BIOLOGICAL CRYSTALLOGRAPHY

ISSN 1399-0047

Novel crystalline phase and first-order phase transitions of human insulin complexed with two distinct phenol derivatives

A. Valmas,[a] K. Magiouf,[a] S. Fili,[a] M. Norrman,[b] G. Schluckebier,[b] D. Beckers,[c] T. Degen,[c] J. Wright,[d] A. Fitch,[d] F. Gozzo,[e] A. E. Giannopoulou,[a] F. Karavassili[a] and I. Margiolaki[a]*

[a]Section of Genetics, Cell Biology and Development, Department of Biology, University of Patras, GR-26500 Patras, Greece, [b]Diabetes Protein Engineering, Novo Nordisk A/S, Novo Nordisk Park, DK-2760 Malov, Denmark, [c]PANalytical B.V., Lelyweg 1, 7602 EA Almelo, The Netherlands, [d]European Synchrotron Radiation Facility, BP-220, F-38043 Grenoble CEDEX 9, France, and [e]Excelsus Structural Solutions SPRL, Belgium. *Correspondence e-mail: imargiola@upatras.gr

Received 9 December 2014
Accepted 21 January 2015

Edited by M. Schiltz, Fonds National de la Recherche, Luxembourg

Keywords: powder diffraction; human insulin; pH variation.

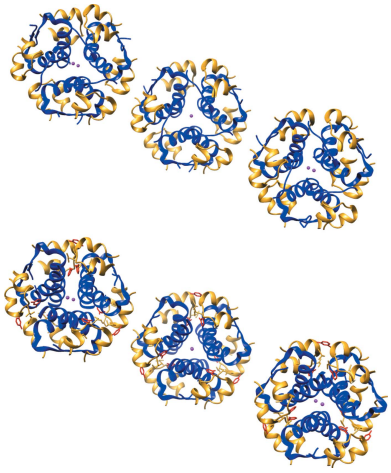
The primary focus of the present work is the study of the effects that two ligands and the crystallization pH have on the crystalline forms of human insulin. For this purpose, human insulin (HI) was co-crystallized with two distinct phenolic derivatives: the organic ligands *meta*-cresol (*m*-cresol) and 4-nitrophenol. The formation of polycrystalline precipitates was then followed by means of structural characterization of the individual specimens in terms of unit-cell symmetry and parameters. In both cases, two different polymorphs were identified *via* X-ray powder diffraction measurements, the first of hexagonal symmetry (*R*3 space group) at higher pH values and the second of monoclinic symmetry (space group $P2_1$) with unit-cell parameters a = 87.4282 (5), b = 70.5020 (3), c = 48.3180 (4) Å, β = 106.8958 (4)°, the latter of which to our knowledge has never been observed before.

1. Introduction

Human insulin (HI) is a peptide hormone of molecular weight 5808 Da. It consists of two chains: A (21 amino acids) and B (30 amino acids). The two disulfide bridges formed between cysteine residues in chains A and B stabilize the monomer, whereas a third disulfide bridge is formed between cysteines in chain A (Sanger & Tuppy, 1951). The hormone is produced by the β-pancreatic cells and is released in the presence of specific stimuli (Itoh & Okamoto, 1980).

Inadequate control of insulin levels is the main cause of diabetes mellitus. As a consequence, insulin is used medically to treat certain forms of this common disease (Banting & Best, 1922). Patients suffering from diabetes type 1 are not capable of producing insulin and thus entirely depend on external insulin provision (Alberti & Zimmet, 1998).

In contrast, those suffering from diabetes type 2 can produce a normal amount of insulin, but still need a daily supplement of external insulin in order to amplify its effect (Alberti & Zimmet, 1998). External insulin is provided *via* hypodermic injections that contain microcrystals or an intermixture of microcrystals and amorphous protein. The gradual dissolution of these crystals leads to a stable augmentation of insulin in the bloodstream (Norrman & Schluckebier, 2007).



Several different insulin polymorphs were originally distinguished depending on the number of zinc ions (two or four) involved in the formation of their hexameric structural motif (Schlichtkrull, 1958; Dunn, 2005). However, since 1989 this distinction has been relinquished. The terms 'T' and 'R' that denote the specific molecular conformation, which leads to the T_6, the $T_3R_3^f$ and the R_6 conformations (Kaarsholm *et al.*, 1989; Norrman, 2007), have been widely used for the characterization of polymorphs to date.

Proteins often form microcrystalline precipitates by batch precipitation (Von Dreele, 2003). The average size of individuals crystallites (∼0.1–1 µm) in such precipitates is usually far too small for single-crystal diffraction studies, but it is in many cases ideal for X-ray powder diffraction (XRPD). In the last ten years, the use of powder data with macromolecules has upgraded from being an ambitious suggestion to being a respectable method. A series of experiments and data analyses have been carried out which have established the validity of the method (Von Dreele, 2003; Margiolaki *et al.*, 2005, 2007; Margiolaki & Wright, 2008). Although there is clearly still much scope for improvement of the phasing and structure-solution methods, the molecular-replacement and structure-refinement methods can already be regarded as sufficiently mature for use by adventurous protein crystallographers. The application of powder diffraction to proteins has in fact been proved to be an effective method for high-throughput polymorph identification and crystal screening (Norrman *et al.*, 2006; Collings *et al.*, 2010; Karavassili *et al.*, 2012) as well as to be suitable for structure solution (Margiolaki *et al.*, 2007) and refinement (Margiolaki *et al.*, 2013).

In the present study, we investigated the effects of two phenol-based ligands, *meta*-cresol (*m*-cresol) and 4-nitrophenol, on the crystallization of HI as a function of pH. To date, XRPD methods have already been employed for the investigation of various interactions, such as the detection of protein–ligand complex formation with varying ligand and ion concentrations (Von Dreele, 2001, 2005). The efficiency of this approach has been confirmed in this study by monitoring the evolution of crystal symmetry and unit-cell parameters with two different types of ligands, *m*-cresol and 4-nitrophenol, and crystallization pH. These organic additives have previously been reported to be bactericidal agents and earlier structural results on HI complexed with phenol, resorcinol or *m*-cresol exist in the literature (Whittingham *et al.*, 1998; Smith *et al.*, 2000; Karavassili *et al.*, 2012). The use of *m*-cresol in particular has been suggested as a more effective germicide than phenol and it is widely used as an antimicrobial preservative in pharmaceutical formulations (Rowe *et al.*, 2005).

