Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Pros and cons of cryocrystallography: should we also collect a room-temperature data set?

Kirsty V. Dunlop, Randall T. Irvin and Bart Hazes*

Department of Medical Microbiology and Immunology, University of Alberta, Canada

Correspondence e-mail: bhazes@ualberta.ca

Received 23 September 2004
Accepted 26 October 2004

PDB References: PAK pilin, 1x6z, r1x6zsf; 1x6x, r1x6xsf; 1x6q, r1x6qsf; 1x6p, r1x6psf; 1x6r, r1x6rsf; 1x6y, r1x6ysf.

High-resolution protein structures are becoming more common owing to the availability of increasingly brilliant synchrotron X-ray sources. However, to withstand the increased X-ray dose the crystals must be held at cryogenic temperatures. To compare the benefit of increased resolution with the drawback of potential temperature-induced changes, three room-temperature and three cryogenic data sets for PAK pilin have been collected at resolutions between 1.8 and 0.78 Å. The results show that although the high-resolution cryogenic structures are more precise and more detailed, they also show systematic deviations from the room-temperature structures. Small but significant differences are even observed in the structural core, whilst more extensive changes occur at the protein surface. These differences can affect biological interpretations, especially because many important biological processes take place at the protein surface. Accordingly, although high-quality cryogenic synchrotron data is extremely valuable to protein crystallography, room-temperature structures are still desirable, especially if the research question involves protein features that are sensitive to temperature-induced changes.

1. Introduction

The goal of macromolecular crystallography is to determine the physiological three-dimensional structure of the macromolecule of interest as accurately as possible. This explains the importance of the resolution of the experimental diffraction data, since it determines to a large extent the accuracy and the level of detail that can be obtained for a given structure. Several technical developments in the past decade have led to improvements in resolution, in particular the construction of third-generation synchrotrons and the adoption of cryocooling of crystals (Teng, 1990). As a result, there are now about 150 structures with a resolution of 1.0 Å or better in the Protein Data Bank (Berman *et al.*, 2000) and six of these have a resolution better than 0.8 Å. In these structures, H atoms can be visualized and N, O and C atoms can be differentiated on the basis of electron density (Betzel *et al.*, 2001; Kuhn *et al.*, 1998). In addition, atoms with low occupancy, such as weakly bound water molecules and sparsely populated alternate side-chain conformations, can be observed (Betzel *et al.*, 2001). If the resolution of the data is exceptionally high, for example crambin at 0.54 Å, electron-density maps start to show the features of non-spherical electron orbitals (Jelsch *et al.*, 2000).

Acquisition of the highest possible resolution data requires a high X-ray dose, which in turn necessitates cryocooling of the crystal. However, despite cryogenic cooling, the very

bright synchrotron beam will still cause radiation damage, leading to increased sample mosaicity, increased temperature factors and decreased diffraction limits (Burmeister, 2000). More importantly, the radiation damage will induce both global and local changes in the protein structure. Global changes include an increase in unit-cell volume (Burmeister, 2000; Leiros *et al.*, 2001; Ravelli & McSweeney, 2000) and rotation and translation of the protein molecules (Leiros *et al.*, 2001; Ravelli & McSweeney, 2000). Local changes include disulfide-bond breakage and loss of carboxyl, hydroxyl and methylthiol groups of acidic, tyrosine and methionine residues, respectively (Burmeister, 2000; Leiros *et al.*, 2001; Ravelli & McSweeney, 2000; Weik *et al.*, 2000). Although this does not normally prevent determination of a high-resolution protein structure, it must be kept in mind that radiation damage may have caused structural artifacts.

Cryogenic techniques greatly reduce radiation damage but unfortunately can also introduce new artifacts resulting directly from temperature effects or indirectly from the addition of cryoprotectants and temperature-induced pH changes. Indeed, decreases in unit-cell volume, changes in molecular packing and perturbation of local structure have been reported (Juers & Matthews, 2001). In a recent theoretical study, Halle (2004) proposed that cryogenic crystal structures are not kinetically trapped in a room-temperature equilibrium state. Because crystal cooling is slow, $\geq$0.1 s (Kriminski *et al.*, 2003), proteins adapt to the changing free-energy landscape during cooling until their conformation is immobilized when the glass transition is reached at around 200 K (Miyazaki *et al.*, 2000). Halle therefore predicts 'substantial cryoinduced shifts of substate populations' where the structure largely reflects the free-energy state at the glass-transition temperature of the solvent. The effect is expected to be most pronounced at the protein surface, a region of critical functional importance. It is unlikely that cryoinduced artifacts can be prevented, as this would require crystal-cooling rates near the picosecond timescale of molecular motions. It is therefore important to determine the nature and extent of temperature-induced changes.

We have previously solved a room-temperature crystal structure of truncated PAK pilin at 1.63 Å resolution (Hazes *et al.*, 2000). A 0.78 Å data set was subsequently collected at 100 K on beamline SBC-19ID at the Advanced Photon Source (Chicago, USA). Distinct differences in protein and hydration structure were noted and only part of these could be ascribed to the increased resolution. We therefore collected additional data sets to determine the roles of resolution, radiation damage and crystal temperature in these differences. In total, three cryogenic and three room-temperature PAK pilin data sets were collected using third- and second-generation synchrotrons and rotating-anode X-ray sources. Our results indicate that cryogenic cooling of crystals introduces systematic changes that are large relative to the expected coordinate error. Moreover, although the changes are most pronounced at the protein surface, even the main-chain atoms of residues forming the main secondary-structural elements are affected.

