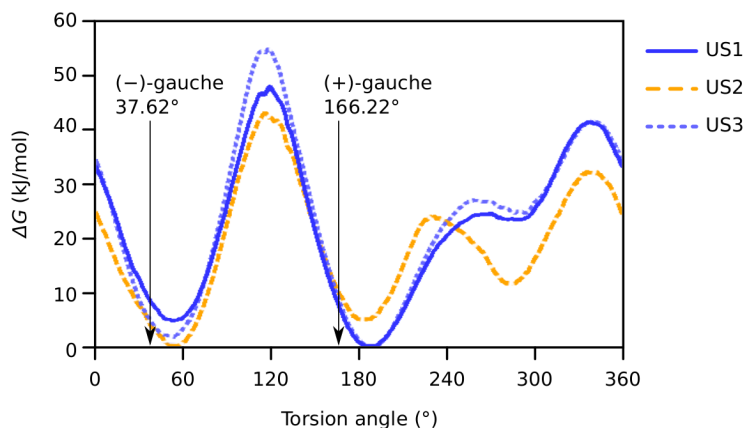
Figure 5.7: Gibbs free energy curves for the rotation of O6 groups measured by the torsion angle C4–C5–C6–O6.[a] **US1:** solid blue, AP setup at 0.0001 GPa. **US2:** dashed orange, HP setup at 0.65 GPa. **US3:** dotted light blue, AP setup at 0.65 GPa.

[a] 0° corresponds to C4–C5 and C6–O6 bonds being eclipsed. The (−)-gauche conformation (O6 outwards the α-CD cavity) corresponds to $\sim$60°. The (+)-gauche conformation (outwards) corresponds to $\sim$180°. Corresponding torsion angles in the crystal structure at 0.65 GPa are given as a reference. For further information on how the angles are defined, see Appx. H.

the relative energies of the minima are inverted. The (−) conformation is the absolute minimum at this conditions, with the (+) conformation 4.9 kJ/mol higher in energy. In addition, the intermediate state is stabilised by 11.5 kJ/mol upon increasing pressure, which provides a route with substantially lower energy barriers (a maximum of 20.1 kJ/mol) through which the conformational change can occur.

The energy curve obtained from **US3** coincides with that from **US1**, which indicates that the change in the energy curve is not caused by the application of pressure, but also by the presence of two additional water molecules, **WC** and **WD**, which were only present in the HP setup employed in **US2**. These observations point to the *sequential, inclusion-driven* mechanism of hydration, as it seems that the disorder of O6_1 is induced by the presence of water.

5.3.2.4 Metadynamics

The 3D relative Gibbs free energy landscape of the unit cell of α-CD form I at 0.65 GPa, determined from the HP setup, cannot be shown on print and is only available electronically. A script has been written to explore the data and obtain 2D slices of the landscape (§5.2.4.4). Fig. 5.8 shows slices parallel to the *bc* plane at different positions in *a*.

Figure 5.8: Slices of the Gibbs free energy landscape of α-CD form I at 0.65 GPa. Read from left to right (across the two pages) and top to bottom with increasing values of a. All slices are parallel to the bc plane, with $b \in [0, b/2]$ and $c \in [0, c]$ (origin at the bottom right corner), at every 0.42 Å (three voxels) in a. Slices at a below 0.00 Å and over 14.80 Å have been included because the calculation extends outside the unit cell due to the use of soft walls to limit the sampled space. The remaining half of the unit cell (along b) can be generated by symmetry. The approximate outline of the two α-CD molecules present in the region (labelled **1** and **2**) have been projected on each slice. Energy levels are colour-coded from 0 to 105 kJ/mol, being blue the lowest energy and red the highest.

$a = -0.84$ Å

$a = -0.42$ Å

$a = 0.00$ Å

$a = 1.68$ Å

$a = 2.10$ Å

$a = 2.52$ Å

$a = 4.20$ Å

$a = 4.63$ Å

$a = 5.05$ Å

$a = 6.73$ Å

$a = 7.15$ Å

$a = 7.57$ Å

$a = 9.25$ Å

$a = 9.67$ Å

$a = 10.09$ Å

$a = 11.77$ Å

$a = 12.19$ Å

$a = 12.61$ Å

$a = 14.30$ Å

$a = 14.72$ Å

$a = 15.14$ Å

$a = 0.42$ Å $\quad\quad$ $a = 0.84$ Å $\quad\quad$ $a = 1.26$ Å

$a = 2.94$ Å $\quad\quad$ $a = 3.36$ Å $\quad\quad$ $a = 3.78$ Å

$a = 5.47$ Å $\quad\quad$ $a = 5.89$ Å $\quad\quad$ $a = 6.31$ Å

$a = 7.99$ Å $\quad\quad$ $a = 8.41$ Å $\quad\quad$ $a = 8.83$ Å

$a = 10.51$ Å $\quad\quad$ $a = 10.93$ Å $\quad\quad$ $a = 11.35$ Å

$a = 13.03$ Å $\quad\quad$ $a = 13.45$ Å $\quad\quad$ $a = 13.88$ Å

$a = 15.56$ Å $\quad\quad$ $a = 15.98$ Å $\quad\quad$ ΔG (kJ/mol)

129
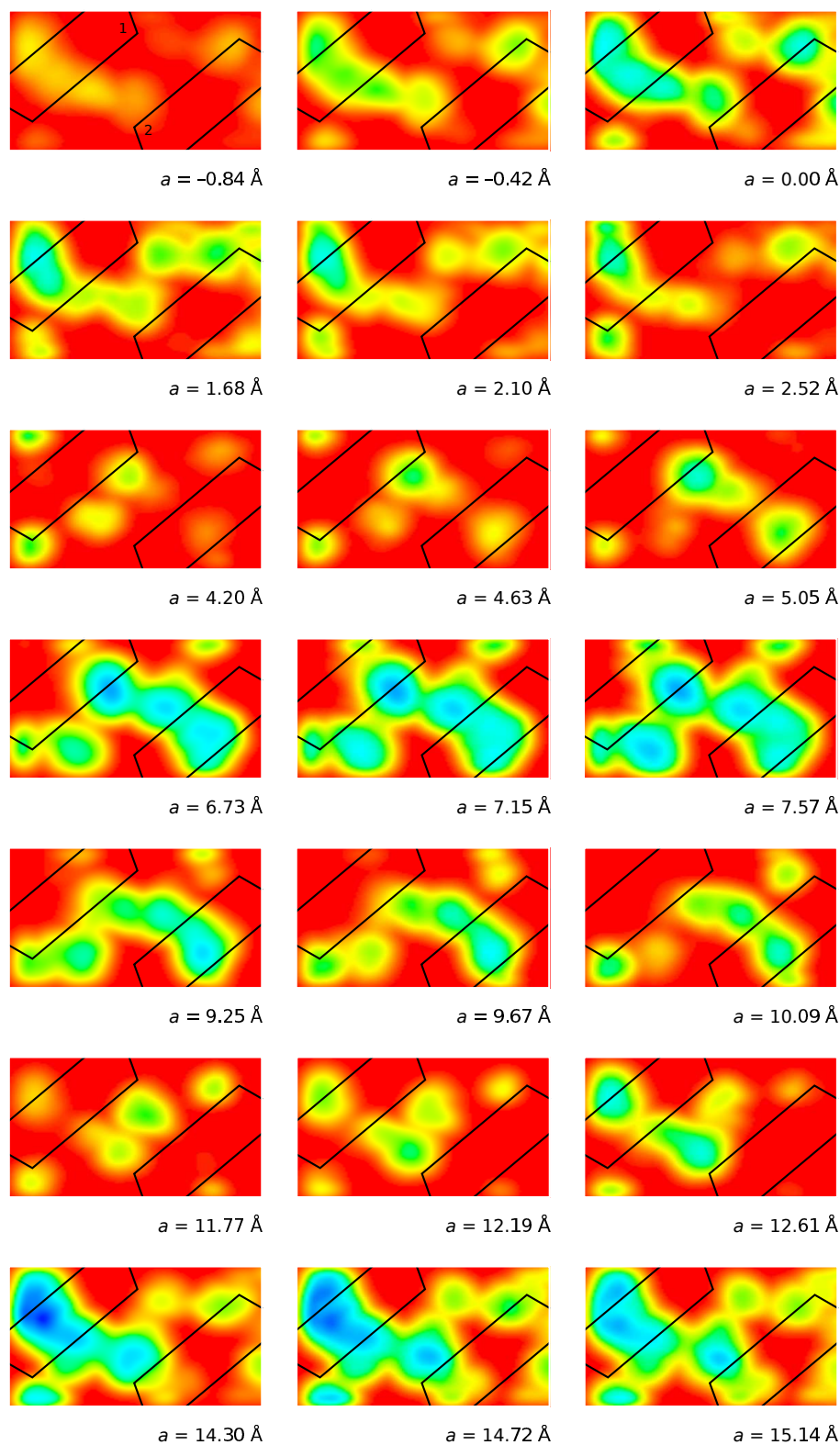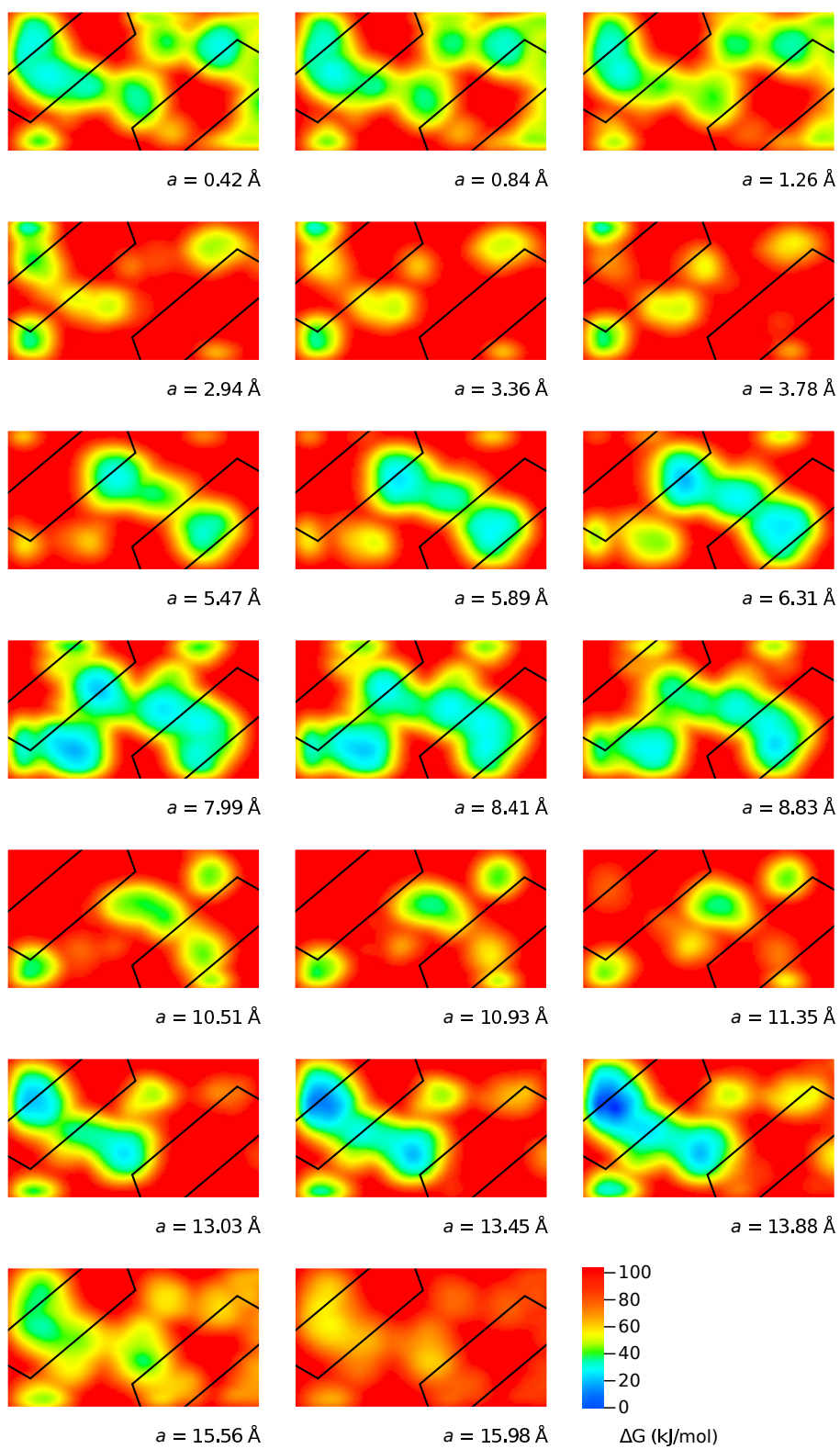
The two symmetry-related α-CD molecules (**1** and **2** in Fig. 5.8) which are almost entirely contained in the space of the unit cell defined in the figure have been also represented.

The energy maps show that the lowest energy regions in the crystal structure, depicted in dark blue, correspond to the interior of the cavities and the intermolecular spaces, were lattice water molecules are located. The highest energies, depicted in red, are associated with regions in which atoms from the α-CD molecule cross through the plotted slice. It can be seen between $a = 5.47$ Å and $a = 9.67$ Å that there is an area, with energies below 50 kJ/mol, connecting the cavities of both macrocycles. A water molecule could potentially follow a path across this area to move between the cavities. Further exploration of the geometry of the path, paying special attention to the arrangement of the O6 groups in its vicinity, shows that O6_1 of molecule **1**, which exhibits (+)-gauche conformation, effectively blocks the path.

Movement of a water molecule from the intermolecular space into an α-CD cavity consists of a destabilising component—breaking the intermolecular interactions involving the water molecule and the group O6_1, which has to rotate to allow for the movement—and an stabilising component—formation of new interactions. Typically, moderate H bonds, as those seen in the structures of α-CD, exhibit energies between 20 and 60 kJ/mol,[28] which are in the same order of magnitude as the process being described. For instance, the *PV* work per α-CD at 0.65 GPa is 32.5 kJ/mol, which could be sufficient to trigger H-bond rearrangement. In the case of the group O6_1, rotation and rearrangement of H bonds have been estimated at 38.6 kJ/mol by umbrella sampling (in two steps, see Fig. 5.7). Other interactions might play a role too, but are likely to contribute in a lesser extent.

5.3.3 High-pressure hydration mechanism

Based on the determination of the crystal structure of α-CD at three different pressures and the results from MD presented above, it is possible to postulate a *sequential, inclusion-driven* mechanism for the hydration of an α-CD form I crystal induced by high pressure, clearly distinct from

the induced-fit mechanism which operates at ambient conditions.[4]

In both mechanisms, α-CD molecules start in the *tense* conformation: the overall shape of the macrocycle resembles a distorted circle, with one glucose ring tilted with respect to the others, and two O6 groups in the less-preferred (+)-gauche conformation (Fig. 5.9a). At ambient pressure, the presence of a guest of suitable size causes the conformation to change from *tense* to *round* (Fig. 5.9b); in certain cases this allows the inclusion of additional water (Fig. 5.9c). Eventually, the water is displaced in favour of the guest, with the α-CD remaining in the *round* conformation.
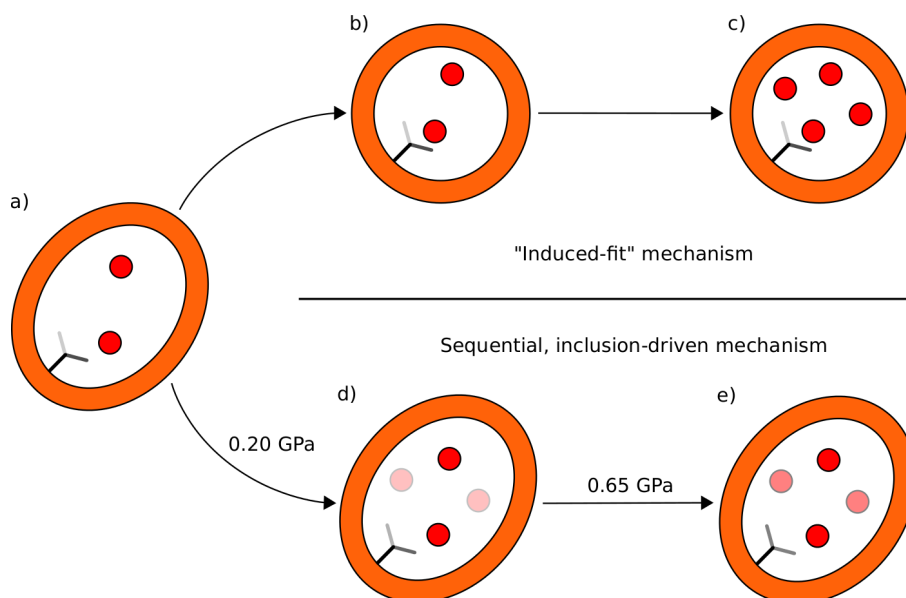
Figure 5.9: Sketch of the two possible inclusion mechanisms operating on α-CD form I. **a)** Initial state, *tense* conformation. **b)** α-CD changes to the *round* conformation in the presence of a suitable guest. **c)** A higher water content in the cavity is possible after the conformational change. **d)** Upon application of pressure, 0.4 additional water molecules are included in the α-CD cavity maintaining the *tense* conformation, albeit the O6_1 group is partially rotated outwards the cavity. **e)** 0.8 water molecules are included at 0.65 GPa, with O6_1 populating equally both (+) and (−) conformations.

If pressure is applied on form I, bulk water molecules are pushed into the crystal. Overall, water inclusion and crystal volume reduction explain why crystals of α-CD dissolve in water upon compression, but given the right conditions (that is, recrystallisation on decreasing pressure) water can be accommodated inside the cavities of the α-CD molecules. The process takes place in the solids state, that is, hydration is a continuous process and the structures reported here at different pressures represent different hydration states of the same structure and not different hydrates. In other words, water inclusion is a gradual process that occurs without distorting the overall packing of the crystal structure or

the conformation of individual α-CD molecules, requiring only the conformational change of the O6_1 group, which allows for the movement of water in the absence of natural channels in the crystal and for larger cavity volumes (Fig. 5.9d,e). Energy calculations from MD simulations also support this proposal.

5.4 Conclusions and outlook

A series of single-crystal X-ray diffraction experiments on α-CD hydrate form I have been undertaken. The crystallographic study shows that a gradual uptake of water occurs upon compression, with the number of water molecules occupying the CD cavity increasing from 2 at ambient pressure to 2.8 at 0.65 GPa. Moreover, it clarifies with respect to previous studies that this continuous process takes place in the solid state and does not require in-situ recrystallization. A change in the conformation of one of the hydroxy groups of the macrocycle (O6) accompanies the change in hydration. By means of MD simulations, a feasible mechanism which explains all the features observed in the crystal structures at different pressures has been proposed. Such a mechanism, which is different from the *induced-fit* mechanism via which α-CD forms inclusion complexes in solution, considers that water is first pushed into the crystal upon application of pressure, and then travels through the crystal thanks to temporary opening of channels via a conformational change of the O6 groups.

This study provides an example of the two radically different inclusion mechanisms that take place either in solution or in the solid state. One of these mechanisms is only accessible when pressure is applied to the system, highlighting the importance of exploring this thermodynamic variable for *getting the complete picture*. In this respect, α-CD is an ideal model system to study these phenomena using both diffraction and MD methods. For instance, pressure-induced denaturation of proteins is believed to occur via a similar process—penetration of water molecules into the protein hydrophobic core[29,30]—although the exact mechanism operating at the molecular level is still not fully understood.

To the best of our knowledge, combined high-pressure MD and diffraction studies have been performed only on a handful of proteins; this is probably because proteins are by nature considerably challenging to study (both computationally and experimentally); in this respect, focussing on smaller, more amenable model systems, such as CDs, may provide information which could be extrapolated to better understand the behaviour of complex biomolecules.

Enough evidence has been provided to support the new *sequential, inclusion-driven* hydration mechanism of α-CD at high pressure; however, further high-pressure studies would be desirable to fully understand the hydration process, as this can shed light on similar molecular-level processes occurring in other materials. The simulations performed thus far provide indirect evidence in support of one particular mechanism. While the inclusion event does happen during the MD simulations it has not been possible to actually isolate it, probably because it occurs in the time frame between the actual outputs: in order to explore the mechanism long simulations are needed (for instance, to ensure convergence and reproducibility); however, given the enormous amount of data that has been produced, frequent outputs are simply unfeasible. As a next step, shorter simulations focussing on the minimal energy path that water molecules can follow inside the crystal should be performed. This will imply to shift the focus from the cavities to the intermolecular spaces, which have not been considered in this study because their water content does not vary. It is likely that the number of water molecules inside said spaces does in fact change, but their movements cannot be observed by X-ray diffraction. Another aspect to consider in future research is the possible correlation between the high-pressure hydration mechanism and the observation that α-CD crystals dissolve in water as pressure is increased. Supposing that the inclusion of water destabilises the crystal, it would be interesting to know if it is possible to include more than 2.8 water molecules per α-CD cavity without destroying the crystal. Understanding the dissolution behaviour might help in understanding if and under which conditions high-pressure complexation of small organic molecules might take place in solution.

References

[1] P. C. Manor and W. Saenger, *J. Am. Chem. Soc.*, 1974, **96**, 3630–3639.

[2] B. Klar, B. Hingerty and W. Saenger, *Acta Cryst.*, 1980, **B36**, 1154–1165.

[3] W. Saenger and T. Steiner, *Acta Cryst.*, 1998, **A54**, 798–805.

[4] W. Saenger, M. Noltemeyer, P. C. Manor, B. Hingerty and B. Klar, *Bioorg. Chem.*, 1976, **5**, 187–195.

[5] K. K. Chacko and W. Saenger, *J. Am. Chem. Soc.*, 1981, **103**, 1708–1715.

[6] W. Saenger, *Nature*, 1979, **279**, 343–344.

[7] K. Lindner and W. Saenger, *Acta Cryst.*, 1982, **B38**, 203–210.

[8] R. Puliti, C. A. Mattia and L. Paduano, *Carbohydr. Res.*, 1998, **310**, 1–8.

[9] D. E. Koshland, *The Enzymes, Vol. I*, Academic Press, 1959, p. 305.

[10] L. Liu and Q. X. Guo, *J. Incl. Phenom. Macrocycl. Chem.*, 2002, **42**, 1–14.

[11] H. C. Georg, K. Coutinho and S. Canuto, *Chem. Phys. Lett.*, 2005, **413**, 16–21.