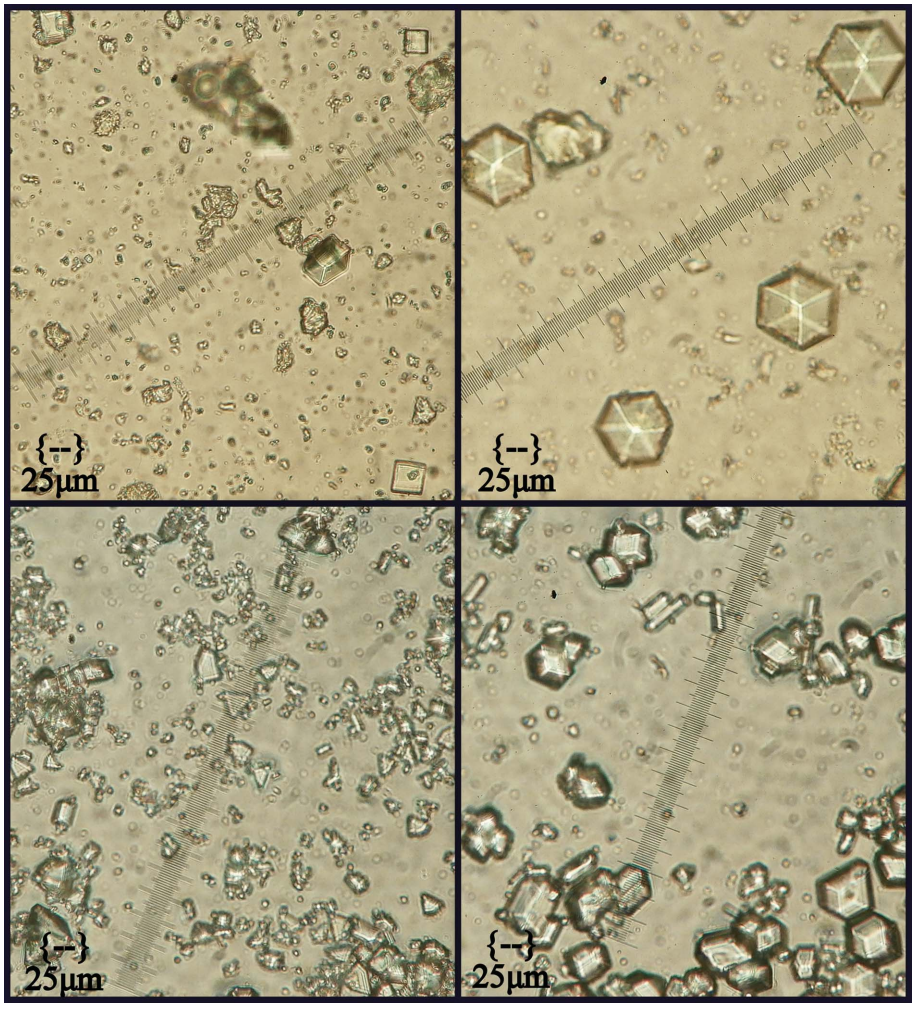


Figure 1
Polycrystalline samples of human insulin co-crystallized with two different ligands: 4-nitrophenol (upper images) corresponding to pH 5.71 (left) and pH 6.41 (right) and *m*-cresol (lower images) corresponding to pH 5.5 (left) and pH 6.5 (right). Two different space groups were observed: monoclinic $P2_{1(y)}$ (left images) and hexagonal $R3$ (right images).

2. Experimental

2.1. Crystallization

Purified HI was provided by Novo Nordisk A/S and crystallization was performed using the salting-out method (Hofmeister, 1888) in batch. In order to verify the reproducibility of our results, three series of crystallization experiments took place for each ligand following the same procedure in terms of protein, ligand and ion concentrations. The stock protein solution was prepared by adding 362.4 mg as-received freeze-dried insulin to 19 ml double-distilled H_2O, along with 2.2 ml 0.01 *M* zinc acetate solution, resulting in a protein concentration of 17.09 mg ml^{-1}. For preparation of the protein mixture, 15 ml protein solution

Table 1
Alterations of the normalized unit-cell parameters of the monoclinic polymorph $P2_{1(\gamma)}$ with increasing pH for each ligand.

Delta variations were calculated between the samples at the lowest and highest pH values observed for the polymorph.

	4-Nitrophenol [$P2_{1(\gamma)}$]					*m*-Cresol [$P2_{1(\gamma)}$]		
	2011, ESRF	2012, ESRF		2013, ESRF		2012, SLS	2012, ESRF	
	Series 1	Series 1	Series 2	Series 2	Series 3	Series 2	Series 1	Series 2
$\Delta a/a_i$ (%)	0.281	0.286	0.135	0.241	0.205	0.548	2.827	0.568
$\Delta b/b_i$ (%)	0.151	0.119	0.105	−0.095	0.248	0.855	0.770	0.638
$\Delta c/c_i$ (%)	−0.065	0.146	−0.063	−0.313	0.019	0.790	0.873	0.463
$\Delta\beta/\beta_i$ (%)	0.106	0.029	0.010	0.248	−0.044	0.470	1.807	0.339
ΔpH	0.420	1.195	0.630	0.565	0.835	1.810	1.720	1.765
$\Delta V/V_i$ (%)	0.308	0.536	0.172	−0.303	0.498	1.941	3.405	1.485
$(\Delta V/V_i)/\Delta$pH (%)	0.730	0.448	0.273	−0.536	0.596	1.072	1.979	0.841

was extracted and placed in a Falcon tube along with either 0.401 ml 2 *M* 4-nitrophenol diluted in DMSO or 0.402 ml 25%(*v*/*v*) ethanol-based *m*-cresol solution. Finally, after ∼5 min of incubation, 0.2 ml 1 *M* sodium thiocyanate was added to the protein mixture. Furthermore, 2 *M* stock buffers of Na_2HPO_4 and KH_2PO_4 were prepared. These solutions were mixed in order to produce a buffer pH range from 4.5 to 8.7 with a step of roughly 0.3 units.

Each sample was produced by placing 1 ml protein mixture and 250 µl pH-buffer mixture in an Eppendorf tube, resulting in a final protein concentration of 13.15 mg ml^{-1}. The concentration of 4-nitrophenol in each sample was 41 m*M*, while that of *m*-cresol was 0.51%(*v*/*v*).

The samples were crystallized at 298 K. After ∼48 h, the solution yielded a polycrystalline material (Fig. 1). The pH of the crystallization solutions was measured before crystallization, as well as after the diffraction experiments, and a small shift towards higher pH values was observed for the majority of the samples. The values for the pH levels quoted throughout the article always refer to the mean values of the above measurements.

2.2. X-ray data collection and processing

Data collection was performed using different instruments and sources to exploit their individual performances and to cross-check the results. We employed the ESRF synchrotron source in France and the SLS synchrotron source in Switzerland as well as laboratory instrumentation (PANalytical X'Pert PRO, suitable for advanced measurements in Debye–Scherrer mode), and the data quality is discussed in the following paragraphs. Powder X-ray diffraction data were collected at room temperature (RT) on the High Resolution Powder Diffraction Beamline ID31 at ESRF (Fitch, 2004) in three consecutive years [2011, λ = 1.29992 (3) Å; 2012, λ = 1.29994 (1) Å; 2013, λ = 1.29989 (3) Å]. Additional measurements were performed on the materials science beamline MS-X04SA at SLS (Willmott *et al.*, 2013) in 2012 [RT, λ = 1.3781 (2) Å] and in our laboratory using the PANalytical X'Pert PRO diffractometer (λ = 1.541874 Å) (Table 1).

All samples prepared for diffraction experiments were loaded into borosilicate glass capillaries of different diameters in order to minimize peak broadening when using position-sensitive detectors and to increase the counting statistics with larger powder volumes when using ultrahigh-resolution detectors. The sample-containing capillaries were centrifuged in order to enhance crystal packing. Excess mother liquor was removed and the samples were sealed to prevent dehydration. The samples were spun during data collection in order to ensure adequate crystallite orientation averaging. For the ultrahigh-resolution synchrotron XRPD data collection, radiation-damage effects owing to the intense synchrotron beam, which are commonly apparent in these kinds of materials as considerable changes in the unit-cell parameters along with gradual increases in peak broadening, were limited by means of capillary translation. Thus, every 4 min the samples were translated by 2 mm, exposing a fresh region of protein powder and thereby increasing the counting statistics without compromising the data quality for structure analysis. The first scans, for which the sample had just been translated to expose fresh material, were combined, and the second scans were subsequently merged into separated profiles. In the case of data collection using the laboratory diffractometer or efficient position-sensitive detectors, no radiation damage was observed even after 24 h or several minutes, respectively, of measurement.

