Acta Cryst. (1998). D**54**, 938–953

Continuous and Discontinuous Changes in the Unit Cell of HIV-1 Reverse Transcriptase Crystals on Dehydration

ROBERT M. ESNOUF,[a,b,c] JINGSHAN REN,[a] ELSPETH F. GARMAN,[a] DON O'N. SOMERS,[d]† CARL K. ROSS,[d] E. YVONNE JONES,[a,b] DAVID K. STAMMERS[d]‡ AND DAVID I. STUART[a,b]*

[a]*The Laboratory of Molecular Biophysics, Rex Richards Building, South Parks Road, Oxford OX1 3QU, England,* [b]*The Oxford Centre for Molecular Sciences, New Chemistry Building, South Parks Road, Oxford OX1 3QT, England,* [c]*The Rega Institute for Medical Research, Katholieke Universiteit Leuven, Minderbroedersstraat 10, B-3000 Leuven, Belgium, and* [d]*Structural Biology Group, The Glaxo Wellcome Research Laboratories, Langley Court, Beckenham, Kent BR3 3BS, England. E-mail: dave@biop.ox.ac.uk*

(*Received 25 November 1997; accepted 19 March 1998*)

Abstract

A crystal form of HIV-1 reverse transcriptase (RT) complexed with inhibitors showed diffraction to a high-resolution limit of 3.7 Å. Instability in the unit-cell dimensions of these crystals was observed during soaking experiments, but the range of this variability and consequent change in lattice order was revealed by a chance observation of dehydration. Deliberately induced dehydration results in crystals having a varicty of unit cells, the best-ordered of which show diffraction to a minimum Bragg spacing of 2.2 Å. In order to understand the molecular basis for this phenomenon, the initial observation of dehydration, the data sets from dehydrated crystals, the crystal packing and the domain conformation of RT are analysed in detail here. This analysis reveals that the crystals undergo remarkable changes following a variety of possible dehydration pathways: some changes occur gradually whilst others are abrupt and require significant domain rearrangements. Comparison of domain arrangements in different crystal forms gives insight into the flexibility of RT which, in turn, may reflect the internal motions allowing this therapeutically important enzyme to fulfill its biological function.

1. Introduction

The reverse transcriptase (RT) of human immunodeficiency virus (HIV) performs a crucial step in the viral life-cycle, transcribing the genomic single-stranded RNA into double-stranded DNA prior to its incorporation into the host-cell genome. Most presently approved anti-AIDS therapies inhibit the normal functioning of this protein (DeClercq, 1995*a*,*b*). Inhibitors of RT fall into two groups: dideoxynucleoside analogues, which (as their metabolically activated triphosphates) cause premature termination of the extending DNA strand, and non-nucleoside inhibitors (NNIs; DeClercq, 1996). Kinetic (Spence *et al.*, 1995) and structural (Esnouf *et al.*, 1995) studies have shown that NNIs, which are almost exclusively HIV-1 specific, function by an allosteric mechanism whereby binding of an NNI in an internal pocket of the RT causes distortion of the polymerase active site. The effectiveness of anti-AIDS therapies is compromised by the selection for drug-resistant viral populations (Larder & Kemp, 1989; Schinazi *et al.*, 1996). Drug resistance and the presence of serious side-effects associated with some therapies (Richman *et al.*, 1987) make the search for new drugs and better ways of combining the use of existing drugs the focus of intensive research, one component of which is structural analysis.

Extensive efforts resulted in the growth of many different crystal forms for HIV-1 RT (*e.g.* Stammers *et al.*, 1990; Unge *et al.*, 1990; Lloyd *et al.*, 1991; Kohlstaedt *et al.*, 1992; Jones *et al.*, 1993; Jacobo-Molina *et al.*, 1993; Stammers *et al.*, 1994; Rodgers *et al.*, 1995), although few showed diffraction to high resolution. The structure was eventually described [for RT in complex with the NNI nevirapine (Merluzzi *et al.*, 1990)] based on data to a minimum Bragg spacing of 3.4 Å (Kohlstaedt *et al.*, 1992), and was later extended to 2.9 Å resolution (Smerdon *et al.*, 1994). The RT heterodimer is organised into nine structural domains (Fig. 1). The p66 subunit (560 residues) contains five domains named fingers, palm, thumb, connection and RNase H (Kohlstaedt *et al.*, 1992), and the smaller p51 subunit (the first 440 residues of the p66) comprises the first four of these domains. The internal structures of equivalent domains in the p66 and p51 subunits are similar, but their relative positioning is different. The arrangement of the p66 domains is open with some resemblance to a right hand, and this structure sits on the more compact p51 subunit (Fig. 1). This domain structure has been present for all

† Current address: Biomolecular Structure Unit, Glaxo Wellcome Research and Development, Medicines Research Centre, Gunnell's Wood Road, Stevenage, Hertfordshire SG1 2NY, England.

‡ Current address: The Laboratory of Molecular Biophysics, Rex Richards Building, South Parks Road, Oxford OX1 3QU, England.

reported crystal forms, but substantial domain rearrangements have been observed [an RT/Fab/DNA complex solved at 3.0 Å resolution (Jacobo-Molina *et al.*, 1993); RT complexed with NNIs other than nevirapine at 2.8–3.0 Å resolution (Ding, Das, Moereels *et al.*, 1995; Ding, Das, Tantillo *et al.*, 1995; Das *et al.*, 1996) and unliganded RT at both 3.2 Å resolution (Rodgers *et al.*, 1995) and 2.7 Å resolution (Hsiou *et al.*, 1996)]. Higher resolution data for fragments of the heterodimer confirm the structure within specific domains [the RNase H domain at 2.4 Å resolution (Davies *et al.*, 1991) and the fingers and palm domains at 2.2 Å resolution (Unge *et al.*, 1994)].

Our attempts to find a well ordered crystal form of HIV-1 RT had yielded crystals of several morphologies (Stammers *et al.*, 1990; Jones *et al.*, 1993), but in no case could high-resolution diffraction be obtained. Subsequently, orthorhombic crystals (space group $P2_12_12_1$, having one heterodimeric RT molecule with bound NNI in each asymmetric unit) were obtained which showed diffraction to 3.7 Å resolution at synchrotron sources (Stammers *et al.*, 1994). Soaking experiments with heavy-metal salt and oligomeric DNA solutions induced changes in the unit-cell dimensions of the crystals and in some cases the high-resolution limit was extended to 3.4 Å. The full extent of this variability was revealed by in-house X-ray measurements on a single crystal mounted in an imperfectly sealed capillary tube. As the crystal gradually dehydrated over a period of several hours, data from a whole series of reduced-volume cells were recorded. Analysis of these data based on Wilson plots (Wilson, 1949) suggested that some of these cell forms might show diffraction to a resolution better than 3.4 Å at a suitable X-ray source. A protocol for deliberate dehydration was devised and the first crystal to be treated in such a manner and exposed at a synchrotron source (BL6-A2 at the Photon Factory, Japan) showed diffraction to 2.5 Å resolution. This limit was subsequently extended to 2.2 Å resolution (Stammers *et al.*, 1994; Ren, Esnouf, Garman *et al.*, 1995).

Despite the difficulties associated with such pleomorphic crystals, structures for a series of RT–NNI complexes and for the unliganded enzyme have been obtained at resolutions of between 2.2 and 3.0 Å (Ren, Esnouf, Garman *et al.*, 1995, Esnouf *et al.*, 1995, 1997; Ren, Esnouf, Hopkins *et al.*, 1995; Hopkins *et al.*, 1996). We examine here the dehydration process central to the success of these structure determinations. Firstly, the data obtained from the crystal in the imperfectly sealed tube are re-examined, showing that dehydration produces a series of cell forms linked by both smooth and abrupt changes in unit-cell dimensions. Secondly, the best models for each cell form obtained by deliberately dehydrating crystals are examined and the changes are related to crystal-packing interactions and domain reorientations. These analyses show that at least six cell forms are possible, arising from a branched dehydration pathway, and that by reorientations of domains the crystals can withstand the removal of one third of their total solvent content.

2. Initial observation of dehydration

2.1. *Materials and methods*

Crystals of HIV-1 RT complexed with the NNI 1051U91 (Hargrave *et al.*, 1991) were grown at pH 5.0 by equilibration of sitting drops against a 6%(*w*/*v*) PEG 3400 reservoir (Stammers *et al.*, 1994). As part of a heavy-atom derivative search, one of these crystals was soaked in a tetrakis(acetoxymercuri)methane (TAMM) solution for 20 h prior to mounting in a quartz capillary tube. Data images from this crystal were collected at

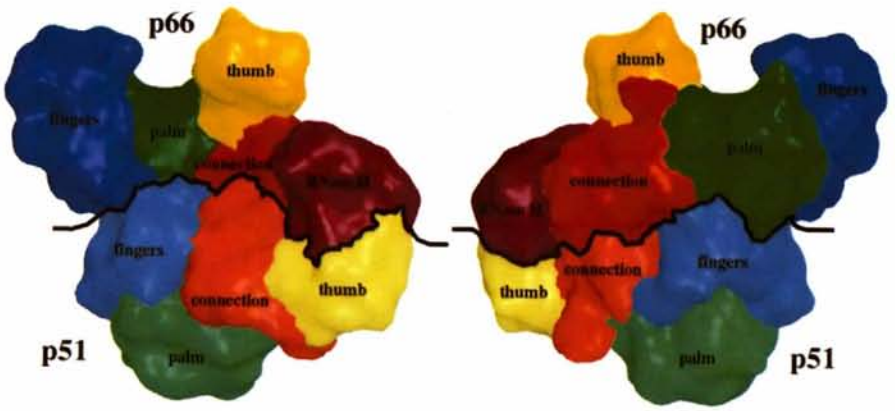

Fig. 1. Domain structure of the RT heterodimer. 'Front' and 'back' views of the structure of RT in complex with 1051U91 (Ren, Esnouf, Garman *et al.*, 1995) show the open structure for the p66 subunit (dark shades) sitting on a more compact base formed by the rearranged domains of the p51 subunit (pale shades of the same hue). For Figs. 1, 7, 9, 10 and 11 the surfaces for the heterodimer, individual domains and the contact patches were calculated using *VOLUMES* (R. M. Esnouf, unpublished program). Figs. 1, 7, 8(*a*), 8(*b*) and 11 were drawn using *BobScript* (Esnouf, 1997), a modified version of *MolScript* (Kraulis, 1991), and rendered with *Raster3D* (Bacon & Anderson, 1988; Merritt & Murphy, 1994).

room temperature using Cu $K\alpha$ radiation from a Rigaku RU200H rotating-anode generator operating at 60 kV and 70 mA and recorded on an 18 cm diameter MAR Research imaging plate placed 189.1 mm from the crystal. Images were collected overnight using 1° oscillations and an exposure time of 15 min. The unusual nature of these images (Fig. 2) was realised the next morning and further images (a total of 81 altogether) were recorded until (apparent) radiation damage became severe. The images showed both increases and decreases in the high-resolution limit of diffraction over time, along with the appearance and disappearance of diffraction due to multiple lattices (Fig. 2). On dismounting the capillary it was observed that the crystal was no longer surrounded by mother liquor and examination revealed a small crack in the tube wall, suggesting dehydration as the cause of this unusual diffraction behaviour.