[12] R. Granero-García, F. J. Lahoz, C. Paulmann, S. Saouane and F. P. A. Fabbiani, *CrystEngComm*, 2012, **14**, 8664–8670.

[13] C. Paulmann, *MARTOOLSCP*, 2007.

[14] Bruker-AXS, *APEX2*, Bruker-AXS, Madison, Wisconsin, USA, 2007.

[15] G. M. Sheldrick, *Acta Cryst.*, 2015, **C71**, 3–8.

[16] C. B. Hübschle, G. M. Sheldrick and B. Dittrich, *J. Appl. Cryst.*, 2011, **44**, 1281–1284.

[17] *GRADE*, http://grade.globalphasing.org.

[18] G. M. Sheldrick, *SADABS v2008-1*, Bruker-AXS, Madison, Wisconsin, USA, 2008.

[19] O. Guvench, S. N. Greene, G. Kamath, J. Brady, R. M. Venable, R. W. Pastor and A. D. MacKerell, *J. Comput. Chem.2*, 2008, **29**, 2543–2564.

[20] O. Guvench, E. Hatcher, R. M. Venable, R. W. Pastor and A. D. MacKerell, *J. Chem. Theory Comput.*, 2009, **5**, 2353–2370.

[21] W. L. Jorgensen and J. D. Madura, *J. Am. Chem. Soc.*, 1983, **105**, 1407–1413.

[22] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781–1802.

[23] M. P. Allen, *Computational Soft Matter: From Synthetic Polymers to Proteins, Lecture Notes*, John von Neumann Institute for Computing, Jülich, 2004, pp. 1–28.

[24] A. Grossfield, *WHAM: the weighted histogram analysis method, version 2.0.9*, http://membrane.urmc.rochester.edu/content/wham.

[25] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Comp. Phys. Comm.*, 2014, **185**, 604–613.

[26] P. Ramachandran and G. Varoquaux, *Comput. Sci. Eng.*, 2011, **13**, 40–51.

[27] C. F. Macrae, I. J. Bruno, J. A. Chisholm, P. R. Edgington, P. McCabe, E. Pidcock, L. Rodriguez-Monge, R. Taylor, J. van de Streek and P. A. Wood, *J. Appl. Cryst.*, 2008, **41**, 466–470.

[28] T. Steiner, *Acta Cryst.*, 2002, **E41**, 49–76.

[29] M. D. Collins, G. Hummer, M. L. Quillin, B. W. Matthews and S. M. Gruner, *Proc. Nat. Acad. Sci. USA*, 2005, **102**, 16668–16671.

[30] P. Das and S. Matysiak, *J. Chem. Phys. B*, 2012, **116**, 5342–5348.

High pressure study of the water–*tert*-butylamine system

6.1 Introduction

Alkylamine hydrates are one of the three types of water inclusion compounds according to Jeffrey's classification.[1] Although closely related to gas hydrates, there is a major distinction: alkylamine guests are bonded to the H-bonded water network through their amine groups. The existence of these interactions causes the structures of this type of compounds to be more complex than that of the gas hydrates, to the extent that a systematic classification is not possible. Overall, alkylamine hydrates resemble gas hydrates with partially open cages, and are usually referred as semi-clathrates.

Among all alkylamines, ^{t}BA has attracted considerable interest, because its 9.75-hydrate, whose crystal structure was determined in 1967,[2] has long been considered to be the only true clathrate in this family of compounds. The structure of this hydrate is in fact the only known ex-

ample of a gas hydrate type VI structure.[3] A second hydrate of ^{t}BA, a 7.25-hydrate, was reported in 1995.[4] It exhibits a complex structure formed by a combination of eight different cages, all of them with open edges and faces distorted with respect to regular polyhedra. The crystal structure of six other phases of the water–^{t}BA system have been recently reported,[5] showing the existence of three other semi-clathrates (7.75-, 11-, and 17-hydrate), two simple hydrates (with 0.25 and 1 water molecules per ^{t}BA molecule) and the solid form of pure ^{t}BA.

A comparison of the densities of all ^{t}BA hydrates shows that a maximum of 1.076 g/cm^3 is expected for a hypothetical 4.32-hydrate. Crystallisation, by means of solvent evaporation or cold gas flow and infrared laser heating techniques, seems to be hindered in the high-density region. In this chapter the crystal structure of two new hydrates are reported, with 5.65 and 5.8 water molecules per ^{t}BA molecule, respectively. Both hydrates were obtained by means of in-situ high-pressure crystallisation. These two new structures provide a further insight into the water–^{t}BA system and contribute to explain how more compact packing arrangements can be achieved in this type of structures.

In addition, a new anhydrate phase of ^{t}BA, characterised by extensive internal disorder, is reported. The low-temperature phase[5] has been termed here form I, while the newly discovered HP phase is labelled form II. The interactions between ^{t}BA molecules and the disorder in the structure have been studied by MD. Such a study shows that form II is an orientational glass,[6,7] a structure in which the position of the molecules within the unit cell are related by symmetry, while their orientations are unrelated.

6.2 Experimental

6.2.1 Screening overview

All samples referred in this chapter were prepared by mixing appropriate volumes of ^{t}BA (Sigma-Aldrich, used as received) with demineralised water to achieve certain molar ratios in 1 mL of solution. The high volume of solution, especially when compared to the size of the sample chamber

of a DAC, is justified by the high volatility of ^{t}BA. The quantities of ^{t}BA needed to prepare the mixtures poorer in the amine would easily evaporate if they were smaller than the ones used here. For the same reason, all mixtures were prepared and kept cold during the screening, to avoid a change in composition due to the evaporation of ^{t}BA.

All the crystallisation experiments were performed in a similar manner. The ^{t}BA:water mixtures were loaded in the DAC and compressed at room temperature until crystallisation occurred. Mixtures richer in ^{t}BA were loaded in the DAC at 263 K, to reduce the volatility of the sample, and brought to room temperature once the DAC was sealed. Subsequently, temperature and pressure cycling were performed to obtain single crystals large enough for diffraction.

In their description of the 9.75-hydrate of ^{t}BA, McMullan et al.[8] indicated that crystals were obtained from a 1:8 molar mixture of ^{t}BA:water. The same mixture yielded crystals of a different phase upon compression, based on unit-cell comparison, as crystal quality was not sufficient for a complete structural determination. Mixtures 1:7 and 1:9 were also tried and consistently yielded crystals of the same material, corresponding to a new phase, a 5.8-hydrate, which tends to crystallise in the form of several long thick needles, all stacked in the same direction. Diffraction from such samples occurs in the form of very broad Bragg spots limited to low resolution only, which makes structure determination challenging. Eventually, a suitable single crystal was grown from a 1:7 sample and the structure could be completely determined.

Another new phase, a 5.65-hydrate, was grown from a 1:6 mixture. The last mixture to be tested was 1:3; however, none of the crystals obtained were suitable for X-ray diffraction due to their small size. Over a certain pressure, single crystals stopped growing and instead new crystallisation occurred in the form of thin needles. Unfortunately, no single crystal could be grown from this material and the identity of this phase remains unknown.

The following sections contain further information on each of the crystals which were measured.

6.2.2   5.65-Hydrate

6.2.2.1   Crystal growth

450 µm
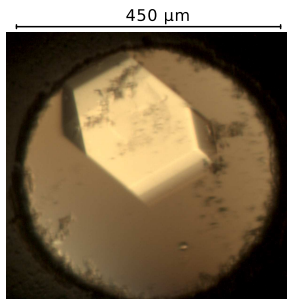
Figure 6.1: Single crystal of ^{t}BA 5.65-hydrate at 0.76 GPa.

A 1:6 mixture of ^{t}BA:water was loaded in a triangular DAC (Appx. I.2) equipped with an Inconel 718 gasket pre-indented to 160 μm thickness with a 450 μm diameter hole. Crystallisation on direct compression was observed at 0.65 GPa. At this pressure crystals melt very fast and grow back very slowly, hence temperature cycling is not appropriate. Pressure was increased to 0.76 GPa to decrease dissolution speed, and a single crystal was obtained by mild temperature cycling in the range 330-350 K (Fig. 6.1).

6.2.2.2   X-ray data collection and processing

X-ray data were collected on said crystal at 0.76 GPa on a Bruker ApexII CCD diffractometer equipped with a Mo microsource ($\lambda = 0.71073$ Å). The data collection strategy consisted of ω and ϕ 0.30° scans with the DAC in three different orientations (rotations of 120° around the beam axis). Exposure time was 40 s. Data were integrated following the strategy outlined in §1.1.3.

6.2.2.3   Structure refinement

The structure was solved using the charge flipping method implemented in SUPERFLIP[9] and subsequently refined by full-matrix least squares against F^2 using SHELXL V2014/7,[10] through the GUI SHELXLE.[11] For the ^{t}BA molecules, non-H atoms were refined anisotropically with their ADPs subject to similarity restraints, unless very disordered, in which case isotropic refinement was preferred. Restraints on bond lengths and angles were used to ensure proper geometry. H atoms of methyl groups were placed in idealised positions and refined riding on their parent atoms. For methyl groups lying on special positions, H atoms were placed similarly, but their coordinates were later constrained to the special position when needed and spare symmetry equivalents were removed; positions were no longer refined but kept riding on the corresponding carbon atom. H atoms were placed only in one of the symmetry-independent

amine groups; positions were located through electron density maps and then kept riding on the nitrogen atom. The rest of the amine groups were disordered and H atoms could not be placed. In water molecules, all oxygen atoms were located through Fourier difference maps and refined anisotropically, unless residing on special positions. Each oxygen atom was assigned four half-occupied H atoms in the direction of the vertices of a tetrahedron, which coincide with the direction of the H bonds in the structure. The positions of the H atoms were refined constrained to move along with the oxygen atoms they are attached to. A final R factor of 6.54% was obtained. For further crystallographic data see Table 6.1.

6.2.3 5.8-Hydrate

6.2.3.1 Crystal growth

A crystal was grown by loading a sample of the 1:7 mixture ^{t}BA:water in a DAC (Appx. I.2) equipped with a 200 μm gasket pre-indented to a thickness of 160 μm and a sample chamber 300 μm in diameter. Upon fast compression, crystallisation occurred at 0.80 GPa in the form of fine needles (Fig. 6.2a). Such material melts easily during temperature cycling, at the same time that single crystals start to form (Fig. 6.2b). Finally, two single crystals were obtained inside the sample chamber at 0.61 GPa (Fig. 6.2c).
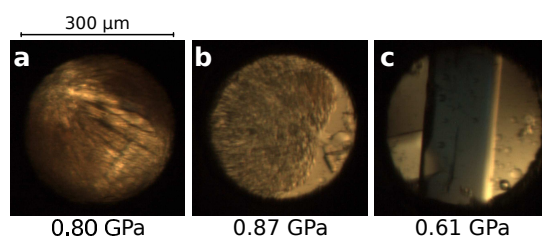
Figure 6.2: Crystals of ^{t}BA 5.8-hydrate. **a)** Thin needles obtained by fast compression. **b)** Needles and small single crystals coexisiting at 0.87 GPa after temperature cycling. **c)** Two single crystals at 0.61 GPa.

6.2.3.2 X-ray data collection and processing

Data were collected on a MAR345 image plate using an INCOATEC microfocus (IμS) source of Ag radiation (0.56085 Å). Data collection strategy consisted of 0.5° ϕ scans at three different orientations of the DAC (rotations of 90 and 120° around the beam axis). Frames collected in this

detector were converted from MAR to SFRM format by using the tools included in APEX2.[12] Data were integrated following the strategy outlined in §1.1.3.

Reflections originating from the two crystals in the sample chamber could be indexed in the frames. The best integration procedure consisted in indexing both crystals and integrating according to the two possible lattices. Subsequently, TWINABS[13] was used to apply absorption correction and to extract only the reflections corresponding to the largest crystal to proceed with the refinement of the structure. Each run was processed individually and all data was finally merged with SORTAV.[14]

6.2.3.3 Structure refinement

The structure was solved using SUPERFLIP[9] and subsequently refined by full-matrix least squares against F^2 using SHELXL v2014/7,[10] through the GUI SHELXLE.[11] For ^{t}BA molecules, non-H atoms were refined isotropically. Restraints on bond lengths and angles were used to ensure proper geometry. H atoms of methyl groups were placed in idealised positions and refined riding on their parent atom. Each nitrogen atom was assigned three partially-occupied H atoms in the directions of the vertices of a tetrahedron, which coincide with the directions of the H bonds in the structure. The positions of the H atoms were refined at first subject to restraints and then constrained to move along with the nitrogen atoms they are attached to, in order to reduce the number of parameters. For water molecules, all oxygen atoms were located through Fourier difference maps and refined anisotropically. Each oxygen atom was assigned four half-occupied H atoms in a similar manner than for the amine groups. A final R factor of 10.49% was obtained. For further crystallographic data see Table 6.1.

6.2.4 Pure *tert*-butylamine form II

6.2.4.1 Crystal growth

Pure ^{t}BA was loaded in a square DAC (Appx. I.1) equipped with a steel gasket 230 μm thick and a sample chamber 280 μm in diameter. Loading

	5.65-hydrate	5.8-hydrate
Crystal data		
Chemical formula	$C_4H_{11}N\cdot5.65H_2O$	$C_4H_{11}N\cdot5.8H_2O$
Molecular mass (g mol^{-1})	174.92	177.63
Temperature (K)	298(2)	298(2)
Pressure (GPa)	0.76(5)	0.61(5)
Crystal system	orthorhombic	monoclinic
Space group	$Cmca$	$P2_1/n$
a (Å)	17.2082(6)	12.5091(13)
b (Å)	15.6395(6)	22.213(3)
c (Å)	30.9066(8)	18.766(2)
β (°)		92.985(3)
V (Å^3)	8317.8(5)	5207.1(10)
Z	32	20
Crystal size (mm)	~0.25×0.20×0.10	~0.30×0.20×0.10
Data collection		
Radiation wavelength (Å)	0.71073	0.56085
Absorption correction	SADABS[15]	TWINABS[13]
T_{min}, T_{max}	0.7439, 0.6459	0.7447, 0.9330
Number of $[I > 2\sigma(I)]$ reflections:		
- measured	15326	7547
- independent	1532	3189
- observed	1101	2071
Completeness (%)	91.3	79.4
$(sin\theta/\lambda)_{max}$ (Å^{-1})	0.30	0.38
R_{int} (%)	3.78	7.21
Refinement		
$R[F^2 > 2\sigma(F^2)]$ (%)	6.54	10.49
$wR(F^2)$ (%)	20.06	20.86
S	1.131	1.111
Number of reflections	1532	3189
Number of parameters	208	386
Number of restraints	31	70
$\Delta\rho_{min}$, $\Delta\rho_{max}$ (e Å^{-3})	−0.244, 0.320	−0.229, 0.370

Table 6.1: Crystallographic data and refinement parameters of ^{t}BA 5.65- and 5.8-hydrate.

was carried out at 263 K to avoid evaporation of the sample. Solidification in the form of poly-crystalline material occurs at pressures above 0.60 GPa (Fig. 6.3a). This new phase (named form II) is very responsive to changes in pressure and temperature, making the usual annealing procedure for crystallisation challenging. Eventually, a single crystal, occupying the entire sample chamber (Fig. 6.3b), was grown at 0.78 GPa. A second smaller crystal grew included into the first one, but could be neglected due to its limited scattering power. Single-crystalline growth of this phase is favoured at temperatures between 310 and 320 K. At lower temperatures the growth is too fast and always results in poly-crystalline material; at higher temperatures the sample melts.

Figure 6.3: Crystals
of pure ^{t}BA. **a)** Poly-
crystalline material at
0.63 GPa. **b)** Single crys-
tals. Crystal **I** is colourless
and occupies the entire
sample chamber; crystal
II is a small inclusion. **c)**
Compressed single crystal
at 1.07 GPa.
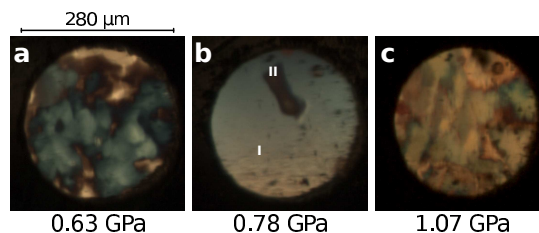
X-ray diffraction data showed that this HP phase of ^{t}BA is very disordered at the molecular level. In an attempt to induce order in the structure, pressure was slowly increased up to 1.07 GPa, but the crystal always transitions to a non-diffracting state, as confirmed by X-ray diffraction (Fig. 6.3c). Isochoric thermal annealing under these conditions is not possible, because at such pressures the crystal does not melt (tested at temperatures up to 400 K).

6.2.4.2 X-ray data collection and processing

Data were collected following the strategy in §6.2.2.2, although with the DAC in a single orientation.

6.2.4.3 Structure refinement

Full-matrix least-squares refinement against F^2 was performed using SHELXL v2014/7,[10] through the GUI SHELXLE,[11] on a solution obtained using SHELXT.[16] The structure, belonging to space group $P2_1/n$, is characterised by a high degree of disorder. The asymmetric unit contains a single molecule of ^{t}BA disordered over four possible orientations; the central carbon atom occupies a fixed position, but it is not possible to discern which one of the four vertices of the tetrahedral molecule is the nitrogen atom. This has been modelled by setting the likelihood of each terminal atom to be a carbon atom at 75% and to be a nitrogen atom at 25%. H atoms were placed in idealised positions for the methyl groups and refined with the riding model. The 75% occupancy of the carbon atoms was applied to the H atoms too. Because nitrogen atoms lay in the exact same positions than carbon atoms, no H atoms were placed on them. Final R factor was 11.73%. For further crystallographic information see Table 6.4.

All atoms were refined anisotropically; although the ADPs of the terminal atoms are over two times larger than that of the central atom, compared as equivalent isotropic displacement parameters. Such elongation, which occurs mainly in the directions normal to the covalent bonds, could be attributed to a certain degree of libration of the molecule, as if describing a sphere. Different options have been considered to model the libration as rotational disorder. In SHELXL, two partially occupied ^{t}BA molecules instead of one, both sharing the same central carbon atom, have been implemented. CRYSTALS[17] offers the option of modelling electron density with more complex figures than ellipsoids;[18] a sphere containing the four terminal atoms was implemented. In both cases there was an increase in the R factor, hence the initial model was maintained, which indicates that full rotational disorder occurs only to a small extent, if at all.

6.2.4.4 Molecular dynamics

MD simulations on ^{t}BA form II have been performed to better understand the disorder and internal structure of this HP phase. For the purpose of the simulations, the disorder has been removed from the refined model, by making one of the four terminal atoms (position 2) a fully occupied nitrogen atom, and the other three atoms fully occupied carbon atoms. Methyl H atoms were replaced by a rigid methyl group, which is computationally less expensive, given that the structure is likely to be more influenced by the entire methyl group rather than by the individual atoms. Amine H atoms were geometrically placed using MERCURY,[19] with bond distances elongated to neutron values, and included in the force field as individual species. The initial setup for the simulations consisted of an arrangement of 24 unit cells (four repetitions along a, two along b and three along c), adding up to 96 ^{t}BA molecules. This arrangement constitutes the simulation box, which is subsequently put in PBC to simulate an infinite crystal and energy minimised prior to each simulation.

All simulations here described were performed in a similar manner using GROMACS (v5.0.3).[20] At first, four different simulations where car-

ried out to compare the performance of two different force fields with pressure. Each simulation was 2.5 ns in length, consisting of 2 fs steps, and was repeated 5 times, starting always from the same setup, to ensure complete sampling. The NPT ensemble was used in all cases, with the simulations performed at either 0.0001 or 0.78 GPa, the latter being the pressure at which form II was crystallised. Two force fields were employed: GROMOS with parameter set 54A7[21] and OPLS-aa (Optimised Potentials for Liquid Simulations).[22] The former is the default option in GROMACS, while the latter was specifically designed to reproduce the behaviour of organic liquids, hence it was expected to better suit the needs of this study. Two simulations were performed at 0.0001 GPa using the two different force fields, and two more, also with different force fields, at 0.78 GPa.

As suggested by Nemkevich et al.,[23] force-field parameters must ensure that the properties derived from simulations agree with properties measured experimentally, which does not necessarily mean that force-field parameters themselves must agree with experimental parameters. Following this idea, the compressibility of ^{t}BA has been derived from appropriate simulations, to ensure consistency with the force field. Two NVT simulations were performed, one using the initial setup described previously, and another one in which the volume of the simulation box of said setup was enlarged by 1%. The compressibility (β) was calculated by measuring the changes in pressure in both simulations (Eq. 6.1). A compressibility of 0.21562 GPa^{-1} was determined in this way. Further details about the simulation can be found in Table 6.2.

$$\beta = -\frac{1}{V_0}\frac{dV}{dP} \sim -\frac{1}{V_0}\frac{\Delta V}{\Delta P} \tag{6.1}$$

After the first tests, the OPLS force field was considered to better reproduce the properties of the system than GROMOS, although a re-parametrisation was needed for further improvement. According to Bernardes and Joseph,[24] it is usually better to use quantum-mechanical charges to derive structural information from a simulation, while the default charges in OPLS are better suited for energy calculations. Although the final aim of these simulations is to obtain FESs, this is done here by measuring structural parameters, hence quantum-mechanical charges were preferred. Charges for each atom of the ^{t}BA molecule were

Simulation length	2.5 ns
Time step	2 fs
Frame output	every 1000 steps (2 ps)
Number of frames	1250
Minimisation	forces below 10.0 kJ/mol/nm
Force field	GROMOS 54A7 and OPLS-aa
PBC	yes
Ensemble	NPT
Temperature control	Nosé-Hoover
Temperature	298 K
Pressure control	Parrinello-Rahman (isotropic)
Pressure	0.0001 or 0.78 GPa
Treatment of interactions	
- Covalent	< 9 Å
- Coulomb and Van der Waals	> 9 Å (PME)
Interactions cutoff	10 Å
Treatment of bonds with H atoms	rigid

Table 6.2: Details of the simulations of pure ^{t}BA.

computed from a MP2/TZVP calculation in GAUSSIAN03.[25] Charges of H atoms belonging to the same type of functional group were averaged; then charges within the same methyl group were added together to make a single entity, as explained previously. Final charges used in the production runs are detailed in Table 6.3. Other parameters of the force field, which mainly control Van der Waals interactions, were left untouched. The reason behind this is that intermolecular interactions in MD simulations, which greatly affect the trajectory, are mainly governed by electrostatic forces, while the main role of Van der Waals interactions is to avoid the molecules clashing.