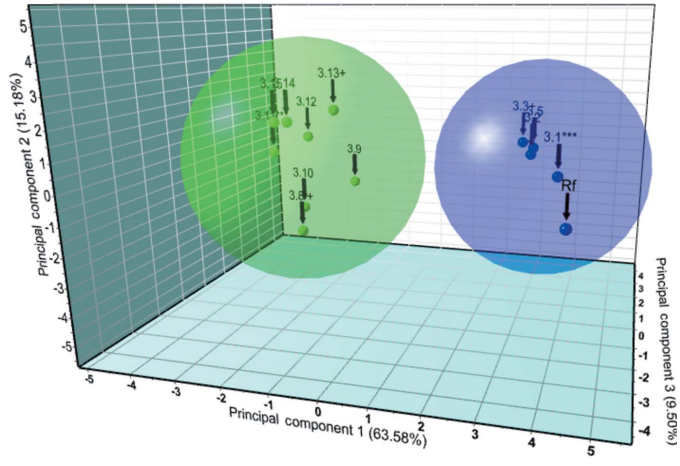


Figure 2
Cluster analysis using synchrotron profiles of human insulin co-crystallized with 4-nitrophenol. Two distinct clusters are observed: the blue cluster contains all data sets belonging to the new monoclinic phase [$P2_{1(\gamma)}$], while the green cluster contains all data sets with *R*3 symmetry ($T_3R_3^f$ conformation). Sphere labels from 3.1 to 3.15 correspond to sample codes N3_1 to N3_15, while 'Rf' indicates the monoclinic pattern which was used as a reference by the algorithm during the analysis. Data sets were collected on ID31 at ESRF [λ = 1.29989 (3) Å, RT].

Table 2
Alterations of normalized unit-cell parameters of both the $T_3R_3^f$ and R_6 polymorphs (*R*3 space group) with increasing pH for each ligand.

Delta variations were calculated between the samples at the lowest and highest pH values observed for the polymorph.

	4-Nitrophenol ($T_3R_3^f$)			*m*-Cresol (R_6)		
	2011, ESRF	2012, ESRF	2013, ESRF	2012, SLS	2012, ESRF	
$\Delta a/a_i$ (%)	−0.113	−0.070	−0.307	0.246	0.004	0.042
$\Delta c/c_i$ (%)	−0.382	−0.119	−0.587	0.446	1.258	0.868
ΔpH	1.235	1.490	1.680	0.860	1.670	0.905
$\Delta V/V_i$ (%)	−0.607	−0.258	−1.196	0.940	1.266	0.953
$(\Delta V/V_i)/\Delta$pH (%)	−0.491	−0.173	−0.712	1.093	0.758	1.053

The patterns were typically indexed using *DASH* (Boultif & Louër, 1991; David *et al.*, 1998) employing the fitted positions of at least the first 20 reflections of the high angular resolution powder diffraction profiles. From the extracted data, we were able to determine the symmetry and unit-cell parameters for all samples. In order to obtain accurate values of the unit-cell parameters and to characterize the peak shape and background coefficients without a structural model, Pawley fits (Pawley, 1981) were performed using *PRODD* (Wright, 2004).

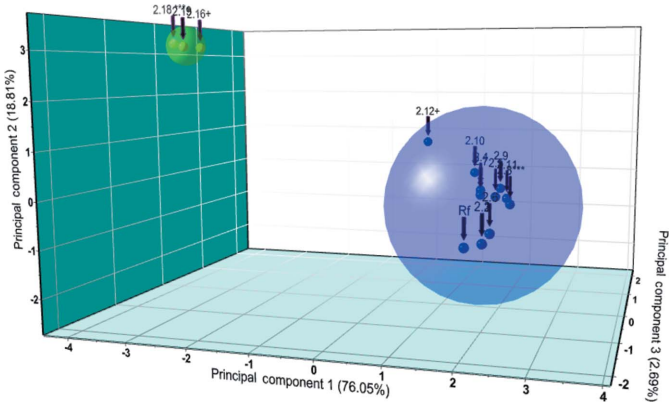


Figure 3
Cluster analysis using synchrotron profiles of human insulin co-crystallized with *m*-cresol. Two distinct clusters were observed: the blue cluster contains all data sets belonging to the new monoclinic phase [$P2_{1(\gamma)}$], while the green cluster contains all data sets with *R*3 symmetry (R_6 conformation). Sphere labels from 2.1 to 2.19 correspond to sample codes C2_1 to C2_19, while 'Rf' indicates the monoclinic pattern which was used as a reference by the algorithm during the analysis. Data sets were collected on ID31 at ESRF [λ = 1.29994 (1) Å, RT].

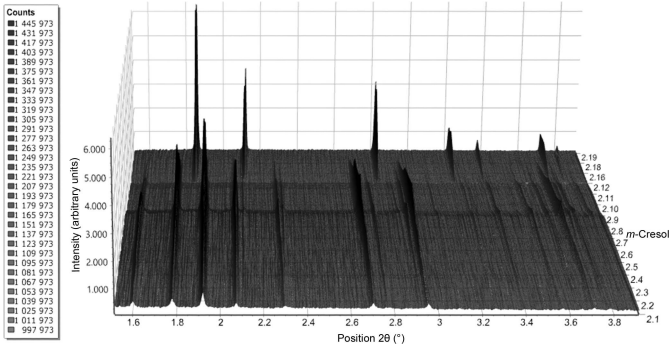


Figure 4
Data sets of human insulin co-crystallized with *m*-cresol collected on ID31 [λ = 1.29994 (1) Å, RT].

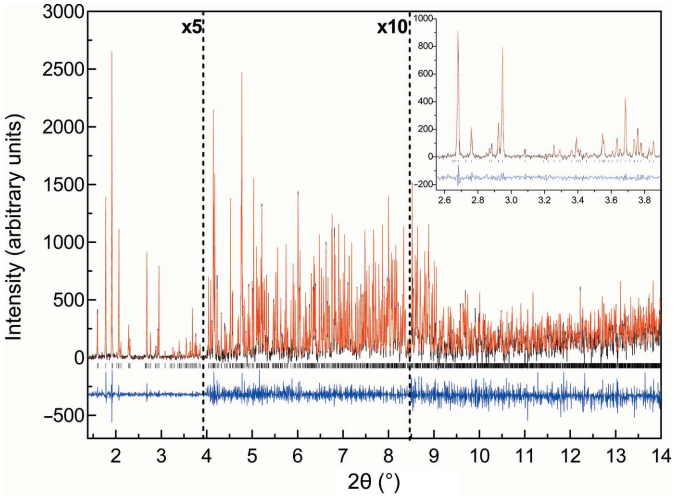


Figure 5
Pawley fit of human insulin co-crystallized with the ligand *m*-cresol (sample code C1_4, pH 6.30, monoclinic crystal system, space group $P2_1$). The data were collected on ID31 at ESRF [RT, λ = 1.299944 (22) Å]. The black, red and lower blue lines represent the experimental data, the calculated pattern and the difference between the experimental and calculated profiles, respectively. The vertical bars correspond to Bragg reflections compatible with the space group ($P2_1$). The profiles have been expanded by a factor of five at Bragg angles larger than 4° and by a factor of ten at Bragg angles larger than 8.5°. The inset corresponds to a magnification of the fit in the selected 2θ range.

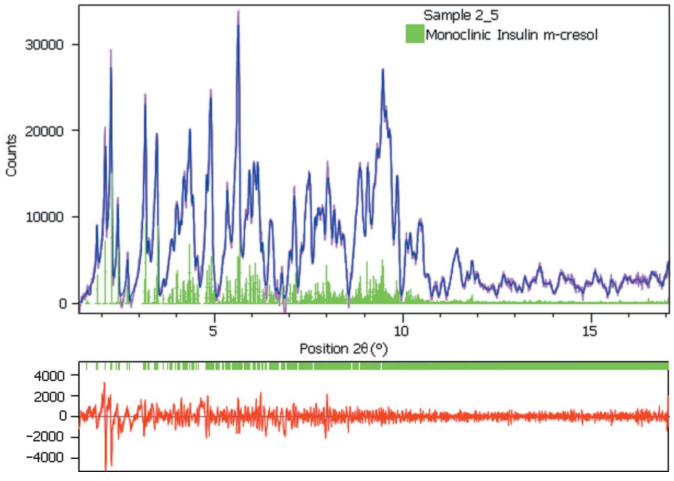


Figure 6
Pawley fit of human insulin co-crystallized with the ligand *m*-cresol (sample code C2_5, pH 5.21, monoclinic crystal system, space group $P2_1$). The data were collected at a wavelength of λ = 1.541874 Å using the X'Pert PRO laboratory diffractometer. The pink, blue and lower red lines represent the experimental data, the calculated pattern and the difference between the experimental and calculated profiles, respectively. The vertical green bars (lower panel) correspond to Bragg reflections compatible with this particular space group ($P2_1$), while the bars of the same colour in the upper panel correspond to the peak positions of our pattern.

