Acta Cryst. (1994). D**50**, 85–92

Comparison of Two Crystal Structures of TGF-β2: the Accuracy of Refined Protein Structures

BY SUN DAOPIN AND DAVID R. DAVIES

Laboratory of Molecular Biology, National Institute of Diabetes, Digestive and Kidney Diseases, National Institute of Health, Bethesda, MD 20892, *USA*

AND MICHAEL P. SCHLUNEGGER AND MARKUS G. GRÜTTER

Department of Biotechnology, Pharmaceuticals Division, Ciba-Geigy, CH-4002 *Basel, Switzerland*

(*Received* 4 *February* 1993; *accepted* 2 *August* 1993)

Abstract

Transforming growth factor-β is a multifunctional cell-growth regulator and is a member of the TGF-β superfamily of cytokines. Each monomer is 112 amino acids long and the mature active form is a 25 kDa homodimer. Recently, the crystal structure of TGF-β2 has been determined independently in two laboratories [Daopin, Piez, Ogawa & Davies (1992). *Science*, **257**, 369–373; Schlunegger & Grütter (1992). *Nature* (*London*), **358**, 430–434] and subsequently refined to higher resolutions [Daopin, Li & Davies (1993). *Proteins Struct. Funct. Genet.* In the press; Schlunegger & Grütter (1993). *J. Mol. Biol.* In the press]. A detailed structural comparison shows that the two structures are nearly identical with the differences mostly located on the mobile regions of the molecule. The r.m.s. differences between the two structures are 0.10 Å for 104 pairs of C^{α} atoms, 0.15 Å for 434 pairs of main-chain atoms, 0.33 Å for 860 out of 890 pairs of protein atoms and a correlation of 90% between the temperature B factors of all protein atoms. Based on a comparison of the water molecules, a B value of 60.0 Å^2 is recommended as the cut off for modeling new waters. The structural identity is striking because in one case the material was expressed *in vivo* in CHO cells whereas in the other case it was expressed in *E. coli* and had to be refolded *in vitro*. The overall coordinate errors are estimated to be 0.21 Å from the Luzzati plot, 0.18 Å from the σ_A plot, 0.24 Å with Cruickshank's equations and 0.25 Å using the empirical method of Perry & Stroud. These estimates are comparable to the r.m.s. structure superposition. The r.m.s. differences correlate very well with the crystallographic B values and the relation is best described with the Cruickshank formula. In addition to the estimation of an overall error, a new application of the Cruickshank formula is presented here to estimate the local errors.

Introduction

There has been a rapid increase in the number of new protein structures determined by X-ray crystallography. As a result, it has become quite common to have structures of homologous proteins or even identical proteins solved simultaneously by separate research groups. For example, the structure of IL-1β has recently been determined independently in three different laboratories as well as by an NMR method (Priestle, Schar & Grütter, 1989; Finzel *et al.*, 1989; Veerapandian *et al.*, 1992; Clore, Wingfield & Gronenborn, 1991; Shaanan *et al.*, 1992) and the structure of a cupredoxin has been solved independently in two laboratories (Adman *et al.*, 1989; Petratos, Banner, Beppu, Wilson & Tsernoglou, 1987). While the comparison between homologous protein structures often provides insight into the functional mechanism of a protein, the comparison between identical structures measures the correctness and accuracy of the structure determinations. Here, we describe a detailed comparison between the two crystal structures of TGF-β2. Since they have identical protein sequences and have been determined to similar resolutions in two independent laboratories, the comparison between them provides an estimate of the accuracy of refined X-ray crystallographic coordinates.

Results and discussion

Structure comparison

1TGI* was a recombinant human protein expressed in CHO cells whereas 1TFG was expressed in *E. coli* and folded *in vitro* (Schlunegger, Cerletti *et al.*, 1992). Both proteins were crystallized in space

* The TGF-β2 crystals and structure described by Daopin *et al.* are referred to as 1TGI whereas those described by Schlunegger & Grütter are referred to as 1TFG, following their PDB entry names.

ISSN 0907-4449

Table 1. *Refinement statistics of* 1TGI *and* 1TFG

Parameter	1TGI	1TFG
Space group	$P3_221$	$P3_221$
Cell dimensions (Å)	$a=b=60.7$	$a=b=60.6$
	$c=75.3$	$c=75.2$
Resolution (Å)	15–1.8	8.0–1.95
Solvent model	*TNT*	—
Individual isotropic *B* factor	Yes	Yes
Protein atoms	890	890
Water molecules	58	84
Refinement method	*TNT*	*TNT*
R factor*	0.173	0.188
R.m.s. bond length (Å)	0.016	0.008
R.m.s. bond angle (°)	2.3	1.98
R.m.s. planar groups (Å)	0.019	0.008

* R factor $= \Sigma_h|F_{obs} - F_{calc}|/\Sigma_h F_{obs}$.

