Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

On the correlation between the main-chain and side-chain atomic displacement parameters (*B* values) in high-resolution protein structures

S. Parthasarathy and M. R. N. Murthy*

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

Correspondence e-mail: mrn@mbu.iisc.ernet.in

Received 23 December 1997
Accepted 6 May 1998

Correlation coefficients (CCs) between the mean atomic displacement parameters (*B* values) of the main-chain and the side-chain atoms in selected high-resolution protein structures are distributed over a broad range (0.5–1.0). The distribution of CCs is found to be related to the mean difference in *B* values of C^{α} and C^{β} atoms. High CCs are also associated with the frequent occurrence of consecutive C^{α} atoms with relatively high *B* values. The distribution of CCs and its relation to the mean difference between the *B* values of C^{α} and C^{β} atoms shows dependence on the package used for refinement (*X-PLOR*, *PROLSQ* or *TNT*). These observations reflect the differences in the way *B*-value constraints are handled in these packages. Further differences are discernible in the distributions for proteins refined using the same package. It is likely that these differences are related to the different refinement protocols or weighting schemes followed by investigators. The resolution of these issues is important for evolving correct strategies for the refinement of the atomic displacement parameters in X-ray diffraction studies of proteins. Furthermore, it may be possible to develop refinement-validation tools by observing the features that are invariant in the distribution of atomic displacement parameters.

1. Introduction

Advances in data-collection techniques, radiation sources and computational methods have resulted in the determination of a large number of protein structures at near atomic resolution (Dauter *et al.*, 1997). These studies provide information on the flexibility and dynamics of protein structures. It is now generally agreed that atomic displacements occurring in protein crystals are comparable to those occurring in solution (Ringe & Petsko, 1986). The information about protein dynamics in X-ray crystallographic studies is contained in the atomic displacement parameters (*B* values). Atomic displacement parameters are also related to the crystal disorder. It is not possible to partition the contributions of dynamics and static disorder to the atomic displacement parameters from diffraction studies conducted at a single temperature. However, careful analysis of the existing data could reveal information on protein dynamics (Peters-Libeu & Adman, 1997). These aspects of protein structure have been studied in less detail compared with conformational aspects. Several methods have been developed to validate the observed conformational characteristics of proteins. In contrast, no readily applicable method is available for validating the reported *B* values. The observed *B* values have been used to derive flexibility indices for amino-acid residues (Karplus & Schulz, 1985; Ragone *et al.*, 1989; Vihinen *et al.*, 1994).

Table 1
PDB codes for the protein structures selected for analysis, grouped into CC ranges.

X denotes *X-PLOR*, *P* denotes *PROLSQ* and *T* denotes *TNT* used for the refinement.

CC range	Number	PDB codes for the structures
0.4–0.5	1	4GCR
0.5–0.6	11	1ATL(*X*), 1CHM, 1GPR(*T*), 1IAE(*X*), 1LEN, 1REG(*X*), 2ER7, 2HMZ(*T*), 3TGI(*P*), 4FGF(*T*), 8TLN(*T*)
0.6–0.7	18	131L(*T*), 1BP2, 1CNS(*X*), 1CSE, 1CSH(*T*), 1FNC(*T*), 1HMT, 1OVA, 1PBE(*P*), 1PGS(*T*), 1TAH(*X*), 2AZA, 2CPL(*X*), 2OLB(*P*), 2POR(*X*), 2SIL, 3DFR, 3GRS
0.7–0.8	22	153L(*T*), 1CCR, 1DUP(*P*), 1GOF(*P*), 1ISC(*X*), 1MOL(*X*), 1NFP(*P*), 1PTX, 1TCA(*X*), 1THV(*T*), 1TRY(*P*), 2ACQ(*X*), 2ALP, 2AYH(*T*), 2CTC(*P*), 2END(*P*), 2NAC(*P*), 2TGL(*T*), 3CLA(*P*), 5TIM(*T*), 8ABP(*P*), 9RNT
0.8–0.9	36	1AMP(*X*), 1ARB(*P*), 1ARV(*X*), 1BAM(*X*), 1CHD, 1DTS(*X*), 1DYR, 1EDE(*T*), 1FKJ(X), 1KPT(*X*), 1LCP(*X*), 1LTS, 1NAR(*P*), 1NHK, 1PDA, 1PHG(*P*), 1REC(*X*), 1RSY(*X*), 1SAT(*X*), 1SBP(*P*), 1SRI(*P*), 1TAG(*X*), 1TYS(*X*), 1XNB(*X*), 1XYZ(*X*), 2CBA, 2CCY, 2GST, 2HTS(*X*), 2MNR, 2PHY(*X*), 2PRK(*P*), 3CHY, 3COX(*X*), 3PTE, 8FAB(*X*)
0.9–1.0	7	1DAA(*X*), 1SNC(*P*), 1TTB, 2CDV, 3SIC(*P*), 4ENL(*P*), RUB(*P*)