2.3. Data clustering

In order to investigate the affinity of our protein samples throughout the pH range of interest, cluster analysis was initially employed for the high-resolution synchrotron data sets. Categorizing data sets is a common approach in several fields of bioscience employing high-throughput methods to screen multiple biochemical variations. Nonetheless, analysis of data derived from these experiments becomes exhausting without an adequate degree of automation (Bruno *et al.*, 2014). This industrial approach of analysis was performed using the multi-functional *HighScore Plus* software (Degen *et al.*, 2014). Measured profiles of HI samples co-crystallized with each of the two ligands were investigated separately. Analysis produced two clusters for both ligands, each containing one of the two different phases observed in our experiments. In the case of 4-nitrophenol, the two clusters correspond to the phases $P2_{1(\gamma)}$ and $R3$ ($T_3R_3^f$ conformation) (Fig. 2), and for *m*-cresol the two clusters correspond to the $P2_{1(\gamma)}$ and $R3$ (R_6 conformation) polymorphs as these were identified at a later stage of our analysis and are described in the following sections (Fig. 3). Clustering allowed us not only to detect the existence of three well separated crystalline phases in our data, even before their identification, but also enhanced the rapidity of the analysis by tagging the most representative data set of each cluster (marked with '***' in Figs. 2 and 3).

3. Results

3.1. Novel monoclinic ($P2_1$) phase γ

In the case where HI was crystallized with *m*-cresol in the pH range 4.50–6.70, a polymorph of monoclinic symmetry

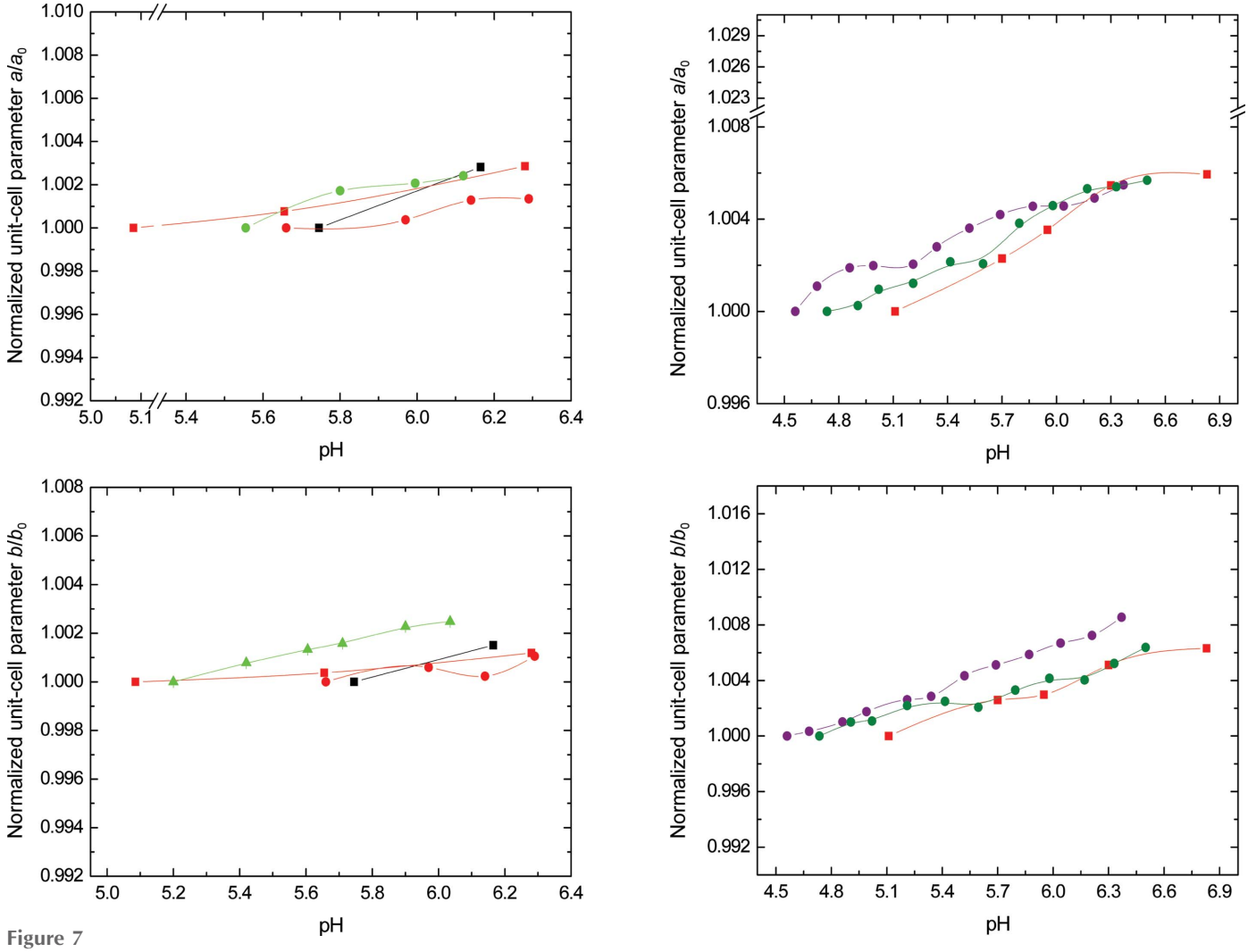


Figure 7
The evolution of the normalized unit-cell parameters *a* and *b* of human insulin co-crystallized with 4-nitrophenol (left) and *m*-cresol (right) with increasing pH for monoclinic space group $P2_1$. The black symbols for the 4-nitrophenol diagrams correspond to the data collected on ID31 at ESRF in November 2011, the red symbols to the data collected at the same facility one year later (ID31 at ESRF, November 2012) and the light green symbols to the data collected on ID31 in November 2013. Squares, crystallization series 1; circles, series 2; triangles, series 3. The purple symbols for the *m*-cresol diagrams correspond to the data collected on the MS-X04SA beamline at SLS (June 2012) and both red and dark green symbols to data collected on ID31 at ESRF (November 2012). Squares, crystallization series 1; circles, series 2. The values are listed in Supplementary Table S1. The lines are guides for the eye.

[space group $P2_1$, referred to in the following as $P2_{1(\gamma)}$] with unit-cell parameters a = 87.0749 (7), b = 70.1190 (5), c = 48.1679 (5) Å, β = 106.7442 (8)° was identified (Fig. 4), which to our knowledge has never been observed before. The precision of these values was verified by performing multiple measurements at different facilities (Tables 1 and 2).

Pawley analysis was satisfactory for all profiles (Figs. 5 and 6), with typical agreement factors of R_{wp} = 10.965% and χ^2 = 0.8036 for the sample crystallized at pH 5.11. The data acquired for the $P2_{1(\gamma)}$ polycrystalline samples extended to a d-spacing of approximately ∼6.8 Å.

Crystallization of HI with 4-nitrophenol in the pH range 5.1–6.3 had essentially the same effect as the crystallization experiments with m-cresol and yielded a phase with monoclinic symmetry [space group $P2_1$, unit-cell parameters a = 87.118 (1), b = 70.9493 (9), c = 48.4967 (9) Å, β = 106.653 (1)°]. Pawley analysis of all data sets was performed in order to extract accurate unit-cell parameters, resulting in noteworthy agreement factors (R_{wp} = 6.033% and χ^2 = 1.2201 for the sample crystallized at pH 5.97). The data obtained from these samples on ID31 extended to a d-spacing of ∼6.9 Å, comparable with the related profiles obtained for HI complexed with m-cresol.