group $P3_221$ with nearly identical cell dimensions of $a = b = 60.6$ and $c = 75.3$ Å (Table 1) and both structures were determined by the MIR (multiple isomorphous replacement) method, although different sets of heavy-atom derivatives were used to provide the initial phases. The refinements of x, y, z and isotropic B factor for each non-H atom were performed with the program package *TNT* at comparable resolutions of 1.8 (1TGI) and 1.95 Å (1TFG). Table 1 lists their refinement statistics. The two structures were superimposed using the program *ALIGN* (Satow, Cohen, Padlan & Davies, 1986), a Needleman & Wunsch (1970) dynamic search structure-alignment program. Only the C^α atoms of 1TGI and 1TFG were used in the calculation of the transformation matrix. Since the structures conform to the same space-group symmetry, the two models are related by the space-group symmetry operators. The exact rotation and translation matrix that superimposes 1TFG onto 1TGI is as follows

$$R = \begin{matrix} -0.50247 & -0.86459 & -0.00083 \\ 0.86459 & -0.50247 & 0.00098, \\ -0.00126 & -0.00022 & 1.0 \end{matrix} \qquad T = \begin{matrix} 30.268 \\ 52.568. \\ 50.223 \end{matrix}$$

The alignment resulted in an agreement better than 0.3 Å among 104 pairs of C^α's with an r.m.s. (root-mean-square) difference between them of 0.10 Å. The r.m.s. difference for all 112 pairs of C^α's is 0.27 Å. Fig. 1 shows the overlay of all the C^α atoms of the two structures. The eight residues for which the C^α positions deviate more than 0.3 Å are Arg9, Asn10, Val11 and Ile92 to Pro96. They are all located in poorly defined electron-density regions and are refined to very high B values (the average B value for residues 9–11 and 91–96 is 79.9 Å^2 in 1TGI and 86.3 Å^2 in 1TFG, see Figs. 5*a* and 5*b*).

Based on the C^α alignment, the deviation among the xyz of the two TGF-β2 structures ranges from 0.01 to 6.5 Å with the biggest differences being located at the two mobile loop regions (residues 9–11 and 91–97). The r.m.s. differences are 0.15 Å for 434 main-chain atoms (0.29 for all 449 main-chain atoms) and 0.33 Å for 860 (a total of 890) main-chain and side-chain atoms. There are 30 atoms that deviate more than 1.8 Å; most of these are side-chain atoms which either have excessively high B values (clipped at an upper limit of 100.0 Å^2) or have been assigned zero occupancy during the refinement. The electron densities surrounding them are so weak that their positions probably reflect more of the modeler's bias rather than the true centroid of the electron densities. They include (i) the side-chain atoms of the two most mobile turns, residues 9–11 and residues 91–97 (the overlay of one of the most mobile regions, residues 91–96, is shown in Fig. 2); (ii) some mobile Lys and Arg side-chain atoms; and (iii) the side chains of Asn69 and Glu71 which differ in their χ_1 angle by about 120°. When these 30 atoms are also included in the calculation, the r.m.s. deviation for all 890 protein atoms is 0.70 Å.

The average B value of the protein atoms is 29.4 Å^2 for 1TGI and 33.1 Å^2 for 1TFG. In addition, a correlation coefficient c between their crystallographic B values is calculated using the following equation

$$c = \frac{\sum_{i=1}^{N}(B_{1,i}-\langle B_1\rangle)(B_{2,i}-\langle B_2\rangle)}{\left[\sum_{i=1}^{N}(B_{1,i}-\langle B_1\rangle)^2 \sum_{i=1}^{N}(B_{2,i}-\langle B_2\rangle)^2\right]^{1/2}}, \qquad (1)$$

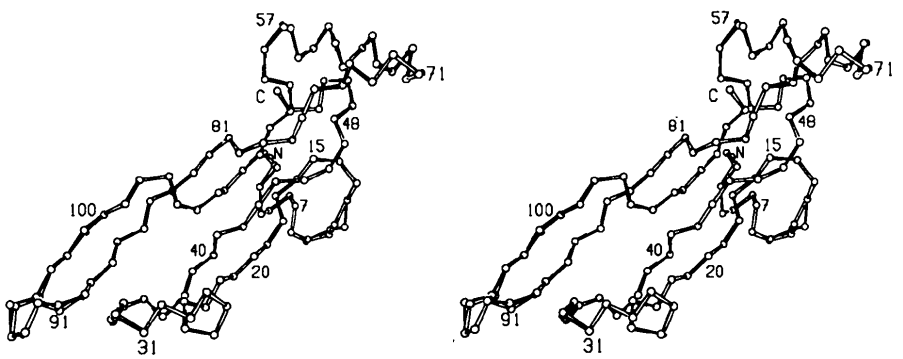

Fig. 1. *ORTEP* drawing of an overlay of the C^α atoms between 1TGI (solid bond) and 1TFG (open bonds) in stereo pairs (Johnson, 1970).

Table 2. *Structure alignment statistics*

		$\langle \Delta r \rangle_{r.m.s.}$* (Å)		
	B-value correlation	Comparison	Cruickshank	Perry & Stroud
C^{α} atoms				
104 pairs of C^{α}'s		0.10	0.17	0.10
All C^{α}'s	0.97	0.27	0.29	0.27
Main-chain atoms				
434 atoms		0.15	0.14	0.20
All atoms	0.91	0.29	0.24	0.29
Main-chain and side-chain atoms				
860 pairs	0.90	0.33	0.34	0.37

* Root-mean-square atomic positional difference.

where N is the number of common atoms in both coordinates sets and $B_{1,i}$ represents the B value for the ith atom in coordinate set 1. $\langle B_1 \rangle$ is the average B value for coordinate set 1, calculated from

$$\langle B \rangle = \sum_{i=1}^{N} B_i / N . \quad (2)$$

The calculated B-value correlation is 0.97 between all the C^{α} atoms and 0.90 between all the protein atoms (Table 2).