Pressure coupling was changed from isotropic to anisotropic, to allow for each dimension of the simulation box to compress independently. This function may fail to reproduce unit-cell volumes at high pressures, but it was a sound option in this particular case, based on the simplicity and homogeneity of the system, to allow for more realistic simulations (see §5.3.2 for further discussion).

After implementing all these changes, the simulations at 0.0001 and 0.78 GPa were repeated. Two other simulations, at 0.40 and 1.20 GPa, were are also performed, to further explore the behaviour of the system.

A Python script was written to analyse the trajectories and produce FESs. The orientation of the ^{t}BA molecules was chosen as an order parameter, in terms of which the trajectories could be expressed. The orientation of each ^{t}BA molecule in each frame is described by means

Table 6.3: MP2/TZVP charges calculated for each atom or group atoms of the ^{t}BA molecule. GROMOS and OPLS charges are given for reference.

MP2/TZVP	
$C_{central}$	0.72710
CH_3	−0.13957
N	−1.03307
H_{amine}	0.36234
GROMOS	
$C_{central}$	0.00000
CH_3	0.00000
N	−0.83000
H_{amine}	0.41500
OPLS	
$C_{central}$	0.18000
CH_3	0.00000
N	−0.90000
H_{amine}	0.36000
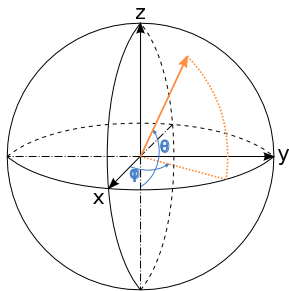
Figure 6.4: Graphic relationship between polar and Cartesian coordinates. x, y, and z are the orthonormal axes corresponding to the monoclinic a, b, and c axes.

of the vector between the central carbon atom and the nitrogen atom (here termed the CN vector). These vectors are expressed in polar coordinates, r being the length, φ the angle contained in the horizontal plane and θ the angle in the vertical plane (Fig. 6.4). Polar coordinates are advantageous over Cartesian, because r is essentially constant—vibrations of covalent bond lengths are not observed in these simulations—and otherwise not related to the orientation, hence the analysis can be performed in terms of φ and θ only. By decreasing the number of geometrical parameters, the representation of FESs also becomes simpler. Notwithstanding this, polar coordinates are used only internally, the final results are presented as polar maps in Cartesian coordinates.

After recording all the CN vectors, a 3D histogram of the distribution of (φ, θ) pairs is built. The reference frame of such histogram is a sphere, based on the definition of the polar coordinates. Sampling over the surface of a sphere suffers from an entropic effect; the likelihood of a certain CN vector increases when moving from the poles towards the equator, because the area available at the pole is smaller than that available near the equator. This effect has been corrected according to Eq. 6.2.

$$f_{corrected} = f\sin(\theta)/2 \tag{6.2}$$

Finally, the histogram is directly converted into a FES, setting the most frequent state as the ground energy state, according to §1.2.3. The 3D spherical FES is plotted in 2D as the polar aspect of an azimuthal equidistant projection, implemented in the MATPLOTLIB module of Python.[26] In this type of projection, a sphere is mapped onto a plane parallel to the equator and tangent to the correspondent pole, in such a way that all distances and directions measured from the pole are correct, although other distances, angles, shapes, and sizes are not preserved. As a result, meridians are projected as straight lines from the centre of the projection and parallels are concentric circles equally spaced. This facilitates measuring positions in the projection by directly reading the axis. Although sizes and shapes are not true in the projection, distortion is relatively small, especially if a single hemisphere is projected at a time, as distortion increases away from the centre.

148

6.2.5 *tert*-Butylammonium carbonate

6.2.5.1 Crystal growth

After completion of the experiments described in §6.2.1, spare ^{t}BA:water samples were stored at 277 K. After a certain period of time, several small transparent crystals had grown in the top half of the vials. Several of those crystals were analysed, to find that they correspond to the carbonate salt of ^{t}BA, whose structure, to the best of our knowledge, has not been reported.

6.2.5.2 X-ray data collection and processing

X-ray diffraction data were collected at 150 K on one of those crystals on a Bruker ApexII CCD diffractometer equipped with a Mo microsource ($\lambda = 0.71073$ Å). A hemisphere data collection strategy was used, with scan steps of 0.5° and exposure times of 60 s per frame. Data were integrated following the strategy outlined in §1.1.3, omitting the corrections related to high-pressure samples.

6.2.5.3 Structure refinement

The structure was solved using SHELXS[27] and subsequently refined by full-matrix least squares against F^2 using SHELXL v2014/7,[10] through the GUI SHELXLE.[11] A final R factor of 3.15% was achieved. For further crystallographic information see Table 6.4.

6.3 Results and discussion

6.3.1 Structure description

The two hydrates of ^{t}BA reported here, with hydration numbers 5.65 and 5.8, are semi-clathrates, as the ^{t}BA molecules are both enclosed and interacting with the caged water network. Such interaction, by means of H bonds between the amine group and water molecules, causes two types of defects: substitution, when an amine group replaces a water molecule from the caged network, and intercalation, when the amine group forms

Table 6.4: Crystallographic data and refinement parameters of pure ^{t}BA and ^{t}BA$^+$ carbonate.

	pure ^{t}BA	^{t}BA$^+$ carbonate
Crystal data		
Chemical formula	C$_4$H$_{11}$N	(C$_4$H$_{12}$N)$_2$CO$_3$
Molecular mass (g mol^{-1})	73.14	208.30
Temperature (K)	298(2)	150(2)
Pressure (GPa)	0.78(5)	0.0001
Crystal system	monoclinic	trigonal
Space group	$P2_1/n$	$R\bar{3}c$
a (Å)	5.6799(4)	6.3122(12)
b (Å)	10.1457(9)	6.3122(12)
c (Å)	9.2680(12)	53.215(11)
β (°)	100.611(8)	90
V (Å^3)	524.95(9)	1836.2(8)
Z	4	6
Crystal size (mm)	~0.15×0.14($\varnothing$)	~0.15×0.10×0.05
Data collection		
Radiation wavelength (Å)	0.71073	0.71073
Absorption correction	Semi-empirical (using I meas.), SADABS [15]	
T_{min}, T_{max}	0.8530, 1.0000	0.6233, 0.7454
Number of $[I > 2\sigma(I)]$ reflections:		
- measured	2529	3926
- independent	288	416
- observed	219	354
Completeness (%)	69.7	97.9
$(sin\theta/\lambda)_{max}$ (Å^{-1})	0.45	0.60
R_{int} (%)	1.60	2.49
Refinement		
$R[F^2 > 2\sigma(F^2)]$ (%)	11.73	3.15
$wR(F^2)$ (%)	19.72	8.13
S	1.514	1.052
Number of reflections	288	3926
Number of parameters	50	25
Number of restraints	10	0
$\Delta\rho_{min}, \Delta\rho_{max}$ (e Å^{-3})	−0.084, 0.160	−0.181, 0.244

H bonds with two neighbouring water molecules, forcing them apart and elongating the cage edge. As a result, cages in these semi-clathrates are considerably distorted from regular polyhedra.

6.3.1.1 5.65-hydrate

^{t}BA 5.65-hydrate crystallises at 0.76 GPa in an orthorhombic unit cell with parameters $a = 17.2082(6)$ Å, $b = 15.6395(6)$ Å, $c = 30.9066(8)$ Å, space group *Cmca*. There are four sites in the asymmetric unit occupied by ^{t}BA molecules, of which only one is fully occupied; the other three lie on special positions, hence containing only a fragment of the ^{t}BA molecule. Each molecule is contained in a different water cage (named **C1** to **C4**), with no empty cages in the structure (Fig. 6.5).

Cage **C1** contains the only fully-occupied ^{t}BA molecule of the asymmetric unit, which does not exhibit disorder. The cage exhibits a feature common to other semi-clathrates: the poles are formed by rather regular faces, although some of the faces along the equator are irregular, and their sides are longer than H-bond distances, due to intercalation defects.

Cage **C2**, and the ^{t}BA molecule it contains, are divided into two symmetric parts by a mirror plane. Similarly to **C1**, there are open edges at the equator due to the presence of amine groups H bonded to water molecules belonging to the cage.

Only a quarter of cage **C3** lies inside the asymmetric unit, the rest of it being generated by a mirror plane and a two-fold rotation axis perpendicular to it. The corresponding ^{t}BA molecule is fully generated by the mirror plane, while the rotation axis creates another entire molecule in a different orientation and position, leading to disorder. All the edges of this cage exhibit lengths corresponding to H bonds, but there are substitutional defects; two water molecules are affected by the disorder of the ^{t}BA and each one can coexist with only one of the two disordered orientations. The lack of intercalation defects allows for a cage with more regular faces.

Cage **C4** resembles **C3**, in that is also generated by a mirror plane and a two-fold axis, although the ^{t}BA molecule inside this cage is only orientationally (and not positionally) disordered because the central carbon atom lies directly on the inversion center—where the rotation axis intersects the mirror plane. In this case both substitutional and intercalation defects are present.

6.3.1.2 5.8-hydrate

The 5.8-hydrate of ^{t}BA crystallises at 0.61 GPa in a monoclinic cell with parameters $a = 12.5091(13)$ Å, $b = 22.213(3)$ Å, $c = 18.766(2)$ Å, $\beta = 92.985(3)°$, space group $P2_1/n$. It is also a semi-clathrate and can be described in terms of five different cages (named **D1** to **D5**), each containing a fully occupied, although sometimes disordered, ^{t}BA molecule (Fig. 6.6).

Figure 6.5: Top: Perspective view along the *a* axis of the asymmetric unit of ^{t}BA 5.65-hydrate. ^{t}BA molecules are represented as grey spheres. H bonds between water molecules are plotted as solid lines, coloured to differentiate the four cages. Cages have been expanded out of the asymmetric unit for clearer representation. **Bottom:** four independent water cages in the structure of ^{t}BA 5.65-hydrate. ^{t}BA molecules are depicted by solids lines, colour-coded according to the atom type (C in dark grey, N in blue, and H in light grey). H bonds between water molecules making up the cages are represented by thick lines, solid at the front and dashed at the back to give a sense of perspective, each cage coloured differently. Cage edges longer than H bond distances are depicted as thin dashed lines. H bonds between amine groups and water molecules are shown as thin red dashed lines. Mirror planes are presented as semi-transparent yellow planes; two-fold rotation axis are shown as green lines; cage vertices corresponding to water molecules affected by substitutional defects are highlighted as orange circles.
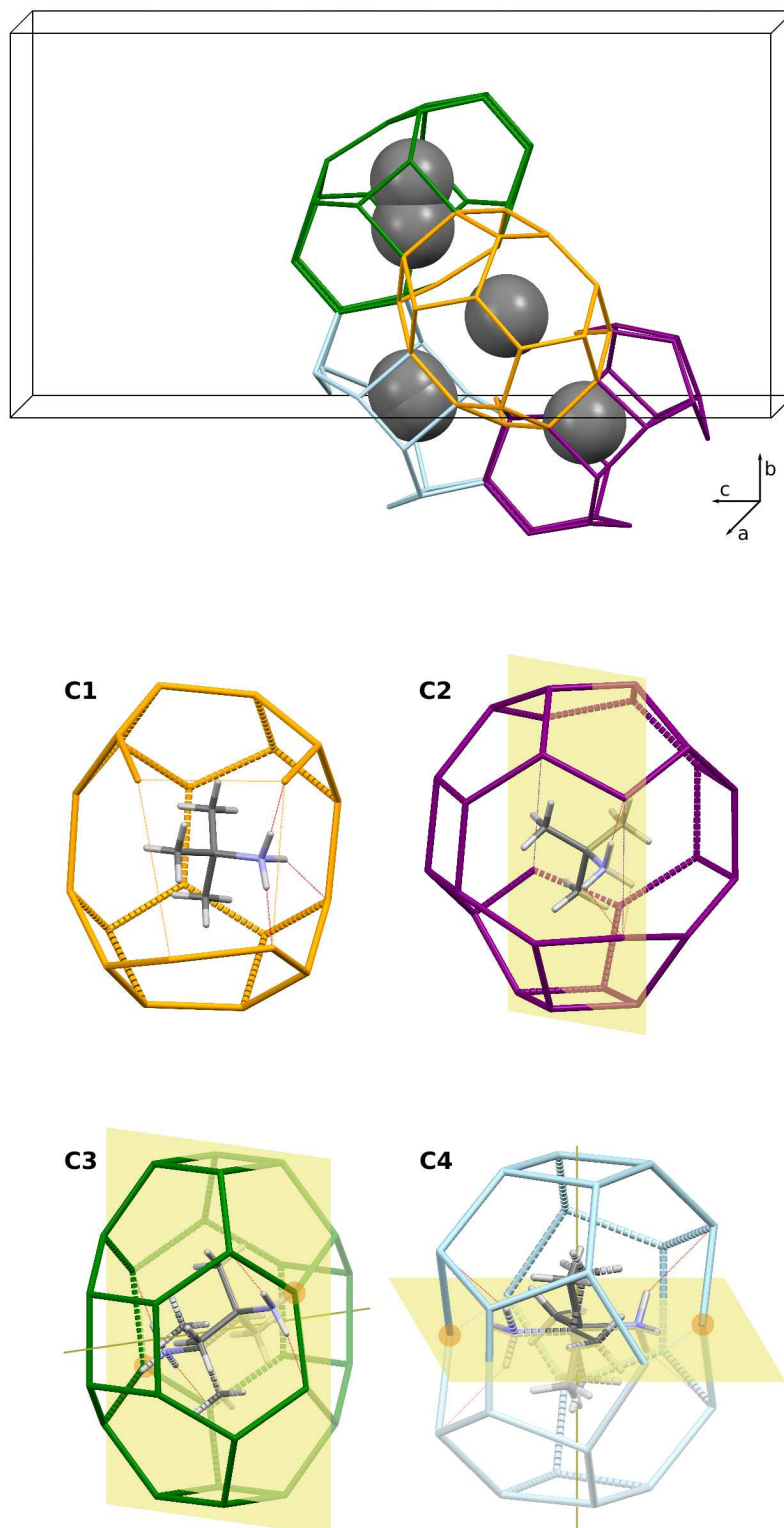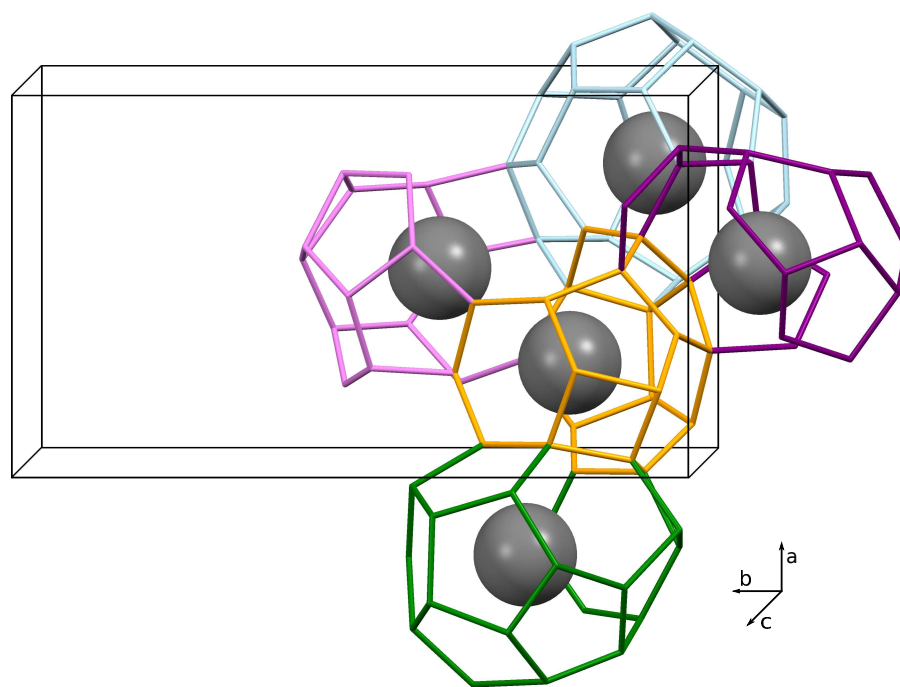
Figure 6.6: Top: Perspective view along the c axis of the asymmetric unit of ^{t}BA 5.8-hydrate. ^{t}BA molecules are represented as grey spheres. H bonds between water molecules are plotted as solid lines, coloured to differentiate the five cages. Cages have been expanded out of the asymmetric unit for clearer representation. **Bottom:** five independent water cages in the structure of ^{t}BA 5.8-hydrate. ^{t}BA molecules are depicted by solids lines, colour-coded according to the atom type (C in dark grey, N in blue, and H in light grey). H bonds between water molecules making up the cages are represented by thick lines, solid at the front and dashed at the back to give a sense of perspective, each cage coloured differently. Cage edges longer than H bond distances are depicted as thin dashed lines. H bonds between amine groups and water molecules are shown as thin red dashed lines.
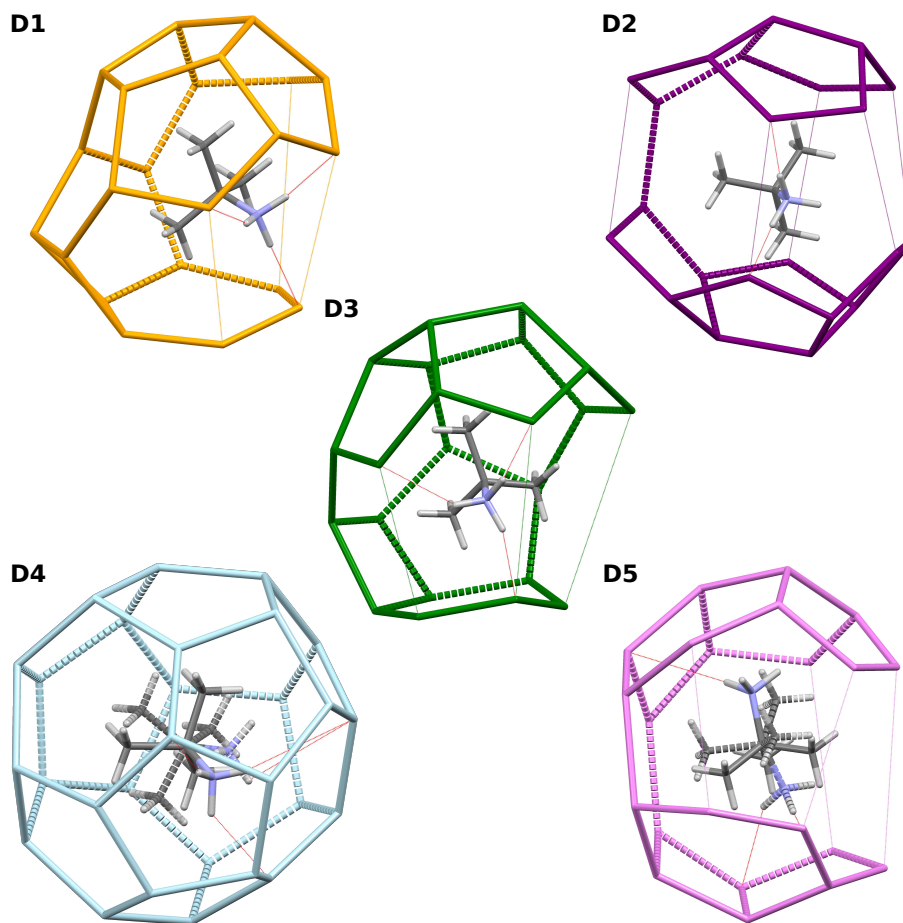
Cage **D1** exhibits a feature similar to many of the cages described for the previous structure: the poles are formed by faces close to a regular polygon, while the cage is open around the equator due to the intercalation of the amine group. This feature is present in all of the cages, with the exception of **D4**.

The effect of intercalation defects in cage **D2** is more dramatic, as there is a single H bond linking both poles of the cage, with the rest of the edges being considerably larger.

Cage **D3** is very similar to **D1**.

Cage **D4** is the largest of all, which allows for a ^{t}BA molecule to be contained inside without causing defects. All edges correspond to H bonds, based on their lengths, although this does not imply that the cage is closed, as not all edges can coexist with the H bonds formed between the amine group and the water network. The ^{t}BA molecule inside this cage is equally disordered between two possible orientations.

Cage **D5** resembles **D2**, as there are only two H bonds linking both poles, although in this case the ^{t}BA molecule is disordered inside the cage. One orientation is populated 80% of the time, while the other only 20%.

6.3.1.3   Pure *tert*-butylamine form II

The structure of ^{t}BA form II, crystallised at 0.78 GPa, has already been introduced in §6.2.4.3. The unit cell, with parameters $a = 5.6799(4)$ Å, $b = 10.1457(9)$ Å, $c = 9.2680(12)$ Å, $\beta = 100.611(8)°$, space group $P2_1/n$, contains four symmetry-equivalent ^{t}BA molecules. Molecules have been described to be disordered between four possible orientations, with an equal likelihood of the nitrogen atom being in each of the four vertices of the tetrahedral geometry. Based on single-crystal X-ray diffraction only, it is not possible to determine whether the molecules are in fact disordered, or if the data simply does not allow to distinguish between carbon and nitrogen atoms. Complete rotational freedom of the ^{t}BA molecules is another possibility, although the refinement does not support it.

The packing can be described by layers of ^{t}BA molecules stacked along the c axis (Fig. 6.7). Each molecule is surrounded by another six in the same layer, and three other above and below. The separation between molecules, measured as distance between the central carbon atoms, varies between 5.342 and 6.329 Å, with an average value of 5.785 Å.
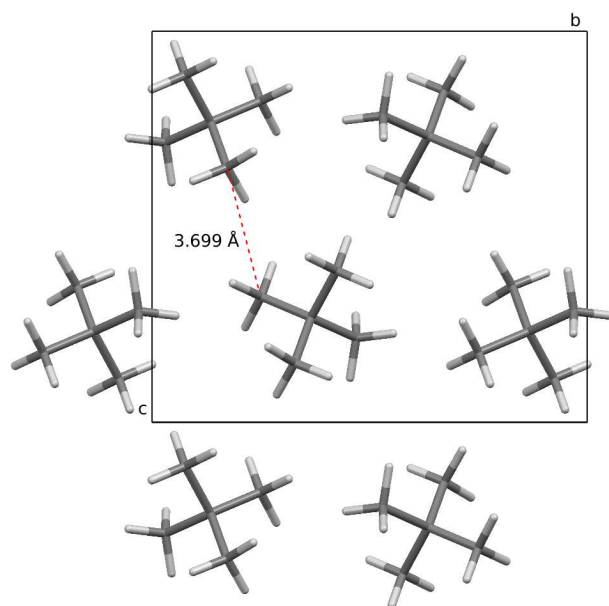
Figure 6.7: Packing arrangement of ^{t}BA form II at 0.78 GPa, view along the a axis. Three layers, parallel to the ab plane, have been depicted. Molecules in the top and bottom layer are aligned with the intermolecular spaces in the middle layer, resembling spheres close packing. The shortest possible H bond in the structure (as nitrogen–nitrogen distance) is indicated by a dashed red line.