Table 1
Batch statistics for the ultrahigh-resolution 0.78 Å data set (data set 1).

Batch	Max. resolution (Å)	No. images	Exposure (s)	Attenuator
1	0.76	500	5	None
2	0.81	530	10	None
3	1.12	240	5	0.75 mm Al
4	1.64	120	5	0.13 mm Ag

2. Materials and methods

2.1. Crystallization, data collection and data processing

Expression, purification and crystallization of truncated PAK pilin (residues 29–144) has been described previously (Hazes *et al.*, 2000). Briefly, crystals were grown using the hanging-drop vapour-diffusion method by equilibrating a 6 µl drop consisting of equal volumes of protein solution (10 mg ml^{-1} in doubly distilled water) and precipitant solution [60% $(NH_4)_2SO_4$, 0.1 *M* HEPES pH 8.2] against 1 ml precipitant solution. Seeding was used to obtain crystals, which were typically 0.3 × 0.3 × 0.2 mm in size. All crystals belong to space group $P4_12_12$.

For the cryogenic data sets, fresh crystals were transferred to a solution consisting of 20% glycerol and 80% precipitant solution and flash-cooled in a nitrogen-gas stream at 100 K. Room-temperature crystals were mounted in quartz capillaries. For data set 5, the crystal used previously to collect cryogenic data set 3 was returned to standard mother liquor and remounted in a quartz capillary. Data set 1 was processed with *d*TREK* (Pflugrath, 1999), whereas all others were integrated with *MOSFLM* (Leslie, 1992) and scaled with *SCALA* (Collaborative Computational Project, Number 4, 1994).

Data set 1 was collected as four consecutive batches (Table 1). During processing, it became clear that significant radiation damage had occurred. The first batch had sustained the least amount of damage and was therefore the preferred source of experimental data. However, owing to saturation of the detector no data below 2.5 Å were included in the first batch. To obtain the final data set, scaling and merging were carried out in a non-standard manner. Firstly, batch 1 was scaled and merged separately. Next, all four batches were processed together and scaled to batch 1. Finally, the program *SFTOOLS* (B. Hazes, unpublished results) was used to create data set 1 by taking all reflections from batch 1 and just adding the missing reflections from the scaled set of all data. This minimized the effect of radiation damage without compromising the data completeness.

2.2. Refinement

The previously published 1.63 Å PAK pilin structure (PDB code 1dzo; Hazes *et al.*, 2000) was used as the starting model for all refinements. To calculate $R_{\rm free}$ statistics, each data set inherited the test set from the 1.63 Å structure. For data exceeding 1.63 Å resolution, a random set of 5% of the new reflections were added to the test set. For all data sets, an initial round of rigid-body refinement was performed with

REFMAC (Murshudov *et al.*, 1997) to position the molecule more accurately in the cell. Subsequently, cycles of *REFMAC* maximum-likelihood positional and *B*-factor refinement were alternated with manual remodeling using *Xfit* (McRee, 1999) until the R and $R_{\rm free}$ factors converged. For the models based on data sets 1 and 2, riding H atoms were introduced, followed by further positional and anisotropic temperature-factor refinement in *REFMAC*. Finally, *SHELXL* (Sheldrick & Schneider, 1997) was used to assign occupancies for side chains with multiple conformations. Water molecules were added using 20 cycles of *ARP*/*wARP* (Perrakis *et al.*, 1999), after which waters without clear density in a $2F_{\rm o} - F_{\rm c}$ density map and waters not within 2.5–3.3 Å of an N or O atom were manually discarded. This semi-automatic water-building method was chosen to minimize bias so that the water molecules in all six structures could be objectively compared. To obtain the final hydration structures, more water molecules were subsequently added based on visual inspection of electron-density maps.

2.3. H-atom density

H atoms were deleted from the final models of high-resolution structures 1 and 2. The program *PDBSET* (Collaborative Computational Project, Number 4, 1994) was used to apply a random Gaussian shift (r.m.s. = 0.1156 Å; maximum shift = 0.2 Å) to the coordinates to remove model bias. After refinement of the resulting model without H atoms, $\sigma_{\rm A}$-weighted difference map Fourier coefficients from *REFMAC* were read by the program *HYDENS* (B. Hazes, unpublished results) to quantify the electron-density signal at riding H-atom positions using a classical Fourier summation.

2.4. Hydration shells

The program *MCFMAN* (B. Hazes, unpublished results) was used to assign hydration shells. The program *LSQKAB* (Collaborative Computational Project, Number 4, 1994) was used to superimpose pairs of structures, after which the programs *GROUPWATER* and *H2ODIST* (B. Hazes, unpublished results) were used to assign and tabulate groups of conserved water positions. Individual water O atoms were considered to correspond to equivalent positions if they lay within a distance of 1 Å of each other. The same procedure was used to analyse the lysozyme structures (PDB codes 3lzt and 4lzt; Walsh *et al.*, 1998).

2.5. Program availability

SFTOOLS is distributed as part of the *CCP*4 package (Collaborative Computational Project, Number 4, 1994). All other unpublished programs are available at http://eagle.mmid.med.ualberta.ca/highlights.html.