There are 58 water molecules in 1TGI with an average B value of 31.8 Å^2 and 84 waters in 1TFG with an average B value of 43.3 Å^2. The positions of 54 of these waters superimpose within 0.6 Å which results in an r.m.s. displacement of 0.24 Å. The 54 common water molecules of 1TFG display much lower B values (with an average of 34.0 Å^2) than the 30 non-common water molecules of 1TFG (with an average B of 60.1 Å^2). Since the existence of these 30 crystallographically built water molecules is less certain than that of the common ones, their averaged B value of 60 Å^2 can be used as a useful cutoff criterion for modeling new water molecules.

It is worth noting that the sources of the two TFG-β2 proteins are very different. While protein for the 1TGI structure was a recombinant human protein expressed in CHO cells, protein for the 1TFG structure was a recombinant human protein expressed in *E. coli* (Schlunegger, Cerletti *et al.*, 1992), isolated from inclusion bodies under denaturing conditions and then refolded *in vitro* to the biologically active molecule. The comparison shows that a protein folded properly *in vitro* adopts the same structure as a naturally folded protein even if, as in this case, a complex disulfide cluster has to be formed.

Accuracy assessments

There are, generally, three different ways to assess errors associated with a coordinate set. The first and the most direct way of estimating errors is by comparing multiple solutions of the same structure. Examples of this category include the comparisons of refined bovine trypsin structures (Chambers & Stroud, 1979) and the structures of cupredoxin from *Alcaligenes faecalis* (Adman *et al.*, 1989). Though the most direct and probably the most accurate way, it is the least frequently used in error assessments because of the lack of independently determined structures in most cases. In the situation when there are no independently determined multiple structure solutions available, a second and often-adopted way of error estimation is to compare the final refined model with a model from earlier or different refinement rounds. This has been used to estimate the errors of the complex of *Streptomyces griseus* protease B with the third domain of the turkey ovomucoid inhibitor (Read, Fujinaga, Sielecki & James, 1983) and the structure of poplar plastocyanin (Guss, Bartunk & Freeman, 1992). The errors obtained this way often tend to be underestimated because the structures from different refinement rounds are not independent determinations. The third and very useful way is to estimate errors using a kind of error function (Stout & Jensen, 1989), such as the Luzzati plot (Luzzati, 1952), the σ_A plot (Read, 1986), Cruickshank's equations (Cruickshank, 1967) and the empirical B-factor equations of Perry & Stroud (Perry *et al.*, 1990). The errors calculated from these error functions are limited to random errors that follow a normal statistical distribution. Systematic errors are reflected only in comparison of independently determined structures.

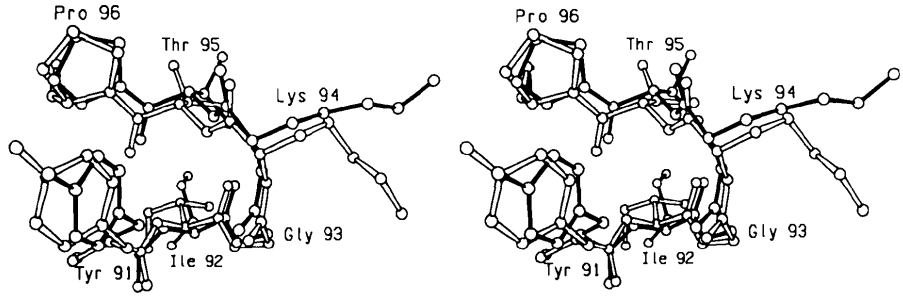

Fig. 2. *ORTEP* overlay of the β-turn region of residues 91–96. This is one of the most mobile regions of the molecule and the r.m.s. agreement between the two structures in this region is 0.84 Å. The solid bond corresponds to the region from 1TGI and in the open bond is the region from 1TFG.

Not only do the two structures (1TGI and 1TFG) correspond to identical protein sequences, but they have also been determined in the same space group at similar resolutions and refined using the same refinement program package (*TNT*). Therefore the type of systematic error associated with conformational differences as a result of the influence of crystal packing or refinement method are not present, though the systematic error from data collection and processing will affect the results of our comparison. This offers an unusual example of two identical but independently determined structures and the results in the structure-comparison section provide an estimate of the error associated with a crystallographically determined structure.

In the following section, four different coordinate-error estimation methods are applied to evaluate the expected random error in the two TGF-β2 coordinates and their results are compared to the structural comparison of the two coordinate sets.