Further analysis shows that the shortest nitrogen–nitrogen interaction is 3.699 Å, allowing only for weak H bonds.[28] Nevertheless, the presence of such H bonds might be sufficient to limit the rotational freedom of the molecules to a certain set of orientations. A deeper study of the disorder in this structure, based on MD simulations, is presented in §6.3.3.

6.3.1.4 *tert*-Butylammonium carbonate

The crystal structure of the unhydrated ^{t}BA$^+$ carbonate has a trigonal unit cell with parameters $a = 6.3122(12)$ Å and $c = 53.215(11)$ Å, space group $R\bar{3}c$. The assymmetric unit contains a sixth of a formula unit: a third of a ^{t}BA$^+$ cation and a sixth of a carbonate anion. The structure consists of sandwich layers—anions surrounded above and below by cations—parallel to the ab face of the unit cell (Fig. 6.8). Each carbon-

Figure 6.8: Layered packing of ^{t}BA$^+$ carbonate, view along the a axis. Layers are parallel to the ab plane and are stacked along the c axis. Only a third of the cell has been represented. The orange areas correspond to regions governed by H bonds or ionic interactions. The blue area corresponds to the region governed by London dispersion forces.
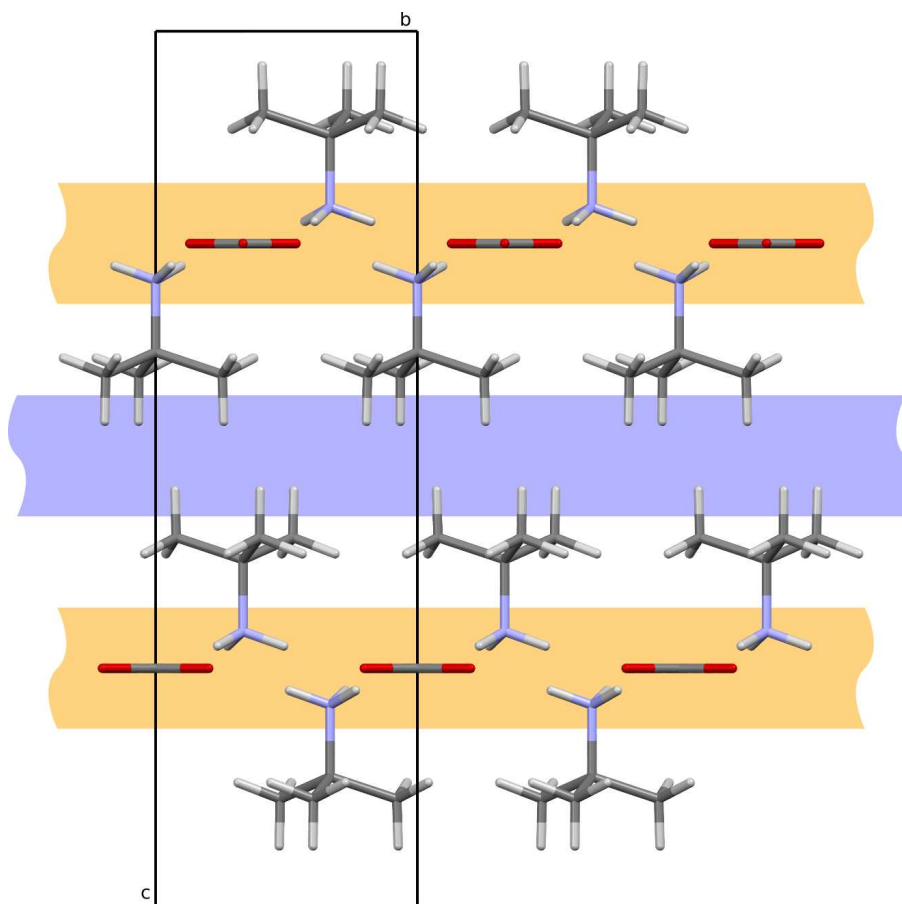
ate anion is linked to six ^{t}BA$^+$ units, two per oxygen atom, through the corresponding amine groups; the relative degree of covalent and ionic character of these links has not been evaluated. The orientation of the cations is such that the carbon–nitrogen bond is perfectly perpendicular to the plane containing the carbonate anions, which is in turn perfectly parallel to the ab plane of the unit cell. Sandwich layers are stacked along the c axis held together by London dispersion forces between the *tert*-butyl groups.

The crystallisation ^{t}BA$^+$ carbonate is remarkable, as crystals grew out of solution, on the top half of vials containing ^{t}BA:water mixtures stored at 277 K, without any source of carbonate intendedly added to the medium. The most reasonable explanation is that ^{t}BA accumulated in the vapour phase inside the close vial, due to its high vapour pressure (0.388 atm at 20°C). Then an acid-base reaction must have occurred with the CO_2 naturally occurring in the air trapped inside the vial.

^{t}BA^{+} carbonate formed as a result of this reaction and crystallised attached to the walls of the vials, favoured by the low temperatures. The natural amount of CO_2 is in fact sufficient to grow small crystals, as observed here. This process is not unknown, as it constitutes the basis for industrial CO_2 removal from gas streams.[29]

6.3.2 Crystalline phases in the water–*tert*-butylamine system

There are eleven known crystalline phases in the water–^{t}BA system, including two anhydrous forms and the three structures reported in this work (Table 6.5). Six of these structures were reported by Dobrzycki et al., along with a relationship for the variation of density with the hydration number of each phase.[5] Such relationship takes the form of an inverted Morse potential:

$$\rho(N) = s - d_e(1 - \mathrm{e}^{-a(N-r_e)})^2 \qquad (6.3)$$

Were ρ is the density of each crystalline phase N (denoted by its hydration number), normalised with respect to ice Ih, s is 1.076 g/cm^3, d_e is 0.163 g/cm^3, a is 0.179, and r_e is 4.32. Values for the four parameters s, d_e, a, and r_e were obtained by fitting the equation to the experimental densities.[5] The validity of the proposed relationship was confirmed by its ability to predict the densities of the *tert*-butanol water clathrates.[5]

An important observation by Dobrzycki et al. is the existence of a region of hydration numbers between 1 and 7¼ for which no crystal structures could be obtained. The maximum density, predicted for a potential 4.32-hydrate, falls in this region, hence structures with hydration numbers between 4 and 5 are expected to show the most efficient packings. It was stated that *"this gap should almost be expected because it represents a fuzzy guest/host transition area without nucleation/crystal growth"*,[5] based on the possible interactions between water molecules in the structure. For the lowest hydration numbers, water molecules cannot interact between them, as they are located in the intermolecular spaces in the packing of ^{t}BA molecules. In theory, as the hydration number in-

Table 6.5: Crystalline phases of the water–^{t}BA system.

N	Structure type	T or P^b	ρ (g/cm³)	Norm. ρ (g/cm³)	Authors
0	Anhydrate	133 K	0.860	0.857	Dobrzycki et al.[5]
0	Anhydrate	0.78 GPa	0.925	–	This work
0.25	Simple hydrate	273 K	0.884	0.884	Dobrzycki et al.[5]
1	Simple hydrate	173 K	0.975	0.975	Dobrzycki et al.[5]
5.65	Semi-clathrate	0.76 GPa	1.117	1.059	This work
5.8	Semi-clathrate	0.61 GPa	1.133	1.086	This work
7.25	Semi-clathrate	248 K	1.038^c	1.041	Stäben and Mootz[4]
7.75	Semi-clathrate	203 K	1.035	1.041	Dobrzycki et al.[5]
9.75	Semi-clathrate^a	271 K	1.005^c	1.010	McMullan et al.[2]
11	Semi-clathrate	250 K	1.004	1.014	Dobrzycki et al.[5]
17	Semi-clathrate	243 K	0.932	0.940	Dobrzycki et al.[5]

^{a} Although formally a semi-clathrate, this structure does not exhibit permanent defects in the water network, hence it resembles a true clathrate.
^{b} Structures for which a temperature is given were obtained at ambient pressure. Structure for which a pressure is given were obtained at ambient temperature.
^{c} From the redeterminations by Dobrzycki et al. rather than the original author.

creases, the water network changes from this 0D state (0.25-hydrate) to chains (1D, mono-hydrate), layers (2D), and finally cages (3D, 7.25- and higher hydrates). In practice, considering the whole continuum, a layered water network has only been observed for *tert*-butanol hydrates.[30] For ^{t}BA, a 2D arrangement has not been observed; the gap in hydration numbers probably exists because the jump from a 2D to a 3D arrangement requires a considerable increase in water molecules.

The densities of the two high-pressure semi-clathrates reported in this work have been normalised (§6.3.2.1) and the relationship between the density and the hydration number has been recalculated using FITYK[31] (Fig. 6.9). The new parameters for Eq. 6.3 are in Table 6.6, along with the original parameters. Overall, only a small change is observed, meaning that the two new structures behave as expected.

Table 6.6: Parameters of Eq. 6.3, as determined originally by Dobrzycki et al. (top) and redetermined in this work (bottom). s and d_e in g/cm³, a and r_e are dimensionless.

s	1.076(11)
d_e	0.163(22)
a	0.179(19)
r_e	4.32(31)
s	1.079(9)
d_e	0.166(22)
a	0.179(21)
r_e	4.31(33)

The use of high pressure shows that it is possible to obtain crystal structures which are closer to the maximum possible density. Nevertheless, the transition between simple hydrates and clathrate-like structures is still unclear. Without neglecting the idea of the stability of the structure associated with the water network proposed by Dobrzycki et al., the 5.65- and 5.8-hydrates point towards a more complex scenario, in which a 3D water network still exists at low hydration numbers, but cannot be interpreted without considering the intercalation of amine groups.

Figure 6.9: Variation of density with the hydration number of the different crystalline phases of the ^{t}BA:water system. Each data point is labelled with the corresponding hydration number; the two HP semi-clathrates are highlighted in red. The orange curve corresponds to the recalculated relationship; the dashed blue curve to the original calculation.[5]

Form II of pure ^{t}BA has not been included in this calculation, because there is not an appropriate reference against which to normalise its density at ambient pressure.

6.3.2.1 Density normalisation

The densities of the structures determined by Dobrzycki et al.[5] (Table 6.5) were normalised to 173 K taking non-deuterated ice Ih as a reference. Unit cell volumes at each temperature of interest were calculated using the V-T equation of state determined by Röttger et al.,[32] normalisation factors f_T were calculated according to Eq. 6.4, and then the densities were normalised according to Eq. 6.5.

A similar procedure was applied to normalise the densities of the high-pressure 5.65- and 5.8-hydrates to ambient pressure. A quadratic expression for the pressure dependence of the density of non-deuterated ice Ih was proposed by Gagnon et al. based on Brillouin spectroscopy data collected at 237.65 K up to 0.28 GPa.[33] Densities were calculated using this relationship, they were converted into volumes afterwards, and all volumes were finally normalised to 173 K:

$$f_T = \frac{V(T)}{V(173)} \quad (6.4)$$

$$\rho(173) = f_T\rho(T) \quad (6.5)$$

$$f_P = \frac{V(P)}{V(0)} \quad (6.6)$$

$$\rho(173) = f_T\rho(T) \quad \Rightarrow \quad \frac{m}{V(173)} = f_T\frac{m}{V(T)} \quad \Rightarrow$$
$$V(173) = \frac{1}{f_T}V(T) \quad (6.7)$$

Normalisation factors calculated according to Eq. 6.6 have been computed using these data (Table 6.7).

Table 6.7: Normalisation factors calculated from data corresponding to non-deuterated ice Ih at 237.65 K.

P (GPa)	ρ (g/cm^3)	V (Å^3)	f_T	T-norm. V (Å^3)	f_P
0.00	0.9228	129.67	1.0076	128.69	
0.61	0.9783	122.32	1.0076	121.40	0.943
0.76	0.9897	120.91	1.0076	119.99	0.932

The normalisation factors in Table 6.7 have been used for all purposes in this study. Nevertheless, acknowledging the inherent risk of extrapolating the polynomial curve from 0.28 to 0.61 and 0.76 GPa, two tests have been performed to ensure that the previous procedure performs appropriately.

A *P-V* equation of state was proposed by Strässle et al. for deuterated ice Ih at 145 K.[34] These data cannot be used to compute the normalisation factors in this study, because the compressibilities of deuterated and non-deuterated ice Ih are slightly different and the isotope effect has not been evaluated at the pressures used in this study.[35] Notwithstanding, the isotope effect is expected to be small and the *P-V* relationship for the deuterated compound takes the form of a Murnaghan equation of state, which can be extrapolated more reliably than a quadratic equation. Normalisation factors were calculated using these data, to check that they do not differ considerably from those calculated previously. This indicates that the extrapolation on the quadratic equation does not introduce large errors. The same procedure as before was used (Eqs. 6.7 and 6.6), although the step involving the densities was not required. The normalisation factors obtained were indeed very close to the ones obtained from non-deuterated ice Ih (Table 6.8). See Fig. 6.10 for a comparison of both equations of state.

As second test, a quadratic equation of the same form as proposed by Gagnon et al. was fitted to 11 data points generated from the Murnaghan *P-V* equation of state. Both the quadratic and Murnaghan equations overlap below 0.5 GPa, but diverge at higher pressures, albeit just slightly (Fig. 6.10). Translating this observation into the data for deuterated ice Ih, it is safe to assume that, in the absence of a more robust equation, the error in the normalisation factors due to the extrapolation of the quadratic equation should be small.

P (GPa)	V (Å^3)	f_T	T-norm. V (Å^3)	f_P
0.00	128.18	1.0035	127.73	
0.61	121.69	1.0035	121.27	0.949
0.76	120.43	1.0035	120.01	0.940

Table 6.8: Normalisation factors calculated from data corresponding to deuterated ice Ih at 237.65 K.
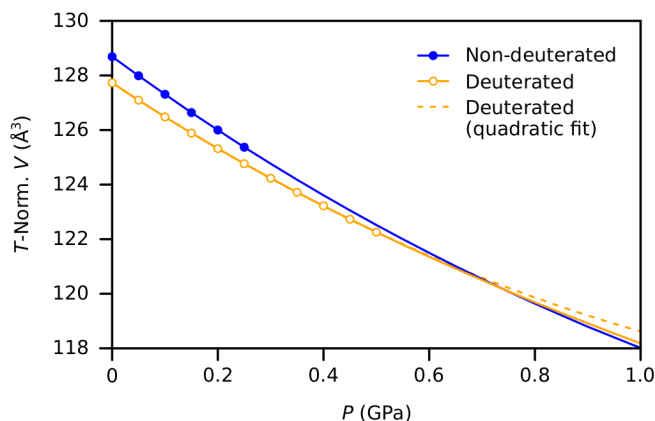
Figure 6.10: Compressibility data of deuterated and non-deuterated ice Ih. Data points are represented in the interval for which data is available (up to 0.28 GPa for non-deuterated ice and up to 0.5 GPa for deuterated ice). The curves are drawn to 1.0 GPa as solid lines to show the crossing point at 0.75 GPa. The quadratic fit of the data for deuterated ice is shown as a dashed line.

6.3.3   Computational interpretation of pure *tert*-butylamine form II

Simulations were initially performed using two different force fields, GROMOS and OPLS. The latter was considered a better choice, and after modifying the atomic charges, was used in the production runs. Each simulation was repeated five times, at the exact same conditions and starting from the same setup. Fig. 6.11 shows a zenithal view of the FESs obtained with the three different force fields, after combining the five repeated simulations. Lighter areas correspond to energy minima, while darker areas should be interpreted as energy barriers. As explained in §6.2.4.4, ^{t}BA molecules are orientationally disordered, but there is no indication that their position inside the unit cell changes significantly.
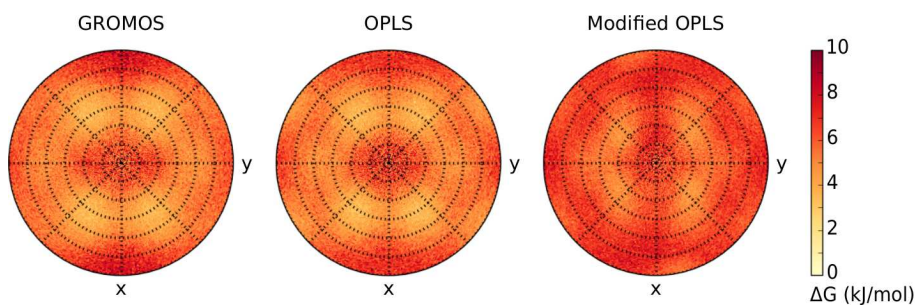
Figure 6.11: FESs of ^{t}BA form II at 0.78 GPa, view along z, calculated from five simulations, using three different force fields, GROMOS, OPLS and modified OPLS.

161

Therefore, it is possible to interpret these graphs as the projection of a hemisphere, with the central carbon atom of a ^{t}BA molecule sitting at the centre of the sphere and the amine being potentially located at any of the low energy regions. The results from GROMOS and OPLS are nearly identical, while the use of the modified OPLS force field suggests a considerably more restricted set of possible orientations. Fig. 6.12 shows similar FESs, obtained from just one simulation instead of five, using either GROMOS or OPLS as a force field. The latter produces a better sampling of the high-energy orientations, while points in the graph obtained using GROMOS are more concentrated at the low-energy regions. Although this difference is compensated by computing the graphs from several simulations, OPLS was chosen as it has the potential advantage of generating more detailed FESs.

Figure 6.12: FESs of ^{t}BA form II at 0.78 GPa, view along z, calculated from a single simulation, using two different force fields, GROMOS and OPLS.
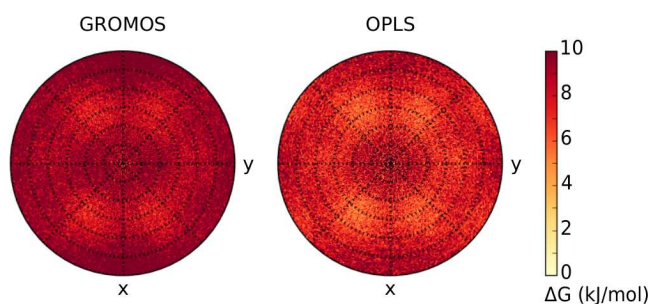
The FES obtained from the simulations at 0.78 GPa using the modified OPLS force field is presented in Fig. 6.13. The results indicate that the amine group is preferentially located in eight different orientations. Energy barriers between different orientations vary between 6 and 9 kJ/mol. Weak H bonds as observed in this structure are expected to exhibit energies in the order of less than 20 kJ/mol,[28] hence dynamic disorder in the solid state cannot be discarded on the basis of these results.

A reconstruction of the eight possible positions in fractional coordinates is shown in Fig. 6.14, superimposed with the four possible positions obtained from X-ray diffraction for comparison. Although each position is represented by a single point, it covers in fact a small area, 15° wide in all directions. Given that the radius of the sphere in which the ^{t}BA molecules are inscribed is 1.47 Å (average length of the carbon–nitrogen
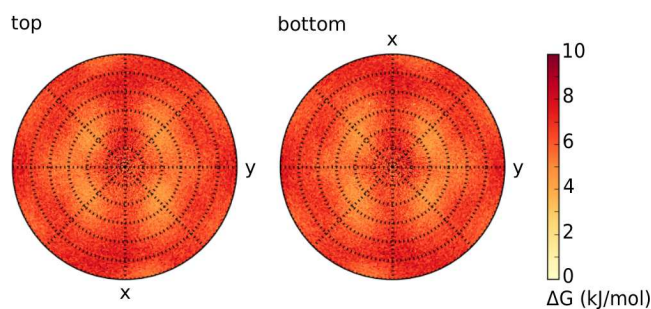
Figure 6.13: FESs of ^{t}BA form II at 0.78 GPa, view along z from above and below, calculated from five simulations using the modified OPLS force field.

bonds), those 15° correspond to a circle perpendicular to the carbon–nitrogen bond 0.768 Å in diameter. Each N atom in the crystal structure can be associated to a different preferred position from the simulation, with distances varying from 0.236 to 0.787 Å, which indicates a good agreement between both structures. The four additional positions predicted by MD probably correspond to the disorder that cannot be modelled from the experimental data. In fact, if all ^{t}BA molecules within the unit cell were to be reoriented as to match one of the eight possible orientations determined from MD, nitrogen-nitrogen distances between neighbouring molecules will range between 2.5 and 3.5 Å, resulting in stronger H bonds.

Figure 6.14: Comparison between the positions of the amine group obtained from X-ray diffraction (in blue) and from MD (in red).

FESs determined from all the molecules inside the simulation box allow to recreate an average structure, but do not inform about its dynamic behaviour. The same analysis script was applied to individual molecules within a simulation; Fig. 6.15 shows FESs for three independent molecules (all the molecules were inspected and show similar behaviour). Differently from previous FESs, sampling from just a single molecule is notably restricted to specific areas, which vary from molecule to molecule. This indicates that each ^{t}BA molecule exhibits a single preferred orientation, with a certain degree of libration (∼15° in

Figure 6.15: FESs of ^{t}BA form II at 0.78 GPa, zenithal view along z, calculated for three different individual molecules of the simulation box.

all directions), that is, there is no dynamic disorder in the structure. Reorientation occurs, as evidenced by spots scattered outside the preferred areas, but can be considered a rare event. Based on this, the structure of ^{t}BA form II can be catalogued as an orientational glass,[6,7] that is, only the central carbon atom of each ^{t}BA molecule is subject to the space group symmetry ($P2_1/n$), while each molecule is orientated differently. This successfully explains the concerns during the refinement expressed in §6.2.4.3.