3. Results and discussion

3.1. Ultrahigh-resolution structure and radiation damage

The crystal structure of PAK pilin has been determined at 1.63 Å resolution using diffraction data collected at room temperature on a rotating-anode X-ray source (Hazes *et al.*, 2000). Subsequently, a 0.78 Å resolution data set was obtained at cryogenic temperature on beamline SBC-19ID of the Advanced Photon Source (Chicago, USA). A data set based on all diffraction images gave significantly improved electron-density maps. However, the R and $R_{\rm free}$ refinement statistics converged at 16.4 and 17.9%, respectively, which is considerably higher than comparable ultrahigh-resolution structures. We anticipated that this was a result of radiation damage and re-refinement against the 'batch 1' data set, a 0.78–2.50 Å data set based on the least-damaged first 500 images (see §2.1), reduced R and $R_{\rm free}$ to 12.7 and 14.1%, respectively. To obtain a complete data set with reduced radiation damage, we took every reflection present in the 0.78–2.5 Å data set and added only the missing reflections from the more damaged data sets (see §2.1). The final model was refined against these data, yielding R and $R_{\rm free}$ factors of 14.2 and 15.4%, respectively.

3.2. Radiation damage to the disulfide bond

In the 1.63 Å PAK pilin structure, a well defined disulfide bond is observed between Cys129 and Cys142 (Fig. 1*a*). In the 0.78 Å structure this disulfide bond is still present, but a second conformation of Cys142 S^{γ} is clearly indicated by the electron density (Fig. 1*b*). Both cysteines were therefore modeled with a second '*B*' conformation. The distance between the sulfurs in the '*B*' conformation is 2.78 Å, which is too long for a disulfide bond and too short for a normal van

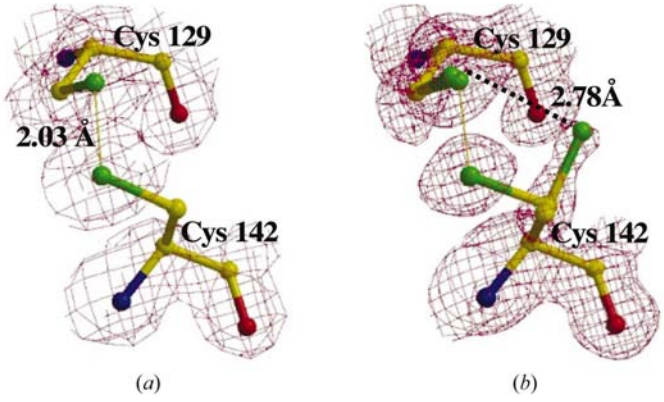

Figure 1
$2F_{\rm o} - F_{\rm c}$ electron density for cysteine residues 129 and 142 contoured at 1.5σ. (*a*) Density for structure 4, based on rotating-anode diffraction data to 1.63 Å resolution, shows a single conformation of the disulfide bond. (*b*) Density for structure 1, based on 0.78 Å diffraction data from a third-generation synchrotron source, shows two conformations for the S atom of Cys142. Cysteines 129 and 142 have both been given dual side-chain conformations to model the intact and broken disulfide, respectively. This yields a disulfide bond with a typical length of 2.03 Å and two S atoms at a distance of 2.78 Å, representative of a disulfide radical (Weik *et al.*, 2002). The figure was produced using *Bobscript* (Esnouf, 1997) and *Raster3D* (Merritt & Murphy, 1994). C, N, O and S atoms are displayed in yellow, blue, red and green, respectively.

Table 2
Diffraction data and refinement statistics.

Data set	1†	2	3	4	5	6
Light source	APS	SRS	Rotating anode	Rotating anode	Rotating anode	Rotating anode
Beamline	19ID	9.6	—	—	—	—
Detector	SBC2 CCD	ADSC Q210	R-AXIS IV	MAR 300	R-AXIS IV	R-AXIS IV
Temperature (K)	100	100	100	293	293	293
Wavelength (Å)	0.65	0.87	1.54	1.54	1.54	1.54
Maximum resolution (Å)	0.78	0.95	1.51	1.63	1.80	1.80
Unit-cell parameters (Å)						
$a = b$	37.24	37.58	37.48	38.11	38.04	37.81
c	147.86	148.56	147.89	149.78	149.58	149.44
Total observations	1274902 (608237)	1198789	165793	110628	167068	312705
Unique observations	107451 (100725)	67341	16090	14500	10953	10797
$R_{\rm merge}$‡ (%)	8.2 (8.0)/35.3 (33.7)	7.4/12.3	5.0/12.2	4.9/19.0	7.9/49.8	10.9/25.1
$I/\sigma(I)$‡	5.47 (3.38)§/0.70 (1.0)¶	21.41/2.46	23.50/4.09	26.70/10.10	11.59/2.50	14.62/7.22
Completeness‡ (%)	89.8 (84.3)/42.1 (42.4)	75.0/4.5††	91.7/16.3††	99.5/89.6	98.6/76.0	94.6/62.1
Wilson B factor (Å^2)	6.8 (6.1)	7.4	12.2	16.8	19.0	15.0
R factor (%)	14.2	11.3	13.8	14	14.8	14.5
$R_{\rm free}$ (%)	15.4	13.1	16.1	15.8	18.5	17.2
Bond length (Å)	0.019	0.007	0.011	0.009	0.020	0.014
Bond angle (°)	1.99	1.34	1.42	1.34	2.43	1.54
Estimated coordinate error	0.01	0.02	0.07	0.07	0.10	0.07