Luzzati plot. The Luzzati plot is commonly used to estimate the mean error associated with a coordinate set (Luzzati, 1952). Assuming the coordinate errors follow a Gaussian distribution and that they are independent of orientation, *i.e.* $\sigma_x = \sigma_y = \sigma_z$, the mean error in x, y, z coordinates can be evaluated from the differences in the observed and calculated structure factors $||F_o| - |F_c||$ or equivalently from the crystallographic R factors in a resolution dependent way, the so called Luzzati plot. When the method is applied to the two refined TGF-β2 structures, it results in an overall error of 0.21 Å for 1TGI and 0.23 Å for 1TFG (Fig. 3). These estimated errors are then related to the root-mean-square difference in the coordinates by

$$\langle \Delta r \rangle = (\sigma_1^2 + \sigma_2^2)^{1/2}, \tag{3}$$

where σ_1 and σ_2 are the estimated errors for the coordinate sets 1 and 2. The calculated r.m.s. difference from (3) is 0.31 Å between the two coordinate sets. It is larger than the 0.15 Å r.m.s. difference among the main-chain atoms and agrees well with the 0.33 Å r.m.s. difference of all the atoms.

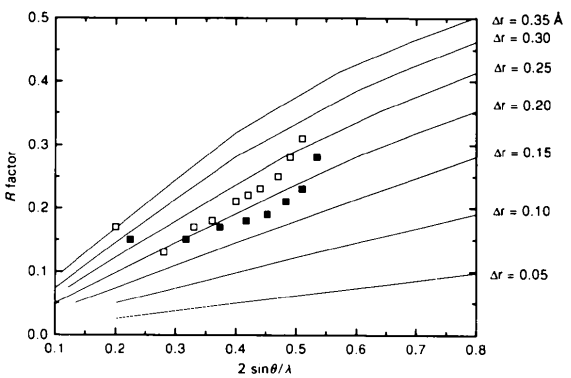

Fig. 3. Luzzati plot showing the refined R factor as a function of resolution for 1TGI (solid squares) and 1TFG (open squares).

Modified Luzzati method: the σ_A plot. When the errors are normally distributed, the Luzzati quantity D [$D = \cos(2\pi \Delta \mathbf{r} \cdot \mathbf{s})$] can be expressed as:

$$D = \exp[-\pi^3 \sigma^2 (\sin\theta/\lambda)^2] \tag{4}$$

where σ is the expected value of the coordinate error (Luzzati, 1952; Read, 1986). Srinivasan & Ramachandran (1965) noted that generally, when the calculated structure factors arise from a partial structure containing P out of a total of N atoms, the probability distribution function of the observed and calculated structure factors needs to be modified as compared to Luzzati's approach by replacing D with a new quantity σ_A defined as (Srinivasan & Ramachandran, 1965):

$$\sigma_A = \left(\sum_j^P f_j^2 \Big/ \sum_j^N f_j^2 \right)^{1/2} D . \tag{5}$$

One way to estimate σ_A, as suggested by Hauptman (1982) is to calculate the quantity in a form of the correlation coefficient between the square of the two normalized structure factors, E_P and E_N:

$$\sigma_A = \left\{ \frac{\sum(|E_P|^2 - \overline{|E_P|^2})(|E_N|^2 - \overline{|E_N|^2})}{[\sum(|E_P|^2 - \overline{|E_P|^2})^2 \sum(|E_N|^2 - \overline{|E_N|^2})^2]^{1/2}} \right\}^{1/2} . \tag{6}$$

In the case of estimation of the errors associated with a refined structure, E_P and E_N in the σ_A expression are replaced with F_c and F_o. Since the average intensities of reflections vary significantly with resolution, the calculation of σ_A should be performed in several resolution shells. When (4) is substituted into (5), a linear relation between $\ln\sigma_A$ and $(\sin\theta/\lambda)^2$ is obtained (Read, 1986):

$$\ln\sigma_A = (1/2)\ln\left(\sum_j^P f_j^2 \Big/ \sum_j^N f_j^2 \right) - \pi^3 \sigma^2 (\sin\theta/\lambda)^2 . \tag{7}$$

Assuming the ratio $(\sum_P f_j^2 / \sum_N f_j^2)$ is constant, (7) results in a straight line in the σ_A plot with the slope of the line relating to the expected error of a coordinate.

While a Luzzati plot estimates the error from comparing a set of theoretical error curves to R factors calculated in resolution shells, a σ_A plot obtains the expected error from the straight-line fitting of the σ_A plot. Since the scale factor between F_o and F_c is cancelled in the σ_A expression (6), the method is relatively insensitive to the scaling compared to a Luzzati plot. In the structure of tonin, a serine protease from rat submaxillary gland, the σ_A

plot resulted in an estimated error comparable to that from the Luzzati plot (Fujinaga & James, 1987). Fig. 4 shows the σ_A plot for 1TGI; the expected coordinate error obtained from the slope of the fitted line is 0.18 Å, and the estimated r.m.s. difference between 1TGI and 1TFG is 0.25 Å [see (3), assuming the expected error of 1TFG is also 0.18 Å].

The calculated r.m.s. differences between the two structures from both the Luzzati plot (0.31 Å) and the σ_A plot (0.25 Å) compare well with the 0.33 Å r.m.s. difference of all the atoms resulting from the coordinate superposition.