Parthasarathy & Murthy (1997, 1998) have shown that amino-acid residues display characteristic *B*-value distributions from which useful parameters regarding the thermal properties of the amino-acid residues can be deduced. Tronrud (1996) has deduced constraints useful in the refinement of atomic displacement parameters by carrying out careful unconstrained refinement of a few structures for which X-ray diffraction data to 1.6 Å are available.

In this communication, we show that the frequency distribution of *B* values at C^{α} atoms, when expressed in units of standard deviation about the mean value (Parthasarathy & Murthy, 1997, 1998), is a useful tool for the examination of thermal parameters of high-resolution structures. We also analyse the correlation between the main-chain and the side-chain *B* values in high-resolution protein structures. The correlation coefficients (CCs) between the mean thermal parameters of these groups are found to be scattered over a broad range. An attempt is made to relate these observations to the methods of refinement and to a variety of protein properties. Analysis with structures refined using standard packages, *viz*. *X-PLOR* (Brünger *et al.*, 1987), *PROLSQ* (Hendrickson & Konnert, 1980) and *TNT* (Tronrud *et al.*, 1987), reveal that the differences in the constraints used in these packages for *B*-value refinement do not drastically alter the scatter in CC values. Within a set of structures refined using a particular package, the observed variations might arise from genuine differences in the dynamics of the protein structures or from non-uniformity of the refinement protocols used by crystallographers. The resolution of this issue will lead to better protein refinement procedures. Furthermore, such analysis may also suggest methods for validating the reported atomic displacement parameters.

2. Methods

2.1. Selection of high-resolution protein structures

95 protein structures, determined at resolutions better than 2 Å and with *R* factors less than 0.2, were selected from the representative set provided by Hobohm & Sander (1994) and available in the November 1996 release of the Protein Data Bank (Bernstein *et al.*, 1977). Table 1 lists the PDB codes for the structures along with their CC range (see §2.3). These 95 structures were further divided into sets based on the program used for their refinement: *X-PLOR*, *PROLSQ* or *TNT*. 30 of the structures were refined using *X-PLOR*, 22 using *PROLSQ* and 14 using *TNT*. The remaining 29 structures were refined using either a combination of these packages or by using some other protocol.

2.2. Distribution of *B* values

The *B* values vary widely from one structure to another. Therefore, to compare different structures, *B* values at C^{α} and C^{β} positions in individual proteins were normalized as $B' = (B - \langle B \rangle)/\sigma(B)$, where *B* is the experimental *B* value and $\langle B \rangle$ and $\sigma(B)$ are the mean and the standard deviation, respectively, of *B* values at C^{α} positions. The observed frequency distributions of the *B* factor sampled in σ bins of size 0.5 were calculated. The observed distribution was fitted (Parthasarathy & Murthy, 1997, 1998) analytically as a summation of two Gaussian functions:

$$f = k_1 \exp[-k_2(B - B_1)^2] + k_3 \exp[k_4(B - B_2)^2] \qquad (1)$$

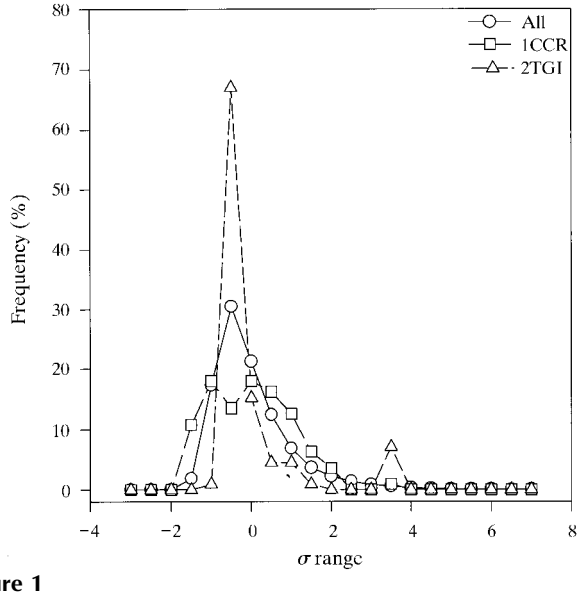

Figure 1
Observed B'-factor frequency distribution in σ bins of size 0.5. ○, overall distribution; □, distribution for rice ferricytochrome *c* (1CCR); △, distribution for transforming growth factor β2 (2TGI).

where k_1, k_2, k_3, k_4, B_1 and B_2 are constants. In order to avoid instability in the least-squares refinement, the distribution was initially fitted to a single Gaussian and the residual was fitted by a second Gaussian.

2.3. Correlation coefficients

For each protein, the correlation coefficient between the main-chain and the side-chain thermal parameters was computed as

$$CC = \frac{\sum(B_{\rm main} - \langle B_{\rm main}\rangle)(B_{\rm side} - \langle B_{\rm side}\rangle)}{[\sum(B_{\rm main} - \langle B_{\rm main}\rangle)^2 \sum(B_{\rm side} - \langle B_{\rm side}\rangle)^2]^{1/2}}. \quad (2)$$

The summation is over all the residues except glycine, which has no non-hydrogen side-chain atom. The distribution of CCs was examined for all 95 structures together and also separately for the structures refined using different packages.

2.4. Frequency distribution of high-*B*-value stretches

For each protein, a 'high' B value at C^{α} atoms was defined as $B > \langle B\rangle + 0.5\sigma(B)$. The occurrence of different lengths of consecutive high B values were counted in each protein. This analysis was carried out separately for the sets of proteins showing low and high CCs between the main-chain and the side-chain atoms.

3. Results

3.1. Distribution of B' factors

The observed frequency of B' factors in σ bins of size 0.5 was counted (Fig. 1). The distribution for all the amino acids over all the proteins, for individual proteins and individual amino acids could be fitted accurately as the summation of two