† Values in parentheses are statistics for the first 500 images (batch 1, 0.76–2.5 Å). ‡ Overall and highest resolution shell statistics are listed separated by a forward slash. The highest resolution shells for data sets 1–6 are 0.78–0.79, 0.95–1.01, 1.51–1.55, 1.63–1.72, 1.8–1.85 and 1.8–1.85 Å, respectively. § Average $I/\sigma(I)$ is around 60 prior to merging, as expected for a strongly diffracting crystal. During merging, $\sigma(I)$ is scaled up to account for larger than expected variations between symmetry-related reflections, which are likely to have arisen at least in part owing to radiation damage. This reduces the average $I/\sigma(I)$ to the low values listed in the table. ¶ Inclusion of data to 0.78 Å was not based on $I/\sigma(I)$, which is unusually low for the reasons explained above. $R_{\rm merge}$, although typically a less reliable statistic, did indicate useful signal to this resolution. This was confirmed by the fact that calculated amplitudes from a model refined with data to 0.85 Å resolution showed a positive correlation with the observed amplitudes to 0.78 Å resolution. †† The low completeness at high resolution arises from the square shape of the detector. At the shortest allowed distance, only the detector corners contributed to the highest resolution reflections.

der Waals interaction. This distance is indicative of a disulfide radical (Carugo & Bordo, 1999). Occupancy refinement in *SHELXL* (Sheldrick & Schneider, 1997) resulted in occupancies of 0.54 for both sulfurs forming the disulfide bond and of 0.38 and 0.29 for the alternate conformations of Cys129 and Cys142, respectively. The reduction in the total occupancy for the S$^{\gamma}$ atoms of Cys129 and Cys142 to 0.83 and 0.92, respectively, suggests that S atoms have been lost owing to radiation damage, a phenomenon that has been reported previously (Burmeister, 2000; Leiros *et al.*, 2001). No damage to the disulfide bond has been observed with any of the rotating-anode data sets, whilst less damage was seen with data set 2, which was collected at a second-generation synchrotron, or when only the first 500 images of data set 1 were used (results not shown).

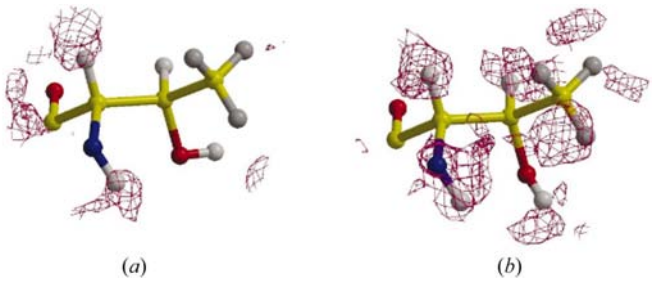

Figure 2
$F_{\rm o} - F_{\rm c}$ electron-density maps for Thr95 in structure 1 contoured at 1.5σ. To remove model bias, the final structure 1 model was re-refined after removal of the H atoms and applying a small random perturbation of the coordinates (see §2). (*a*) Map based on all data shows density for only two H atoms. (*b*) Map based on just the data in batch 1, which has less radiation damage, shows six H atoms. The figure was produced using *Bobscript* (Esnouf, 1997) and *Raster3D* (Merritt & Murphy, 1994). H, C, N and O atoms are displayed in white, yellow, blue and red, respectively.

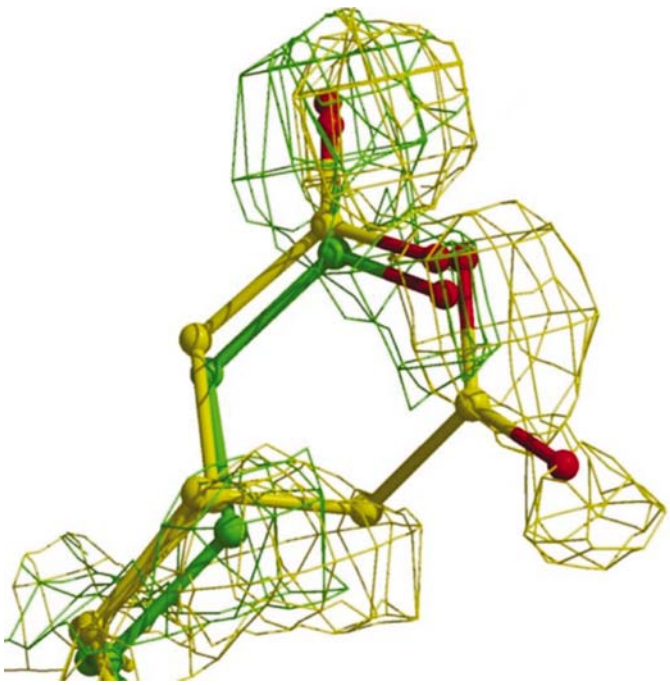

Figure 3
$2F_{\rm o} - F_{\rm c}$ density for Glu27 contoured at 1σ. Density for structure 3 (yellow) shows two Glu27 side-chain conformations. Density for structure 4 (green) shows only one side-chain conformation. Structures 3 and 4 are representative of all cryocooled and all room-temperature structures, respectively. The figure was produced using *Xfit* (McRee, 1999) and *Raster3D* (Merritt & Murphy, 1994).