2.2. *Data processing*

Originally, blocks of images were indexed using *IMSTILLS* and *REFIX* (*MOSFLM* v. 5.20; A. G. W. Leslie, unpublished program) and spot intensities were integrated with *MARXDS* (Kabsch, 1988). For the present analysis, the images were reprocessed using *DENZO* (Otwinowski, 1993). The changes in crystal orientation and unit-cell dimensions during data collection were sometimes quite abrupt and successful processing required that indexing was performed independently for several images spread through the data set. Contiguous batches of images were processed, going both forwards and backwards through time from these reference images. This protocol gave consistent unit-cell parameters and crystal orientations for most images regardless of the choice of reference images. The exceptions were images 3–5 (*e.g.* Fig. 2*b*) and 25–29 (*e.g.* Fig. 2*e*), during which there were jumps in the unit-cell dimensions and diffraction from multiple lattices was evident. The best images were used to obtain estimates for the experimental parameters (*e.g.* crystal-to-detector distance) and then all images were reprocessed keeping these parameters fixed in order to give a better indication of relative changes to the unit cell.

Since changes between consecutive images could be substantial, it was impossible to estimate the mosaic spread of the crystal by comparison between different images in post-refinement programs (Winkler *et al.*, 1979) such as *POSTREF* (Collaborative Computational

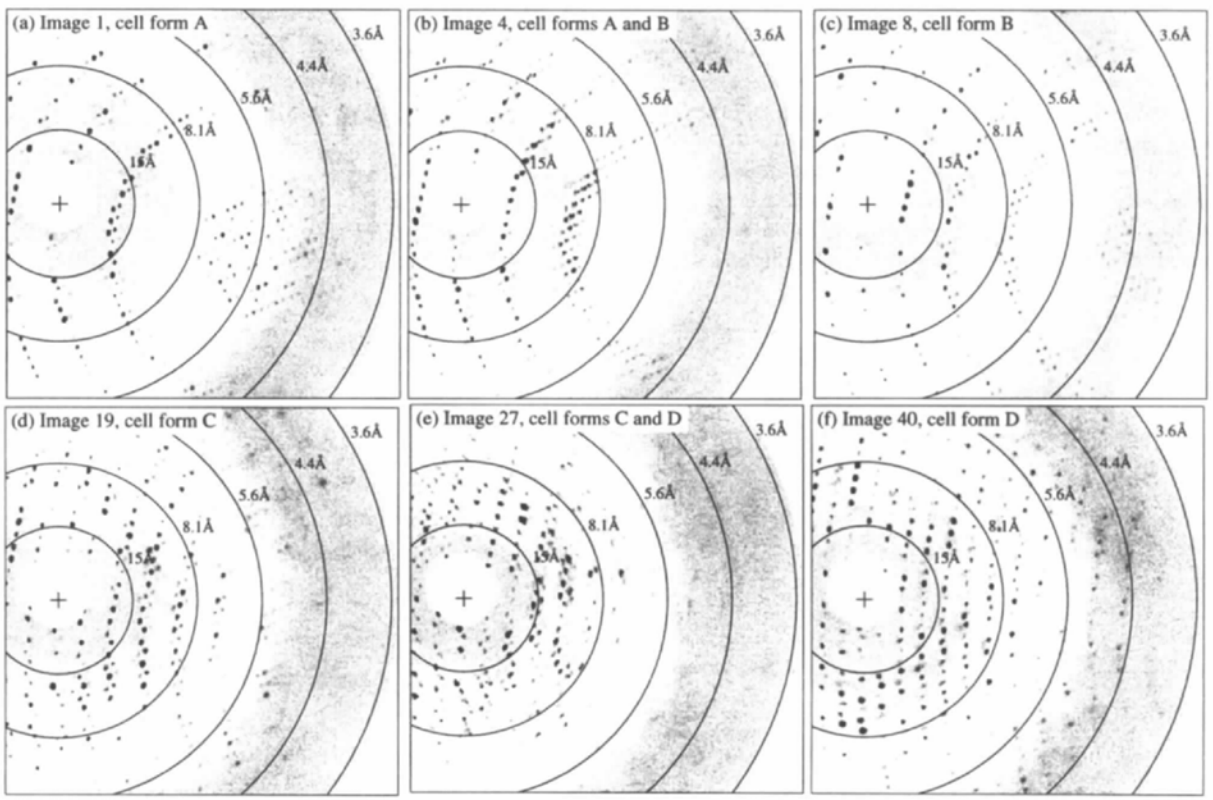

Fig. 2. Diffraction images from the crystal mounted in an imperfectly sealed capillary tube. Images are numbered and labelled with the cell forms contributing to the diffraction. The time between consecutive images is approximately 20 min. Figure produced using *PSIMAGE* (R. M. Esnouf, unpublished program).

Project, Number 4, 1994) and *SCALEPACK* (Otwinowski, 1993). Thus, we are limited to estimates based on the histogram produced by *DENZO* of mean intensity for partially recorded reflections.

For the periods of discontinuous unit-cell change around images 4 and 27 (Figs. 2*b* and 2*e*) it was possible to process images for two lattices separately by processing towards the discontinuity using the images to either side. Thus, unit-cell and mosaic-spread estimates were obtained for all 81 images, although the last few images were so poor that these estimates have little meaning.

2.3. *Results*

Initially (Fig. 2*a*), the crystal had unit-cell dimensions $a \simeq 147$, $b \simeq 112$ and $c \simeq 79$ Å (cell form *A*), but during the exposure of image 4 (Fig. 2*b*) there was a sudden reduction in the *a* dimension. For image 8 (Fig. 2*c*), the

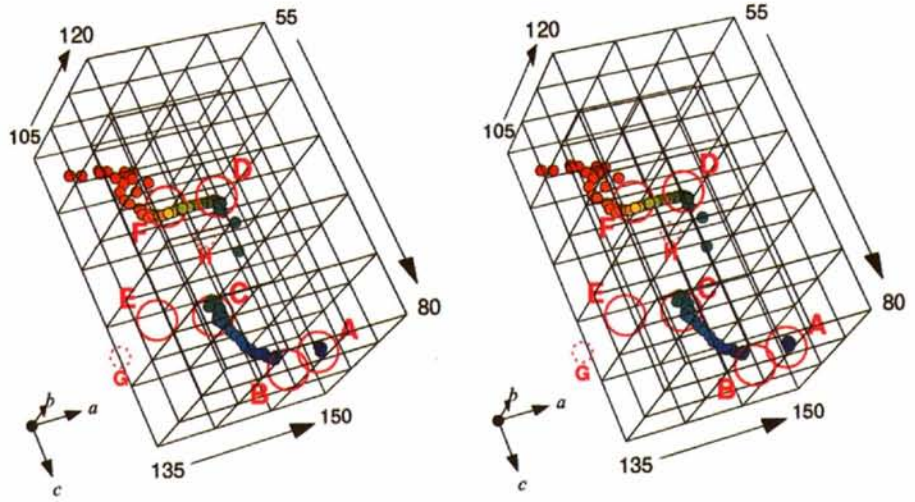

Fig. 3. Stereo diagram showing the unit-cell dimensions of the RT crystal obtained from each image during dehydration. Each point represents one image with the *a*, *b*, and *c* unit-cell dimensions of the crystal (in Å) defining its coordinate in three dimensions. The colour of the points changes from blue to red in order of image number. Magenta circles define the positions of the six well characterized cell forms (*A*–*F*) and two further possible forms (*G* and *H*). Figs. 3, 5(*a*), 5(*b*), 9 and 10 were produced by modifying the output of *BobScript* (Esnouf, 1997).

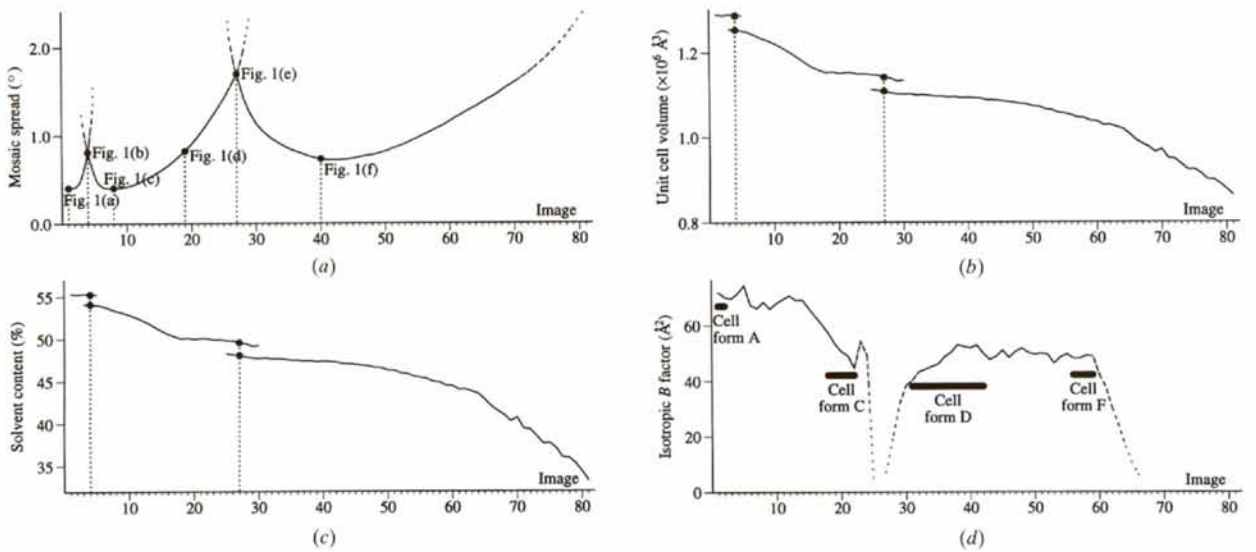

Fig. 4. Properties of the RT crystal during dehydration. (*a*) The estimates for the mosaic spread of the crystal based on the histogram of mean intensity for partially recorded reflections for each image. (*b*) The unit-cell volume obtained from each image. For the region of discontinuous change around image 27 it was possible to index separately reflections arising from each lattice by processing towards the change from either side of it. (*c*) The solvent content of the crystal for each image calculated using the expression [100 − 123 (molecular weight in unit cell/unit-cell volume)]% (Matthews, 1968). (*d*) The isotropic *B* factor for data measured during dehydration. Due to the small number of reflections recorded on each image, reflections (at resolutions higher than 6 Å) from batches of five images centred on each image in turn were sorted into resolution bins and the *B* factor estimated from a plot of the logarithm of mean reflection intensity against resolution ($\sin^2\theta/\lambda^2$). Around image 27 and after image 60 there are very few data at sufficiently high resolution and the *B*-factor estimates have no significance. Periods during which the crystal unit-cell dimensions roughly correspond to well defined cell forms are indicated.