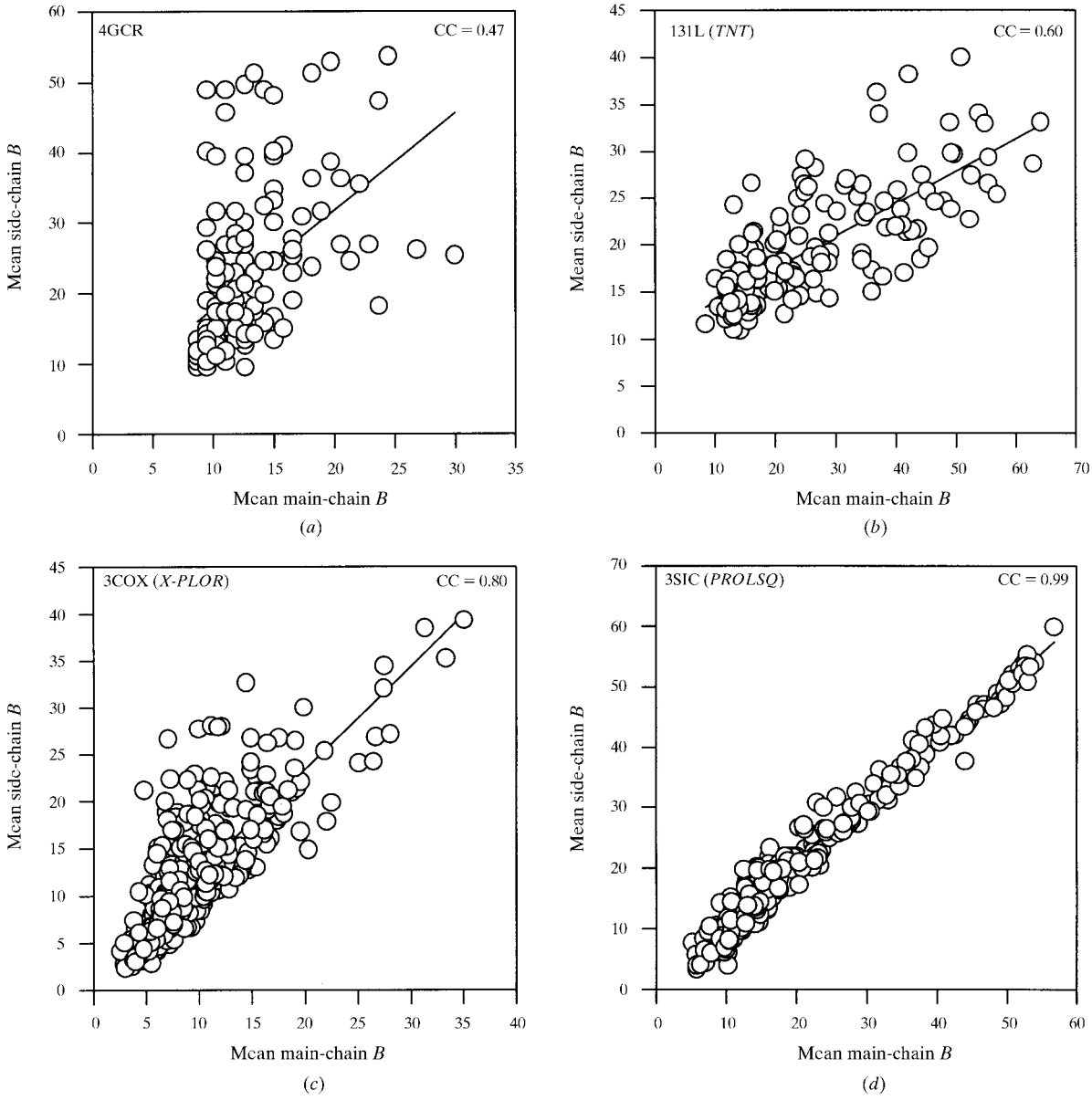

Figure 2
Plots illustrating the correlation between the mean B values of the main-chain atoms and the side-chain atoms. (*a*) 4GCR (γ-B crystallin), (*b*) 131L (T4 lysozyme mutant), (*c*) 3COX (cholesterol oxidase), (*d*) 3SIC (subtilisin–inhibitor complex).

Gaussian functions (1). The average k_1, k_2, k_3, k_4, B_1 and B_2 values over all the proteins are 27.59, 0.82, 5.86, 11.72, 0.27 and 7.72, respectively. The maximum in the distribution is usually around $B' = -0.57$. Also, the B_1 and B_2 values are within 10% for 93 and 92 structures, respectively, out of the 95 used in the analysis. However, in a few cases, such as rice ferricytochrome *c* (1CCR; Ochi *et al.*, 1983) and transforming growth factor β2 (2TGI; Daopin *et al.*, 1992), the fitting of the two Gaussians was not good. The observed distribution for these proteins is very different from the overall distribution (Fig. 1), suggesting possible problems in the refinement of these structures or unusual dynamics (thermal properties) of the polypeptide chains (see §3.7).