Estimation from the Cruickshank formula. The limitation of a Luzzati plot and a σ_A plot is that both methods only estimate the overall coordinate error. It has been previously proposed that the r.m.s. differences between the coordinates correlate with the local crystallographic B values (Chambers & Stroud, 1979; Guss, Harrowell, Murata, Norris & Freeman, 1986; Perry *et al.*, 1990). Generally, there is a close correlation between the accuracy of atomic positions and their B values (Figs. 5*a*, 5*b*). The existence of such a correlation mostly reflects the fact that a well defined electron-density region corresponds to high accuracy and low B values whereas a poorly defined electron-density map region results in low accuracy and high B values. This correlation provides a means of estimating the accuracy of a structure based on the refined crystallographic B values.

Cruickshank derived a set of equations to estimate the coordinate errors from the errors in a difference Fourier map (Cruickshank, 1949, 1954, 1967). Assuming the electron density ρ_o and the difference density as a result of error $(\rho_o - \rho_c)$ surrounding an atom in the unit cell are spherically distributed, then the estimated standard deviation of a coordinate for a type i atom with temperature factor B can be expressed as

$$\sigma_i(x) = \frac{\frac{1}{V}\frac{2\pi}{a}\left[\sum_{hkl} h^2(|F_o| - |F_c|)^2\right]^{1/2}}{-\frac{1}{V}\frac{4\pi^2}{a^2}\sum_{hkl}\frac{m}{2}h^2 f_0 \exp(-B_i \sin^2\theta/\lambda^2)}. \quad (8)$$

The same formula applies for $\sigma(y)$ and $\sigma(z)$ with proper substitution of h, k, and l. And

$$\sigma_i(r) = [\sigma_i(x)^2 + \sigma_i(y)^2 + \sigma_i(z)^2]^{1/2}, \quad (9)$$

is the estimated atomic coordinate error for atom i. The average residue error is given as

$$\sigma_{\text{res}}(r) = \left[\sum_i^n \sigma_i(r)\right]/n, \quad (10)$$

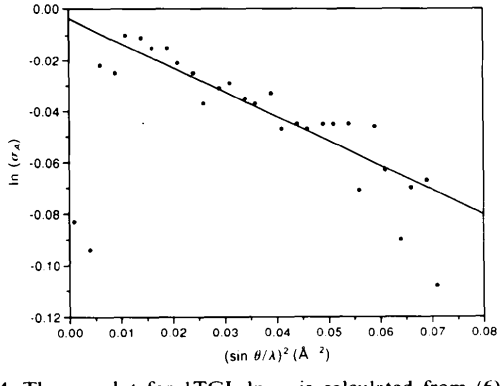

Fig. 4. The σ_A plot for 1TGI. $\ln\sigma_A$ is calculated from (6) in 29 resolution shells between 15.0 and 1.8 Å. A linear regression (with the two lowest resolution points and the one highest resolution point omitted) results in a slope of -0.958 Å^2 and the expected error $\sigma = 0.18$ Å.

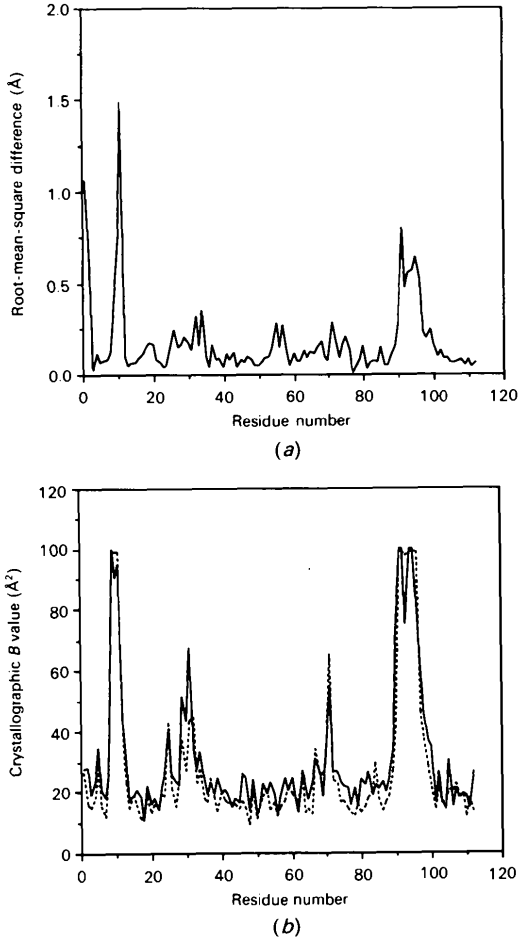

Fig. 5. (*a*) Plot of the structural differences between C$^\alpha$ atoms of 1TGI and 1TFG *versus* residue number. (*b*) B values of C$^\alpha$ atoms in 1TGI (solid line) and 1TFG (dotted line) are plotted *versus* the residue number. The resemblance between (*a*) and (*b*) shows that the structural displacements are highly correlated with the atomic temperature B factors.

where the sum is over all atoms in a residue and n is number of atoms in the residue.