Another aspect of the structure addressed in §6.2.4.1 concerns the compression of a crystal of ^{t}BA over 0.78 GPa, which inevitably leads to loss of crystal quality. In order to check the effect of compression, two additional simulations were performed at 0.40 and 1.20 GPa. FESs, averaged for all molecules over five simulations, obtained at the four different pressures, are reported in Fig. 6.16. At ambient pressure all orientations are equally likely, that is, the material is in its liquid state, as expected. As pressure increases, preferred orientations become more and more restricted. At 1.20 GPa the eight preferred orientations for the amine group have become four (half of them are visible in the graphs, the other half are visible in the remaining hemisphere, which is symmetric and has not been represented here) and are slightly shifted with respect to the lower pressures. Despite such preferred orientations being lower in energy than those at lower pressure, there is a large connected region at ~5 kJ/mol around them, which indicates that the degree of orientational disorder is even larger than before. The most likely explanation is that at this pressure ^{t}BA molecules get closer, and due to increasing repulsive interactions not all molecules can reorient into the most favourable orientation. This increased degree of internal disorder might also explain why it was not possible to observe X-ray diffraction from samples compressed over 1 GPa.

6.4   Conclusions and outlook

The crystal structures of three new solid phases of ^{t}BA have been determined, increasing the total number of known phases in the water–^{t}BA
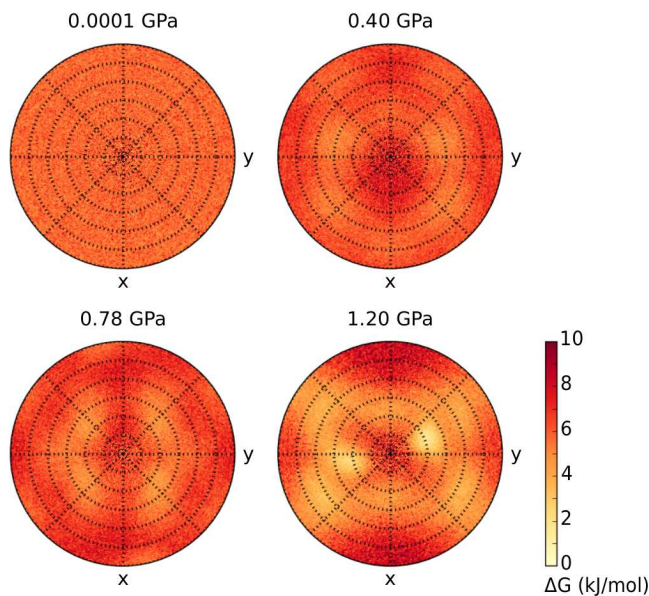
system to ten. Two of these phases are semi-clathrate structures, in which water molecules form an irregular network of cages that host ^{t}BA molecules. The importance of these semi-clathrates stems from the fact that they exhibit the highest density among all the phases, very close to the predicted maximum density of the system. On the basis of experiments conducted at ambient pressure and low temperature, it was previously considered that crystal growth in this high-density region was hindered due to the impossibility of forming a 2D water network which could host the ^{t}BA molecules. By means of in-situ high-pressure crystallisation it has been shown that crystallisation in said region is in fact possible, considerably reducing the size of the gap in which a 2D arrangement of water molecules might appear.

The structure of a new polymorph of pure ^{t}BA has also been reported, showing that cooling and compression of the neat substance lead to two different phases. The HP form II has been characterised as an orientational glass. Single-crystal X-ray diffraction allows to determine the overall symmetry of the crystal, that is, the position of the molecules in the unit cell, but information on molecular orientation cannot be derived due to extensive disorder. MD simulations have been successfully employed to identify the most likely orientations and intermolecular contacts, although this information cannot be directly input in the

refinement, because the average unit cell obtained from the simulated structure is considerably more detailed than what can be determined from the experimentally-measured electron density. Rather than simply averaging all positions, it should be possible to model an arrangement of ^{t}BA molecules in which the most favourable intermolecular interactions are present, which could be a better match to the experimental data. For this, it would be necessary to extract from the simulated trajectory all possible pairs of molecules and calculate the energy of the interaction between the molecules in each pair. The PIXEL method for energy calculations would be particularly well suited for this problem, as it determines not only the total energy of the interaction, but also the magnitude of the Coulombic, polarisation, dispersion, and repulsion contributions. It is expected in this case that the dispersion and repulsion contributions to the intermolecular contacts might be as important as the Coulombic contribution.

In order to fully understand the water–^{t}BA system, a complete systematic screening of the crystallisation by means of pressure of mixtures of different composition should be carried out, as in the present work certain mixtures have been crystallised but could not be characterised. The use of powder X-ray diffraction and Raman spectroscopy might contribute towards this end. Adding more information to the phase diagram, in the form of a density curve in this case, would provide a further insight into the formation of the amine water clathrates which could potentially be used to improve computational models.

References

[1] G. A. Jeffrey, *J. Inclusion Phenom.*, 1984, **1**, 211–222.

[2] R. K. McMullan, G. A. Jeffrey and T. H. Jordan, *J. Chem. Phys.*, 1967, **47**, 1229–1234.

[3] G. A. Jeffrey, *Inclusion Compounds, Vol. I*, Academic Press, 1984, pp. 135–190.

[4] D. Stäben and D. Mootz, *J. Incl. Phenom. Mol.*, 1995, **22**, 145–154.

[5] Ł. Dobrzycki, P. Taraszewska, R. Boese, M. K. Cyrański and S. A. Cirkel, *Angew. Chem. Int. Ed.*, 2015, **54**, 10138–10144.

[6] S. C. Abrahams and W. N. Lipscomb, *Acta Cryst.*, 1952, **5**, 93–99.

[7] J. C. W. Folmer, R. L. Withers, T. R. Welberry and J. D. Martin, *Phys. Rev. B*, 2008, **77**, 1–10.

[8] R. K. McMullan, T. H. Jordan and G. A. Jeffrey, *J. Chem. Phys.*, 1967, **47**, 1218–1222.

[9] L. Palatinus and G. Chapuis, *J. Appl. Cryst.*, 2007, **40**, 786–790.

[10] G. M. Sheldrick, *Acta Cryst.*, 2015, **C71**, 3–8.

[11] C. B. Hübschle, G. M. Sheldrick and B. Dittrich, *J. Appl. Cryst.*, 2011, **44**, 1281–1284.

[12] Bruker-AXS, *APEX2*, Bruker-AXS, Madison, Wisconsin, USA, 2007.

[13] G. M. Sheldrick, *TWINABS v2008-1*, Bruker-AXS, Madison, Wisconsin, USA, 2008.

[14] R. H. Blessing, *Cryst. Rev.*, 1987, **1**, 3–58.

[15] G. M. Sheldrick, *SADABS v2008-1*, Bruker-AXS, Madison, Wisconsin, USA, 2008.

[16] G. M. Sheldrick, *Acta Cryst.*, 2015, **71**, 3–8.

[17] P. W. Betteridge, J. R. Carruthers, R. I. Cooper, K. Prout and D. J. Watkin, *J. Appl. Cryst.*, 2003, **36**, 1487–1487.

[18] L. Schröder, D. J. Watkin, A. Cousson, R. I. Cooper and W. Paulus, *J. Appl. Cryst.*, 2004, **37**, 545–550.

[19] C. F. Macrae, I. J. Bruno, J. A. Chisholm, P. R. Edgington, P. McCabe, E. Pidcock, L. Rodriguez-Monge, R. Taylor, J. van de Streek and P. A. Wood, *J. Appl. Cryst.*, 2008, **41**, 466–470.

[20] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindah, *SoftwareX*, 2015, **1-2**, 19–25.

[21] N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark and W. F. Van Gunsteren, *Eur. Biophys. J.*, 2011, **40**, 843–856.

[22] W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.

[23] A. Nemkevich, H. B. Bürgi, M. A. Spackman and B. Corry, *Phys. Chem. Chem. Phys*, 2010, **12**, 14916–14929.

[24] C. E. S. Bernardes and A. Joseph, *J. Phys. Chem. A*, 2015, **119**, 3023–3034.

[25] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Naka-jima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Broth-ers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Sal-vador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *GAUS-SIAN09 revision A.02*, Gaussian Inc., Wallingford CT, 2009.

[26] J. D. Hunter, *Comput. Sci. Eng.*, 2007, **9**, 99–104.

[27] G. M. Sheldrick, *Acta cryst.*, 2008, **A64**, 112–22.

[28] T. Steiner, *Acta Cryst.*, 2002, **E41**, 49–76.

[29] A. L. Kohl and R. B. Nielsen, *Gas Purification*, Gulf Publishing Company, 1997.

[30] D. Mootz and D. Stäben, *Z. Naturforsch. B*, 1993, **48**, 1325–1330.

[31] M. Wojdyr, *J. Appl. Cryst.*, 2010, **43**, 1126–1128.

[32] K. Röttger, A. Endriss, J. Ihringer, S. Doyle and W. F. Kuhs, *Acta Cryst.*, 1994, **B50**, 644–648.

[33] R. E. Gagnon, H. Kiefte, M. J. Clouter and E. Whalley, *J. Chem. Phys.*, 1988, **89**, 4522–4528.

[34] T. Strässle, S. Klotz, J. S. Loveday and M. Braden, *J. Phys. Condens. Matter*, 2005, **17**, S3029–S3033.

[35] A. Klapproth, E. Goreshnik, D. Staykova, H. Klein and W. F. Kuhs, *Can. J. Phys.*, 2003, **81**, 503–518.

Appendices

A   List of structural sets

The following list contains all structural sets extracted from the CSD in which at least one determination has been recorded at high pressure. Reference code words and numbers have been separated to avoid unnecessary repetitions. The reference code *NNNNNN00* stands for *NNNNNN*; the reference code number *00* is given only to avoid confusion.

Structural set 1

HP	BOLDIP	22 23 24 25 26
AP	BOLDIP	00 10 11 12 13 14 15 16 17 18 19 20 21

Structural set 2

HP	CLMETH	01 02 03 04
AP	CLMETH	00

Structural set 3

HP	FOXNUB	11 12 13 14 15 16 17 18
AP	FOXNUB	00 01 02 03 04 05 06 07 08 09 10

Structural set 4

HP	MBRMET	11 12 13 14 15
AP	MBRMET	10

Structural set 5

HP	NALCYS	19 20 21 22 23 24 25 26 27 28
AP	NALCYS	00 01 02 10 11 12 13 14 15 16 17 18

Structural set 6

HP	ABULIT	03 04 05 06 07
AP	ABULIT	00 01 02

Structural set 7

HP	ABUMIT	01
AP	ABUMIT	00

Structural set 8

HP	ABUMOZ	01
AP	ABUMOZ	00

Structural set 9

HP	ACETAC	08 09
AP	ACETAC	00 01 02 03 07

Structural set 10

HP	ACRLAC	03 04
AP	ACRLAC	00 01 02

Structural set 11

HP	ADAMAN	05
AP	ADAMAN	00 01 02 03 06 07 08

Structural set 12

HP	ADMNTB	02
AP	ADMNTB	00 01 03 04

Structural set 13

HP	AFUDEL	02 03 04 05 06 07 08 09 10
AP	AFUDEL	00 01 11 12

structural set 14

HP	AWIHOE	11 12
AP	AWIHOE	00 01 02 03 04 05 06 07 08 09 10

Structural set 15

HP	BAZGOY	05
AP	BAZGOY	00 01 06

structural set 16

HP	BEDMIG	13 14 15 16 17 18
AP	BEDMIG	00 01 02 03 04 05 06 07 08 09 10 11 12

Structural set 17

HP	BENZEN	08 09 10 11 12 13 15 16 17
AP	BENZEN	00 01 02 18

Structural set 18

HP	BOQCUF	12 13
AP	BOQCUF	00 01 02 03 04 05 06 07 08 09 10 11

Structural set 19

| HP | BOQQUT | 03 04 05 06 07 08 |
| AP | BOQQUT | 00 01 02 |

Structural set 20

| HP | BROFRM | 05 06 |
| AP | BROFRM | 00 |

Structural set 21

| HP | BUQGUQ | 02 03 |
| AP | BUQGUQ | 00 01 |

Structural set 22

| HP | BZDMAZ | 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 |
| AP | BZDMAZ | 00 01 02 03 04 |

Structural set 23

| HP | CARZIF | 03 04 05 06 07 08 09 |
| AP | CARZIF | 00 01 02 |

Structural set 24

HP	CILHIO	07 08
AP	CILHIO	00 03 04 05 06 09 10 12 13 14 15 16 17 19 20 21 22 23 24 26 27 28 31 32 35 36 37 38 40 41 42 43 44
	DATREV	01

Structural set 25

| HP | CLFORM | 01 02 |
| AP | CLFORM | 00 03 |

Structural set 26

| HP | CLPHOL | 14 |
| AP | CLPHOL | 02 10 11 12 13 |

Structural set 27

| HP | COKCEL | 00 |

Structural set 28

| HP | CTMTNA | 01 02 06 |
| AP | CTMTNA | 00 03 04 05 07 08 09 |

Structural set 29

| HP | CUKCIU | 03 |
| AP | CUKCIU | 00 01 02 |

Structural set 30

| HP | DANTEN | 05 06 07 08 09 10 11 |
| AP | DANTEN | 00 01 02 03 04 12 13 14 15 16 |

Structural set 31

| HP | YISBIM | 00 01 |

Structural set 32

| HP | DBRMET | 11 |
| AP | DBRMET | 10 |

Structural set 33

| HP | DCLBEN | 10 |
| AP | DCLBEN | 00 01 02 03 04 05 06 07 08 09 11 |

Structural set 34

| HP | DCLMET | 11 12 |
| AP | DCLMET | 10 |

Structural set 35

| HP | DECWAJ | 02 06 07 08 09 10 11 12 13 14 |
| AP | DECWAJ | 00 01 03 04 05 |

Structural set 36

| HP | DIMETH | 11 13 14 15 |
| AP | DIMETH | 01 02 03 04 05 06 10 12 |

Structural set 37

| HP | DLALNI | 09 10 11 12 13 14 15 16 17 18 19 20 21 |
| AP | DLALNI | 00 01 02 03 04 05 06 07 08 |

Structural set 38

| HP | DLSERN | 19 20 21 22 23 24 25 26 27 28 29 30 |
| AP | DLSERN | 00 02 11 12 13 14 15 16 17 18 31 32 33 |

Structural set 39

| HP | DMETSO | 04 05 06 |
| AP | DMETSO | 00 01 |

Structural set 40

| HP | DOKJAP | 02 |
| AP | DOKJAP | 00 01 |

Structural set 41

| HP | EFUMAU | 03 |
| AP | EFUMAU | 00 01 02 |

Structural set 42

HP	ETANOL	01
AP	ETANOL	00

Structural set 43

HP	ETEYAE	00

Structural set 44

HP	ETEYEI	00

Structural set 45

HP	ETEYIM	00

Structural set 46

HP	FACRIK	02 03 04
AP	FACRIK	00 01 05 06 07 08 09

Structural set 47

HP	FIGYID	01
AP	FIGYID	00

Structural set 48

HP	FOHCEK	00 01 02 03
AP	FOHCEK	04

Structural set 49

HP	FORMAC	04
AP	FORMAC	00 01

Structural set 50

HP	FORMAM	03 04
AP	FORMAM	00 01 02

Structural set 51

HP	GEYREH	02
AP	GEYREH	00 01

Structural set 52

HP	GLUTAM	04 05 06 07 08 09
AP	GLUTAM	00 01 02 03

Structural set 53

HP	GLUTAS	03 04 05
AP	GLUTAS	02 06

Structural set 54

HP	GLYCIN	35 36 37 38 39 40 41 42 43 44
		45 46 47 48 49 50 51 52 53 54
		55 56 57 58 59 60 61 62 63 64
		65 66 67 68 72 73 74 75 76 77
		78 79
AP	GLYCIN	00 01 02 03 04 05 06 07 08 09
		10 11 15 16 17 18 19 20 21 22
		23 24 25 26 27 28 29 30 31 32
		33 34 69 70 71 80 81 82 83 84
		85 86 87 88 89 90 91 92 93 94
		95 96 97 98 99
	DOLBIR	00 01 02 03 04 05 06

Structural set 55

HP	GOFPUN	00 01 02

Structural set 56

HP	GUHHAS	01 02 03
AP	GUHHAS	00

Structural set 57

HP	GUMMUW	04 05 06 07 08 09
AP	GUMMUW	00 01 02 03 10 11

Structural set 58

HP	GHEGHUX	00

Structural set 59

HP	HIXFIX	00
AP	HIXFIX	01 02 03 05

Structural set 60

HP	IMAZOL	15 16 17 18 19 20 21 22 23 24
		25 26 27 29 30 31 32 33
AP	IMAZOL	01 02 03 04 05 06 10 13 14

Structural set 61

HP	IODOFO	03 04
AP	IODOFO	00 02 05

Structural set 62

HP	IZOXOL	00 01 02 03 04 05 06 07 08
AP	IZOXOL	09

Structural set 63

HP KECYBU 07 08

AP KECYBU 00 02 03 05 06 11 12 13 14 15
16 17 18

Structural set 64

HP KETVEK 03

AP KETVEK 00 01 02

Structural set 65

HP LALNIN 25 26 27 28 29 30 31 32 33

AP LALNIN 00 01 03 12 13 14 15 16 17 18
19 20 22 23 24 52 53 54 55

Structural set 66

HP LCYSTI 16 17 18 19

AP LCYSTI 10 11 12 13 14 15

Structural set 67

HP LCYSTN 24 25 26 27

AP LCYSTN 02 12 21 22 23 28 29 30 31 32
33 34 35 36

Structural set 68

HP LENBOV 01 02 03

AP LENBOV 04

Structural set 69

HP LETGIA 07 08 09 10

AP LETGIA 00 01 02 03 04 05 06

Structural set 70

HP MCBENZ 00

AP MCBENZ 01 02

Structural set 71

HP MECHLF 13 14 15

AP MECHLF 02 03 04 10 11 12

Structural set 72

HP MESEFV 05 06 11

AP MESEFV 00 04 13 14

Structural set 73

HP MIGPAU 01 03 04 05 06 07

AP MIGPAU 00 02

Structural set 74

HP MIMETH 12 13 14

AP MIMETH 10

Structural set 75

HP NAGHOT 00 02 03

AP NAGHOT 01

Structural set 76

HP NAPHTA 12 19 20 21 22

AP NAPHTA 00 02 04 05 06 07 08 09 10 11
15 16 17 18 23 24 25 26 27 28
29 30 31 32 33 34 35 36 37 38

Structural set 77

HP NEBGUX 01 02 03 04 05

Structural set 78

HP NIMRIK 01 02 03 04 05

AP NIMRIK 00 06 07 08

Structural set 79

HP NIRJAA 02 03 04 05

AP NIRJAA 00 01

Structural set 80

HP NIRJEE 02 03 04 05 06 07

AP NIRJEE 00 01

Structural set 81

HP NIRJII 02 03 04 05 06 07

AP NIRJII 00 01

Structural set 82

HP NTROMA 02 03 04 05 06 07 08 09 11 12

AP NTROMA 00 01

Structural set 83

HP NUXJUM 02 03 04

AP NUXJUM 00 01

Structural set 84

HP OCRSOL 02

AP OCRSOL 00 01

Structural set 85

HP OCUJOM 00 02

AP OCUJOM 01

QIFNIE 00

Structural set 86

HP OHIJAR 00 01 02 03 04

Structural set 87

HP OMISIM 18

175

Structural set 88

HP   POGJOM   00 01 02

Structural set 89

HP   POHYIV   00

AP   POHYIV   01

Structural set 90

HP   PRONAC   02 03

AP   PRONAC   00 01

Structural set 91

HP   PYRDNO   17 18 19 20 21

AP   PYRDNO   10 11 16

Structural set 92

HP   PYRENE   08 09

AP   PYRENE   00 01 02 03 04 05 07

Structural set 93

HP   PYRZOL   28 29 30 31 32 33 34 35 36 37
                39 40

AP   PYRZOL   00 01 02 04 05 06 07 08 09 10
                11 12 13 14 15 16 17 18 19 20
                21 23 24 25 26 27 38