3.2. Correlation coefficients

Fig. 2 shows representative examples of scatter plots illustrating the correlation between the mean *B* values of the main-chain and the side-chain atoms for individual proteins. The lowest, highest and mean values of CC are 0.47 (for γ-crystallin; 4GCR), 0.99 (for subtilisin–inhibitor complex, 3SIC) and 0.76, respectively. For 61% of the structures studied, the CC values fall between 0.7 and 0.9. The contrast between the plots for 4GCR (Fig. 2*a*) and 3SIC (Fig. 2*d*) is striking. Although many proteins have high CCs, the range of values observed for CCs is large. The number of protein structures in different CC ranges for all the proteins and for those refined using *X-PLOR*, *PROLSQ* and *TNT* are shown as histograms in Fig. 3. The most frequent CC range for all structures is 0.8–0.9 (36 out of 95). The distribution in different CC ranges for proteins refined using *X-PLOR* resembles closely the distribution for all the structures. Structures refined with *PROLSQ* have a similar trend, although the number of structures in the CC ranges 0.7–0.8 and 0.8–0.9 are comparable. On the contrary, the structures refined using *TNT* are spread more uniformly over the CC range 0.5–0.8. Thus, the program used for refinement appears to influence the spread of CCs. The very high CC value observed in a few cases (for example, 3SIC, subtilisin–inhibitor complex, Fig. 2*d* and Table 1), could be a result of incomplete refinement.

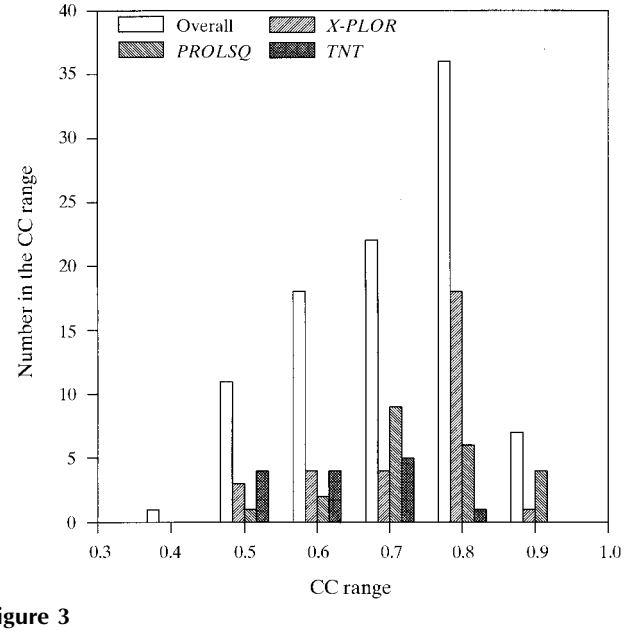

Figure 3
Histogram illustrating the number of structures in different ranges of CC.

3.3. Dependence of correlation coefficients on experimental temperature

The structure of hen egg-white lysozyme has been determined to a resolution of 1.7 Å at different temperatures: 95, 120, 180, 250, 280 and 295 K (Kurinov & Harrison, 1995). The CCs for all these structures are around 0.57 and vary between 0.53 and 0.63 (Table 2). It is clear that the experimental temperature does not have a significant effect on the correlation. The maximum correlation is observed for the inter-