The evolution of the normalized unit-cell parameters and volume for all samples, as extracted from the synchrotron data, are presented in Figs. 7, 8 and 9, respectively. This is the

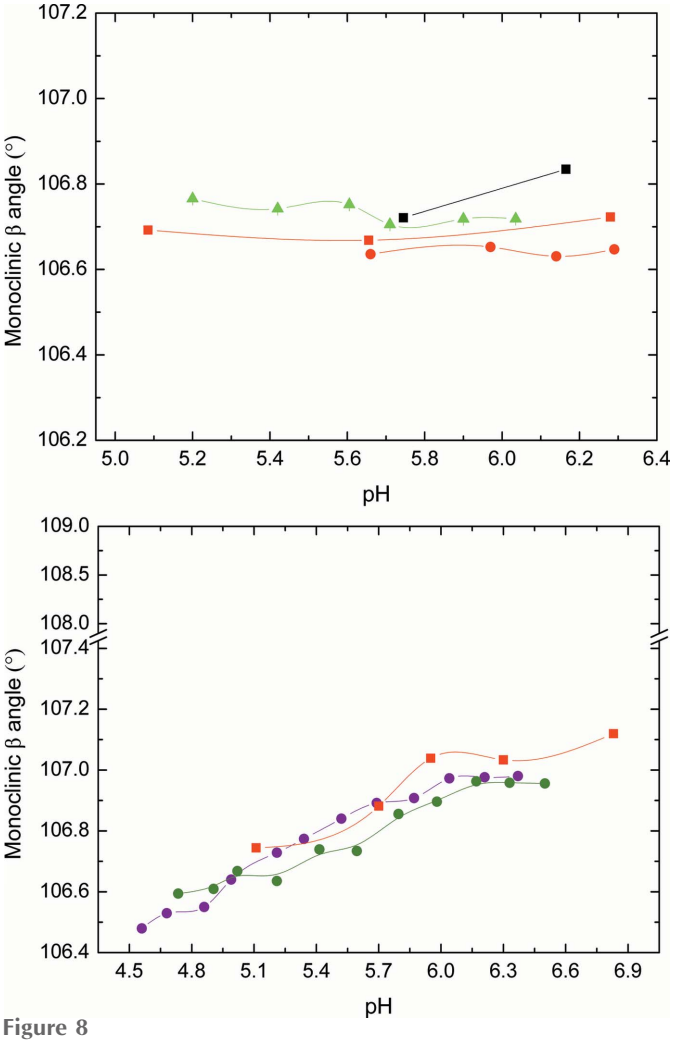


Figure 8
The evolution of the monoclinic β angle of human insulin co-crystallized with 4-nitrophenol (upper panel) and m-cresol (lower panel) with increasing pH for the monoclinic phase $P2_{1(\gamma)}$. Upper panel: the black symbols correspond to values extracted from the data collected on ID31 at ESRF in November 2011, the red symbols to the data collected at the same facility one year later (ID31 at ESRF, November 2012) and the light green symbols to the data collected on ID31 in November 2013. Squares, crystallization series 1; circles, series 2; triangles, series 3. Lower panel: the purple symbols correspond to the values extracted from the data collected on the MS-X04SA beamline at SLS (June 2012) and both red and dark green symbols to data collected on ID31 at ESRF (November 2012). Squares, crystallization series 1; circles, series 2. The values are listed in Supplementary Tables S1 and S2. The lines are guides for the eye.

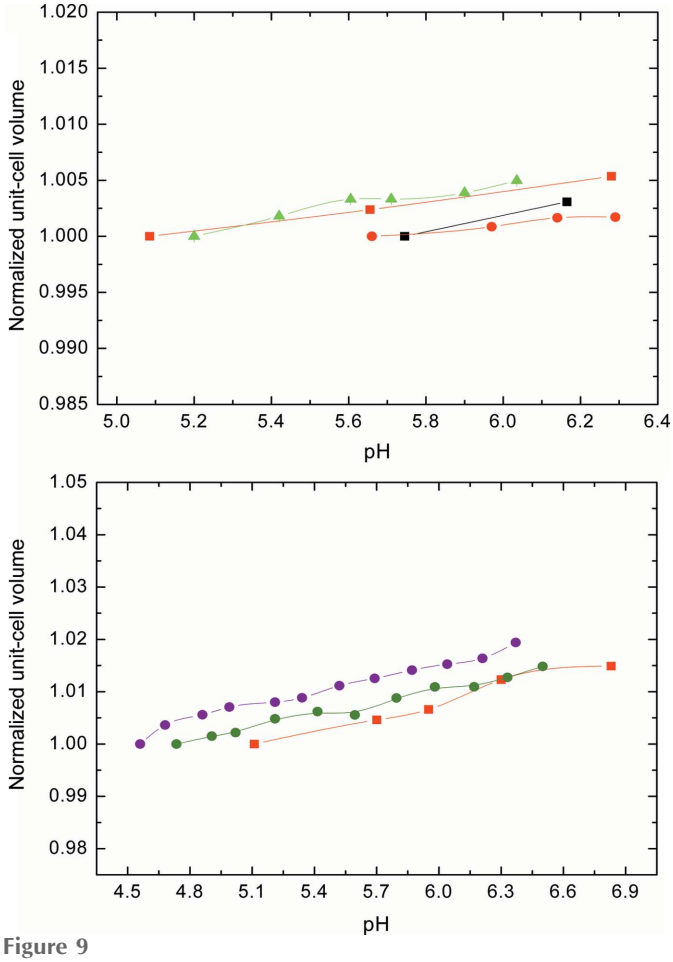


Figure 9
The evolution of the normalized unit-cell volume V of human insulin co-crystallized with 4-nitrophenol (upper panel) and m-cresol (lower panel) with increasing pH for the monoclinic polymorph $P2_{1(\gamma)}$. Upper panel: the black symbols correspond to the values extracted from the data collected on ID31 at ESRF in November 2011, the red symbols to the data collected at the same facility one year later (ID31 at ESRF, November 2012) and the light green symbols to the data collected on ID31 in November 2013. Squares, crystallization series 1; circles, series 2; triangles, series 3. Lower panel: the purple symbols correspond to the values extracted from the data collected on the MS-X04SA beamline at SLS (June 2012) and both red and dark green symbols to data collected on ID31 at ESRF (November 2012). Squares, crystallization series 1; circles, series 2. The values are listed in Supplementary Tables S1 and S2. The lines are guides for the eye.

first report of this specific crystallographic phase of human insulin.

3.2. Hexagonal (*R*3) phase

In the case of co-crystallization of HI with 4-nitrophenol in the pH range 6.2–8.1 or with *m*-cresol in the pH range 6.7–8.6, we observed a hexagonal symmetry (space group *R*3; see Supplementary Tables S3 and S4 for unit-cell parameters)

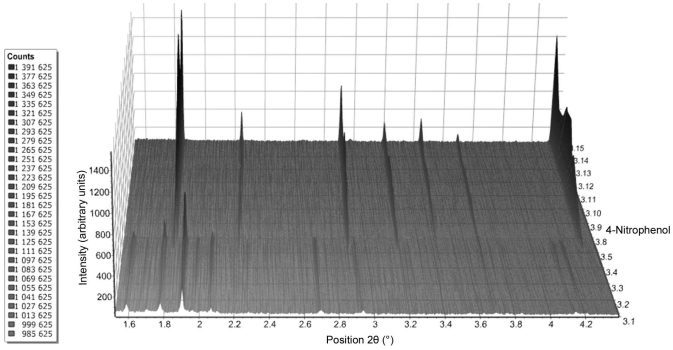


Figure 10
Data sets for human insulin co-crystallized with 4-nitrophenol collected on ID31 [λ = 1.29989 (3) Å, RT].