Structural set 94

HP   QAMTUU   01

AP   QAMTUU   00

Structural set 95

HP   QAMVEG   01

AP   QAMVEG   00

Structural set 96

HP   QAMVEH   01 02

AP   QAMVEH   00

Structural set 97

HP   QESGOL   00 01

Structural set 98

HP   QOSCUX   00

Structural set 99

HP   QQQAUG   10 11 12 13 14 15 16 17

AP   QQQAUG   00 01 02 03 04 05 06 07 08 09
                18

Structural set 100

HP   QQQBNG   02

AP   QQQBNG   00 01

Structural set 101

HP   QQQCIV   06 07 08

AP   QQQCIV   00 01 03 04 05

Structural set 102

HP   RITNOY   00 01

AP   RITNOY   02

Structural set 103

HP   RIVDOQ   02 03

AP   RIVDOQ   00 01

Structural set 104

HP   ROMJUZ   01 03 04

AP   ROMJUZ   00 02

Structural set 105

HP   SACBAA   00 01

Structural set 106

HP   SALMID   03 04 05 06 07

AP   SALMID   00 01 02

Structural set 107

HP   SALOXM   04 05 06 07 08 09

AP   SALOXM   00 01 02 03 10

Structural set 108

HP   SUCROS   18 19 20 21 22 23 24 25 26 27
                28 29 30 31 32 33 34 35 36 37
                38 39 40

AP   SUCROS   00 01 03 04 08 11 12 13 14 15
                16 17

Structural set 109

HP   SUVGEW   00

Structural set 110

HP   SUVGIA   00 01 02 03

Structural set 111

HP   TAURIN   07

AP   TAURIN   01 04 05 06 08 09 10

Structural set 112

HP   TIVWOL   00

AP   TIVWOL   01

Structural set 113

HP UHUTOH   00

Structural set 114

HP UREAXX   24

AP UREAXX   00 01 02 03 04 06 07 08 09 10
                11 12 13 14 15 22 23 25

Structural set 115

HP UTIXAX   01 02

AP UTIXAX   00

Structural set 116

HP VOBJEB   01 02

AP VOBJEB   00

Structural set 117

HP VOFVAN   22 23 24 25 26

AP VOFVAN   00 10 20 21

Structural set 118

HP VUHTOI   00 01 02

Structural set 119

HP VUHVEA   00 01

Structural set 120

HP WANMUU   01 02

AP WANMUU   00

Structural set 121

HP XAJMUS   00

Structural set 122

HP XAXCOQ   00 01

AP XAXCOQ   02 03

Structural set 123

HP XUDVOH   05

AP XUDVOH   00 01 02 03 04

Structural set 124

HP YARZUN   01 02 03

AP YARZUN   00

Structural set 125

HP YIHHON   03 04 05 06 07 08 09 10 11

AP YIHHON   00 02 12 13 14 15 16 17 18 19
                20

Structural set 126

HP YIQGAH   00 01 02 03

Structural set 127

HP YOLDAF   02 03 00 01

Structural set 128

HP YOMXUT   04 05 06 07

AP YOMXUT   03

Structural set 129

HP ZEFJAX   00 01

Structural set 130

HP ZEFPAD   00 01 02 03

Structural set 131

HP ZEFPUX   00 01

Structural set 132

HP ZZZPRO   02

AP ZZZPRO   00 01 05 06 07 08

Structural set 133

HP CITSOP   02 03 04 05 06 07

AP CITSOP   00 01

Structural set 134

HP BALNIN   03 04 05

AP BALNIN   00 01 02

Structural set 135

HP CERKIU   02

AP CERKIU   00 01

Structural set 136

HP IBPRAC   05 07 08 09 10 11 12 13 14 15

AP IBPRAC   00 01 02 03 04 06 16 17 18 19

Structural set 137

HP KOWYEA   03 04 05 06 07 08 09 10 11 12
                13 14

AP KOWYEA   00 01 02

Structural set 138

HP KUBMAX   02 03

AP KUBMAX   00 01

Structural set 139

HP KUDMON   01 02

AP KUDMON   00

Set anhydrate-hydrate 1

Structural set 140

HP BISMEV   04 06 07 08 09 10

AP BISMEV   00 01 02 03 05 11 12 13

Structural set 141

HP LIFNOE   00

Structural set 142

AP YAKWAJ   00 01

Set anhydrate-hydrate 2

Structural set 143

HP BOBVIY   04 05 06 07

AP BOBVIY   00 01 02 03 08 09 10 11 12 13
14 15 16

Structural set 144

HP EVUQAO   00 01

Set anhydrate-hydrate 3

Structural set 145

HP CYSTEA   00

Structural set 146

HP CYSTAC   03 04 05 06 07 08

AP CYSTAC   00 01 02

Set anhydrate-hydrate 4

Structural set 147

HP EJIQEU   01 02

AP EJIQEU   00

Structural set 148

AP AMTETZ   00 01

Set anhydrate-hydrate 5

Structural set 149

HP ETDIAM   14 15 16 17 18 19 20

AP ETDIAM   01 10 12 13

Structural set 150

AP SOWSUS   00

Structural set 151

AP GIXDAS   00

Set anhydrate-hydrate 6

Structural set 152

HP GLYGLY   15 16 17 18

AP GLYGLY   00 01 02 03 04 05 06 07 08 13
14 19 20 21 22 23 24 25 26 28

Structural set 153

AP POTPET   00 01 02 03

Set anhydrate-hydrate 7

Structural set 154

HP HXACAN   09 10 11 12

AP HXACAN   00 01 02 03 04 05 06 07 08 13
14 15 16 17 18 19 20 21 22 23
24 25 26 27 28 29 30 31 32 33
34

Structural set 155

AP HUMJEE   00

Structural set 156

HP WAFNAT   00

Structural set 157

AP XOMWOL   00

Set anhydrate-hydrate 8

Structural set 158

AP VENLUW   00

Structural set 159

HP IFIZIG   02 03 04 05 06 07 08 09

AP IFIZIG   00 01

Set anhydrate-hydrate 9

Structural set 160

AP NELPUP   00

Structural set 161

HP IMEGIR   01 02 03 04 05 06 07 08 09

AP IMEGIR   00

Set anhydrate-hydrate 10

Structural set 162

HP   WEMWEQ   02 03 04 05 06

AP   WEMWEQ   00 01 07 08 09 10 11 12 13 14
                    15

Structural set 163

HP   KICCOO   03 04 05 06 07 08 09 10 11 12

AP   KICCOO   00 01 02

Set anhydrate-hydrate 11

Structural set 164

HP   LSERIN   11 12 13 14 15 16 21 26 41 42
                   43 44 45 46 47 48 49

AP   LSERIN   01 02 03 10 17 18 19 20 27 28

Structural set 165

HP   LSERMH   11 12 13 15 16

AP   LSERMH   10 18 19

Set anhydrate-hydrate 12

Structural set 166

AP   OXALAC   00 01 02 03 04 05 06 07 08 09
                    10 11

Structural set 167

HP   OXACDH   36 37

AP   OXACDH   00 01 03 04 06 07 09 11 12 13
                    14 15 16 17 18 19 21 22 23 24
                    25 26 35 39

Set anhydrate-hydrate 13

Structural set 168

HP   PASWUD   00

Structural set 169

AP   PASXAK   00 01

Structural set 170

HP   PASXIS   01 02 03

AP   PASXIS   00

Set anhydrate-hydrate 14

Structural set 171

HP   PHENOL   11

AP   PHENOL   01 02 03 10

Structural set 172

AP   PHOLHH   00

Set anhydrate-hydrate 15

Structural set 173

HP   PIWXEY   02 03 04 05 06

AP   PIWXEY   00 01

Structural set 174

AP   LAJGUA   00

Set anhydrate-hydrate 16

Structural set 175

HP   PUBMUU   23

AP   PUBMUU   00 01 02 03 04 05 06 07 08 09
                    10 11 12 13 14 15 16 17 18 19
                    20 21 22

Structural set 176

AP   PUBMII   00 01

Structural set 177

AP   FEFRIS   00

Set anhydrate-hydrate 17

Structural set 178

HP   PYRDNA   03 06

AP   PYRDNA   00 01

Structural set 179

AP   PYRTHA   00 01

Set anhydrate-hydrate 18

Structural set 180

HP   THIOUR   17 18 19

AP   THIOUR   00 01 02 03 04 05 06 08 09 10
                    11 12 13 14

Structural set 181

AP   NEXSUG   00 01

Structural set 182

AP   NEXSIU   00 01

Set anhydrate-hydrate 19

Structural set 183

HP   VATSAK   01

AP   VATSAK   00 02 03

Structural set 184

AP   LEBJUX   00

Structural set 185

AP   LEBKEI   00

Set anhydrate-hydrate 20

Structural set 186

AP   QIMKIG   00 01 02 03 04 05

Structural set 187

AP   QIMKOM   00 01 02

Structural set 188

HP   YUZTET   00

Set anhydrate-hydrate 21

Structural set 189

AP   GAMBUT   01 02 03 04 10

Structural set 190

HP   ZIPWOM   00

AP   ZIPWOM   01

Set anhydrate-hydrate 22

Structural set 191

HP   LOWFUA   00

HP   GLUCSA   11 12 13 14 15 16 17 18 19 20
                       21 22 23 24

AP   GAMBUT   00 02 03 10 25

Structural set 192

AP   GLUCMH   00 01

Set anhydrate-hydrate 23

Structural set 193

HP   NEDMIS   02 03 04 05 06 07

AP   NEDMIS   00 01

Structural set 194

AP   TOKYIC   00

B   Compressibility of lattice water

V_{cell}, V_H, and $V_{H,void}$ for all hydrates with at least an HP determination.

IN-SITU FORMED HYDRATES													
Ref. code	Phase	T (K)	P (GPa)	Z	$V_{cell}(T)$ (Å³)	$V_{cell}(293)$ (Å³)	V_Z (Å³)	N	$V_{H,wat}(T)$ (Å³)	$V_{H,nwat}(T)$ (Å³)	$V_{H,wat}(293)$ (Å³)	$V_{H,nwat}(293)$ (Å³)	$V_{H,void}(293)$ (Å³)
Piracetam													
YAKWAJ	1H	150	0.0001	2	396.238	401.621	200.810	1	22.78	168.91	23.09	171.20	6.52
YAKWAJ01	1H	100	0.0001	2	393.225	400.435	200.217	1	22.60	167.60	23.01	170.67	6.53
LIFNOE	2H	293	0.60	2	414.974	414.974	207.487	2	22.32	157.33	22.32	157.33	7.67
									19.62		20.17		
Paracetamol													
HUMJEE	1H	150	0.0001	4	804.177	815.102	203.775	1	21.99	172.04	22.29	174.38	7.11
XOMWOL	3H	150	0.0001	8	2089.640	2118.028	264.753	3	20.51	185.56	20.79	188.08	10.63
									22.09		22.70		
									21.94		22.55		
Gaba													
ZIPWOM01	1H	150	0.0001	8	1169.055	1184.937	148.117	1	20.33	120.06	20.61	121.69	5.82
ZIPWOM	1H	296	0.44	8	1151.612	1151.284	143.910	1	20.07	118.19	20.06	118.16	5.69
Gabapentin													
QIMKOM	1H I	153	0.0001	4	1025.190	1038.825	259.706	1	21.20	227.99	21.48	231.02	7.20
QIMKOM01	1H I	293	0.0001	4	1043.111	1043.111	260.778	1	21.92	231.76	21.92	231.76	7.10
QIMKOM02	1H II	293	0.0001	8	2074.345	2074.345	259.293	1	21.77	230.45	21.77	230.45	7.07
YUZTET	7H	293	0.90	2	737.485	737.485	368.743	7	21.01	207.56	21.01	207.56	14.32
									21.53		22.13		
									20.05		20.61		
									19.68		20.23		
									20.22		20.78		
									20.96		21.54		
									20.00		20.56		
dabcoHI													
EVUQAO	1H α	296	0.5	40	9794.255	9791.464	244.787	1	22.48	212.15	22.47	212.09	10.22
EVUQAO01	1H β	296	1.4	4	856.679	856.435	214.109	1	19.06	185.91	19.05	185.86	9.20
Thiourea													
NEXSUG01	0.67H	296	0.70	12	1154.473	1154.144	96.179	0.67	20.93	77.88	20.92	77.86	4.30
NEXSUG	0.67H	296	0.95	12	1132.561	1132.238	94.353	0.67	20.46	76.61	20.45	76.59	4.06
NEXSIU01	1H	296	0.60	4	425.817	425.696	106.424	1	20.92	80.06	20.91	80.04	5.47
NEXSIU	1H	296	1.20	4	411.966	411.849	102.962	1	20.11	77.51	20.10	77.49	5.37
dabco2HBr													
PASXAK	0.5H α	296	0.0001	8	2030.353	2029.774	253.722	0.5	22.84	231.47	22.83	231.40	10.90
PASXIS	1H α	296	0.0001	8	2080.335	2079.742	259.968	1	23.32	224.54	23.31	224.48	12.18
PASXIS02	1H γ	296	1.50	2	450.385	450.257	225.128	1	21.12	193.32	21.11	193.26	10.75
PASXIS03	1H δ	296	0.78	2	490.151	490.011	245.006	1	21.61	212.36	21.60	212.30	11.10

COMPRESSED HYDRATES

Ref. code	Phase	T (K)	P (GPa)	Z	$V_{cell}(T)$ (Å³)	$V_{cell}(293)$ (Å³)	V_Z (Å³)	N	$V_{H,wat}(T)$ (Å³)	$V_{H,nwat}(T)$ (Å³)	$V_{H,wat}(293)$ (Å³)	$V_{H,nwat}(293)$ (Å³)	$V_{H,void}(293)$ (Å³)
						Cysteic acid							
CYSTAC	1H	295	0.0001	4	699.199	699.066	174.767	1	21.81	147.11	21.81	147.08	5.88
CYSTAC01	1H	295	0.0001	4	699.199	699.066	174.767	1	21.90	146.83	21.90	146.80	6.07
CYSTAC02	1H	293	0.0001	4	698.386	698.386	174.597	1	21.83	146.75	21.83	146.75	6.02
CYSTAC03	1H	293	0.20	4	693.060	693.060	173.265	1	21.79	145.49	21.79	145.49	5.98
CYSTAC04	1H	293	1.20	4	665.239	665.239	166.310	1	20.43	139.98	20.43	139.98	5.90
CYSTAC05	1H	293	2.80	4	633.205	633.205	158.301	1	18.93	133.58	18.93	133.58	5.79
CYSTAC06	1H	293	4.50	4	611.870	611.870	152.968	1	18.05	129.18	18.05	129.18	5.74
						Phenyl alanine							
IFIZIG	1H I	295	0.0001	2	576.668	576.558	288.279	1	24.25	256.13	24.25	256.08	7.95
IFIZIG01	1H I	293	0.0001	2	578.216	578.216	289.108	1	24.38	256.85	24.38	256.85	7.88
IFIZIG02	1H I	293	0.20	2	571.119	571.119	285.560	1	23.83	253.89	23.83	253.89	7.84
IFIZIG03	1H I	293	1.00	2	545.135	545.135	272.568	1	22.39	242.42	22.39	242.42	7.76
IFIZIG04	1H II	293	2.50	4	974.611	974.611	243.653	1	18.14	218.63	18.14	218.63	6.88
IFIZIG05	1H II	293	4.00	4	936.253	936.253	234.063	1	17.11	210.14	17.11	210.14	6.81
						Alaninium oxalate							
IMEGIR	1H	300	0.0001	4	860.043	859.471	214.868	1	21.00	185.28	20.99	185.16	8.72
IMEGIR01	1H	293	0.30	4	839.975	839.975	209.994	1	20.46	180.92	20.46	180.92	8.61
IMEGIR02	1H	293	0.50	4	830.109	830.109	207.527	1	20.11	178.83	20.11	178.83	8.59
IMEGIR03	1H	293	1.10	4	805.899	805.899	201.475	1	19.26	173.77	19.26	173.77	8.44
IMEGIR04	1H	293	1.50	4	791.348	791.348	197.837	1	18.83	170.60	18.83	170.60	8.41
IMEGIR05	1H	293	2.40	4	745.952	745.952	186.488	1	17.26	161.12	17.26	161.12	8.11
IMEGIR06	1H	293	3.10	4	725.179	725.179	181.295	1	16.79	156.48	16.79	156.48	8.02
IMEGIR07	1H	293	3.60	4	713.457	713.457	178.364	1	16.67	153.78	16.67	153.78	7.91
IMEGIR08	1H	293	4.30	4	700.812	700.812	175.203	1	16.26	151.06	16.26	151.06	7.88
						Betaine							
KICCOO	1H	295	0.0001	8	1428.353	1428.082	178.510	1	21.99	150.92	21.99	150.89	5.63
KICCOO02	1H	293	0.0001	8	1424.266	1424.266	178.033	1	21.99	150.42	21.99	150.42	5.62
KICCOO03	1H	293	0.10	8	1404.275	1404.275	175.534	1	21.57	148.37	21.57	148.37	5.59
KICCOO04	1H	293	1.60	8	1284.940	1284.940	160.618	1	19.22	135.98	19.22	135.98	5.42
KICCOO05	1H	293	2.90	8	1227.902	1227.902	153.488	1	18.07	130.08	18.07	130.08	5.34
KICCOO06	1H	293	4.00	8	1190.086	1190.086	148.761	1	17.35	126.13	17.35	126.13	5.28
KICCOO07	1H	293	4.90	8	1163.267	1163.267	145.408	1	16.83	123.33	16.83	123.33	5.25
						Serine							
LSERMH10	1H	295	0.0001	4	554.179	554.074	138.518	1	23.04	109.84	23.04	109.82	5.66
LSERMH18	1H	97	0.0001	4	542.418	552.518	138.129	1	22.26	107.74	22.67	109.75	5.71
LSERMH19	1H	120	0.0001	4	542.049	550.958	137.739	1	22.25	107.65	22.62	109.42	5.70
LSERMH11	1H	293	1.00	4	522.922	522.922	130.731	1	21.24	104.03	21.24	104.03	5.46
LSERMH12	1H	293	2.60	4	494.272	494.272	123.568	1	19.49	98.75	19.49	98.75	5.33
LSERMH13	1H	293	3.70	4	476.027	476.027	119.007	1	18.43	95.35	18.43	95.35	5.23
LSERMH15	1H	293	4.50	4	463.213	463.213	115.803	1	17.67	93.01	17.67	93.01	5.12

	COMPRESSED HYDRATES (continue)												
Ref. code	Phase	T (K)	P (GPa)	Z	$V_{cell}(T)$ (Å³)	$V_{cell}(293)$ (Å³)	V_Z (Å³)	N	$V_{H,wat}(T)$ (Å³)	$V_{H,nwat}(T)$ (Å³)	$V_{H,wat}(293)$ (Å³)	$V_{H,nwat}(293)$ (Å³)	$V_{H,void}(293)$ (Å³)
Oxalic acid													
OXACDH01	2H	100	0.0001	2	245.612	250.115	125.058	2	20.99	73.80 20.99	21.37	75.15 21.57	6.96
OXACDH04	2H	295	0.0001	2	255.390	255.341	127.671	2	21.94	76.66 21.94	21.94	76.65 22.55	6.54
OXACDH06	2H	100	0.0001	2	245.135	249.630	124.815	2	20.97	73.61 20.97	21.35	74.96 21.55	6.95
OXACDH11	2H	295	0.0001	2	255.390	255.341	127.671	2	21.88	76.86 21.88	21.88	79.00 21.88	4.91
OXACDH15	2H	100	0.0001	2	245.951	250.461	125.230	2	21.02	73.90 21.02	21.41	75.25 21.61	6.96
OXACDH16	2H	100	0.0001	2	245.951	250.461	125.230	2	21.02	73.91 21.02	21.41	75.27 21.61	6.95
OXACDH17	2H	130	0.0001	2	247.333	251.163	125.581	2	21.09	74.45 21.09	21.42	75.60 21.68	6.88
OXACDH18	2H	130	0.0001	2	247.333	251.163	125.581	2	21.09	74.45 21.09	21.42	75.60 21.68	6.88
OXACDH19	2H	170	0.0001	2	248.134	251.033	125.517	2	21.23	74.56 21.23	21.48	75.43 21.82	6.79
OXACDH21	2H	170	0.0001	2	248.134	251.033	125.517	2	21.21	74.59 21.21	21.46	75.46 21.80	6.80
OXACDH22	2H	225	0.0001	2	250.974	252.595	126.298	2	21.48	75.44 21.48	21.62	75.93 22.08	6.67
OXACDH23	2H	225	0.0001	2	250.974	252.595	126.298	2	21.47	75.47 21.47	21.61	75.96 22.07	6.66
OXACDH24	2H	295	0.0001	2	255.386	255.337	127.669	2	21.95	76.68 21.95	21.95	76.67 22.56	6.50
OXACDH26	2H	15	0.0001	2	242.683	249.092	124.546	2	20.74	72.84 20.74	21.29	74.76 21.32	7.18
OXACDH35	2H	295	0.0001	2	255.035	254.987	127.493	2	21.96	76.51 21.96	21.96	76.50 22.57	6.47
OXACDH39	2H	293	0.0001	2	247.270	247.270	123.635	2	21.17	74.24 21.17	21.17	74.24 21.76	6.47
OXACDH36	2H	295	1.60	2	226.518	226.475	113.237	2	19.11	68.28 19.11	19.11	68.27 19.64	6.22
OXACDH37	2H	295	3.60	2	206.399	206.360	103.180	2	17.16	62.36 17.16	17.16	62.35 17.64	6.04

C   Volume-pressure diagrams

Calculation of V_{excl} for all hydrates with at least an HP determination.

IN SITU-FORMED HYDRATES										
Ref. code	**Phase**	**T (K)**	**P (GPa)**	**Z**	**$V_{cell}(T)$ (Å³)**	**$V_{cell}(293)$ (Å³)**	**V_Z (Å³)**	**N**	**$V_{H,wat}$ (Å³)**	**V_{excl} (Å³)**
Piracetam										
BISMEV03	I	295	0.0001	4	723.361	723.224	180.806	0	0	180.806
BISMEV	II	295	0.0001	2	348.508	348.442	174.221	0	0	174.221
BISMEV06	II	293	0.45	2	334.013	334.013	167.007	0	0	167.007
BISMEV02	III	295	0.0001	4	684.160	684.030	171.008	0	0	171.008
BISMEV04	IV	293	0.40	4	642.199	642.199	160.550	0	0	160.550
BISMEV07	V	293	0.70	2	325.534	325.534	162.767	0	0	162.767
BISMEV08	V	293	0.90	2	314.262	314.262	157.131	0	0	157.131
BISMEV09	V	293	2.50	2	292.826	292.826	146.413	0	0	146.413
BISMEV10	V	293	4.00	2	280.206	280.206	140.103	0	0	140.103
YAKWAJ01	1H	100	0.0001	2	393.225	400.435	200.217	1	22.37	177.85
LIFNOE	2H	293	0.60	2	414.974	414.974	207.487	2	41.86	165.62
Paracetamol										
HXACAN01	I	295	0.0001	4	776.272	776.125	194.031	0	0	194.031
HXACAN09	I	295	1.00	4	711.734	711.599	177.900	0	0	177.900
HXACAN10	I	295	2.00	4	672.012	671.884	167.971	0	0	167.971
HXACAN11	I	295	3.00	4	651.428	651.304	162.826	0	0	162.826
HXACAN12	I	295	4.00	4	621.881	621.763	155.441	0	0	155.441
HXACAN33	II	293	0.0001	8	1500.185	1500.185	187.523	0	0	187.523
HXACAN29	III	300	0.0001	8	1501.411	1500.413	187.552	0	0	187.552
HUMJEE	1H	150	0.0001	4	804.177	815.102	203.775	1	22.37	181.41
WAFNAT	2H	293	1.10	4	854.296	854.296	213.574	2	40.04	173.53
XOMWOL	3H	150	0.0001	8	2089.640	2118.028	264.753	3	67.10	197.65
GABA										
GAMBUT10	I	295	0.0001	4	561.147	561.040	140.260	0	0	140.260
GAMBUT04	II	295	0.0001	16	2187.059	2186.643	136.665	0	0	136.665
ZIPWOM	1H	296	0.44	8	1151.612	1151.284	143.910	1	21.27	122.64
ZIPWOM01	1H	150	0.0001	8	1169.055	1184.937	148.117	1	22.37	125.75
Gabapentin										
QIMKIG01	II	293	0.0001	4	918.086	918.086	229.522	0	0	229.522
QIMKIG05	III	293	0.0001	8	1879.865	1879.865	234.983	0	0	234.983
QIMKIG04	IV	293	0.0001	4	933.391	933.391	233.348	0	0	233.348
QIMKOM01	1H I	293	0.0001	4	1043.111	1043.111	260.778	1	22.37	238.41
QIMKOM02	1H II	293	0.0001	8	2074.345	2074.345	259.293	1	22.37	236.93
YUZTET	7H	293	0.90	2	737.485	737.485	368.743	7	142.53	226.22
Parabanic acid										
PARBAC02	I	295	0.0001	4	441.756	441.672	110.418	0	0	110.418
BINWEC	1.5H	293	0.40	8	1062.489	1062.489	132.811	1.5	31.72	101.09