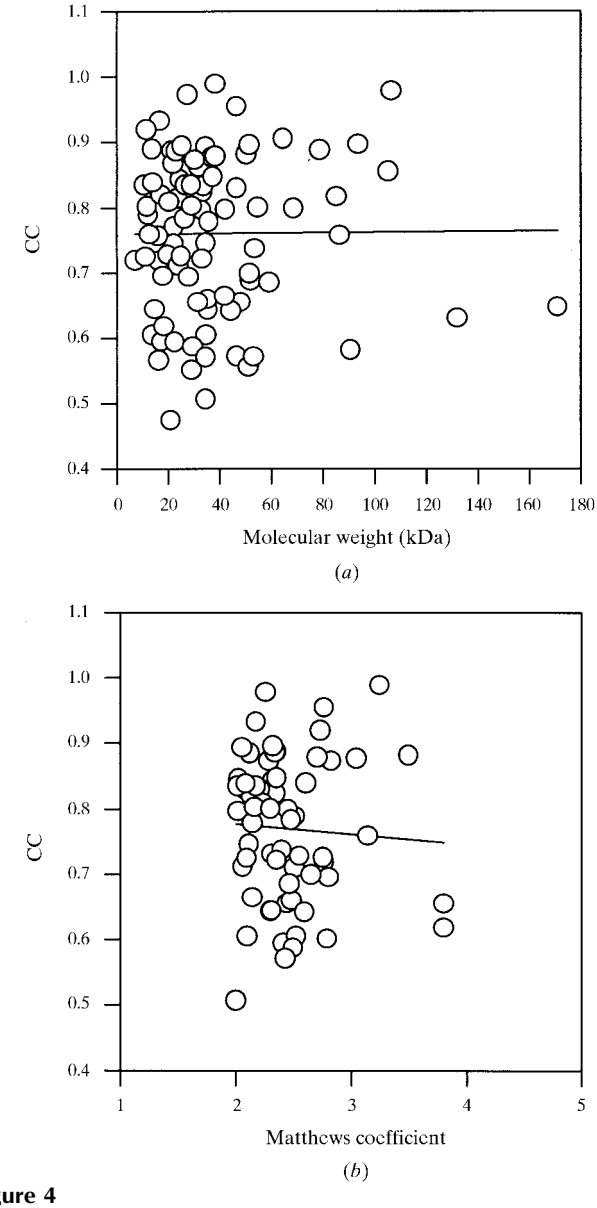

Figure 4
Scatter plot of CC values and (*a*) protein molecular weights, (*b*) Matthews coefficients.

Table 2
CC values for lysozyme structures determined at different temperatures.

Temperature (K)	CC value	PDB code
95	0.5995	1LSF
120	0.5648	1LSA
180	0.6125	1LSB
250	0.6302	1LSC
280	0.5784	1LSD
295	0.5257	1LSE

mediate temperature of 250 K. At lower temperatures, freezing of different but distinct conformations of flexible segments of side chain might reduce the CC. At higher temperatures, differential increase in B values of the main-chain and the side-chain atoms might lead to reduced correlation.

3.4. Dependence of correlation coefficients on protein molecular weights and solvent content

The fraction of residues inaccessible to solvent in a folded protein is likely to depend on the protein size. In small proteins, most side chains may be accessible to the solvent. The degree to which the thermal parameters of the side-chain atoms could be different from those of the main-chain atoms might depend on their solvent accessibility. Therefore, it would not be surprising if a correlation exists between the CCs and the molecular weight, assuming that globular proteins of large molecular weight also have a large proportion of residues buried. Fig. 4(*a*) shows a plot of CC against molecular weight. No significant dependence of CC on molecular weight is observed. Another parameter to which CCs are likely to be related is the Matthews coefficient (Matthews, 1968). A low