unit-cell dimensions were $a \simeq 143$, $b \simeq 112$ and $c \simeq 79$ Å (cell form *B*), corresponding to the cell form from which diffraction had been previously observed to a resolution limit of 3.4 Å. This proved to be merely the 'tip of the iceberg', however, and a steady reduction in the unit-cell volume occurred until a more stable unit cell emerged around image 19 (Fig. 2*d*) having dimensions $a \simeq 141$, $b \simeq 111$ and $c \simeq 73.5$ Å (cell form *C*) and showing diffraction to the edge of the image plate (3.5 Å resolution). This stability was short-lived and the images around number 27 (Fig. 2*e*) showed poor diffraction arising from multiple lattices. Remarkably, the crystal re-annealed into yet another stable cell form, and by image 40 (Fig. 2*f*) it was again showing diffraction to 3.5 Å resolution, for a unit cell of dimensions $a \simeq 142$, $b \simeq 116$ and $c \simeq 66.5$ Å (cell form *D*). Steady dehydration continued and beyond image 60 this was accompanied by a steady degradation in the quality of diffraction. The most dehydrated cell giving respectable diffraction was not initially classed as a distinct cell form, but it has since been regularly observed for crystals flash-cooled to 100 K. This unit cell has dimensions $a \simeq 138.5$, $b \simeq 115$ and $c \simeq 66$ Å (cell form *F*). The other well characterized unit cell, $a \simeq 137$, $b \simeq 109.5$ and $c \simeq 72$ Å (cell form *E*), is also observed after flash-cooling, but was not observed for this crystal. From image 60 onwards, continuing dehydration rather than radiation damage is likely to have been the main cause of the degradation of diffraction quality.

The effect of dehydration on the unit cell can be shown as a three-dimensional scatter plot (Fig. 3), which shows two periods of steady dehydration and two abrupt changes in unit-cell dimensions. The same pattern is observed in the estimates for the mosaic spread of the crystal (Fig. 4*a*). Sharp increases in the mosaic spread accompany abrupt changes in cell form, but as the new cell form anneals the mosaic spread decreases substantially before further dehydration and radiation damage increase the disorder again. Dehydration reduces the unit-cell volume from 1.29×10^6 Å^3 (cell form *A*; image 1) to 1.05×10^6 Å^3 (cell form *F*; image 56), yet the crystal still gives good diffraction (Fig. 4*b*). This 18% reduction in volume corresponds to a reduction in solvent content from 56 to 45% (Fig. 4*c*), or a loss of 34% of the solvent molecules from the original crystal. By image 81 the cell volume is 0.86×10^6 Å^3, giving a

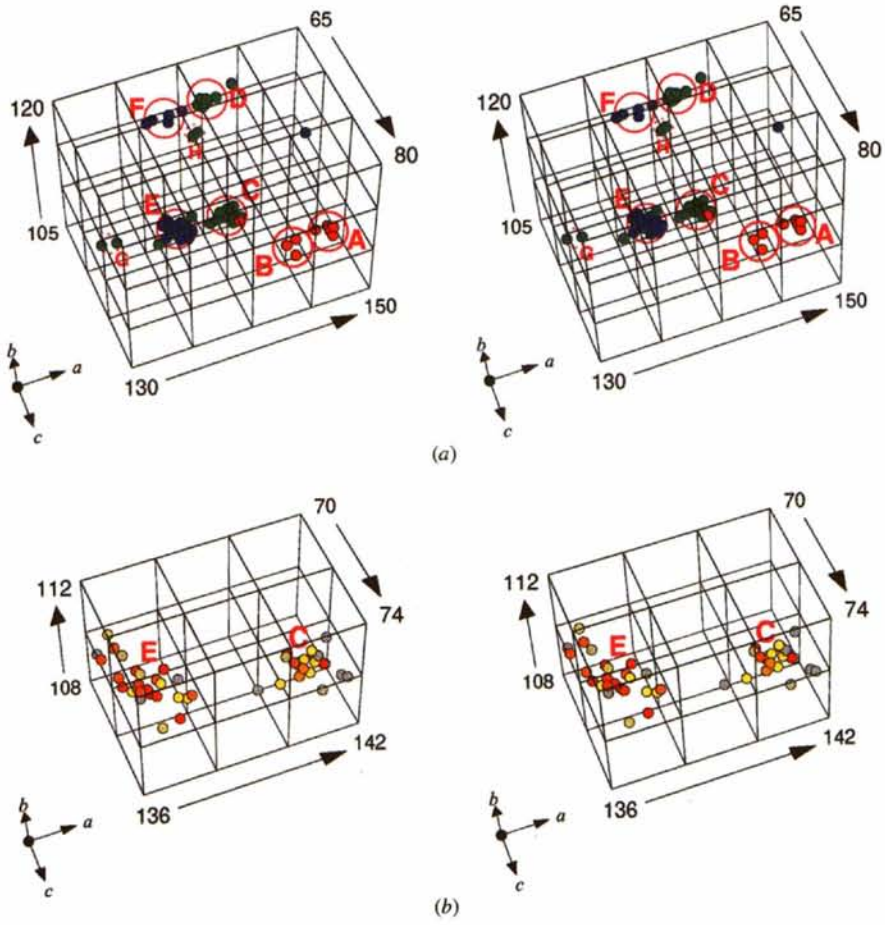

Fig. 5. Stereo diagrams showing the unit-cell dimensions for data sets collected from these RT crystals. Each point represents one data set with the *a*, *b* and *c* unit-cell dimensions (in Å) defining its coordinate in three dimensions. (*a*) All 78 data sets are shown with points coloured according to the crystal preparation (red, no special treatment or soaking in solutions of heavy-atom salts or DNA oligomers; green, soaking in concentrated PEG 3400 solutions; blue, soaking in concentrated PEG 3400 solutions followed by flash-cooling to cryogenic temperatures). Magenta circles define the positions of the six well characterized cell forms (*A*–*F*) and two further possible cell forms (*G* and *H*). (*b*) Data sets from crystals of cell forms *C* and *E* with points coloured according to the high-resolution limit of the diffraction from grey (worse than 3.5 Å) *via* yellow to red (better than 2.5 Å).

solvent content of 33% and corresponding to a total loss of 60% of solvent molecules. However, such a high level of dehydration is clearly not consistent with a well ordered crystal structure and results in poor diffraction.

Wilson plots (Wilson, 1949) from batches of five consecutive images were used to estimate the isotropic *B* factor at different times during the dehydration process (Fig. 4*d*). The *B* factor decreases from 70 Å^2 for the initial cell form, form *A*, to less than 50 Å^2 for the dehydrated cell forms *C*, *D* and *F*. It was this observation that prompted experiments in deliberate dehydration, since, while the dramatic crystal shrinkage is interesting in itself, its importance for us is the effect on the crystal-lattice order and hence the possibility of obtaining diffraction to higher resolution.

3. Deliberately induced dehydration

3.1. *Materials and methods*

Since the crystals grew in drops containing 6%(*w*/*v*) PEG 3400 (Stammers *et al.*, 1994), experiments were performed to establish their behaviour and stability in more concentrated PEG 3400 solutions. Crystals transferred between wells of increasing concentration were found to remain stable in buffered solutions of up to 46%(*w*/*v*) PEG 3400. This solution dehydrated most crystals to cell form *C* and some to form *D*. Our current protocol involves transfers over a period of 3 d between wells with the PEG concentration increasing in steps of 5% (Stammers *et al.*, 1994). The crystals are stable in the 46% solution and, indeed, it seems that storage in this solution gives them time to re-anneal, resulting in a higher percentage of crystals subsequently yielding useful data. Experiments using flow cells to increase the PEG concentration more gradually only achieved a similar success rate, so we prefer the simpler transfer protocol. Given the variability in unit-cell dimensions, it is important to obtain complete data sets from single crystals, for example by using cryo-crystallography. Fortunately, the concentrated PEG solution acts as a satisfactory cryo-protectant and the crystals can be spared the further osmotic shock of adding a specific cryo-protectant.

Data have been collected from crystals of RT (both unliganded and complexed with a wide variety of inhibitors), some of which have been subjected to further soaking procedures. Some data were collected using

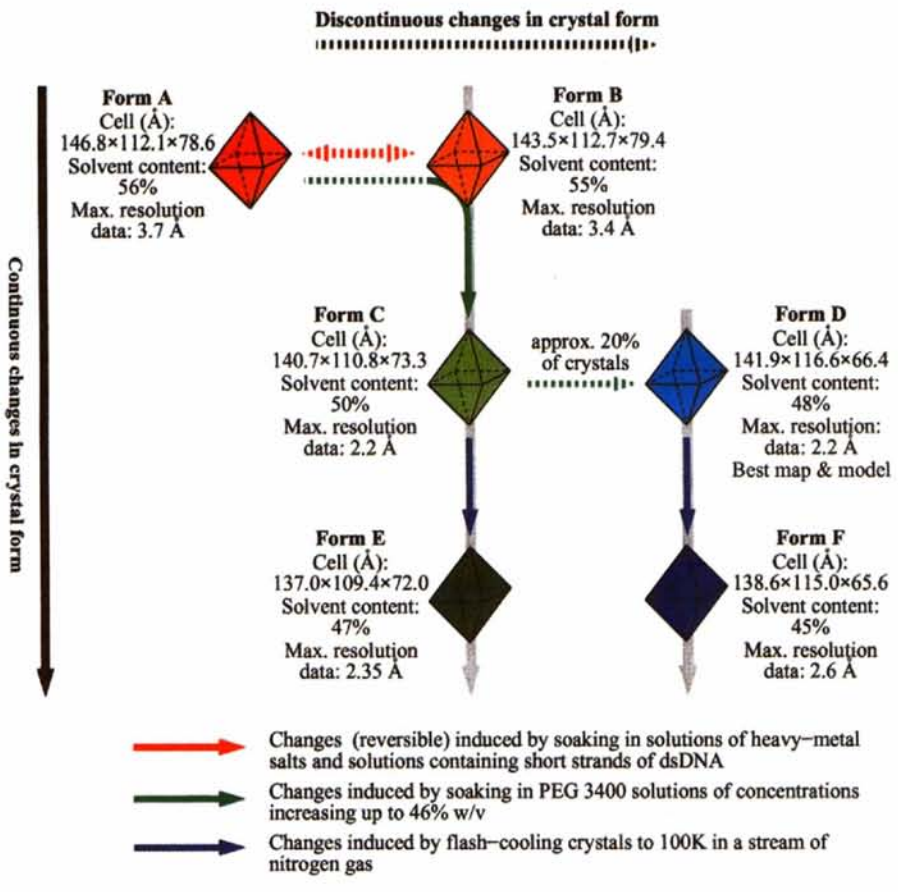

Fig. 6. Schematic diagram of the dehydration pathways followed by these RT crystals. Continuous conversions between cell forms follow vertical paths, discontinuous jumps are connected horizontally. Arrows show the effects of different treatments of the crystals using the colour scheme of Fig. 5(*a*). The average unit-cell dimensions, the solvent content and the highest resolution to which diffraction data have been observed are shown for each cell form.