Ref. code	Phase	T (K)	P (GPa)	Z	$V_{cell}(T)$ (Å³)	$V_{cell}(293)$ (Å³)	V_Z (Å³)	N	$V_{H,wat}$ (Å³)	V_{excl} (Å³)
					IN SITU-FORMED HYDRATES (continue)					
					tert-Butylamine					
	I	133	0.0001	12	1694.1	1719.9	143.3	0	0	143.3
	0.25H	173	0.0001	32	4668.0	4721.2	147.5	0.25	5.59	141.9
	1H	173	0.0001	4	620.7	627.8	156.9	1	22.37	134.6
	7.25H	203	0.0001	32	10421.7	10510.8	328.5	7.25	162.17	166.3
	7.75H	225	0.0001	16	5461.3	5496.6	343.5	7.75	173.35	170.2
	9.75H	218	0.0001	16	6577.1	6624.0	414.0	9.75	218.09	195.9
	11H	250	0.0001	4	1795.7	1803.0	450.8	11	246.05	204.7
	17H	243	0.0001	8	5409.4	5435.1	679.4	17	380.25	299.1
	II	293	0.78	4	525.0	525.0	131.2	0	0	131.2
	5.65H	293	0.76	32	8317.8	8317.8	259.9	5.65	116.49	143.4
	5.8H	293	0.76	20	5207.1	5207.1	260.4	5.8	119.58	140.8
					dMBzIm					
	α	296	0.11	12	2469	2468	206	0	0	206
	α	296	0.22	12	2443	2442	204	0	0	204
	β	296	0.28	12	2410	2409	201	0	0	201
	β	296	0.33	12	2398	2397	200	0	0	200
	β	296	0.64	12	2333	2332	194	0	0	194
	0.5H	296	0.0001	16	3274.5	3273.6	204.6	0.5	10.82	193.8
	0.5H	296	0.54	16	3115	3114	195	0.5	10.49	184
	0.5H	296	0.70	16	3059	3058	191	0.5	10.39	181
	0.5H	296	1.10	16	2980.3	2979	186	0.5	10.16	176
	0.5H	296	1.49	16	2899.1	2898.3	181.1	0.5	9.93	171.2
					dabcoHI					
BOBVIY08	IV	400	0.0001	1	220.355	218.115	218.115	0	0	218.115
BOBVIY01	V	300	0.0001	1	232.927	232.772	232.772	0	0	232.772
BOBVIY09	VI	296	0.0001	2	434.963	434.839	217.420	0	0	217.420
BOBVIY04	VII	296	0.20	4	867.428	867.181	216.795	0	0	216.795
BOBVIY05	VIII	296	0.20	4	857.709	857.465	214.366	0	0	214.366
BOBVIY06	IX	296	0.50	4	837.640	837.401	209.350	0	0	209.350
BOBVIY07	IX	296	1.00	4	805.759	805.529	201.382	0	0	201.382
EVUQAO	1H α	296	0.50	40	9794.255	9791.464	244.787	1	21.03	223.76
EVUQAO01	1H β	296	1.40	4	856.679	856.435	214.109	1	19.96	194.15
					Thiourea					
THIOUR13	V	295	0.0001	4	360.155	360.087	90.022	0	0	90.022
THIOUR17	VI	295	0.28	4	347.342	347.276	86.819	0	0	86.819
THIOUR18	VI	295	0.38	12	1020.543	1020.349	85.029	0	0	85.029
THIOUR19	VI	295	0.97	12	973.683	973.498	81.125	0	0	81.125
NEXSUG01	0.67H	296	0.70	12	1154.473	1154.144	96.179	0.67	13.93	82.25
NEXSUG	0.67H	296	0.95	12	1132.561	1132.238	94.353	0.67	13.73	80.63
NEXSIU01	1H	296	0.60	4	425.817	425.696	106.424	1	20.91	85.52
NEXSIU	1H	296	1.20	4	411.966	411.849	102.962	1	20.19	82.77

					IN SITU-FORMED HYDRATES (continue)					
Ref. code	Phase	T (K)	P (GPa)	Z	$V_{cell}(T)$ (Å^3)	$V_{cell}(293)$ (Å^3)	V_Z (Å^3)	N	$V_{H,wat}$ (Å^3)	V_{excl} (Å^3)
					dabco2HBr					
PASWUD	β	296	0.0001	4	999.174	998.889	249.722	0	0	249.722
PASXAK	0.5H α	296	0.0001	8	2030.353	2029.774	253.722	0.5	10.82	242.901
PASXAK01	0.5H β	296	0.0001	4	1028.137	1027.844	256.961	0.5	10.82	246.140
PASXIS	1H α	296	0.0001	8	2080.335	2079.742	259.968	1	21.64	238.326
PASXIS01	1H β	296	0.48	4	989.198	988.916	247.229	1	21.05	226.178
PASXIS02	1H γ	296	1.50	2	450.385	450.257	225.128	1	19.85	205.282
PASXIS03	1H δ	296	0.78	2	490.151	490.011	245.006	1	20.69	224.316

COMPRESSED HYDRATES										
Ref. code	Phase	T (K)	P (GPa)	Z	$V_{cell}(T)$ (Å³)	$V_{cell}(293)$ (Å³)	V_Z (Å³)	N	$V_{H,wat}$ (Å³)	V_{excl} (Å³)
L-Cysteic acid										
CYSTEA	I	295	0.0001	4	631.180	631.060	157.765	0	0	157.765
CYSTAC02	1H	293	0.0001	4	698.386	698.386	174.597	1	21.64	152.95
CYSTAC03	1H	293	0.20	4	693.060	693.060	173.265	1	21.39	151.87
CYSTAC04	1H	293	1.20	4	665.239	665.239	166.310	1	20.19	146.12
CYSTAC05	1H	293	2.80	4	633.205	633.205	158.301	1	18.41	139.89
CYSTAC06	1H	293	4.50	4	611.870	611.870	152.968	1	16.69	136.28
CYSTAC07	1H	293	5.80	4	595.468	595.468	148.867	1	15.48	133.38
CYSTAC08	1H	293	6.80	4	589.954	589.954	147.489	1	14.61	132.87
(S)-4-Sulfo-L-phenylalanine										
VENLUW	I	293	0.0001	2	526.627	526.627	263.314	0	0	263.314
IFIZIG	1H I	295	0.0001	2	576.668	576.558	288.279	1	21.64	266.64
IFIZIG02	1H I	293	0.20	2	571.119	571.119	285.560	1	21.39	264.17
IFIZIG03	1H I	293	1.00	2	545.135	545.135	272.568	1	20.43	252.14
IFIZIG04	1H II	293	2.50	4	974.611	974.611	243.653	1	18.73	224.92
IFIZIG05	1H II	293	4.00	4	936.253	936.253	234.063	1	17.18	216.88
DL-Alaninium oxalate										
NELPUP	I	293	0.0001	4	793.931	793.931	198.483	0	0	198.483
IMEGIR	1H	300	0.0001	4	860.043	859.471	214.868	1	21.64	193.23
IMEGIR01	1H	293	0.30	4	839.975	839.975	209.994	1	21.27	188.72
IMEGIR02	1H	293	0.50	4	830.109	830.109	207.527	1	21.03	186.50
IMEGIR03	1H	293	1.10	4	805.899	805.899	201.475	1	20.31	181.16
IMEGIR04	1H	293	1.50	4	791.348	791.348	197.837	1	19.85	177.99
IMEGIR05	1H	293	2.40	4	745.952	745.952	186.488	1	18.84	167.65
IMEGIR06	1H	293	3.10	4	725.179	725.179	181.295	1	18.10	163.20
IMEGIR07	1H	293	3.60	4	713.457	713.457	178.364	1	17.58	160.78
IMEGIR08	1H	293	4.30	4	700.812	700.812	175.203	1	16.88	158.32

Ref. code	Phase	T (K)	P (GPa)	Z	$V_{cell}(T)$ (Å³)	$V_{cell}(293)$ (Å³)	V_Z (Å³)	N	$V_{H,wat}$ (Å³)	V_{excl} (Å³)
COMPRESSED HYDRATES (continue)										
Betaine										
WEMWEQ	I	295	0.0001	4	611.612	611.496	152.874	0	0	152.874
WEMWEQ02	I	293	0.80	4	577.640	577.640	144.410	0	0	144.410
WEMWEQ03	I	293	1.40	4	558.889	558.889	139.722	0	0	139.722
WEMWEQ04	I	293	2.20	4	540.905	540.905	135.226	0	0	135.226
WEMWEQ05	I	293	3.20	4	522.441	522.441	130.610	0	0	130.610
WEMWEQ06	I	293	4.10	4	509.535	509.535	127.384	0	0	127.384
KICCOO	1H	295	0.0001	8	1428.353	1428.082	178.510	1	21.64	156.87
KICCOO03	1H	293	0.10	8	1404.275	1404.275	175.534	1	21.52	154.02
KICCOO04	1H	293	1.60	8	1284.940	1284.940	160.618	1	19.73	140.88
KICCOO05	1H	293	2.90	8	1227.902	1227.902	153.488	1	18.31	135.18
KICCOO06	1H	293	4.00	8	1190.086	1190.086	148.761	1	17.18	131.58
KICCOO07	1H	293	4.90	8	1163.267	1163.267	145.408	1	16.31	129.10
L-Serine										
LSERIN10	I	295	0.0001	4	448.776	448.691	112.173	0	0	112.173
LSERIN41	I	293	0.10	4	442.267	442.267	110.567	0	0	110.567
LSERIN42	I	293	0.30	4	443.614	443.614	110.904	0	0	110.904
LSERIN43	I	293	0.50	4	436.602	436.602	109.151	0	0	109.151
LSERIN44	I	293	0.80	4	427.346	427.346	106.837	0	0	106.837
LSERIN45	I	293	1.00	4	427.399	427.399	106.850	0	0	106.850
LSERIN12	I	293	1.40	4	419.181	419.181	104.795	0	0	104.795
LSERIN46	I	293	2.10	4	410.549	410.549	102.637	0	0	102.637
LSERIN13	I	293	2.90	4	404.488	404.488	101.122	0	0	101.122
LSERIN47	I	293	3.10	4	407.435	407.435	101.859	0	0	101.859
LSERIN48	I	293	4.00	4	395.111	395.111	98.778	0	0	98.778
LSERIN14	I	293	4.10	4	395.750	395.750	98.938	0	0	98.938
LSERIN26	I	293	4.20	4	392.807	392.807	98.202	0	0	98.202
LSERIN15	I	293	4.80	4	390.090	390.090	97.523	0	0	97.523
LSERIN03	II	295	0.0001	8	899.918	899.747	112.468	0	0	112.468
LSERIN49	II	293	5.00	4	374.873	374.873	93.718	0	0	93.718
LSERMH10	1H	295	0.0001	4	554.179	554.074	138.518	1	21.64	116.88
LSERMH11	1H	293	1.00	4	522.922	522.922	130.731	1	20.43	110.30
LSERMH12	1H	293	2.60	4	494.272	494.272	123.568	1	18.63	104.94
LSERMH13	1H	293	3.70	4	476.027	476.027	119.007	1	17.48	101.53
LSERMH15	1H	293	4.50	4	463.213	463.213	115.803	1	16.69	99.11
Oxalic acid										
OXALAC06	α	295	0.0001	4	313.884	313.824	78.456	0	0	78.456
OXALAC07	β	298	0.0001	2	157.020	156.945	78.473	0	0	78.473
OXACDH11	2H	295	0.0001	2	255.390	255.341	127.671	2	43.28	84.39
OXACDH36	2H	295	1.60	2	226.518	226.475	113.237	2	39.47	73.77
OXACDH37	2H	295	3.60	2	206.399	206.360	103.180	2	35.16	68.02

AMBIENT-PRESSURE HYDRATES										
Ref. code	Phase	T (K)	P (GPa)	Z	$V_{cell}(T)$ (Å³)	$V_{cell}(293)$ (Å³)	V_Z (Å³)	N	$V_{H,wat}$ (Å³)	V_{excl} (Å³)
Ethylenediamine										
ETDIAM14	I α	296	0.15	2	182.918	182.866	91.433	0	0	91.433
ETDIAM15	I β	296	0.30	2	168.946	168.898	84.449	0	0	84.449
ETDIAM16	I β	296	1.10	2	161.644	161.598	80.799	0	0	80.799
ETDIAM17	I γ	296	1.50	4	321.419	321.327	80.332	0	0	80.332
ETDIAM18	III	296	0.20	2	179.477	179.426	89.713	0	0	89.713
ETDIAM19	III	296	0.30	2	173.214	173.165	86.582	0	0	86.582
ETDIAM20	III	296	0.50	2	172.526	172.477	86.238	0	0	86.238
SOWSUS	1H	268	0.0001	4	480.569	481.710	120.428	1	21.64	98.79
GIXDAS	2H	123	0.0001	4	555.789	564.765	141.191	2	43.28	97.91
Glycylglycine										
GLYGLY14	α	293	0.0001	4	578.692	578.692	144.673	0	0	144.673
GLYGLY15	α	293	1.40	4	532.441	532.441	133.110	0	0	133.110
GLYGLY16	α	293	3.00	4	511.844	511.844	127.961	0	0	127.961
GLYGLY17	α	293	3.70	4	499.272	499.272	124.818	0	0	124.818
GLYGLY18	α	293	4.70	4	492.861	492.861	123.215	0	0	123.215
GLYGLY01	β	295	0.0001	8	1159.746	1159.526	144.941	0	0	144.941
GLYGLY08	γ	295	0.0001	4	583.783	583.672	145.918	0	0	145.918
POTPET03	1.5H	295	0.0001	8	1391.156	1390.892	173.861	1.5	32.46	141.40
Phenol										
PHENOL02	I	295	0.0001	6	826.108	825.951	137.659	0	0	137.659
PHENOL11	II	293	0.16	6	756.419	756.419	126.070	0	0	126.070
PHOLHH	0.5H	295	0.0001	8	1132.346	1132.131	141.516	0.5	10.82	130.70
Cyclohexane-1,2-diamine										
PIWXEY	I	220	0.0001	2	341.921	344.292	172.146	0	0	172.146
PIWXEY02	I	296	0.36	2	346.825	346.726	173.363	0	0	173.363
PIWXEY03	I	296	0.52	2	329.784	329.690	164.845	0	0	164.845
PIWXEY04	I	296	0.65	2	324.818	324.725	162.363	0	0	162.363
PIWXEY05	I	296	1.19	2	311.288	311.199	155.600	0	0	155.600
PIWXEY06	I	296	2.04	2	291.039	290.956	145.478	0	0	145.478
LAJGUA	1H	100	0.0001	8	1552.126	1580.584	197.573	1	21.64	175.93
Pyridine										
PYRDNA01	I	153	0.0001	16	1784.225	1807.955	112.997	0	0	112.997
PYRDNA06	I	295	1.23	16	1665.516	1665.200	104.075	0	0	104.075
PYRDNA03	II	293	1.10	4	413.962	413.962	103.491	0	0	103.491
PYRTHA10	3H	223	0.0001	8	1507.175	1517.198	189.650	3	67.10	122.55

Ref. code	Phase	T (K)	P (GPa)	Z	$V_{cell}(T)$ (Å³)	$V_{cell}(293)$ (Å³)	V_Z (Å³)	N	$V_{H,wat}$ (Å³)	V_{excl} (Å³)
AMBIENT-PRESSURE HYDRATES (continue)										
α-**D-Glucose**										
GLUCSA10	I	295	0.0001	4	765.919	765.773	191.443	0	0	191.443
GLUCSA13	I	295	0.27	4	718.518	718.381	179.595	0	0	179.595
GLUCSA11	I	295	0.88	4	733.265	733.126	183.281	0	0	183.281
GLUCSA12	I	295	1.12	4	722.884	722.747	180.687	0	0	180.687
GLUCSA14	I	295	1.80	4	707.518	707.384	176.846	0	0	176.846
GLUCSA15	I	295	2.60	4	685.682	685.552	171.388	0	0	171.388
GLUCSA16	I	295	3.21	4	675.414	675.286	168.821	0	0	168.821
GLUCSA23	I	295	4.00	4	657.831	657.706	164.427	0	0	164.427
GLUCSA24	I	295	4.20	4	651.831	651.707	162.927	0	0	162.927
GLUCMH11	1H	295	0.0001	2	430.674	430.592	215.296	1	21.64	193.65
MDMA										
NEDMIS01	I	293	0.0001	4	1217.768	1217.768	304.442	0	0	304.442
NEDMIS02	I	293	1.14	4	1129.311	1129.311	282.328	0	0	282.328
NEDMIS03	I	293	2.20	4	1053.087	1053.087	263.272	0	0	263.272
NEDMIS04	I	293	3.84	4	970.304	970.304	242.576	0	0	242.576
NEDMIS05	I	293	4.42	4	952.540	952.540	238.135	0	0	238.135
TOKYIC	1H	120	0.0001	4	1313.181	1334.763	333.691	1	21.64	312.05
tert-**Butanol**										
VATSAK02	II	220	0.0001	18	2382.977	2399.503	133.306	0	0	133.306
VATSAK01	II	293	0.85	18	2155.243	2155.243	119.736	0	0	119.736
VATSAK	IV	295	0.0001	6	874.408	874.242	145.707	0	0	145.707
LEBJUX	2H	123	0.0001	2	362.308	368.159	184.080	2	43.28	140.80
LEBKEI	7H	123	0.0001	4	1275.806	1296.410	324.103	7	151.50	172.61
1H-Tetrazole-5-amine										
EJIQEU	I	298	0.0001	4	337.289	337.129	84.282	0	0	84.282
EJIQEU01	I	298	4.80	4	283.494	283.359	70.840	0	0	70.840
AMTETZ01	1H	295	0.0001	4	455.978	455.891	113.973	1	21.64	92.33
Hexanitrohexa-Azaisowurtzitane										
PUBMUU01	*β*	295	0.0001	4	1465.981	1465.702	366.426	0	0	366.426
PUBMUU06	*γ*	295	0.0001	4	1517.408	1517.120	379.280	0	0	379.280
PUBMUU05	*ε*	295	0.0001	4	1432.021	1431.749	357.937	0	0	357.937
PUBMUU23	*ζ*	293	3.30	4	1279.097	1279.097	319.774	0	0	319.774
PUBMII01	0.25H	295	0.0001	8	2985.274	2984.707	373.088	0.25	5.59	367.50
FEFRIS	0.75H	295	0.0001	8	3021.867	3021.293	377.662	0.57	12.75	364.91

D List of available determinations of β-cyclodextrin form I

Ref. code	T (K)^a	R (%)	N	Reference
BCDEXD	RT	7.1	12	K. Lindner and W. Saenger, *Angew. Chem. Int. Ed.*, 1978, **17**, 694.
BCDEXD01	120	13	12	J. J. Stezowski and J. M. Maclennan, *ACA Ser. 2*, 1980, **7**, 24a.
BCDEXD02	RT	7.2	12	T. Fujiwara et al., *J. Chem. Soc. Jpn.*, 1983, 181.
BCDEXD03	RT	6.1	12.26	T. Steiner and G. Koellner, *J. Am. Chem. Soc.*, 1994, **116**, 5122.
BCDEXD04	RT	5.9	11.56	T. Steiner and G. Koellner, *J. Am. Chem. Soc.*, 1994, **116**, 5122.
BCDEXD05	RT	5.9	11.89	T. Steiner and G. Koellner, *J. Am. Chem. Soc.*, 1994, **116**, 5122.
BCDEXD10	RT	7.3	12	K. Lindner and W. Saenger, *Carbohydr. Res.*, 1982, **99**, 103.
BCDEXD11	RT	—	12	J. A. Hamilton et al., *Acta. Cryst.*, 1968, **B24**, 1560.
BUVSEQ	RT	8.0	11	T. Fujiwara et al., *J. Chem. Soc. Jpn.*, 1983, 181.
BUVSEQ01	RT	6.0	11.19	T. Steiner and G. Koellner, *J. Am. Chem. Soc.*, 1994, **116**, 5122.
BUVSEQ04	100	—	11	C. Kurokawa et al., *Supramo. Chem.*, 2004, **16**, 381.
POBRON	RT	6.8	9.35	T. Steiner and G. Koellner, *J. Am. Chem. Soc.*, 1994, **116**, 5122.
POBRON01	RT	7.49	9	L. Damodharan et al., *Mol. Cyst. Liq. Cryst.*, 2004, **423**, 17.
POBRON02	RT	—	—	J. Szejtli and Z. Budai, *Acta Chim. Acad. Sci. Hung.*, 1977, **94**, 383.
POBRUT	RT	—	1	T. Steiner and G. Koellner, *J. Am. Chem. Soc.*, 1994, **116**, 5122.
POBRUT01	RT	—	1	T. Steiner and G. Koellner, *J. Am. Chem. Soc.*, 1994, **116**, 5122.
POBRUT02	RT	—	1	T. Steiner and G. Koellner, *J. Am. Chem. Soc.*, 1994, **116**, 5122.
POBRUT03	RT	—	1	T. Steiner and G. Koellner, *J. Am. Chem. Soc.*, 1994, **116**, 5122.
SUJDAD	110	3.2	10.41	R. W. Seidel and B. B. Koleva, *Acta Cryst.*, 2009, **E65**, o3162.
TUSHUL	RT	7.6	7.5	J.-Y. Li et al., *Carbohydr. Res.*, 2010, **345**, 685.

^{a} RT stands for *room temperature*, between 283 and 303 K according to the CSD.

E   List of all the structures included in the XPAC calculation

$P2_1$ structures used in the XPAC calculation with unit-cell parameters, guest, and hydration number (N). Structures with the reference code in bold are those in which the a axis is longer than the c axis. In all the cases the b axis is the shortest axis, with the exception of the reference code in italics.