3.3. H-atom density

In addition to the very local effect of radiation damage on the disulfide bond, we also observed a global effect on electron-density quality. This was most obvious when looking for weak signals such as experimental density for H atoms. As H atoms have only one electron, their detection is highly susceptible to radiation-damage-induced noise in the electron-density map. When we evaluated the experimental electron density at riding H-atom positions in a σ_A-weighted $F_o - F_c$ difference map (see §2.3), we could only detect 0.76% of the H atoms above 2σ when using all data (Fig. 2*a*). However, if the difference map was based on the 0.78–2.5 Å data set with less radiation damage, then 27.4% of the riding H atoms had experimental density above 2σ and 5.7% had density stronger than 3σ. The latter difference map also showed density for one polar H atom, the hydroxyl H atom of Thr95 (Fig. 2*b*). In the 0.78 Å crystal structure of *Bacillus lentus* subtilisin, 14% of all riding H atoms had a signal above 3σ (Kuhn *et al.*, 1998). In hindsight, our attempt to push resolution by using a high X-ray dose probably decreased our ability to detect H atoms owing to increased radiation damage.

3.4. Temperature-induced structural changes

A comparison of the 1.63 and 0.78 Å structures showed differences in multiple side-chain conformations, small differences in general atomic coordinates and a much larger number of bound water molecules in the high-resolution structure. To determine the contributions of resolution, radiation damage and temperature to these differences, we collected four more data sets. All data sets are based on fresh crystals except for data set 5. To determine whether cryo-induced changes are reversible, we chose to re-use the crystal that gave cryogenic data set 3 for the room-temperature data set 5. Details of the data-collection and refinement statistics for all six data sets are given in Table 2.

3.5. Structural variation in the main chain

The truncated PAK pilin contains residues 29–144 of the native protein preceded by four residues, 25–28, derived from the expression vector. The main-chain structure of all models is basically identical, but superimposing pairs of structures using the program *LSQKAB* (Collaborative Computational Project, Number 4, 1994) shows interesting deviations (Table 3). For the room-temperature structures, the r.m.s. deviations between residues that form the main secondary-structural elements (core residues) are a little less than 0.05 Å and this increases to between 0.07 and 0.10 Å when all residues are used for the analysis. For the cryogenic structures these numbers range from 0.05 to 0.08 Å for the core residues and 0.09 to 0.13 Å when considering all residues. Except for the comparison of structures 1 and 2, these values are not significantly different from the expected coordinate error (Table 2) and the higher values for the cryogenic structures may reflect lower reproducibility of the cryogenic cooling step. Significantly higher values are obtained when comparing the room-temperature structures with the cryogenic structures:

Table 3
Structural deviations between pairs of PAK structures.

Pairs of PAK structures were superimposed using *LSQKAB* (Collaborative Computational Project, Number 4, 1994) and r.m.s. deviations for all superimposed main-chain atoms, including C^{β}, are listed. Results for superimposing all residues are given below the diagonal. Results for superimposing only the core residues, those that form the α-helix and the main four β-strands, are given above the diagonal. Values in bold highlight comparisons between structures determined at different temperatures.

Temperature (K)		100	100	100	293	293	293
	Structure	1	2	3	4	5	6
100	1	—	0.079	0.069	**0.232**	**0.222**	**0.214**
100	2	0.130	—	0.047	**0.192**	**0.188**	**0.178**
100	3	0.088	0.107	—	**0.204**	**0.195**	**0.185**
293	4	**0.267**	**0.238**	**0.229**	—	0.041	0.049
293	5	**0.257**	**0.232**	**0.219**	0.092	—	0.047
293	6	**0.239**	**0.221**	**0.204**	0.069	0.103	—

0.18 to 0.23 Å for the core residues and 0.20 to 0.27 Å for all residues (values in bold in Table 3). The increase is not simply a consequence of the typical cryoinduced 1–2% contraction of proteins (Juers & Matthews, 2001), because computational contraction of the room-temperature structures improved the r.m.s. values by only ~0.04 Å (with an optimal contraction factor of 1%). Clearly, cryoinduced shrinking of the protein does not explain the large r.m.s. values between room-temperature and cryocrystal structures. The high r.m.s. values also do not appear to arise from radiation damage. Although structure 1 (highest radiation dose) has the highest r.m.s. values, even structure 3 (with no observable radiation damage) shows high r.m.s. values when compared with the room-temperature structures. Therefore, any radiation-damage effect must be small. Instead, we conclude that cryogenic cooling causes small but significant systematic changes that are not just rigid-body rotation, translation or scaling effects. The effects are larger in loop regions, but can be readily detected in even the core residues of the protein. These cryoinduced changes are reversible, as structure 5 behaves like a typical room-temperature structure even though a cryocooled crystal was thawed to collect this data set.

3.6. Structural variation in the side chains

In our comparison of the 1.63 and 0.78 Å structures, we noticed changes in side-chain conformations that were not readily explainable by increased resolution or the presence of radiation damage. To determine whether this reflects crystal-to-crystal variation or a more systematic cryoinduced change, we compared the six PAK structures. As Cys129 and Cys142 have a conformational change that is clearly induced by radiation damage, these residues were not considered in this analysis. We analyzed 13 residues with dual side-chain conformations in at least one of the six structures and one residue, Gln136, with a single side-chain conformation that was not identical in all structures. The results are summarized in Table 4. Four out of 13 residues with multiple conformations (Ser52, Thr84, Met104 and Lys112) display the same conformations in all six structures. Two residues (Arg30 and Lys110) have dual side-chain conformations in only one of the six

Table 4
Multiple side-chain conformations of the six PAK structures.

The letters *A*, *B* and *C* indicate distinct side-chain rotamers.