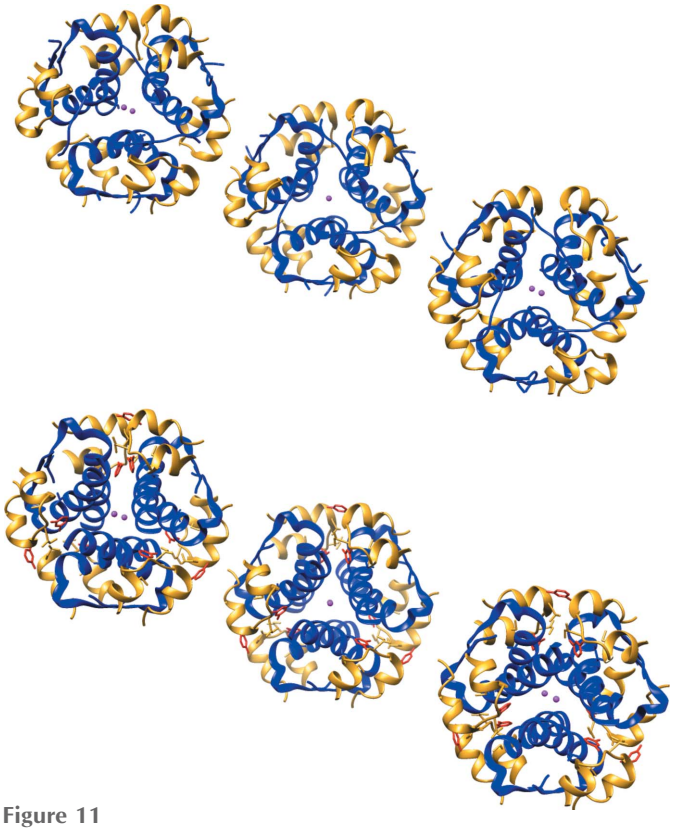


Figure 11
Packing of the insulin hexamers inside the hexagonal unit cell (view parallel to the *c* axis). Upper panel: hexamers with $T_3R_3^f$ conformation (PDB entry 1trz). Lower panel: hexamers with R_6 conformation (PDB entry 1ev3). In both cases there is a HI dimer in the asymmetric unit and three HI hexamers in the unit cell. Zinc ions are coloured purple and *m*-cresol molecules red. The graphical representation was prepared using *UCSF Chimera* (Pettersen *et al.*, 2004).

consisting of three protein hexamers per unit cell. Data analysis of profiles of HI complexed with 4-nitrophenol (Fig. 10) collected on ID31 at ESRF, combined with those collected at MS-X04SA at SLS and using the X'Pert PRO laboratory diffractometer, led to the identification of the hexagonal *R*3 symmetry with unit-cell parameters a = 80.721 (1), c = 37.8039 (5) Å, γ = 120.000°. These values indicate that HI co-crystallized with this ligand acquires the $T_3R_3^f$ (Ciszak & Smith, 1994) conformation (Fig. 11, upper panel). Pawley analysis of all data sets was performed in order to extract accurate unit-cell parameters, resulting in satisfactory fits with typical agreement factors of R_{wp} = 14.553% and

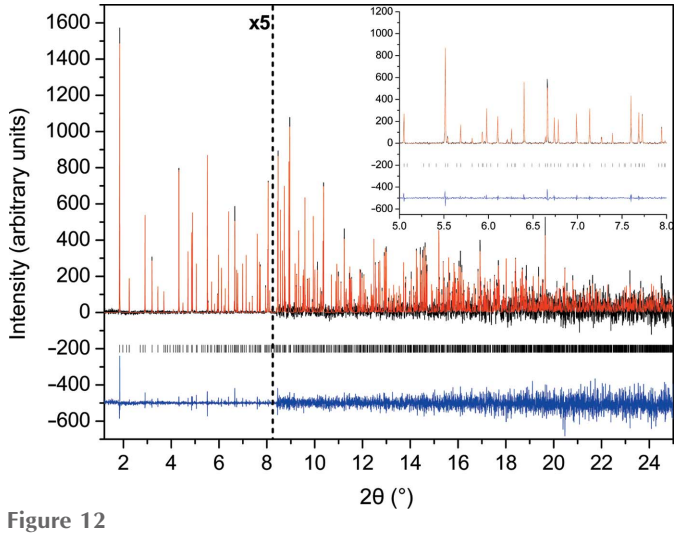


Figure 12
Pawley fit of human insulin co-crystallized with the ligand 4-nitrophenol (sample code N3_10, pH 6.41, hexagonal crystal system, space group *R*3). The data were collected on ID31 at ESRF [RT, λ = 1.29989 (3) Å]. The black, red and lower blue lines represent the experimental data, the calculated pattern and the difference between the experimental and calculated profiles, respectively. The vertical bars correspond to Bragg reflections compatible with this particular space group (*R*3). The profiles have been expanded by a factor of five at Bragg angles larger than 8°. The inset corresponds to a magnification of fit in the selected 2θ range.

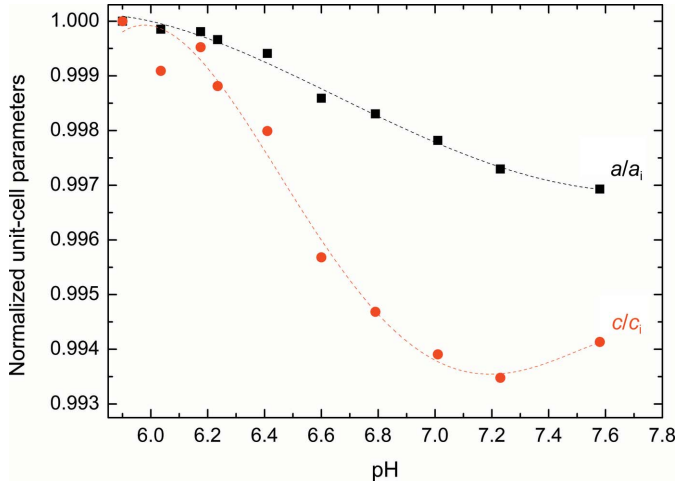


Figure 13
Evolution of the normalized *a* and *c* unit-cell parameters of human insulin co-crystallized with 4-nitrophenol [crystallization series 3, data collected on ID31, λ = 1.29989 (5) Å] with increasing pH for the *R*3 polymorph.

χ^2 = 1.0147 for the sample crystallized at pH 6.41 (Fig. 12). The data collected on ID31 for these samples extended to a resolution of 3.6 Å. The evolution of the normalized unit-cell parameters *a* and *c* with pH is illustrated in Fig. 13.

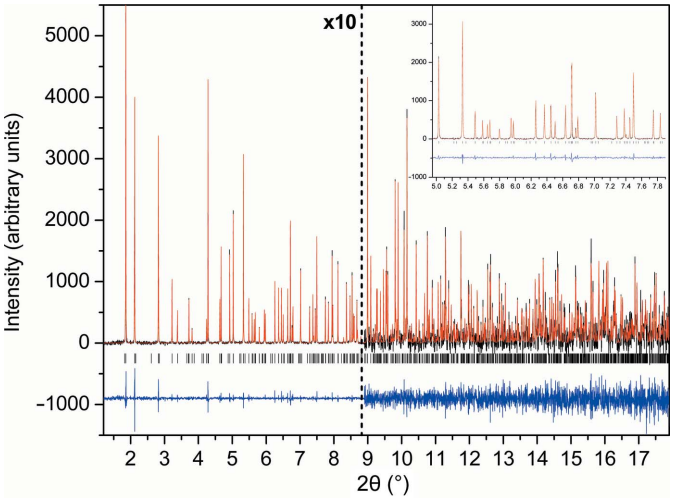


Figure 14
Pawley fit of human insulin co-crystallized with the ligand *m*-cresol [sample code C2_19, pH 8.15, hexagonal crystal system, space group *R*3 (R_6 conformation)]. The data were collected on ID31 at ESRF [RT, λ = 1.299944 (22) Å]. The black, red and lower blue lines represent the experimental data, the calculated pattern and the difference between the experimental and calculated profiles, respectively. The vertical bars correspond to Bragg reflections compatible with this particular space group (*R*3). The profiles have been expanded by a factor of five at Bragg angles larger than 9°. The inset corresponds to a magnification of the fit in the selected 2*θ* range.