Table 1. *Best models of RT–NNI complexes in each cell form*

The bound inhibitor and selected refinement statistics for models of RT–NNI complexes refined against the highest resolution data sets obtained for each cell form.

Cell form	Inhibitor bound to RT	Resolution range (Å)	Number of reflections	*R* factor†	PDB code	Reference
A	1051U91	12–3.7	11087	0.344	‡	(See text)
B	1051U91	12–3.4	15557	0.346	‡	(See text)
C	1051U91	25–2.2	48144	0.214	1RTH	Ren, Esnouf, Garman *et al.* (1995)
D	Nevirapine	25–2.2	49347	0.186	1VRT	Ren, Esnouf, Garman *et al.* (1995)
E	MKC-442	25–2.55	34612	0.197	1RT1	Hopkins *et al.* (1996)
F	9-Cl-TIBO	25–2.6	27108	0.224	1REV	Ren, Esnouf, Hopkins *et al.* (1995)

† R factor = $\sum |F_{obs} - F_{calc}| / \sum F_{obs}$. ‡ See supplementary information, which provides the data necessary to generate these coordinates.

rotating-anode sources (and an experimental setup similar to that described above), but, due to the weak diffraction from these crystals, the majority of data were collected at synchrotron sources. The experimental setups have been described previously (Ren, Esnouf, Garman *et al.*, 1995; Esnouf *et al.*, 1995, 1997; Ren, Esnouf, Hopkins *et al.*, 1995; Hopkins *et al.*, 1996). Nearly 1000 crystals have been examined, but most do not show high-resolution diffraction from single lattices. To date, we have obtained 78 data sets having high-resolution limits better than 5 Å. Most data sets were processed using *DENZO*, although *MARXDS* has also been used.

3.2. *Results*

Analysis of the unit cells for the 78 data sets shows that most can be classified into one of the six cell forms *A*–*F* (Fig. 5*a*). The outliers suggest that under appropriate conditions two further cell forms, *G* ($a \simeq 130.5$, $b \simeq 115$ and $c \simeq 75.5$ Å) and *H* ($a \simeq 140$, $b \simeq 116$ and $c \simeq 69$ Å), can be obtained, although we have no evidence that these are well defined forms capable of diffraction to high resolution. Crystals soaked in solutions of heavy-atom salts or DNA oligomers (red circles in Fig. 5*a*) adopt cell forms *A* or *B* (with one exception) and the conversion between these cell forms is reversible (data not shown). Dehydration protocols using concentrated PEG solutions (green circles) result mainly in cell form *C*, with about 20% of crystals undergoing further dehydration to form *D*. This treatment has occasionally resulted in crystals of cell forms *E*, *G* and *H*. When dehydrated crystals are cooled rapidly to 100 K (blue circles), crystals in cell forms *E* and *F* result. Since (i) our initial observation of dehydration showed continuous shrinkage from cell form *D* to form *F*, (ii) the relative change in unit-cell dimensions between forms *C* and *E* is similar to that between forms *D* and *F* and (iii) the relative frequency of observation of forms *E* and *F* is similar to that for forms *C* and *D*, we infer that flash-cooling crystals of cell form *C* results in form *E*.

We can detect no correlation between the unit-cell dimensions and the presence or nature of a bound inhibitor, yet within each cell form significant unit-cell variation is observed. Some of this variation arises from the experimental conditions, while the unit cell cannot be accurately defined for incomplete data sets from poorly diffracting crystals. The variation may also be linked to how well each crystal has re-annealed, implying that there are ideal unit cells for each cell form and that crystals having unit-cell dimensions close to these ideals might show diffraction to higher resolution. For the two most commonly observed cell forms there is some evidence for an ideal unit cell in one case (form *C*), while for form *E* the better diffracting crystals are associated with a restricted range of unit-cell dimensions rather than a single ideal (Fig. 5*b*).

From our observations on these orthorhombic crystals of HIV-1 RT we can construct a scheme for the dehydration pathways they follow (Fig. 6). Although it is a discrete jump, conversion between cell forms *A* and *B* is reversible. Once a crystal has converted to cell form *B*, steady dehydration produces form *C*, from which a further minor change results in form *E*. Approximately 20% of cell form *C* crystals change to a different dehydration pathway. This jump, to cell form *D*, results in a significant simultaneous increase in the *b* dimension and decrease in the *c* dimension. Having reached this pathway, smooth dehydration to cell form *F* is possible and beyond that further dehydration appears to cause a breakdown in the crystal order.

4. Structures from different cell forms

4.1. *Structure determinations*

The overall structure of HIV-1 RT does not depend significantly upon the presence or nature of a bound NNI, but for this analysis we restrict ourselves to the highest resolution structures for RT–NNI complexes in each cell form (Table 1), ignoring the structure for unliganded RT in cell form *E* determined to 2.35 Å resolution (Esnouf *et al.*, 1995).

Table 2. *Comparison of atomic B factors between domains in each cell form*

Mean *B* factors for the Cα atoms in each cell form and the mean *B* factors for Cα atoms in each domain relative to the mean for each cell form. To reduce the effect of disordered regions of the protein, only residues which are present in all cell forms (922 residues) are used for the calculations. The number of disordered residues for each domain in each cell form is shown in parentheses.

	Ratio of mean *B* factor to mean *B* factor for whole molecule					
RT domain	Form *A*	Form *B*	Form *C*	Form *D*	Form *E*	Form *F*
p66 fingers	2.2	2.1	1.5	1.4 (3)	1.6	1.3 (3)
p66 palm	0.7	0.8	0.9	0.9	1.1	1.0
p66 thumb	1.1	1.3	1.2	1.0	1.0	0.9
p66 connection	0.4	0.7	0.8	0.8	0.7	0.9
p66 RNase H	0.9 (17)	0.8 (17)	0.7 (17)	1.0 (32)	0.6 (21)	0.9 (32)
p51 fingers	0.6 (3)	0.6 (3)	0.8 (3)	0.7 (4)	1.0 (5)	0.8 (1)
p51 palm	1.5 (15)	1.2 (15)	1.3 (15)	1.4 (18)	1.5 (20)	1.2 (15)
p51 thumb	1.0	0.9	0.9	1.0	0.9	1.0
p51 connection	0.7 (2)	0.7 (2)	0.8 (2)	0.9 (17)	0.6	0.9 (12)
Mean *B* factor for whole molecule (Å^2)	69	76	57	48	32	63
Standard deviation of *B* factor (Å^2)	42	41	27	24	18	23

Table 3. *Crystal contacts between heterodimers in each cell form*

The number of residues involved in crystal contacts (defined as intermolecular atomic distances less than 4 Å) in each cell form. Contacts are named in pairs: the '*a*' contact in one RT molecule interacts with the '*b*' contact in the other molecule. The Fig. 7 colours are similarly paired into pale and dark shades of the same hue. The symmetry operations relate each contact molecule to an untransformed molecule. The total number of residues involved in crystal contacts takes into account those residues which make multiple intermolecular contacts.

		Number of residues contributing to each crystal contact						
Contact name	Colour used in Fig. 7	Form *A*	Form *B*	Form *C*	Form *D*	Form *E*	Form *F*	Symmetry operation
1*a*	pale cyan	1	4	6	3	8	3	$(\frac{1}{2}-x, -1-y, \frac{1}{2}+z)$
1*b*	dark cyan	1	2	5	3	6	3	$(\frac{1}{2}-x, -1-y, -\frac{1}{2}+z)$
2*a*	pale magenta	1	1	7	9	9	11	$(x, y, 1+z)$
2*b*	dark magenta	1	1	6	11	10	12	$(x, y, -1+z)$
3*a*	pale yellow	9	7	13	15	12	16	$(-x, -\frac{1}{2}+y, \frac{1}{2}-z)$
3*b*	dark yellow	14	11	19	13	19	15	$(-x, \frac{1}{2}+y, \frac{1}{2}-z)$
4*a*	pale red	22	25	43	33	44	29	$(\frac{1}{2}-x, -y, \frac{1}{2}+z)$
4*b*	dark red	22	26	40	34	45	31	$(\frac{1}{2}-x, -y, -\frac{1}{2}+z)$
5*a*	pale green	–	–	1	4	3	3	$(-x, -\frac{1}{2}+y, \frac{3}{2}-z)$
5*b*	dark green	–	–	1	3	3	3	$(-x, \frac{1}{2}+y, \frac{3}{2}-z)$
6*a*	pale blue	–	–	2	–	7	3	$(-\frac{1}{2}+x, -\frac{1}{2}-y, 1-z)$
6*b*	dark blue	–	–	2	–	6	2	$(\frac{1}{2}+x, -\frac{1}{2}-y, 1-z)$
7*a*	pale orange	–	–	–	3	1	3	$(-\frac{1}{2}+x, -\frac{1}{2}-y, -z)$
7*b*	dark orange	–	–	–	4	3	4	$(\frac{1}{2}+x, -\frac{1}{2}-y, -z)$
Totals	–	71	77	143	135	172	138	–

Data collection and structure refinement for the complexes in cell forms *C* to *F* have been described previously (Ren, Esnouf, Garman *et al.*, 1995; Ren, Esnouf, Hopkins *et al.*, 1995; Hopkins *et al.*, 1996). Details of the collection and processing of data from cell forms *A* and *B* will be described elsewhere. Structures for RT in cell forms *A* and *B* were obtained from the refined form *C* model using the following protocol. The model was superposed onto a molecular-replacement solution obtained with a less accurate search structure and then refined against the data: first as a single rigid body, then as two rigid subunits (p66 and p51) and finally as nine rigid domains. All refinement was performed using *X-PLOR* (Brünger, 1992). Anisotropic scaling of the data against the intermediate model was followed by *B*-factor refinement with the atoms grouped into the nine domains. The positions of the domains were confirmed manually by calculating a series of domain-omit maps. Throughout the following discussion it should be borne in mind that the data quality limits the accuracy of the models for cell forms *A* and *B*.

4.2. *Atomic B factors*

Our original observation of shrinkage indicated that the overall *B* factor was reduced by dehydration (Fig. 4*d*) and this is confirmed by the mean *B* factors for Cα atoms in the different cell forms (Table 2). The relative atomic *B* factors for each domain (Table 2) show that the p66 fingers is the most consistently mobile, followed by the p51 palm. The p66 thumb is very mobile in the less dehydrated cell forms, and close crystal contacts with a symmetry-related p51 thumb (see below) result in similar atomic *B* factors for both thumb domains. The lowest *B* factors are associated with the p66 connection and the p51 fingers and connection domains, which form

the 'core' of the heterodimer (Fig. 1). The mobility of the RNase H is low for cell forms *C* and *E* and high for cell forms *D* and *F*. As with the p66 thumb, crystal contacts play a major role in determining the mobility of this domain. However, the C-terminal part of the RNase H is disordered in all cell forms and so the relative *B* factors do not reflect the true mobility of the whole domain.