Compared to the methods of Luzzati and Read, Cruickshank's error estimation (8) allows one to estimate the error of a given atom type in a temperature B factor dependent way. It has been used to estimate the overall coordinate errors of SGPB (*Streptomyces griseus* protease B) (Read *et al.*, 1983) and a refined porcine pepsin structure (Sielecki, Fedorov, Boodkoo, Andreeva & James, 1990) and resulted in an estimated mean error comparable to that from the Luzzati plot. Using the (8), (9) and (3), Fig. 6 shows the results of comparison between the r.m.s. structural superpositions and the theoretical curves obtained from the Cruickshank formula. The agreements between estimates based on Cruickshank's equations and the observed structural deviations are quite good throughout the entire range of B values. Fig. 6 also illustrates that the r.m.s. differences for low B atoms (B values less than 40.0 Å^2) are less than 0.5 Å and those for high B atoms (B values greater than 60.0 Å^2) are generally larger than 1.0 Å.

In most cases, where multiple solutions of one structure are not available, (8) and (9) give the estimated coordinate errors of a given coordinate set. The estimated overall coordinate error for 1TGI is 0.24 Å using the two equations and the corresponding overall r.m.s. difference is estimated to be 0.34 Å which agrees well with the structural superposition for all the atoms.

Since positional errors correlate with their B values, different regions of the molecule will, in general, differ in their coordinate errors as well. The main difference between using (8) *versus* the Luzzati plot is that it is a real-space estimate and directly relates errors to an atomic property, namely the local atomic B values, whereas a Luzzati plot or a σ_A plot estimate the error of $(|F_o| - |F_c|)$'s as a function of the reciprocal-space quantity $\sin(\theta)/\lambda$. This unique property of a real-space error estimation allows one to estimate regional accuracies beside an overall accuracy, *e.g.* the accuracy of a catalytic domain, a substrate or ligand binding region or protein interior *versus* surface. Table 2 shows the estimated r.m.s. differences for two subsets of atoms, namely the C^α atoms and main-chain atoms, using the Cruickshank formula. A more useful way to represent the local distribution of the coordinate error is to plot the average estimated error of each residue with respect to the sequence number. To achieve this, the expected errors of the atoms in each residue are first calculated using the Cruickshank formula and then averaged using (6) within each residue. Fig. 7(*a*) shows this type of estimated residue error for TGF-β2 (1TGI) plotted against its sequence number. It clearly shows that the two mobile loop regions, residues 9–11 and 91–97, have largest coordinate errors among all the residues. When compared with the residue-averaged B-factor plot (Fig. 7*b*), the error plot (Fig. 7*a*) reveals a striking resemblance with the B-factor plot (Fig. 7*b*), indicating a strong dependence of $\sigma_i(x)$ on B in (8).

Chambers & Stroud's empirical B-factor method. Several other schemes (utilizing the strong correlation between the positional variation and B factor) have been proposed in the past to use B factors in

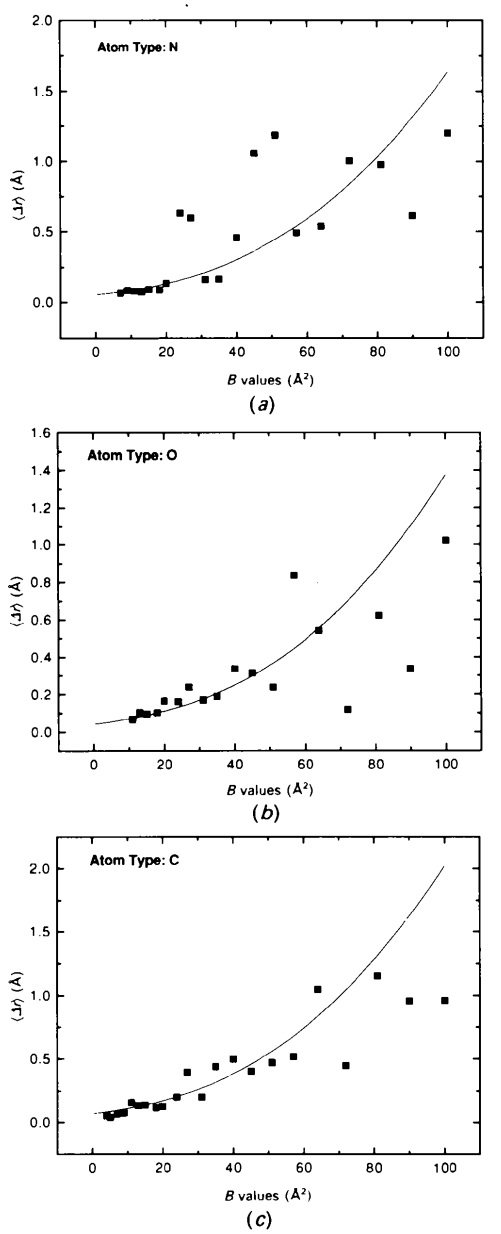

Fig. 6. Comparison between the observed r.m.s. structure deviations and the theoretical distribution curves calculated from the Cruickshank formula (8) for (*a*) all N atoms, (*b*) all O atoms and (*c*) all C atoms.

accuracy assessments. Adman and co-workers applied an inverse B-factor weighting in the structure comparison of a cupredoxin to evaluate the associated coordinate error (Adman *et al.*, 1989). The most direct efforts to correlate the positional errors and their B factors are from the empirical formula of Chambers & Stoud (1979) and Perry *et al.* (1990). From the comparisons of refined trypsin structures and refined thymidylate synthases, Perry *et al.* have empirically related the coordinate errors to the averaged B values by a quadratic equation (Perry *et al.*, 1990)

$$\sigma(B) = (3/4)rR(aB^2 + bB + c) \qquad (11)$$

where r is the resolution in Å, R is the crystallographic R factor and $a = 0.0015$, $b = -0.0203$ and $c = 0.359$ are three curve-fitting constants from the trypsin comparisons.