Ref. code	a (Å)	b (Å)	c (Å)	β (°)	V (Å^3)	guest	N
ARUXIU	15.1711(3)	10.16870(10)	20.9862(4)	110.92	3024.13	formic acid	7.7
ARUXOA	15.263(4)	10.157(2)	21.044(5)	110.698(4)	3051.81	acetic acid	7.7
BCDEXD03	21.283(5)	10.322(1)	15.092(5)	112.41(1)	3065.07		12.26
BCDEXD04	21.161(7)	10.254(1)	15.110(7)	111.91(2)	3041.83		11.56
BCDEXD05	21.233(5)	10.294(1)	15.103(4)	112.22(1)	3055.96		11.89
BCDEXD10	21.29(2)	10.33(1)	15.10(2)	112.3(5)	3072.51		12
BOBPEN	21.25(3)	10.28(2)	15.30(2)	113.25(9)	3070.86	hydrogen iodide	8
BOBPIR	21.03(3)	10.11(2)	15.33(2)	111.02(9)	3042.47	methanol	6.5
BUVSEQ01	21.080(7)	10.197(1)	15.131(6)	111.58(2)	3024.47		11.19
DEBGOG	15.356(3)	10.101(2)	21.287(3)	112.81(1)	3043.63	benzyl alcohol	5
DEWCOY	15.4244(16)	10.1574(11)	20.557(2)	110.074(2)	3025.04	benzamide	6
DIRVOP	15.285(3)	10.345(2)	20.118(2)	102.14(1)	3109.98	hexamethylenetetramine	6
GUXZOO	15.1220(4)	10.3350(2)	20.9150(3)	109.575(3)	3079.8	m-aminophenol	7.5
HAQNUK	15.274(12)	10.178(8)	20.961(16)	110.251(7)	3057.14	1,4-dioxane	7
IZALEB	15.0678(9)	10.2126(6)	20.9653(13)	111.7473(12)	2996.55	nicotinic acid	7.65
KIJSEC	15.1977(7)	10.1763(5)	20.6943(6)	109.239(4)	3021.76	succinic acid	7
KOBRIC	15.223(5)	10.578(3)	20.204(6)	108.37(7)	3087.64	potassium hydroxide	8
KUTKOZ	21.199(12)	9.973(3)	15.271(8)	110.87(3)	3016.74	1,4-butanediol	6.25
LADBAV	15.15(2)	10.343(15)	20.93(3)	109.69(4)	3087.89	pentanedioic acid	4.2
LADQOX	20.82(1)	10.366(6)	14.968(9)	110.12(4)	3033.26	diaqua-dichloro-copper(II)	8.9
MIFHAK	15.231(3)	10.068(3)	21.117(3)	110.16(5)	3039.81	squaric acid	6.65
NUFVEQ	20.8314(9)	10.2937(6)	15.2155(8)	110.733(2)	3051.4	pyrazinamide	5.5
NUFVIU	20.95040(10)	10.2493(2)	15.1424(3)	108.7630(10)	3078.69	piperazine	7.2
PIJGIY	21.212(1)	10.021(3)	15.208(1)	111.47(3)	3008.38	ethylene glycol	8
PIJGOE	21.322(6)	9.954(1)	15.251(6)	111.20(1)	3017.8	glycerol	7.2
POBRON	20.857(6)	10.158(1)	15.140(6)	110.94(1)	2995.79		9.35
POBRON01	15.107(7)	10.242(3)	20.911(8)	110.27(3)	3035.1		9
POVSIC	20.042(6)	10.378(2)	15.123(9)	102.30(2)	3073.32	trans-cyclohexane-1,4-diol	5.4
SUJDAD	20.8353(4)	9.93970(10)	15.2043(3)	110.630(2)	2946.85		10.41
TUSHUL	15.1667(5)	10.1850(3)	20.9694(7)	110.993(2)	3024.2		7.5
UPULEX	20.7890	10.2084	15.1091	110.8250	2997.01	4-aminobenzoic acid	10
VACZIJ	15.1550(3)	10.2850(2)	20.9056(5)	109.86	3064.74	dimethylsulfoxide	7.35
VOJLIQ	15.2196(10)	10.2279(7)	20.9321(14)	110.9270(10)	3043.45	p-aminobenzoic acid	5
XEGTAF	10.601(4)	26.402(9)	15.586(6)	93.145(6)	4355.76	6'-Methoxycinchonan-9-ol	16.5
YIYSII	21.310	9.987	15.247	111.85	3011.8	diethanolamine	6.4
YIYSOO	21.451	10.014	15.240	111.25	3051.12	1,5-pentanediol	6.2
ZIGZIY	20.988(9)	10.092(2)	15.223(10)	110.30(6)	3024.13	but-2-yne-1,4-diol	7
β-CD–DMF	15.500(3)	9.794(2)	21.689(5)	90.138(3)	3292.5(12)	dimethylformamide	12.5

F   Detailed description of the disorder modelled in the high-pressure setup of α-cyclodextrin used in the equilibration as an infinite crystal

The setup consists of ten unit cells and 40 α-CD molecules (four molecules per unit cell). α-CD molecules in the setup are labelled from 1 to 40. The primary hydroxyl group of the glucose ring 1 (O6_1) exhibits (–)-gauche conformation (pointing outwards the CD cavity) in half the α-CD molecules, and (+) conformation (pointing inwards the cavity) in the other half. The 20 α-CD molecules which contain a (–) 06_1 group can also contain water molecules in the positions **WC** and **WD**; 32 out of the 40 available positions are occupied (16 **WC** and 16 **WD**).

α-CD index	O6_1 group conformation	WC occupied	WD occupied	α-CD index	O6_1 group conformation	WC occupied	WD occupied
1	(–)	yes	yes	21	(–)	yes	no
2	(–)	yes	no	22	(+)	-	-
3	(–)	yes	yes	23	(–)	yes	yes
4	(–)	yes	yes	24	(–)	yes	yes
5	(+)	-	-	25	(–)	no	yes
6	(+)	-	-	26	(+)	-	-
7	(–)	yes	yes	27	(–)	yes	yes
8	(–)	yes	no	28	(+)	-	-
9	(+)	-	-	29	(+)	-	-
10	(–)	yes	no	30	(+)	-	-
11	(–)	yes	yes	31	(+)	-	-
12	(–)	yes	yes	32	(+)	-	-
13	(+)	-	-	33	(+)	-	-
14	(–)	yes	yes	34	(–)	no	yes
15	(+)	-	-	35	(+)	-	-
16	(–)	no	yes	36	(+)	-	-
17	(–)	yes	yes	37	(+)	-	-
18	(+)	-	-	38	(–)	no	yes
19	(+)	-	-	39	(–)	yes	yes
20	(+)	-	-	40	(+)	-	-

G Detailed description of the disorder modelled in the high-pressure setup of α-cyclodextrin used in the equilibration as a finite crystal surrounded by water

See Appx. F for a detailed description on how the disorder is modelled. Then main difference here is a double number of α-CD molecules.

α-CD index	O6_1 group conformation	WC occupied	WD occupied	α-CD index	O6_1 group conformation	WC occupied	WD occupied
1	$(-)$	yes	yes	41	$(+)$	-	-
2	$(+)$	-	-	42	$(+)$	-	-
3	$(-)$	no	yes	43	$(-)$	yes	yes
4	$(-)$	yes	yes	44	$(-)$	yes	no
5	$(-)$	yes	yes	45	$(+)$	-	-
6	$(-)$	yes	yes	46	$(+)$	-	-
7	$(-)$	yes	yes	47	$(-)$	yes	yes
8	$(-)$	no	yes	48	$(+)$	-	-
9	$(-)$	no	yes	49	$(+)$	-	-
10	$(-)$	yes	no	50	$(-)$	yes	no
11	$(+)$	-	-	51	$(+)$	-	-
12	$(-)$	yes	yes	52	$(-)$	no	yes
13	$(-)$	yes	yes	53	$(+)$	-	-
14	$(+)$	-	-	54	$(+)$	-	-
15	$(-)$	yes	yes	55	$(+)$	-	-
16	$(-)$	yes	yes	56	$(-)$	yes	yes
17	$(+)$	-	-	57	$(+)$	-	-
18	$(-)$	yes	yes	58	$(-)$	yes	yes
19	$(-)$	yes	yes	59	$(-)$	yes	yes
20	$(-)$	yes	yes	60	$(+)$	-	-
21	$(+)$	-	-	61	$(+)$	-	-
22	$(+)$	-	-	62	$(+)$	-	-
23	$(+)$	-	-	63	$(+)$	-	-
24	$(+)$	-	-	64	$(-)$	yes	no
25	$(+)$	-	-	65	$(-)$	yes	no
26	$(-)$	yes	yes	66	$(+)$	-	-
27	$(+)$	-	-	67	$(-)$	yes	yes
28	$(+)$	-	-	68	$(+)$	-	-
29	$(+)$	-	-	69	$(-)$	yes	yes

(continue)							
α-CD index	O6_1 group conformation	WC occupied	WD occupied	α-CD index	O6_1 group conformation	WC occupied	WD occupied
30	(+)	-	-	**70**	(−)	no	no
31	(−)	yes	yes	**71**	(+)	-	-
32	(+)	-	-	**72**	(+)	-	-
33	(+)	-	-	**73**	(−)	yes	yes
34	(−)	yes	yes	**74**	(−)	no	yes
35	(+)	-	-	**75**	(+)	-	-
36	(−)	yes	yes	**76**	(−)	yes	yes
37	(+)	-	-	**77**	(+)	-	-
38	(−)	yes	yes	**78**	(−)	yes	yes
39	(+)	-	-	**79**	(−)	yes	no
40	(−)	no	yes	**80**	(+)	-	-

H   Complete results from the simulations of α-cyclodextrin infinite crystals

In the main text, the conformation of the O6 groups is always defined as (+)- or (−)-gauche, based on the value of the torsion angle O5–C5–C6–O6, as described in Fig. 3.6. The code used to analyse the MD trajectories uses instead the torsion angle C4–C5–C6–O5. According to this definition, when an O6 group is pointing outwards the α-CD cavity, the torsion angle is smaller than $100°$ and the conformation is (+)-gauche. When it is pointing inwards, the angle is larger $100°$ and the conformation becomes *trans*-gauche. In the following, the conformation is given first according to the definition of the analysis script, and in parenthesis according to the conventional definition, followed by the terms *in* or *out* for clarity. Torsion angle values are given only according to the definition employed in the script.

Simulation	AP setup at 0.0001 GPa				
O6 group	**Conf.**	**Torsion angle (°)**		**Length**	**Occupancy**
		μ	σ	**(ns)**	**(%)**
	+ (−, out)	−	−	−	−
O6_1	*trans* (+, in)	188.19	7.91	2375.392	98.97
	Intermediate	281.23	7.38	0.088	−
	Other			24.568	1.02
	+ (−, out)	46.12	7.33	2368.156	98.67
O6_2	*trans* (+, in)	194.81	9.11	18.724	0.78
	Intermediate	−	−	−	−
	Other			13.168	0.55
	+ (−, out)	−	−	−	−
O6_3	*trans* (+, in)	197.81	9.25	2355.482	98.14
	Intermediate	253.96	10.48	0.686	0.03
	Other			43.880	1.83
	+ (−, out)	45.13	7.50	2376.842	99.03
O6_4	*trans* (+, in)	165.02	7.33	2.312	0.10
	Intermediate	0.00	−	−	−
	Other			20.894	0.87
	+ (−, out)	55.79	7.73	2283.798	95.16
O6_5	*trans* (+, in)	−	−	−	−
	Intermediate	−	−	−	−
	Other			116.250	4.84
	+ (−, out)	49.68	6.87	2382.820	99.28
O6_6	*trans* (+, in)	−	−	−	−
	Intermediate	−	−	−	−
	Other			17.228	0.72

Simulation	HP setup at 0.65 GPa				
O6 group	**Conf.**	**Torsion angle (°)**		**Length**	**Occupancy**
		μ	**σ**	**(ns)**	**(%)**
	+ (−, out)	52.70	8.57	1921.918	48.05
O6_1	*trans* (+, in)	189.96	8.34	2050.088	51.25
	Intermediate	292.21	8.89	3.286	0.08
	Other			24.788	0.62
	+ (−, out)	43.47	7.86	3973.134	99.33
O6_2	*trans* (+, in)	183.69	9.22	3.092	0.08
	Intermediate	229.32	18.27	0.492	0.01
	Other			23.362	0.58
	+ (−, out)	60.81	11.61	74.548	1.86
O6_3	*trans* (+, in)	195.35	11.99	3854.434	96.36
	Intermediate	257.43	9.93	33.720	0.84
	Other			37.378	0.93
	+ (−, out)	45.99	7.50	3970.082	99.25
O6_4	*trans* (+, in)	−	−	−	−
	Intermediate	−	−	−	−
	Other			29.998	0.75
	+ (−, out)	55.58	8.73	3863.250	96.58
O6_5	*trans* (+, in)	167.71	14.57	5.392	0.13
	Intermediate	231.94	9.11	99.094	2.48
	Other			32.344	0.81
	+ (−, out)	49.48	7.05	3969.966	99.25
O6_6	*trans* (+, in)	186.63	14.74	3.290	0.08
	Intermediate	265.45	23.30	3.144	0.08
	Other			23.680	0.59

Simulation	AP setup at 0.65 GPa				
O6 group	**Conf.**	**Torsion angle (°)**		**Length**	**Occupancy**
		μ	σ	**(ns)**	**(%)**
	+ (−, out)	73.91	8.46	267.128	11.13
O6_1	*trans* (+, in)	188.77	8.04	2110.498	87.94
	Intermediate	287.23	5.86	0.166	0.01
	Other			22.256	0.93
	+ (−, out)	45.12	7.55	2276.858	94.87
O6_2	*trans* (+, in)	196.38	9.71	109.410	4.56
	Intermediate	−	−	−	−
	Other			13.780	0.57
	+ (−, out)	−	−	−	−
O6_3	*trans* (+, in)	196.85	10.35	2355.584	98.15
	Intermediate	227.27	4.03	4.650	0.19
	Other			39.814	1.66
	+ (−, out)	44.74	7.62	2381.860	99.24
O6_4	*trans* (+, in)	−	−	−	−
	Intermediate	−	−	−	−
	Other			18.188	0.76
	+ (−, out)	55.00	8.32	2381.016	99.21
O6_5	*trans* (+, in)	137.66	6.77	0.232	0.01
	Intermediate	229.80	7.77	2.270	0.09
	Other			16.530	0.69
	+ (−, out)	50.19	6.99	2339.602	97.48
O6_6	*trans* (+, in)	170.90	7.59	0.130	0.01
	Intermediate	229.73	45.76	44.020	1.83
	Other			16.296	0.68

Simulation	HP setup at 0.0001 GPa				
O6 group	**Conf.**	**Torsion angle (°)**		**Length**	**Occupancy**
		μ	**σ**	**(ns)**	**(%)**
	+ (−, out)	55.33	9.04	1796.190	44.90
O6_1	*trans* (+, in)	184.74	8.48	2158.454	53.96
	Intermediate	276.53	8.17	11.606	0.29
	Other			33.830	0.85
	+ (−, out)	47.79	8.20	3956.594	98.91
O6_2	*trans* (+, in)	191.60	11.35	18.864	0.47
	Intermediate	239.27	11.89	0.178	—
	Other			24.444	0.61
	+ (−, out)	60.33	9.76	175.530	4.39
O6_3	*trans* (+, in)	197.89	9.86	3721.044	93.02
	Intermediate	254.05	11.22	55.110	1.38
	Other			48.396	1.21
	+ (−, out)	48.15	7.83	3969.238	99.23
O6_4	*trans* (+, in)	0.00	—	—	—
	Intermediate	0.00	—	—	—
	Other			30.842	0.77
	+ (−, out)	58.92	8.13	3926.342	98.16
O6_5	*trans* (+, in)	195.01	12.96	31.468	0.79
	Intermediate	282.07	6.57	1.428	0.04
	Other			40.842	1.02
	+ (−, out)	49.61	7.10	3964.592	99.11
O6_6	*trans* (+, in)	212.28	33.25	12.674	0.32
	Intermediate	275.47	10.22	0.792	0.02
	Other			22.022	0.55

I Technical description of the diamond anvil cells

I.1 Square cell

Culet size (μm)	550
Window opening angle (°)	90
Diamond cut	Ahsbahs
Baking plate	Steel
Diamond fixation	Screws
Dimensions (mm)	60×60×21
Used for	Crystal **A** of β-CD form II (§3.2.1)
	Pure ^{t}BA form II (§6.2.4.1)

I.2   Triangular cell

Culet size (μm)	600 (Model 1) 800 (Model 2)
Window opening angle (°)	90
Diamond cut	Bhoeler-Almax
Baking plate	Tungsten carbide
Diamond fixation	Glued
Dimensions (mm)	43×37×20
Used for	(Model 1) Crystal **B** of β-CD form II (§3.2.1) (Model 1) ^{t}BA 5.80-hydrate (§6.2.3.1) (Model 2) α-CD compression study (§5.2.1) (Model 2) ^{t}BA 5.65-hydrate (§6.2.2.1)

I.3   90/100° cell

Culet size (μm)	700
Window opening angle (°)	90/100
Diamond cut	Ahsbahs (90° side)
	Flat (100° side)
Baking plate	Tungsten carbide
Diamond fixation	Glued
Dimensions (mm)	40($\varnothing$)×18
Used for	Crystal **C** of β-CD form III (§3.2.1)
	α-CD at 0.20 GPa (§5.2.1)

This DAC has been custom designed by Hans Ahsbahs (Margburg) in 2011. Differently from the other two models, is asymmetric. One side is similar to the previous ones, with a steel body, tungsten carbide backing plate with an opening angle of 90°, and a glued Bhoeler-Almax diamond. The other side is equipped with a flat diamond, which has a larger support area, hence it can be mounted on a backing plate with a wider opening angle of 100°. The main advantage of this cell is that it allows for better completeness during data collection.

Data quality can be further improved by removing the diffraction and reducing the absorption and shading from the gasket. In practice, this is done by drilling a large hole in the gasket ($\sim$600 μm for 700 μm culets), filling it with epoxy resin, and then drilling the resin to the desired size of the sample chamber (Fig. A.1). In this way, the metallic gasket provides structural support for the cell without interacting with the X-ray beam.
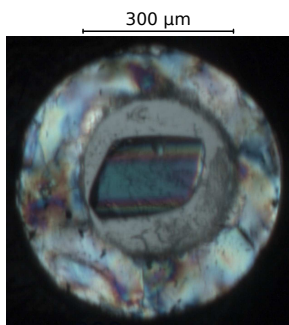

Figure A.1: Epoxy ring placed inside the hole drilled on a brass gasket (in black).

I.3.1   Optimising data collection with the 90/100° cell

The software available to correct for the shadowing effect observed when collecting diffraction data using a DAC considers that both sides of the

cell are symmetric. Under this assumption, it is not possible to take full advantage of the 90/100° DAC, because correcting for a window opening of 100° does not remove completely the shadows cast by the 90° side, while correcting for a window opening of 90° removes the extra data collected thanks to the wider opening of 100° on the other side. A Python program called Sombra[1] (from the Spanish word for shadow), was written to address this situation.

Sombra was designed under the same principles than Shade[2], and works on the output of the latter. Shade takes as input a raw file output by Saint[3] and calculates the angles of incidence and diffraction of each reflection from the corresponding direction cosines, the metric tensor of the unit cell, the orientation matrix of the crystal, and the orientation in which the DAC is mounted on the diffractometer. These angles are used to remove reflections which lye in the shadowed areas, but the complete list can also be directly output to apply the correction with Sombra. In addition, correlation coefficients, which can be used to asses the quality of reflections that are close to the edge of the shadowed area, are passed from the raw file onto Shade's output.

Sombra takes as input the output from Shade and the user is asked to input the opening angles of each side of the DAC (as half-opening angles), referred as the side which faces the incoming beam (B) and the side which faces the detector (D). Angles smaller than the opening of the windows can be introduced for a more conservative result, especially if problematic reflections are expected on the edges of the shadowed regions. The program then removes any reflection which fulfils any of the following criteria:

- The angle of incidence is larger than B.

- The angle of incidence is larger than $B - 2°$ and its correlation coefficient is larger than 0.3.

- The angle of diffraction is larger than D.

---

[1]Source code available on request.

[2]S. Parsons, Shade, The University of Edinburgh, Edinburgh, United Kingdom, 2005.

[3]Bruker-axs, Saint+ Integrating Engine v8.18c, Bruker-axs, Madison, Wisconsin, USA, 2011.

- The angle of diffraction is larger than $D - 2°$ and its correlation coefficient is larger than 0.3.

A new RAW file is finally produced, containing only the accepted reflections.

The performance of SOMBRA has been tested on data collected on a sample of α-CD form I at 0.20 GPa (Ch. 5). Data were collected in four independent runs. The first run (**R1**) was collected with the 100° window facing the incoming beam, while the other three (**R2**, **R3**, and **R4**) were collected with the 90° window facing the beam. All runs were first integrated in SAINT with a cutoff for the window opening of 100°, to remove only the smallest shadow possible. Three post-integration shadow corrections were applied:

- **a)** Using SHADE with a cutoff of 86°.

- **b)** Using SHADE with a cutoff of 96°.

- **c)** Using SOMBRA with a cutoff of 86 and 96° for the narrow and the wide window, respectively.

The number of reflections remaining after each shadow correction, with respect to the total number of integrated reflections, for each run, has been compared (Table A.1). Runs like **R1** allow to collect more frames, but the shadowed region in those frames is larger. On the other hand, runs like **R2-4** provide fewer frames, but with smaller shadows on them. In terms of integrated reflections, **R1** yields more reflections than **R2-4**. In terms of corrected reflections, **R1** yields more reflections when using SHADE, although the situation is inverted when using SOMBRA. Ultimately, the number of unique reflections is more important. It is assumed here that SOMBRA performs a complete removal of all shadows, and it is interesting to realise that the number of unique reflections after the corrections coincide with the one obtained by applying SHADE using as cutoff the opening of the window facing the detector. This means that, at least in this particular case, runs like **R2-4** yield more unique reflections than runs like **R1**, although the difference is small—3010 against 2861 reflections, respectively.

Run	Shadow corrections	Number of reflections		
		Integrated	Corrected	Unique
R1	Shade 86°		9163	2861
	Shade 96°	10698	10240	3042
	Sombra 86/96°		9515	2861
R2	Shade 86°		8858	
	Shade 96°	9934	9622	
	Sombra 86/96°		9621	
R3	Shade 86°		8850	2795
	Shade 96°	9914	9621	3010
	Sombra 86/96°		9621	3010
R4	Shade 86°		8791	
	Shade 96°	9845	9553	
	Sombra 86/96°		9552	

Table A.1: Number of reflections at different points of the data processing.

Based on these results, both Shade and Sombra perform equally well, although it is of paramount importance to properly choose the cutoffs for the correction. If using Shade, the cutoff should correspond to the opening of the window facing the detector, which implies that scans of the front and back side of the DAC must be integrated separately (§1.1.3). Sombra can correct for both sides at the same time; the appropriate cutoff for each side, along with the orientation of the windows, should be given. Although not implemented yet, it is possible to modify the program to apply the correction to several scans that have been integrated together. Integration of several scans together is desirable, although not always possible, to perform more accurate absorption corrections (§1.1.3).