The variation in the atomic *B* factors over the heterodimer reduces as the degree of dehydration increases (Table 2). This can be seen as part of a gradual change from motions associated with a free molecule to more restricted crystal modes. Thus, the *B* factors for RT in cell form *A* may give the best guide to the internal flexibility that is required for RT to fulfill its enzymatic function.

4.3. *Crystal contacts*

All crystal contacts (intermolecular atomic distances less than 4 Å) were calculated for each cell form. These contacts were sorted by residue and symmetry operation to show the distribution of crystal contacts over the surface of RT (Table 3; Fig. 7). Unsurprisingly, there are many more crystal contacts in the dehydrated cell forms, with 172 residues (out of 1000) making contacts with other copies of RT in cell form *E*. For cell forms *D* and *F*, part of the RNase H domain and the C terminus of the p51 subunit are disordered; exclusion of these residues from the models is the main reason for the apparent reduction in the number of contact residues for these cell forms. The increased disorder reflects the loss of intermolecular hydrogen-bonding opportunities between the two domains (see below).

In cell forms *A* and *B*, each RT molecule makes crystal contacts with eight other molecules, four of these contacts being relatively minor (Fig. 7). The significant contacts fall into two pairs (3*a* and 3*b*, coloured pale and dark yellow; 4*a* and 4*b*, coloured pale and dark red). The 3*a* and 3*b* contacts are between the RNase H domain of one molecule and the p66 palm of another. The 4*a* and 4*b* pair result in intermolecular contacts between the

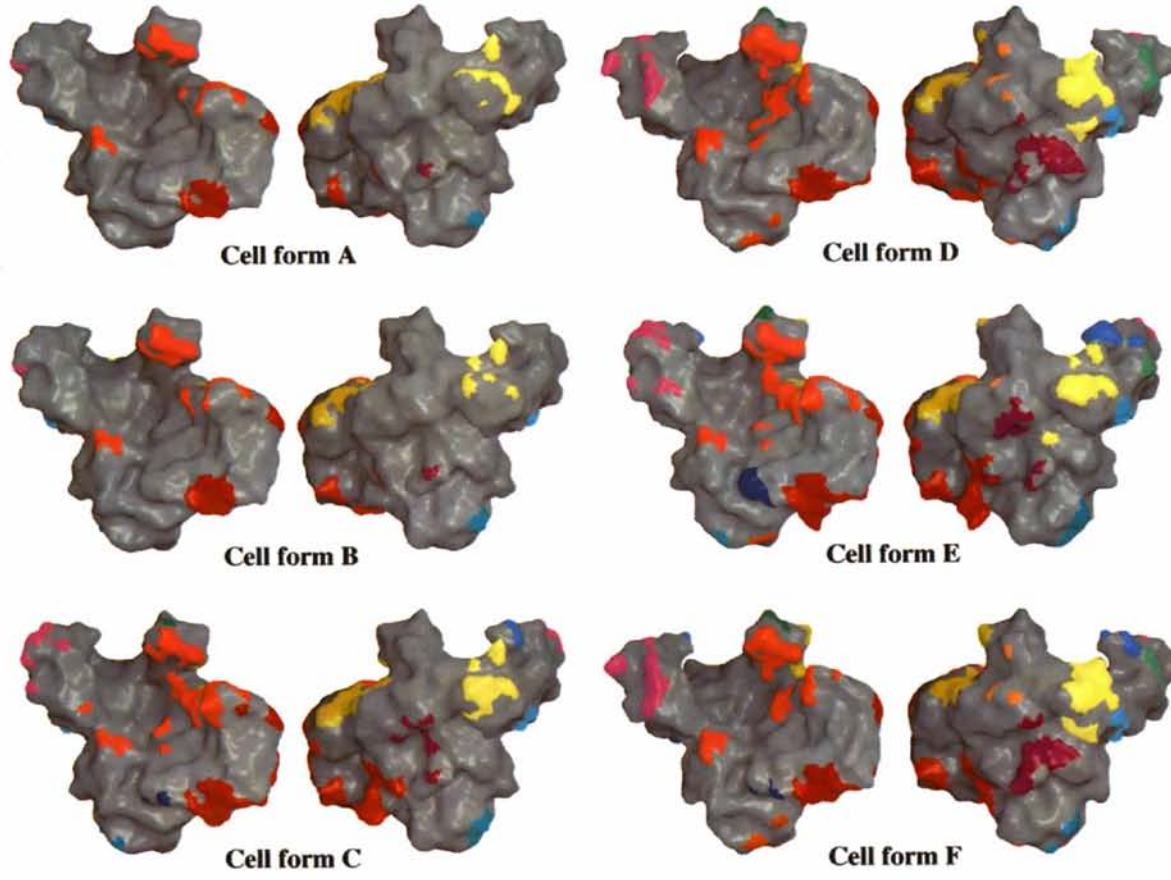

Fig. 7. 'Front' and 'back' views showing areas of crystal contacts on the surface of the RT heterodimer for each cell form. Grey areas of the surface make no crystal contacts. The coloured contact patches can be related to specific crystallographic symmetry operations using Table 3, the pale shade of each hue contacting the corresponding dark shade in a symmetry-related molecule, and *vice versa*.

Table 4. *Intermolecular hydrogen bonding between heterodimers in each cell form*

Intermolecular hydrogen bonds are listed either if they are common stabilizing features of at least three different cell forms or if they involve pairs of main-chain atoms. Donor atoms are chosen to be on the central molecule and the contact molecules are defined in Table 3. For each hydrogen bond there is a second hydrogen bond with the acceptor atom on the central molecule and the donor atom on the complementary contact molecule (Table 3). Due to the data quality for cell forms *A* and *B*, hydrogen bonds for these cell forms cannot be accurately identified.

Donor atom				Acceptor atom			Cell forms					
Subunit	Residue	Atom	Contact	Subunit	Residue	Atom						
Hydrogen bonds which are common to at least three cell forms												
p66	199 R	Nε	3*a*	p66	523 E	Oε1	*A*	–	*C*	*D*	*E*	–
p66	199 R	Nη2	3*a*	p66	511 D	O	–	*B*	*C*	*D*	*E*	*F*
p66	199 R	Nη2	3*a*	p66	523 E	Oε2	*A*	–	*C*	–	*E*	–
p66	258 Q	Nε2	4*a*	p51	273 G	O	–	–	–	*D*	*E*	*F*
p66	258 Q	Nε2	4*a*	p51	305 E	Oε1	*A*	*B*	*C*	*D*	*E*	*F*
p66	284 R	Nε	4*a*	p51	308 E	Oε2	*A*	*B*	*C*	*D*	*E*	–
p66	285 G	N	4*a*	p51	309 I	O	*A*	*B*	*C*	*D*	*E*	*F*
p66	476 K	Nζ	4*a*	p51	433 P	O	*A*	*B*	*C*	–	*E*	–
p66	527 K	Nζ	3*b*	p66	224 E	Oε2		–	*C*	*D*	*E*	*F*
p51	315 H	N	4*b*	p66	285 G	O	–	*B*	*C*	*D*	*E*	*F*
p51	315 H	Nδ1	4*b*	p66	291 E	Oε1	*A*	*B*	*C*	*D*	*E*	*F*
p51	437 A	N	4*b*	p66	470 T	O	*A*	–	*C*	–	*E*	–
Hydrogen bonds involving pairs of main-chain atoms												
p66	285 G	N	4*a*	p51	309 I	O	*A*	*B*	*C*	*D*	*E*	*F*
p66	464 Q	N	4*b*	p51	15 G	O	–	–	*C*	–	*E*	–
p66	468 T	N	4*a*	p51	439 T	O	–	–	–	–	*E*	–
p66	468 T	N	4*a*	p51	440 F	O	–	–	–	–	*E*	–
p66	470 T	N	4*a*	p51	437 A	O	–	–	*C*	–	–	–
p5	17 D	N	4*a*	p66	461 R	O	–	–	–	*D*	–	*F*
p51	17 D	N	4*a*	p66	462 G	O	–	–	*C*	–	–	–
p51	315 H	N	4*b*	p66	285 G	O	–	*B*	*C*	*D*	*E*	*F*
p51	437 A	N	4*b*	p66	470 T	O	*A*	–	*C*	–	*E*	–
p51	439 T	N	4*b*	p66	468 T	O	–	–	–	–	*E*	–

p66 and p51 thumb domains and also between the RNase H and p51 fingers domains. These significant crystal contacts are well preserved during the dehydration process and hydrogen bonding is an important stabilizing element. The contact between thumb domains is the major determinant of the unit-cell *c* dimension, the most variable of the unit-cell dimensions.

With increasing dehydration, up to six more molecules come into contact with each RT heterodimer. The 3*a*, 3*b*, 4*a* and 4*b* contacts remain the most significant, but one further pair (2*a* and 2*b*, coloured pale and dark magenta in Fig. 7) becomes increasingly significant. This last interaction (between p66 fingers and p51 fingers domains) forms a contact surface approximately perpendicular to the *z* axis, limiting further shrinkage of the unit-cell *c* dimension. For cell forms with the smallest *c* dimensions (forms *D* and *F*) this interaction is nearly as significant as the 3*a* and 3*b* interactions, suggesting no opportunity for further shrinkage.

4.4. *Hydrogen bonding*

The program *HBPLUS* (McDonald & Thornton, 1994) was used to identify intermolecular hydrogen bonds for each cell form. As with crystal contacts, the number of intermolecular hydrogen bonds increases with the degree of dehydration, although the trend is again complicated by disorder in the RNase H domain and at the C terminus of the p51 subunit (cell form *A*, 24 bonds; form *B*, 26 bonds; form *C*, 46 bonds; form *D*, 48 bonds; form *E*, 72 bonds; form *F*, 48 bonds). The RNase H is the domain most involved in hydrogen bonding and, perhaps reflecting their mobility, the exposed p66 fingers and p51 palm domains are relatively uninvolved.

Many intermolecular hydrogen bonds are between mobile side-chain atoms or are not preserved during dehydration. Those which are preserved for multiple cell forms (Table 4) all involve the 3*a*, 3*b*, 4*a* and 4*b* contact molecules and involve either the exposed tip of the p66 thumb or the RNase H. Intermolecular hydrogen bonding between the p66 thumb and the p51 thumb of the 4*a* contact molecule (Fig. 8*a*) appears to be more significant in positioning the p66 thumb than the peptide links between the p66 thumb and the rest of the p66 subunit (see below). The tightness of the interaction is reflected in the correlation of atomic *B* factors for these domains (Table 2). All completely conserved hydrogen bonds arise from this interaction.