When (11) is applied to the two refined TGF-β structures, it results in an estimated overall coordinate error of 0.25 for 1TGI and 0.28 Å for 1TFG and hence the estimated overall standard deviation in the comparison of the two is 0.37 Å. The estimated errors for the corresponding C^α and main-chain atoms are listed in the last column of Table 2. Fig. 8 shows the predicted r.m.s. using the empirical B-factor method and the observed r.m.s. from the structure superposition as a function of B factors. The method predicts errors that agree well with those observed at a low B range ($B < 40.0$ Å^2) and tends to overestimate the errors at a high B range because of the quadratic function. The actual errors for atoms with relatively large B factors ($B > 40.0$ Å^2) show a much smaller increase than a quadratic dependence on their B factors would suggest. The poor agreement at the high B end is perhaps not surprising, since the choice of the function and parameters resulted from the trypsin comparisons with only atoms that have B factors less than 40.0 Å^2 included. It is also conceivable that a quadratic function describes errors better for low B atoms than it does for high B atoms, since the use of stereochemical constraints in refinements restrains the positional movements less for low B atoms and more for high B atoms where electron densities are weak and many of the positional shifts are dampened.

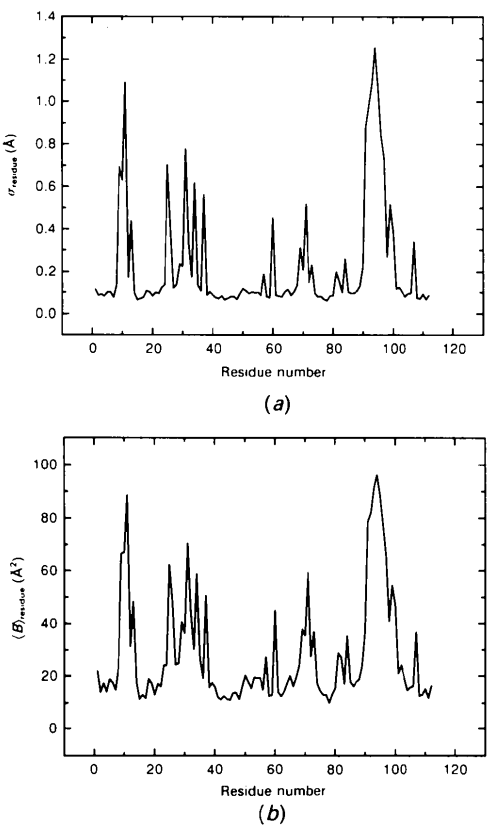

Fig. 7. (*a*) Local-error distribution plot. The errors estimated using the Cruickshank equation for different atom types are averaged within each residue using (10) to give the residue-averaged error. (*b*) Residue-averaged B plot.

Summary

The results of the structure comparison between 1TGI and 1TFG show that the two structural solutions of TGF-β2 are largely identical with the r.m.s. deviation in *xyz* about 0.3 Å and a correlation between B values of 90%. Most of the large differences are located on the surface in the most mobile regions of the molecule (Figs. 5*a* and 5*b*).

The comparison also shows that, in general, the r.m.s. agreement between 1TGI and 1TFG becomes worse as the B value increases. The two structures agree very well between those atoms with B values less than 40.0 Å^2 and much less well for the atoms with B values higher than 40.0 Å^2, suggesting that the regions of structure with B values less than 40.0 Å^2 can be regarded as crystallographically

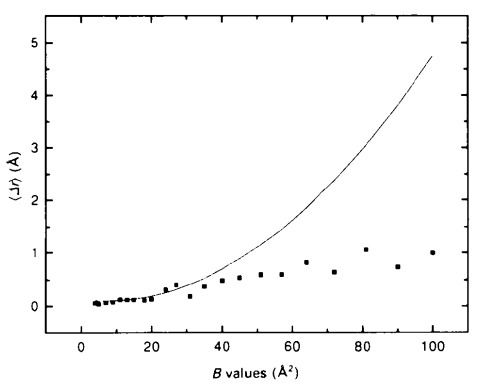

Fig. 8. Comparison between the observed r.m.s. structure deviations (shown as squares) and the empirical estimations (shown as a curve) from Perry & Stroud's equation (11).

reliably determined regions. Since the average error in the coordinates exceeds 1 Å for the atoms with B values higher than 60.0 Å^2 (Fig. 6) and the average B values for the water molecules observed only in 1TFG is 60.1 Å^2, this refined B value of 60.0 Å^2 or slightly lower can be regarded as a useful cut off criterion for building water molecules.