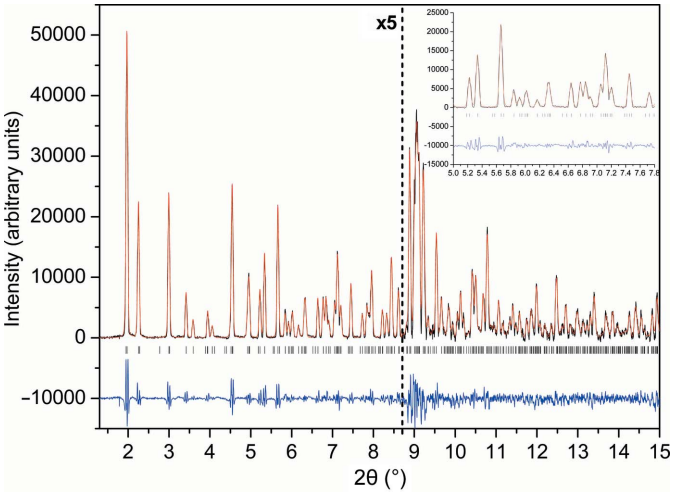


Figure 15
Pawley fit of human insulin co-crystallized with the ligand *m*-cresol [sample code C2_16, pH 7.09, hexagonal crystal system, space group *R*3 (R_6 conformation)]. The data were collected on MS-X04SA at SLS [RT, λ = 1.37807 (20) Å]. The black, red and lower blue lines represent the experimental data, the calculated pattern and the difference between the experimental and calculated profiles, respectively. The vertical bars correspond to Bragg reflections compatible with this particular space group (*R*3). The profiles have been expanded by a factor of five at Bragg angles larger than 8.7°. The inset corresponds to a magnification of the fit in the selected 2*θ* range.

When *m*-cresol is used as a ligand at pH 6.7–8.6, HI acquires a hexagonal *R*3 symmetry. Indexing was performed employing high-resolution profiles collected on ID31 (Fig. 14), combined with data collected at SLS (Fig. 15) and the laboratory diffractometer, resulting in the following unit-cell parameters: *a* = 80.0644 (6), *c* = 40.8396 (3) Å, *γ* = 120.000°. These values indicate that HI co-crystallized with this ligand acquires the R_6 (Smith *et al.*, 2000) conformation (Fig. 11, lower panel). Pawley analysis of all data sets resulted in satisfactory agreement factors (R_{wp} = 6.589% and χ^2 = 1.3147 for the sample crystallized at pH 8.15). The data collected on ID31 for these samples extended to a resolution of 3.7 Å (Fig. 14).

In both cases, increasing the pH in the ranges mentioned above resulted in minor alterations of the unit-cell parameters (Figs. 7, 8 and 9 and Supplementary Tables S1 and S2) and no indication of a first-order phase transition. In the case of 4-nitrophenol, this specific variation in pH resulted in a slight decrease in the unit-cell volume of $\Delta V/V_i$ = −0.71% (2013 measurements at ESRF) and in the case of *m*-cresol there was an increase of $\Delta V/V_i$ = 0.75% (2012 measurements at ESRF). At intermediate pH values of 5.9 and 6.04 (samples 3.6 and 3.7; Supplementary Tables S1 and S3) of 4-nitrophenol, the $P2_{1(\gamma)}$ and *R*3 phases co-exist (Fig. 16). Consequently, the transition from the first crystal type to the second causes a decrease in the unit-cell volume of about $\Delta V[P2_{1(\gamma)} \rightarrow R3]/V[P2_{1(\gamma)}]$ = −25.53%, while for HI complexed with *m*-cresol the decrease is almost 4% smaller $\{\Delta V[P2_{1(\gamma)} \rightarrow R3]/V[P2_{1(\gamma)}]$ = −21.89%}.

4. Conclusions

The main goal of this project was the characterization *via* synchrotron and laboratory XRPD of the polymorphism of HI co-crystallized with phenol-based ligands and the determination of the unit-cell parameters for each crystalline phase, a target that was thoroughly achieved, as well as the structural determination of the adoption of different crystal symmetries by the complex. A series of phase transitions were observed with increasing pH variation using the phenolic derivatives *m*-cresol and 4-nitrophenol, enriching the already detailed phase diagram of HI (Norrman, 2007).

Our results indicate several phase modifications regarding unit-cell parameters and/or crystal symmetry. Besides the previously known conformations deposited in the Protein

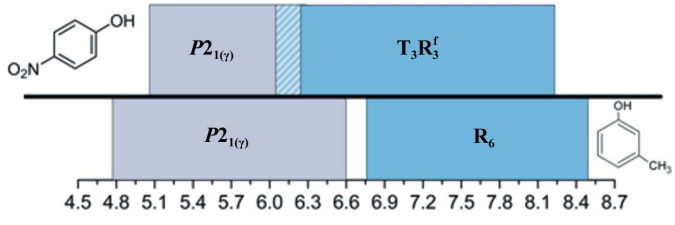


Figure 16
Phase diagram of human insulin polymorphs with the pH variation as indicated by the numeric range. The upper panel represents the phases of insulin co-crystallized with 4-nitrophenol, whereas the lower panel represents those with *m*-cresol. The dashed box corresponds to the intermediate region where crystals co-exist in both phases.

Data Bank (PDB entry 1ev6; Smith *et al.*, 2000) and the monoclinic phase $P2_{1(\alpha)}$ reported previously without any structural characterization (Karavassili *et al.*, 2012), a new polymorph of potential physicohemical importance has been identified (space group $P2_1$) as well as a first-order transition. In the present study, we report the unit-cell parameters of all identified forms. The complete structural model of the novel $P2_{1(\gamma)}$ polymorph has been derived by combining traditional single-crystal and emerging XRPD methods and will be presented in a forthcoming publication (Karavassili *et al.*, in preparation).

As opposed to the isosymmetrical polymorph $P2_{1(\alpha)}$, with surprisingly large *a* and *b* axes [a = 114.0228 (8), b = 335.43 (3), c = 49.211 (6) Å, β = 101.531 (8)°], the unit-cell parameters of the new polymorph $P2_{1(\gamma)}$ are remarkably smaller, approaching the already known range of dimensions adopted by other known monoclinic cells (Derewenda *et al.*, 1989; Smith *et al.*, 2000). $P2_{1(\gamma)}$ is observed in a wider pH range (from 5 to ∼6.3) in comparison to $P2_{1(\alpha)}$, which appears within a pH range with slightly tighter boundaries (from 5 to 5.7). The significantly smaller unit-cell parameters of the $P2_{1(\gamma)}$ polymorph could be explained by tighter crystal packing with associated enhanced stability.

We note that these new monoclinic polymorphs (α and γ) appear only in acidic environments (pH range 5–6.3) using ligands such as phenol and specific derivatives (in this study *m*-cresol and 4-nitrophenol). On the other hand, the already known phase $P2_{1(\beta)}$, with unit-cell parameters a = 61.247, b = 61.739, c = 47.467 Å, β = 111.32° (PDB entry 1ev6), was previously detected at a pH of ∼7 when insulin was co-crystallized with *m*-cresol, resorcinol (Smith *et al.*, 2000) or phenol (Derewenda *et al.*, 1989).

Comparison of the two monoclinic forms $P2_{1(\beta)}$ and $P2_{1(\gamma)}$ indicates, besides the larger unit-cell parameters observed for the second form, an unusual crystal packing for the new polymorph. The known $P2_{1(\beta)}$ form (PDB entry 1ev6) consists of six molecules per asymmetric unit and 48% solvent content. According to a Matthews coefficient calculation (Matthews, 1968; Kantardjieff & Rupp, 2003), the $P2_{1(\gamma)}$ polymorph contains 12 molecules per asymmetric unit and 39% solvent content (Matthews coefficient = 2.03 Å^3 Da^{-1}). This large difference reveals a much denser crystal packing in the case of the $P2_{1(\gamma)}$ form, which could be of great pharmacological importance.