Intermolecular hydrogen bonds between pairs of main-chain atoms (Table 4) involve the 4*a* and 4*b* contact molecules. Two bonds link the p66 and p51 thumb domains (Fig. 8*a*); the rest result from contacts involving the RNase H domain in dehydrated cell forms. Hydrogen bonds link the p51 fingers to the N-terminal end of strand β19 from the RNase H five-stranded β-sheet (Fig. 8*b*). For cell form *C* there are two hydrogen

bonds, but for cell forms *D* and *F* relative rotation of molecules causes the loss of one hydrogen bond and the other 'ratchets' one residue back along the RNase H main chain to the hairpin between strands β18 and β19. For cell forms *C* and *E*, the C-terminal end of strand β19 forms intermolecular hydrogen bonds to the C terminus of the p51 subunit (part of the p51 connection domain). In cell form *E*, three hydrogen bonds with the last four residues of the p51 subunit (residues 437–440) create a β-strand corresponding to strand β17 in the p66 subunit. In cell form *D*, an 8 Å movement of β19 relative to the C terminus of the p51 subunit results in the space previously occupied by the terminal residues being filled by helix α15 from the RNase H domain. Intermolecular hydrogen bonding to strand β19 is observed for structures determined by other groups, where twofold crystallographic symmetry leads to interactions between pairs of β19 strands (Jäger *et al.*, 1994).

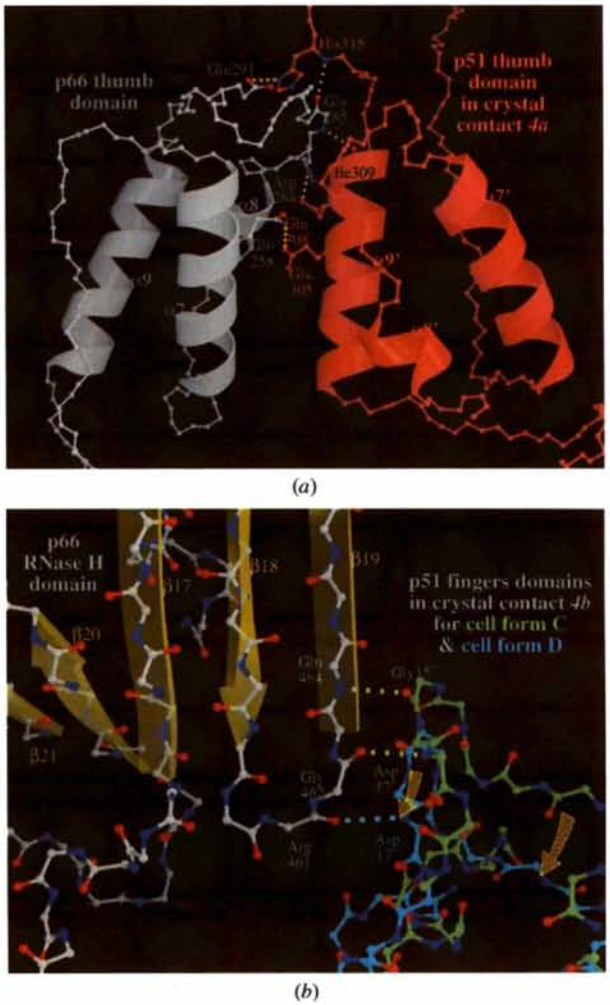

Fig. 8. Intermolecular hydrogen bonding. (*a*) Conserved hydrogen bonds between the tip of the p66 thumb (white) and the p51 thumb of the 4*a* contact molecule (pink). The protein secondary structure is shown with ball-and-stick models for the residues involved in hydrogen bonds. Hydrogen bonds are shown by dashed yellow lines. (*b*) Hydrogen bonds involving main-chain atoms between the RNase H domain (atom-coloured ball-and-stick main-chain trace with the β-sheet structure highlighted by yellow arrows) and the p51 fingers of the 4*b* contact molecule for cell forms *C* and *D*. The two fingers domains are shown by atom-coloured ball-and-stick main-chain traces (cell form *C*, green carbon atoms; cell form *D*, light-blue C atoms) and hydrogen bonds are shown by dashed lines of the appropriate colour. The repositioning of the p51 fingers relative to the RNase H domain on dehydration is shown by yellow arrows.

5. Changes in structure during dehydration

5.1. *Reorientation of the heterodimer*

The reduction in unit-cell volume on dehydration is achieved by a combination of reorientation of the heterodimer as a whole and movements of individual domains. To analyse this, the models for cell forms *B* to *F* were superposed onto the model for cell form *A* using an automated procedure to identify the best conserved structure on which to base each superposition (Table 5). In line with the results of the *B*-factor analysis (Table 2) and previous reports (Rodgers *et al.*, 1995), the two connection domains and the p51 fingers were the main contributors to the structural core of the molecule. Furthermore, part of the p66 palm is largely invariant between cell forms *A*, *B*, *C* and *E* and, conversely, part of the p51 palm is invariant between cell forms *A*, *D* and *F*.

The rotations relating RT in each cell form are most clearly shown by considering the disposition of the unit cells with respect to the superposed structures (Fig. 9). The cell form groups (columns in Fig. 6) relate to the orientation of the heterodimer: cell forms *B*, *C* and *E* show a similar 3–4° rotation of RT with respect to cell form *A*, while cell forms *D* and *F* show a 9–10° rotation with respect to form *A*, but around a different axis.

5.2. *Translation of the p66 thumb domain*

The cell form changes $A \rightarrow B$, $C \rightarrow E$ and $D \rightarrow F$ primarily involve similar 3.5 Å decreases in the *a* dimensions of the unit cells, suggesting that the cell forms are grouped in pairs. However, the initial observation of dehydration (Fig. 3) and the orientation of the heterodimer (Table 5; Fig. 9) group cell form *B* with forms *C* and *E*. The unexpectedly large unit-cell *c* dimension for cell form *B* (Fig. 6) is a result of the position of the p66 thumb, whose contact with the p51 thumb of the 4*a* contact molecule is the principal determinant of this dimension. Conversion from cell form *A* to form *B* involves a rotation of the heterodimer but little internal change, possibly explaining the reversibility of this conversion. Under pressure of further dehydration, the p66 thumb translates, roughly along the *z* axis, by 3.5 Å (mean change in Cα position for residues 284–291), owing to a slight movement of the sheet connecting this domain to the p66 palm (strands β9, β10 and β11). Crystal symmetry results in a

Table 5. *Superposition of RT models for cell forms B–F onto the cell form A model*

Superpositions were performed using the program *SHP* (Stuart *et al.*, 1979) with stringent criteria for defining conserved structure. These criteria identified about one third of each pair of structures which could be superposed with a root-mean-square (r.m.s.) difference in Cα positions of approximately 0.5 Å. The table shows the number of residues from each domain used in the superposition, the r.m.s. difference and the rotation for each pair of models. The small r.m.s. errors for cell forms *B* and *C* partly reflect the restricted refinement procedures which had to be used for cell forms *A* and *B*.

		Number of residues used for superposition				
Structural unit	Total number of residues	Form *B* on *A*	Form *C* on *A*	Form *D* on *A*	Form *E* on *A*	Form *F* on *A*
p66 fingers	114	7	1	28	15	9
p66 palm	128	46	60	12	70	20
p66 thumb	69	13	0	0	27	0
p66 connection	114	85	58	72	46	79
p66 RNase H	135	20	0	9	0	12
p51 fingers	114	81	92	92	94	91
p51 palm	128	0	31	53	27	54
p51 thumb	69	26	8	13	0	12
p51 connection	129	89	119	73	70	74
Whole molecule	1000	367	369	352	349	351
R.m.s. difference in Cα atom positions (Å)	—	0.21	0.36	0.42	0.54	0.47
Rotation of RT (°)	—	3.6	3.0	9.4	4.1	9.7

c-dimension shrinkage of twice this translation. Equivalent slices through RT in each cell form (Fig. 10) show that near the p66 connection, the molecular surface for form *B* (orange) is similar to those for forms *C* (lime green) and *E* (dark green), whilst at the tip of the p66 thumb the form *B* molecular surface matches closely that for form *A*.

5.3. *Reorientation of individual domains*

An analysis of individual domain orientations for each cell form has to consider orientations both with respect to the crystal unit cell and with respect to the rest of the heterodimer [*i.e.* by comparing the superposed structures (Table 5)]. Furthermore, the orientations can be compared with the initial cell form (form *A*) and with the preceding cell form on the dehydration pathway. Thus, for each domain in each of the cell forms *B* to *F*, four rotations are calculated, presented as 2 × 2 arrays in Table 6. These data can be interpreted using the following rules.

(i) If a number on the bottom of an array is large then the preceding step in the dehydration pathway caused a large change in domain orientation.

(ii) If a number on the left of an array is bigger than its counterpart on the right then the domain reorientation is largely accounted for by reorientation of the heterodimer.

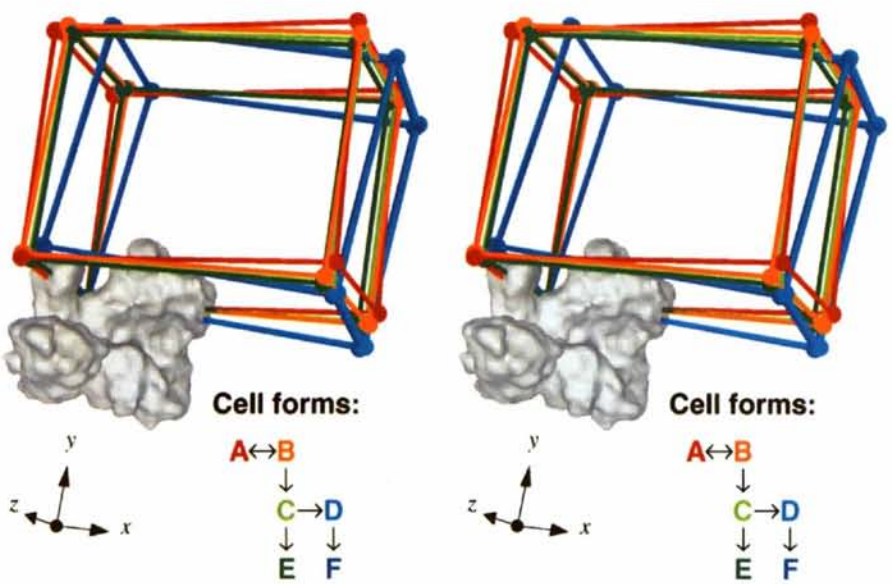

Fig. 9. Stereo diagram showing the reorientation of the RT heterodimer on dehydration relative to the unit cell. To highlight small changes in orientation, the models for RT in each cell form were superposed onto the cell form *A* model (shown by the grey molecular surface) and just the relative unit-cell edges are shown for each cell form using the colour scheme of Fig. 6, summarized at the bottom of the figure.