Four different analytical error-estimation methods are applied here to assess the coordinate errors in the refined TGF-β2 structures. Among them, the two real-space B-factor based estimation methods can be used not only to assess an overall coordinate error but also to assess errors in local regions of a structure, although the quadratic function of Perry & Stroud, when compared to the Cruickshank formula, appears to overestimate errors for high B-value atoms.

References

ADMAN, E. T., TUELEY, S., BRAMSON, R., PETRATOS, K., BANNER, D., TSERNOGLOU, D., BEPPU, T. & WATANABE, H. (1989). *J. Biol. Chem.* **264**, 87–99.

CHAMBERS, J. L. & STROUD, R. M. (1979). *Acta Cryst.* B**35**, 1861–1874.

CLORE, G. M., WINGFIELD, P. T. & GRONENBORN, A. M. (1991). *Biochemistry*, **30**, 2315–2323.

CRUICKSHANK, D. W. J. (1949). *Acta Cryst.* **2**, 65–82.

CRUICKSHANK, D. W. J. (1954). *Acta Cryst.* **7**, 519.

CRUICKSHANK, D. W. J. (1967). *International Tables for X-ray Crystallography*, Vol. 2, edited by J. S. KASPER & K. LONSDALE, pp. 318–340. Birmingham: Kynoch Press. (Present distributor Kluwer Academic Publishers, Dordrecht.)

DAOPIN, S., LI, M. & DAVIES, D. R. (1993). *Proteins Struct. Funct. Genet.* In the press.

DAOPIN, S., PIEZ, K. A., OGAWA, Y. & DAVIES, D. R. (1992). *Science*, **257**, 369–373.

FINZEL, B. C., CLANCY, L. L., HOLLAND, D. R., MUCHMORE, S. W., WATENPAUGH, K. D. & EINSPAHR, H. M. (1989). *J. Mol. Biol.* **209**, 779–691.

FUJINAGA, M. & JAMES, M. N. G. (1987). *J. Mol. Biol.* **195**, 373–396.

GUSS, J. M., BARTUNK, H. D. & FREEMAN, H. C. (1992). *Acta Cryst.* B**48**, 790–811.

GUSS, J. M., HARROWELL, P. R., MURATA, M., NORRIS, V. A. & FREEMAN, H. C. (1986). *J. Mol. Biol.* **192**, 361–387.

HAUPTMAN, H. (1982). *Acta Cryst.* A**38**, 289–294.

JOHNSON, C. K. (1970). *ORTEP. A Fortran Thermal-Ellipsoid Plot Program For Crystal Structure Illustrations.* Oak Ridge National Laboratory, Tennessee, USA.

LUZZATI, P. V. (1952). *Acta Cryst.* **5**, 802–810.

NEEDLEMAN, S. B. & WUNSCH, C. D. (1970). *J. Mol. Biol.* **48**, 443–453.

PERRY, K. M., FAUMAN, E. B., FINER-MOORE, J. S., MONTFORT, W. R., MALEY, G. F., MALEY, F. & STROUD, R. M. (1990). *Proteins Struct. Funct. Genet.* **8**, 315–333.

PETRATOS, K., BANNER, D. W., BEPPU, T., WILSON, K. S. & TSERNOGLOU, D. (1987). *FEBS Lett.* **218**, 209–214.

PRIESTLE, J. P., SCHAR, H.-P. & GRÜTTER, M. G. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 9667–9671.

READ, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

READ, R. J., FUJINAGA, M., SIELECKI, A. R. & JAMES, M. N. G. (1983). *Biochemistry*, **22**, 4420–4433.

SATOW, Y., COHEN, G. H., PADLAN, E. A. & DAVIES, D. R. (1986). *J. Mol. Biol.* **190**, 593–604.

SCHLUNEGGER, M. P., CERLETTI, N., COX, D. A., MCMASTER, G. K., SCHMITZ, A. & GRÜTTER, M. G. (1992). *FEBS Lett.* **303**, 91–93.

SCHLUNEGGER, M. P. & GRÜTTER, M. G. (1992). *Nature (London)*, **358**, 430–434.

SCHLUNEGGER, M. P. & GRÜTTER, M. G. (1993). *J. Mol. Biol.* In the press.

SHAANAN, B., GRONENBORN, A. M., COHEN, G. H., GILLILAND, G. L., VEERAPANDIAN, B., DAVIES, D. R. & CLORE, G. M. (1992). *Science*, **257**, 961–964.

SIELECKI, A. R., FEDOROV, A. A., BOODKOO, A., ANDREEVA, N. S. & JAMES, M. N. G. (1990). *J. Mol. Biol.* **214**, 143–170.

SRINIVASAN, R. & RAMACHANDRAN, G. N. (1965). *Acta Cryst.* **19**, 1008–1014.

STOUT, G. H. & JENSEN, L. H. (1989). *X-ray Structure Determination: A Practical Guide*, ch. 17, pp. 379–399. New York: Macmillan.

VEERAPANDIAN, B., GILLILAND, G. L., RAAG, R., SVENSSON, A. L., MASUI, Y., HIRAI, Y. & POULOS, T. L. (1992). *Proteins Struct. Funct. Genet.* **12**, 10–23.