	Glu27	Arg30	Ser31	Glu49	Ser52	Arg53	Val57	Thr84	Lys88	Thr99	Met104	Lys110	Lys112	Gln136
1	*AB*	*A*	*A*	*AB*	*AB*	*AB*	*AB*	*AB*	*AB*	*A*	*AB*	*AB*	*AB*	*A*
2	*AB*	*A*	*A*	*A*	*AB*	*AB*	*AB*	*AB*	*AB*	*A*	*AB*	*A*	*AB*	*A*
3	*AB*	*A*	*A*	*AB*	*AB*	*AB*	*AB*	*AB*	*AB*	*A*	*AB*	*A*	*AB*	*A*
4	*A*	*AB*	*AB*	*B*	*AB*	*A*	*BC*	*AB*	*C*	*AB*	*AB*	*A*	*AB*	*B*
5	*A*	*A*	*AB*	*B*	*AB*	*A*	*BC*	*AB*	*C*	*AB*	*AB*	*A*	*AB*	*B*
6	*A*	*A*	*AB*	*B*	*AB*	*A*	*BC*	*AB*	*C*	*AB*	*AB*	*A*	*AB*	*B*

Table 5
Incomplete side chains for the six PAK structures.

Atom names indicate the last visible atom in the side chain. √ indicates that the side chain is complete.

	Arg30	Lys68	Lys88	Lys110	Lys128	Glu135
1	√	√	√	C^{δ}	√	√
2	√	C^{ε}	√	C^{δ}	√	√
3	√	√	√	C^{δ}	√	√
4	C^{δ}	√	C^{γ}	√	C^{γ}	C^{γ}
5	C^{β}	C^{δ}	C^{γ}	C^{γ}	C^{γ}	C^{β}
6	C^{β}	C^{δ}	C^{γ}	C^{γ}	C^{γ}	C^{β}

structures. This may reflect individual differences between crystals. In all other cases, side-chain conformations appear to be strictly correlated with temperature (Table 4). Clearly, radiation damage does not play a role in side-chain conformation as structure 3 behaves the same as structures 1 and 2. It is also unlikely that resolution explains the difference as structure 3 is based on data with a resolution that is only slightly higher than structure 4. More importantly, residues Ser31 and Thr99 have dual conformations at room temperature but not in the cryogenic structures and residues Val75, Lys88 and Gln136 have different conformations in the room-temperature and cryogenic structures. These observations indicate actual systematic differences as a function of temperature. An example of different side-chain conformations is given for Glu27 in Fig. 3.

Six residues were modeled with incomplete side chains in at least one of the six structures owing to a lack of electron density. These are all long hydrophilic surface residues, including four of the 11 lysine residues. The data presented in Table 5 show that the room-temperature crystal structures have a significantly larger number of incomplete side chains than the cryostructures. The lower resolution of the room-temperature structures may contribute to the inability to observe weakly occupied side-chain conformations. However, the fact that the number of incomplete side chains in structure 3, which has only slightly higher resolution than the room-temperature structures, is the same as the ultrahigh-resolution structure 1 suggests that temperature again plays a dominant role. Differential occupation of conformational states depends on both the free-energy difference of the states and the temperature, as described by the Boltzmann distribution law. As a consequence, for a given free-energy difference, higher temperature will increase structural disorder. This can lead to a situation where atoms can no longer be detected in the electron-density map. Tables 4 and 5 both show that the side-chain conformations of structure 5 match those of the other room-temperature structures, indicating that cryoinduced effects on side-chain conformations are also reversible.

3.7. Differences in hydration structure

Water molecules are important for many properties of proteins, including molecular recognition, enzyme activity, stability and folding. Unfortunately, protein hydration is often the most difficult aspect of protein structure to determine by X-ray diffraction owing to several technical problems. Firstly, protein–protein interactions in crystal-packing interfaces affect the hydration structure of the protein surfaces near those interfaces. Secondly, water molecules tend to be more difficult to model than protein atoms as their electron-density signal is normally weaker owing to a broader positional distribution (Carugo, 1999). Thirdly, their refinement is more difficult since without covalent bonding there are no strong bond-length and bond-angle restraints to help define the atomic position. The severity of these last two issues is resolution-dependent and the completeness of the hydration model in crystal structures therefore decreases at lower resolution. Indeed, Carugo & Bordo (1999), who studied 873 room-temperature and 33 low-temperature protein structures at a wide range of resolutions, concluded that the number of modeled water molecules depended primarily on the resolution of the structure, while the temperature of the crystal during data collection had only a minor effect. In contrast, in a study of the hydration of several 1.8 Å bovine β-trypsin structures, Nakasako (1999) reported that cryogenic structures contained 1.5–2.1 times more visible bound water molecules than room-temperature structures. In the analysis of our six structures, we observed that hydration levels of cryogenic structures were 1.9–2.7 times that of the room-temperature structures (Table 6). This is even notable when comparing structures 3 and 4, which are of comparible resolution. Resolution may have a small effect, as hydration level correlates with resolution for the cryogenic structures. However, our results, combined with the observations for bovine β-trypsin (Nakasako, 1999) and lysozyme (see below), indicate that cryocooling of crystals leads to a large increase in the number of detectable water molecules, with a smaller effect for resolution. Resolution may be more important for lower resolution structures, such as those included in the study by Carugo & Bordo (1999).

Table 6
Number of water molecules and water-shell distribution for the six PAK structures.

The percentage of water molecules in each shell is given in parentheses.