Table 6. *Changes in orientations of individual domains during dehydration*

For the models for cell forms *B*–*F*, the rotation of each domain is calculated, both with respect to the cell form *A* model (top row) and with respect to the model for the preceding cell form on the dehydration pathway (bottom row). Rotations are calculated both between the pairs of crystal structures (left column) and between the pairs of crystal structures after superposition of their core structures (Table 5) onto the cell form *A* model (right column).

	Relative rotations of individual domains of RT (°)																			
Cell forms	p66 fingers		p66 palm		p66 thumb		p66 conn.		p66 RNase H		p51 fingers		p51 palm		p51 thumb		p51 conn.	
B on *A*	2	2	3	2	4	2	4	1	4	2	4	1	3	1	4	2	4	1
B on *A*	2	2	3	2	4	2	4	1	4	2	4	1	3	1	4	2	4	1
C on *A*	7	5	3	3	3	0	3	2	5	3	5	2	3	2	3	2	3	1
C on *B*	5	4	3	2	3	3	2	1	4	4	2	1	1	1	3	3	2	2
D on *A*	10	4	5	5	6	12	10	1	11	6	10	1	9	3	5	6	11	2
D on *C*	13	8	8	4	5	12	7	1	9	7	8	1	7	2	4	6	9	2
E on *A*	7	4	4	1	3	3	4	3	6	4	5	1	3	5	3	3	3	3
E on *C*	1	1	1	2	2	3	1	1	1	1	2	1	3	4	1	1	0	2
F on *A*	11	4	7	4	5	11	11	1	12	6	10	1	9	4	6	5	12	2
F on *D*	2	2	2	2	1	2	1	1	2	2	1	0	1	1	1	1	1	0

(iii) If a number on the right of an array is bigger than its counterpart on the left then the domain preserves its orientation with respect to the unit cell despite a reorientation of the heterodimer as a whole.

For the abrupt change from cell form *A* to form *B* there is a small rotation of the heterodimer as a whole, but little internal structural change. On further dehydration to cell form *C*, the translation of the p66 thumb (see above) is accompanied by significant (5°) reorientations of the p66 fingers and RNase H domains. The tips of these domains are brought closer to the p66 thumb in a motion akin to a slight closing of the p66 'hand'. For dehydration to cell form *E*; there is little change to the form *C* conformation. The biggest change, to the p51 palm, results from the development of a crystal contact with the 7*b* contact molecule (Table 3; dark orange in Fig. 7).

About 20% of cell form *C* crystals undergo a dramatic change to form *D*. The overall molecular rotation associated with this conversion (Table 5) is also accompanied by large rotations of individual domains, resulting in a 'squashing' of the top part of the heterodimer (Fig. 7). The p66 thumb orientation is little altered by the change while the p66 fingers and RNase H domains rotate by more than the body of the molecule (Table 6) producing a significant opening of the p66 'hand'. Strong intermolecular hydrogen bonding between the p66 and p51 thumb domains restricts the change in orientation of the p51 thumb. Cell form *F* has an open-handed domain conformation virtually indistinguishable from that of form *D*.

Over the six cell forms, the p66 fingers and RNase H domains show the greatest variability in position. The p66 thumb is only connected to the rest of the heterodimer by a flexible link and it interacts more strongly with the p51 thumb of the 4*a* contact molecule. As these symmetry-related molecules rotate in opposite directions during dehydration, the two thumb domains retain their orientations to a greater extent than other domains (Table 6), allowing major changes in crystal packing to occur without destroying the crystal lattice. Thus, the flexibility of the p66 thumb linkage to the rest of the p66 subunit and the conserved inter-thumb intermolecular hydrogen bonds (Fig. 8*a*) appear to be the crucial factors underlying the remarkable plasticity exhibited by these HIV-1 RT crystals.

6. Discussion

Changes in the hydration state of protein crystals and their potential importance for structure solution have been recognized since the very early days of protein crystallography (Bernal *et al.*, 1938). Studies on the

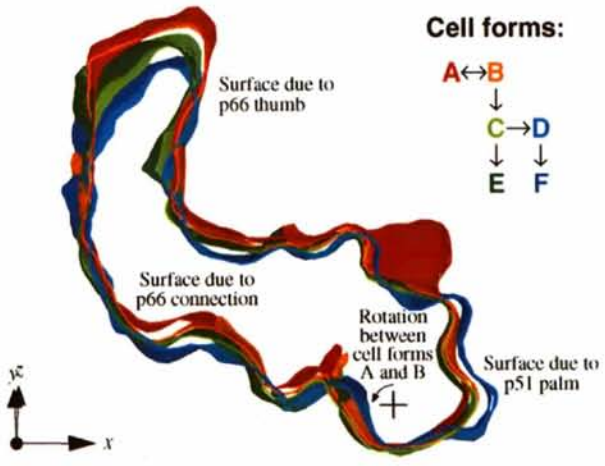

Fig. 10. A slice through the RT heterodimer for each cell form. The surfaces (using the colour scheme of Fig. 6, summarized on the right of the figure) show a 3.5 Å wide slice from the p66 thumb (top) to the p51 palm (right) running roughly along the interfaces between the p66 palm and connection domains, and the p51 fingers and connection domains (Fig. 1).

variable composition of horse met-hemoglobin crystals (Perutz, 1946) led to the development of apparatus for controlling the humidity around crystals (Huxley & Kendrew, 1953) [apparatus which was developed further and miniaturized (Pickford *et al.*, 1993) in the search for well ordered HIV-1 RT crystals]. From these early studies to the recent example of NF–κB P52 homodimer–DNA complex crystals (Cramer & Müller, 1997), the few descriptions of crystal pleomorphism have focused on the effects of discrete changes in unit-cell dimensions on the lattice order. Our studies have not only allowed a description of crystals showing an apparent continuum of possible unit-cell dimensions, but also the analysis of the structures for different dehydration states is the first attempt to interpret this phenomenon on a detailed molecular basis.

Subjecting these HIV-1 RT crystals to dehydration pressure can result in the removal of one third of the solvent from the crystal accompanied by an increase in the order of the crystal lattice, leading to the potential for diffraction to much higher resolution (Fig. 6). On the molecular level, reorientation of the heterodimer as a whole divides the six cell forms into three groups which are interconnected by discontinuous changes (*i.e.* form *A*; forms *B*, *C* and *E*; forms *D* and *F*). Reorientations of domains, particularly the opening and closing of the p66 'hand' allow for the continuous changes within each cell form group. However, the most important crystal contact (between the p66 thumb and the p51 thumb of the 4*a* contact molecule) is preserved across all cell forms, and this is mainly due to the flexible linkage between the p66 thumb and the rest of the p66 subunit.

A slice through crystals of the cell forms lying at either extreme of the hydration spectrum (cell forms *A* and *F*) shows how a 6° rotation of RT with respect to the *z* axis leads to a 13 Å reduction in the *c* dimension of the unit cell (Fig. 11). The rotation is prevented from going further by the crystal contact developed between the p66 and p51 fingers domains (marked *B* in Fig. 11; coloured pale and dark magenta in Fig. 7). This contact also appears to force open the p66 'hand' in cell forms *D* and *F*. Fortunately, the crystal contacts do not directly affect the polymerase active site in any of our cell forms (marked by an *A* in Fig. 11) although the space between heterodimers is not sufficient to allow binding of short DNA oligomers.

RT may need to be very flexible in order both to bind the template initially and to hold on to the template and primer stands during the translocation step of processive transcription, the impulse for the latter presumably arising from the cleavage of the pyrophosphate group from each added nucleoside. The template and primer strands primarily interact with the p66 fingers and thumb domains (Jacobo-Molina *et al.*, 1993) and so the observed flexibility of these domains is unsurprising. The extent of this flexibility is shown by the structures

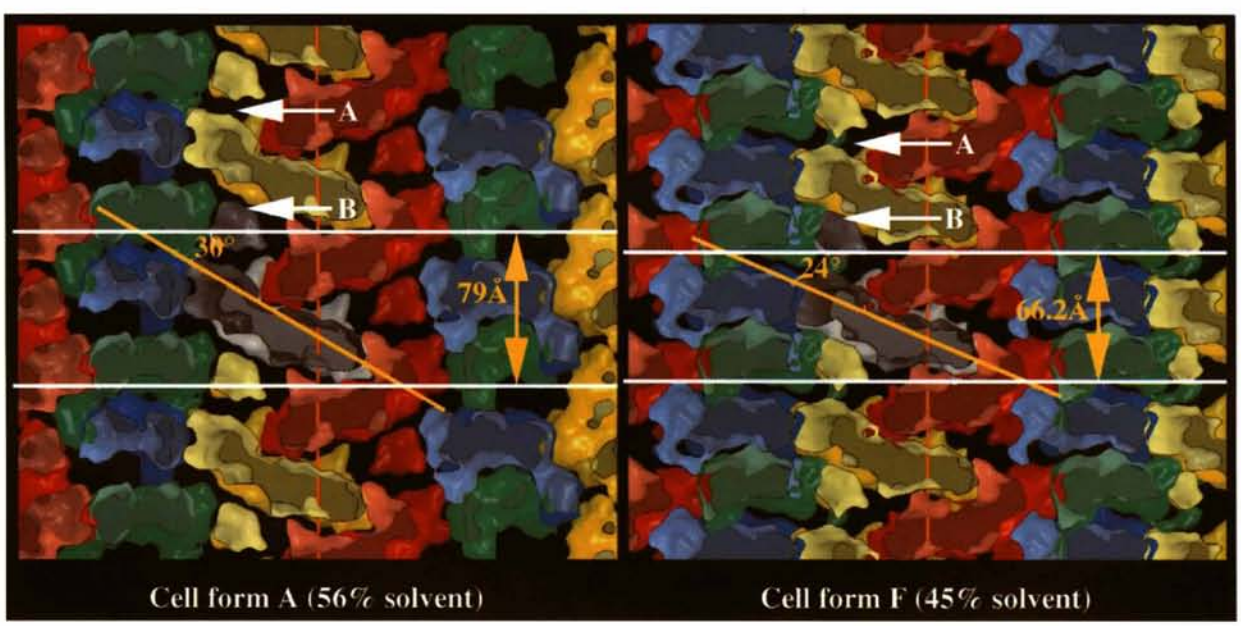

Fig. 11. Slices through the crystal structure in cell forms *A* (left) and *F* (right), showing the extremes of the dehydration pathway. The slices are 25 Å thick and are centred on the crystallographic twofold screw axis parallel to the *z* axis (orange line). To compensate for rotation of the heterodimer, the view for cell form *F* is rotated by 11° about this axis relative to that for form *A*. Colours indicate the relation of each molecule to the white molecule using one of the twofold screw axes followed by unit-cell translations (yellow, no rotation; green, axis parallel to *x*; blue, axis parallel to *y*; red, axis parallel to *z*). Lines show how the unit-cell *c* dimension is related to the orientation of the body of the heterodimer. The sites marked *A* indicate the binding cleft around the polymerase active site. The sites marked *B* show how the p66 fingers of the white molecule (below) and the p51 fingers of the yellow molecule (above) come together on dehydration, thus preventing further shrinkage.

analysed here, as well as by structures from other laboratories [where the p66 thumb has been observed to be folded down against the rest of the molecule (Rodgers *et al.*, 1995; Hsiou *et al.*, 1996)]. We conclude that (especially for the p66 thumb) crystal packing is the major factor determining the orientation of these domains, at least within the range we observe. Whilst data from dehydrated crystals allow more precise definition of the atomic structure for each domain, the lower resolution structures from crystals with higher solvent content better reflect the 'free-molecule' mobility of these domains.