On moving through the neutral pH region and reaching the weakly basic region, a first-order transition occurs. The monoclinic symmetry transforms quickly into a hexagonal symmetry that is stable over a wide pH range (approximately 6.5–8.5). In this case, the packing of hexamers inside the crystal is less dense: the *R*3 polymorphs (R_6 or $T_3R_3^f$ conformation) of both ligands consist of three hexamers per unit cell and approximately 40% solvent content (PDB entries 1ev3 and 1trz; Smith *et al.*, 2000; Ciszak & Smith, 1994). According to the literature, the R_6 and $T_3R_3^f$ insulin conformations correspond to a slight modification of the *c* crystallographic axis, leading to a more dense packing for the $T_3R_3^f$ crystals (Norrman, 2007).

Distinct crystalline polymorphs of biological macromolecules are often associated with different physicochemical properties and/or biological activity. The additional interhexamer interactions that may occur owing to the very dense packing of several polymorphs can increase the stability and thus extend the life of crystalline insulin formulations. Such an interpretation could possibly be a key point in the development of a new generation of insulin-based microcrystalline compounds for the treatment of diabetes. Pharmaceutical products containing crystals with high protein concentration could lead to a minimization of injection times. If one injection per day or two could be sufficient to provide the necessary amount of insulin, stored in crystals and thus slowly released into the bloodstream, it will be a life-quality improvement of great importance for millions of patients.

Our results demonstrate that systematic screening of crystallization conditions in combination with synchrotron and laboratory XRPD yields an exact and unambiguous picture of the crystallization behaviour of insulin even around its pI (∼5.9), where its solubility is lowest and the growth of macroscopic crystals suitable for single-crystal X-ray structure determination is less likely to succeed. We believe that this kind of systematic approach further extends the applicability of powder diffraction methods for systematic macromolecular crystal screening in a wide range of crystallization conditions.

Acknowledgements

We would like to thank the ESRF for provision of beamtime at the ID31 beamline. We would like also to thank the SLS for provision of beamtime at the MS-X04SA beamline, and Nicola Casati and Antonio Cervellino for assistance with data collection and useful advice, as well as PANalytical for instrumentation and software support and Novo Nordisk A/S for the provision of human insulin and their valuable advice on crystallization procedures. IM is grateful to the UNESCO L'Oreal foundations for the award of the International Fellowship for Women in Life Sciences (2010). This research has been co-financed by the following grants: the European Union (European Social Fund) in collaboration with Greek State, under the 'ARISTEIA II' Action (MIS Code 4659) of the 'Operational Program Education And Lifelong Learning', the European Union (European Regional Development Fund – ERDF) and Greek national funds through the Operational Program 'Regional Operational Programme' of the National Strategic Reference Framework (NSRF), Research Funding Program: Support for Research, Technology and Innovation Actions in Region of Western Greece (Karatheodoris Foundation), the EU FP7 REGPOT CT-2011-285950 'SEEDRUG' project, the International Atomic Energy Agency (CRP code F12024) and the COST Action (CM1306). Finally, the company NanoMEGAS supported this work.

References

Alberti, K. G. & Zimmet, P. Z. (1998). *Diabet. Med.* **15**, 539–553.
Banting, F. G. & Best, C. H. (1922). *J. Lab. Clin. Med.* **7**, 251–266.
Boultif, A. & Louër, D. (1991). *J. Appl. Cryst.* **24**, 987–993.

Bruno, A. E., Ruby, A. M., Luft, J. R., Thomas, D., Grant, T. D., Seetharaman, J., Montelione, G. T., Hunt, J. F., Edward, H. & Snell, E. H. (2014). *PLoS One*, **9**, e100782.
Ciszak, E. & Smith, G. D. (1994). *Biochemistry*, **33**, 1512–1517.
Collings, I., Watier, Y., Giffard, M., Dagogo, S., Kahn, R., Bonneté, F., Wright, J. P., Fitch, A. N. & Margiolaki, I. (2010). *Acta Cryst.* D**66**, 539–548.
David, W. I. F., Shankland, K., Shankland, K. & Shankland, N. (1998). *Chem. Commun.*, pp. 931–932.
Degen, T., Sadki, M., Bron, E., König, U. & Nénert, G. (2014). *Powder Diffr.* **29**, S13–S18.
Derewenda, U., Derewenda, Z., Dodson, E. J., Dodson, G. G., Reynolds, C. D., Smith, G. D., Sparks, C. & Swenson, D. (1989). *Nature (London)*, **338**, 594–596.
Dunn, M. F. (2005). *Biometals*, **18**, 295–303.
Fitch, A. N. (2004). *J. Res. Natl Inst. Stand. Technol.* **109**, 133–142.
Hofmeister, F. (1888). *Arch. Exp. Pathol. Pharmakol.* **24**, 247–260.
Itoh, N. & Okamoto, H. (1980). *Nature (London)*, **283**, 100–102.
Kaarsholm, N. C., Ko, H.-C. & Dunn, M. F. (1989). *Biochemistry*, **28**, 4427–4435.
Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1805–1871.
Karavassili, F., Giannopoulou, A. E., Kotsiliti, E., Knight, L., Norrman, M., Schluckebier, G., Drube, L., Fitch, A. N., Wright, J. P. & Margiolaki, I. (2012). *Acta Cryst.* D**68**, 1632–1641.
Margiolaki, I., Giannopoulou, A. E., Wright, J. P., Knight, L., Norrman, M., Schluckebier, G., Fitch, A. N. & Von Dreele, R. B. (2013). *Acta Cryst.* D**69**, 978–990.
Margiolaki, I. & Wright, J. P. (2008). *Acta Cryst.* A**64**, 169–180.
Margiolaki, I., Wright, J. P., Fitch, A. N., Fox, G. C. & Von Dreele, R. B. (2005). *Acta Cryst.* D**61**, 423–432.
Margiolaki, I., Wright, J. P., Wilmanns, M., Fitch, A. N. & Pinotsis, N. (2007). *J. Am. Chem. Soc.* **129**, 11865–11871.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Norrman, M. (2007). PhD thesis. Lund University, Sweden.
Norrman, M. & Schluckebier, G. (2007). *BMC Struct. Biol.* **7**, 83.
Norrman, M., Ståhl, K., Schluckebier, G. & Al-Karadaghi, S. (2006). *J. Appl. Cryst.* **39**, 391–400.
Pawley, G. S. (1981). *J. Appl. Cryst.* **14**, 357–361.
Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
Rowe, R. C., Sheskey, P. J. & Owen, S. C. (2005). Editors. *Handbook of Pharmaceuticals Excipients*, 5th ed. London: Pharmaceutical Press.
Sanger, F. & Tuppy, H. (1951). *Biochem. J.* **49**, 481–490.
Schlichtkrull, J. (1958). *Insulin Crystals.* Copenhagen: Munksgaard.
Smith, G. D., Ciszak, E., Magrum, L. A., Pangborn, W. A. & Blessing, R. H. (2000). *Acta Cryst.* D**56**, 1541–1548.
Von Dreele, R. B. (2001). *Acta Cryst.* D**57**, 1836–1842.
Von Dreele, R. B. (2003). *Methods Enzymol.* **368**, 254–267.
Von Dreele, R. B. (2005). *Acta Cryst.* D**61**, 22–32.
Whittingham, J. L., Edwards, D. J., Antson, A. A., Clarkson, J. M. & Dodson, G. G. (1998). *Biochemistry*, **37**, 11516–11523.
Willmott, P. R. *et al.* (2013). *J. Synchrotron Rad.*. **20**, 667–682.
Wright, J. P. (2004). *Z. Kristallogr.* **219**, 791–802.