Structure	1	2	3	4	5	6
Shell 1	154 (69%)	151 (71%)	139 (71%)	90 (88%)	72 (88%)	84 (86%)
Shell 2	65 (29%)	56 (26%)	50 (25%)	12 (12%)	9 (11%)	14 (14%)
Shell 3	5 (2%)	5 (2%)	8 (4%)	0	1 (1%)	0
Total	224	213	197	102	82	98

Table 7
Structurally equivalent water molecules among pairs of PAK structures.

The numbers of equivalent waters are listed above the diagonal. Percentages are listed below the diagonal. The four values represent the results for shell 1/shell 2/shell 3/total.

	1	2	3	4	5	6
1	—	138/46/3/187	130/37/2/169	80/10/0/90	63/9/0/72	79/11/0/90
2	91/82/60/88	—	138/39/2/179	84/9/0/93	67/8/0/75	81/11/0/92
3	94/74/25/86	99/78/25/91	—	77/7/0/84	60/6/0/66	74/10/0/84
4	89/83/0/88	93/75/0/91	86/58/0/82	—	69/7/0/76	81/11/0/92
5	88/100/0/88	93/89/0/91	83/67/0/80	96/78/0/93	—	68/7/0/75
6	94/79/0/94	96/79/0/94	88/71/0/86	96/92/0/94	94/78/0/91	—

To extend our analysis, we grouped the water molecules into shells. First-shell waters have at least one direct hydrogen bond to a protein atom. Second-shell waters have at least one direct hydrogen bond to a first-shell water and third-shell waters only make hydrogen bonds to second-shell waters. Table 6 shows that for all structures the majority of waters (69–88%) are located in the first shell and a lesser number in the second shell (11–29%); waters are rarely located in the third shell (0–4%). Given that cryocooling had a large effect on the number of detectable waters, the effect on first- and second-shell waters was analyzed. This comparison shows that cryocooling causes a much greater increase in second-shell water numbers (4.9-fold) compared with first-shell water numbers (1.8-fold). To determine whether these hydration effects occur for other proteins, we compared two lysozyme structures (PDB codes 3lzt and 4lzt). 3lzt was refined against cryogenic data to 0.92 Å resolution and 4lzt was based on room-temperature data to 0.95 Å resolution. Both data sets were derived from crystals with the same unit cell and space group. The number of water molecules observed in the cryogenic structure was 1.84 times that of the room-temperature structure and the percentage of waters in shells 1, 2 and 3 was 88, 11 and <1% for the room-temperature structure and 77, 21 and 2% for the cryogenic structure. These numbers are very similar to our observations for PAK pilin and once more demonstrate that cryocooling has a large impact on the extent of hydration in crystal structures.

After showing that cryocooling significantly increases the number of modeled water molecules, we wished to determine whether the room-temperature water positions were conserved at cryogenic temperature. The results are listed in Table 7. A very high percentage of water positions are conserved between all six structures (80–94%), suggesting that there is a well defined subset of bound waters shared by all structures. This subset largely corresponds to the waters observed in the room-temperature structures. Within this subset, the extent of conservation between all room-temperature structures depends on the water shell. Conservation is higher for first-shell waters, averaging 92%, than for second-shell waters, averaging 79%. Waters that are observed in all room-temperature structures are to a very large extent also present in the cryogenic structures; only three water molecules shared by all room-temperature structures are missing from the cryogenic structures. In contrast, 73 waters that are found in all cryogenic structures are not present in any of the room-temperature structures (Fig. 4).

Three distinct effects may contribute to cryoinduced changes in hydration.

(i) The higher resolution of cryogenic data allows the detection of more poorly defined waters. However, the equal resolution of the lysozyme structures and the fact that we see most cryoinduced changes already in structure 3 at 1.51 Å resolution suggests that this effect is small.

(ii) Reduced thermal motion and/or a smaller entropic penalty for ordering water at lower temperature may reduce positional disorder and raise the electron density to detectable levels. Interestingly, the average *B* factor for the cryogenic

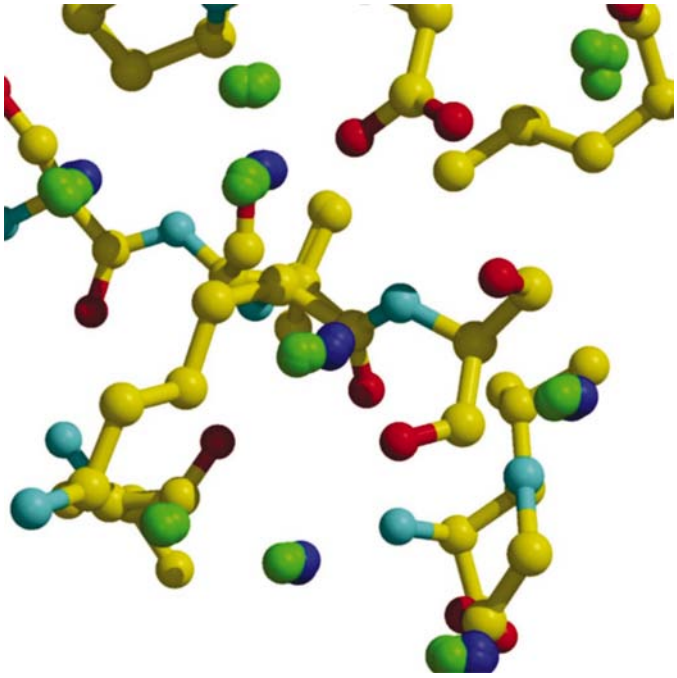

Figure 4
Ball-and-stick model showing superimposed cryocooled (green) and room-temperature (dark blue) water molecules for all six PAK structures within a small section of the PAK protein. The structure shows that waters observed in room-temperature structures are also present in cryocooled structures but not *vice versa*. The figure was made using *Xfit* (McRee, 1999) and *Raster3D* (Merritt & Murphy, 1994).

waters that are not observed in room-temperature structures is 35.72 Å^2. This is higher than for the cryogenic waters that are also present at room temperature (28.47 Å^2), indicating that the latter are more strongly localized hydration sites.