Unfortunately, since dehydration does not always lead to a well defined unit-cell end-point, problems due to multiple lattices and data anisotropy are common. Thus, it is crucial to record complete data sets from the small proportion of crystals which show good diffraction from a single lattice. However, variability in unit-cell dimensions can be used to advantage, and was crucial to our original high-resolution structure determination (Ren, Esnouf, Garman *et al.*, 1995), since it provided suitable data for inter-crystal-form real-space electron-density averaging. Changes in unit-cell dimensions during soaking experiments or flash-cooling are neither unique to this crystal form nor to RT. Provided that the changes result in data sets having some phase independence, then cross-crystal form averaging may be a more generally applicable technique of structure refinement (Rao *et al.*, 1995; Esnouf *et al.*, 1996).

We wish to thank the staff at the EMBL, DESY, Hamburg, the SRS, Daresbury, the Photon Factory, Tsukuba and the ESRF and EMBL, Grenoble, for their assistance with the collection of synchrotron data. We also thank Karl Harlos for help with in-house data collection, Richard Bryan for computing facilities, Steven Lee for help with the preparation of figures and Iris Geens for help preparing the manuscript. RME is now supported by the Belgian Nationaal Fonds voor Wetenschappelijk Onderzoek (Grant G.3304.96) and the 'Onderzoeksfonds van de Katholieke Universiteit Leuven'. EYJ is a Royal Society University Research Fellow and DIS is supported by the MRC. The Oxford Centre for Molecular Sciences is supported by the BBSRC and MRC and this project has received the long-term support of the MRC AIDS-Directed Programme.

References

Bacon, D. J. & Anderson, W. F. (1988). *J. Mol. Graph.* **6**, 219–220.

Bernal, J. D., Fankuchen, I. & Perutz, M. (1938). *Nature (London)*, **141**, 523–524.

Brünger, A. T. (1992). *X-PLOR. Version* 3.1. *A System for X-ray Crystallography and NMR*. New Haven, USA: Yale University Press.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Cramer, P. & Müller, C. W. (1997). *FEBS Lett.* **405**, 373–377.

Das, K., Ding, J., Hsiou, Y., Clark, A. D. Jr, Moereels, H., Koymans, L., Andries, K., Pauwels, R., Janssen, P. A. J., Boyer, P. L., Clark, P., Smith, R. H. Jr, Kroeger Smith, M. B., Michejda, C. J., Hughes, S. H. & Arnold, E. (1996). *J. Mol. Biol.* **264**, 1085–1100.

Davies, J. F. II, Hostomska, Z., Hostomsky, Z., Jordan, S. R. & Matthews, D. A. (1991). *Science*, **252**, 88–95.

DeClercq, E. (1995*a*). *Clin. Microbiol. Rev.* **8**, 200–239.

DeClercq, E. (1995*b*). *J. Med. Chem.* **38**, 2491–2517.

DeClercq, E. (1996). *Rev. Med. Virol.* **6**, 97–117.

Ding, J., Das, K., Moereels, H., Koymans, L., Andries, K., Janssen, P. A. J., Hughes, S. H. & Arnold, E. (1995). *Nature Struct. Biol.* **2**, 407–415.

Ding, J., Das, K., Tantillo, C., Zhang, W., Clark, A. D. Jr, Jessen, S., Lu, X., Hsiou, Y., Jacobo-Molina, A., Andries, K., Pauwels, R., Moereels, H., Koymans, L., Janssen, P. A. J., Smith, R. H. Jr, Kroeger Koepke, M., Michejda, C. J., Hughes, S. H. & Arnold, E. (1995). *Structure*, **3**, 365–379.

Esnouf, R., Ren, J., Jones, Y., Stammers, D. & Stuart, D. (1996). *Macromolecular Refinement*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 153–161. Warrington: Daresbury Laboratory.

Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 132–134.

Esnouf, R. M., Ren, J., Hopkins, A. L., Ross, C. K., Jones, E. Y., Stammers, D. K. & Stuart, D. I. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 3984–3989.

Esnouf, R., Ren, J., Ross, C., Jones, Y., Stammers, D. & Stuart, D. (1995). *Nature Struct. Biol.* **2**, 303–308.

Hargrave, K. D., Proudfoot, J. R., Grozinger, K. G., Cullen, E., Kapadia, S. R., Patel, U. R., Fuchs, V. U., Mauldin, S. C., Vitous, J., Behnke, M. L., Klunder, J. M., Pal, K., Skiles, J. W., McNeil, D. W., Rose, J. M., Chow, G. C., Skoog, M. T., Wu, J. C., Schmidt, G., Engel, W. W., Eberlein, W. G., Saboe, T. D., Campbell, S. J., Rosenthal, A. S. & Adams, J. (1991). *J. Med. Chem.* **34**, 2231–2241.

Hopkins, A. L., Ren, J., Esnouf, R. M., Willcox, B. E., Jones, E. Y., Ross, C., Miyasaka, T., Walker, R. T., Tanaka, H., Stammers, D. K. & Stuart, D. I. (1996). *J. Med. Chem.* **39**, 1589–1600.

Hsiou, Y., Ding, J., Das, K., Clark, A. D. Jr, Hughes, S. H. & Arnold, E. (1996). *Structure*, **4**, 853–860.

Huxley, H. E. & Kendrew, J. C. (1953). *Acta Cryst.* **6**, 76–80.

Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D. Jr, Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P., Hizi, A., Hughes, S. H. & Arnold, E. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 6320–6324.

Jäger, J., Smerdon, S. J., Wang, J., Boisvert, D. C. & Steitz, T. A. (1994). *Structure*, **2**, 869–876.

Jones, E. Y., Stuart, D. I., Garman, E. F., Griest, R., Phillips, D. C., Taylor, G. L., Matsumoto, O., Darby, G., Larder, B., Lowe, D., Powell, K., Purifoy, D., Ross, C. K., Somers, D., Tisdale, M. & Stammers, D. K. (1993). *J. Cryst. Growth*, **126**, 261–269.

Kabsch, W. (1988). *J. Appl. Cryst.* **21**, 916–924.

Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. (1992). *Science*, **256**, 1783–1790.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Larder, B. A. & Kemp, S. D. (1989). *Science*, **246**, 1155–1158.
Lloyd, L. F., Brick, P., Mei-Zhen, L., Chayen, N. E. & Blow, D. M. (1991). *J. Mol. Biol.* **217**, 19–22.
McDonald, I. K. & Thornton, J. M. (1994). *J. Mol. Biol.* **238**, 777–793.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Merluzzi, V. J., Hargrave, K. D., Labadia, M., Grozinger, K., Skoog, M., Wu, J. C., Shih, C.-K., Eckner, K., Hattox, S., Adams, J., Rosenthal, A. S., Faanes, R., Eckner, R. J., Koup, R. A. & Sullivan, J. L. (1990). *Science*, **250**, 1411–1413.
Merritt, E. A. & Murphy, M. E. P. (1994). *Acta Cryst.* D**50**, 869–873.
Otwinowski, Z. (1993). *Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
Perutz, M. F. (1946). *Trans. Faraday Soc. B*, **42**, 187–195.
Pickford, M. G., Garman, E. F., Jones, E. Y. & Stuart, D. I. (1993). *J. Appl. Cryst.* **26**, 465–466.
Rao, Z., Belyaev, A. S., Fry, E., Roy, P., Jones, I. M. & Stuart, D. I. (1995). *Nature (London)*, **378**, 743–747.
Ren, J., Esnouf, R., Garman, E., Somers, D., Ross, C., Kirby, I., Keeling, J., Darby, G., Jones, Y., Stuart, D. & Stammers, D. (1995). *Nature Struct. Biol.* **2**, 293–302.
Ren, J., Esnouf, R., Hopkins, A., Ross, C., Jones, Y., Stammers, D. & Stuart, D. (1995). *Structure*, **3**, 915–926.
Richman, D. D., Fischl, M. A., Grieco, M. H., Gottlieb, M. S., Volberding, P. A., Laskin, O. L., Leedom, J. M., Groopman, J. E., Mildvan, D., Hirsch, M. S., Jackson, G. G., Durack, D. T., Nusinoff-Lehrman, S. & the AZT Collaborative Working Group (1987). *N. Engl. J. Med.* **317**, 192–197.
Rodgers, D. W., Gamblin, S. J., Harris, B. A., Ray, S., Culp, J. S., Hellmig, B., Woolf, D. J., Debouck, C. & Harrison, S. C. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 1222–1226.
Schinazi, R. F., Larder, B. A. & Mellors, J. W. (1996). *Int. Antivir. News*, **4**, 95–107.
Smerdon, S. J., Jäger, J., Wang, J., Kohlstaedt, L. A., Chirino, A. J., Friedman, J. M., Rice, P. A. & Steitz, T. A. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 3911–3915.
Spence, R. A., Kati, W. M., Anderson, K. S. & Johnson, K. A. (1995). *Science*, **267**, 988–993.
Stammers, D. K., Powell, K. L., Larder, B. A., Darby, G., Purifoy, D. J. M., Tisdale, M., Lowe, D. M., Stuart, D. I., Jones, E. Y., Taylor, G. L., Garman, E. F., Griest, R. & Phillips, D. C. (1990). *Use of X-ray Crystallography in the Design of Antiviral Agents*, edited by W. G. Laver & G. M. Air, pp. 309–319. San Diego: Academic Press.
Stammers, D. K., Somers, D. O'N., Ross, C. K., Kirby, I., Ray, P. H., Wilson, J. E., Norman, M., Ren, J. S., Esnouf, R. M., Garman, E. F., Jones, E. Y. & Stuart, D. I. (1994). *J. Mol. Biol.* **242**, 586–588.
Stuart, D. I., Levine, M., Muirhead, H. & Stammers, D. K. (1979). *J. Mol. Biol.* **134**, 109–142.
Unge, T., Ahola, H., Bhikhabhai, R., Backbro, K., Lövgren, S., Fenyo, E. M., Honigman, A., Panet, A., Gronowitz, J. S. & Strandberg, B. (1990). *AIDS Res. Hum. Retroviruses*, **6**, 1297–1303.
Unge, T., Knight, S., Bhikhabhai, R., Lövgren, S., Dauter, Z., Wilson, K. & Strandberg, B. (1994). *Structure*, **2**, 953–961.
Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.
Winkler, F. K., Schutt, C. E. & Harrison, S. C. (1979). *Acta Cryst.* A**35**, 901–911.