(iii) Changes in the free-energy landscape owing to structural rearrangements or direct temperature effects on the force field may change the location or existence of water-binding sites. This does not explain the systematic cryoinduced increase in hydration level but could account, for instance, for the observation of the three water molecules that are only found in the room-temperature structures.

4. Conclusions

Brilliant synchrotron X-ray sources combined with cryocooling have made it possible to solve protein structures from crystals that were previously considered either to be too small or to diffract too weakly. Even for crystal structures that can be solved with conventional X-ray sources, the use of cryogenic data collection at a high-brilliance synchrotron source significantly improves resolution and data quality. However, it is important to remain aware of potential drawbacks. The occurrence of radiation damage to specific chemical groups, in particular disulfide bonds, is now well documented. In contrast, the effect of cryocooling has received much less attention, perhaps because it is less conspicuous and only becomes clear in direct comparisons between equivalent structures solved at ambient and cryogenic temperatures. Our work using three data sets at each temperature demonstrates that there are indeed consistent cryoinduced structural effects. There are small coordinate shifts throughout the protein, but these are not likely to impact on biological interpretations. More importantly, discrete changes are observed in the hydration structure and the conformation of side chains, especially surface-exposed side chains that already display some disorder. In a few cases, bound waters and side-chain conformations seen at room temperature no longer exist in the cryogenic structures. However, the predominant effect is a reduced disorder at cryogenic temperature that results in fewer disordered side chains and more extensive hydration. Since most biological processes occur at the protein surface, it is important to be aware that the physiological side-chain conformation and hydration structure may differ from the cryogenic crystal structure. Basically, the interpretation of protein surface features in a cryogenic structure needs the same caution as has traditionally been applied to the interpretation of surface features in crystal-packing interfaces. It is our opinion that in the great majority of cases the benefits of cryogenic synchrotron diffraction data outweigh the drawbacks and often there may be no choice. However, our results indicate that for crystals with sufficient diffraction power it is advisable to collect a room-temperature data set also.

This research was funded by the Canadian Institutes of Health Research grants to BH (MOP-42448) and RI (MOP-38004). We thank Dr Norma Duke for assistance during data collection at the Advance Photon Source.

References

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Betzel, C., Gourinath, S., Kumar, P., Kaur, P., Perbandt, M., Eschenburg, S. & Singh, T. P. (2001). *Biochemistry*, **40**, 3080–3088.
Burmeister, W. P. (2000). *Acta Cryst.* D**56**, 328–341.
Carugo, O. (1999). *Protein Eng.* **12**, 1021–1024.
Carugo, O. & Bordo, D. (1999). *Acta Cryst.* D**55**, 479–483.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Esnouf, R. (1997). *J. Mol. Graph.* **15**, 133–138.
Halle, B. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 4793–4798.
Hazes, B., Sastry, P. A., Hayakawa, K., Read, R. J. & Irvin, R. T. (2000). *J. Mol. Biol.* **299**, 1005–1017.
Jelsch, C., Teeter, M. M., Lamzin, V., Pichon-Pesme, V., Blessing, R. H. & Lecomte, C. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 3171–3176.
Juers, D. H. & Matthews, B. W. (2001). *J. Mol. Biol.* **311**, 851–862.
Kriminski, S., Kazmierczak, M. & Thorne, R. E. (2003). *Acta Cryst.* D**59**, 697–708.
Kuhn, P., Knapp, M., Soltis, S. M., Ganshaw, G., Thoene, M. & Bott, R. (1998). *Biochemistry*, **37**, 13446–13452.
Leiros, H. K., McSweeney, S. M. & Smalas, A. O. (2001). *Acta Cryst.* D**57**, 488–497.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Merritt, E. A. & Murphy, M. E. P. (1994). *Acta Cryst.* D**50**, 869–873.
Miyazaki, Y., Matsuo, T. & Suga, H. (2000). *J. Phys. Chem. B*, **104**, 8044–8052.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Nakasako, M. (1999). *J. Mol. Biol.* **289**, 547–564.
Perrakis, A., Morris, R. M. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.22.
Pflugrath, J. W. (1999). *Acta Cryst.* D**55**, 1718–1725.
Ravelli, R. B. & McSweeney, S. M. (2000). *Structure Fold. Des.* **8**, 315–328.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **276**, 319–343.
Teng, T.-Y. (1990). *J. Appl. Cryst.* **23**, 387–391.
Walsh, M. A., Schneider, T. R., Sieker, L. C., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1998). *Acta Cryst.* D**54**, 522–546.
Weik, M., Berges, J., Raves, M. L., Gros, P., McSweeney, S., Silman, I., Sussman, J. L., Houee-Levin, C. & Ravelli, R. B. (2002). *J. Synchrotron Rad.* **9**, 342–346.
Weik, M., Ravelli, R. B., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., Gros, P., Silman, I., Kroon, J. & Sussman, J. L. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 623–628.