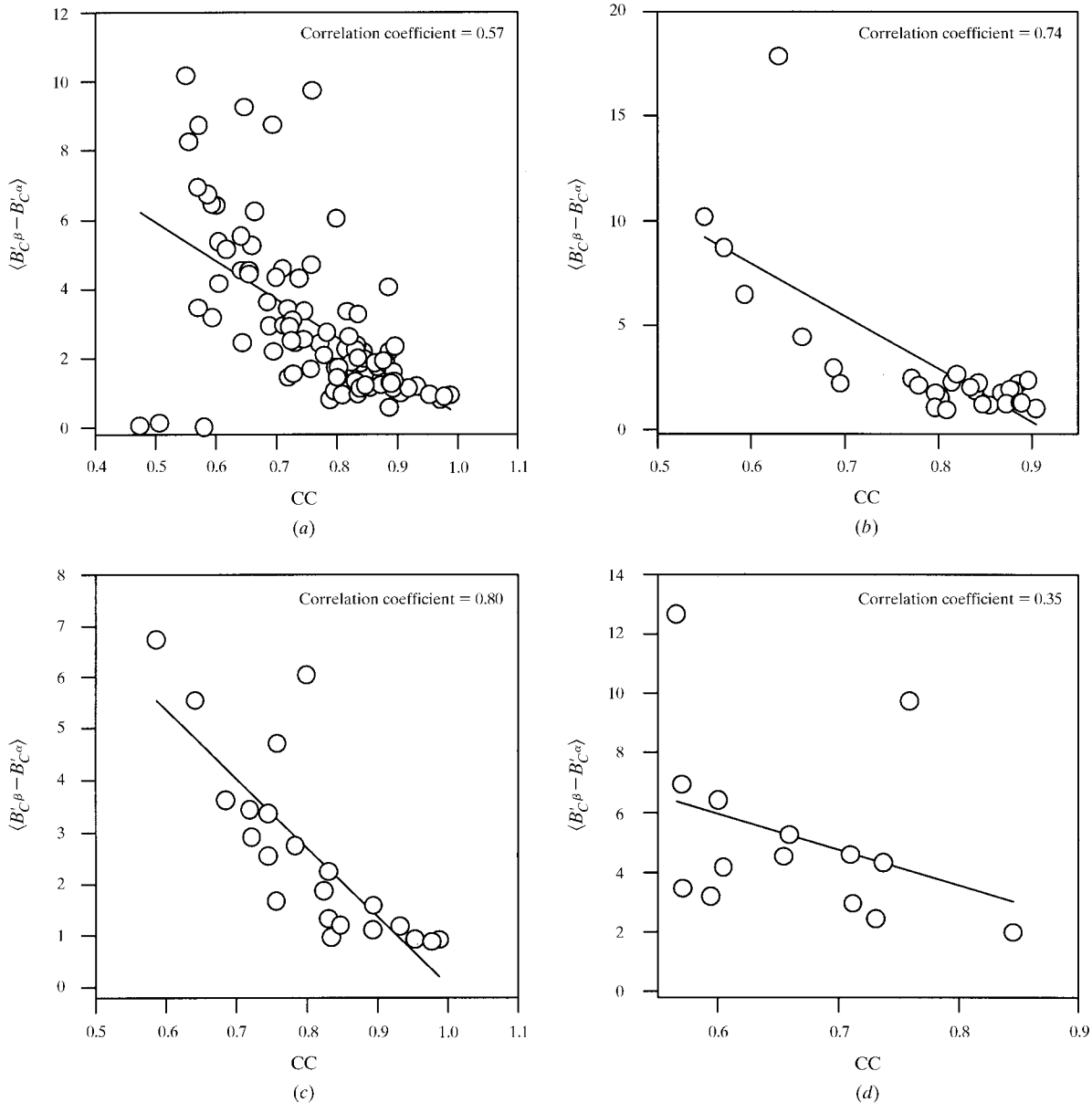

Figure 5
Plots illustrating the dependence of the CC values and mean difference in the B' factors of C^α and C^β atoms. (*a*) All structures, (*b*) structures refined using *X-PLOR*, (*c*) structures refined using *PROLSQ*, (*d*) structures refined using *TNT*.

Matthews coefficient represents a closely packed structure where the thermal parameters are less likely to be independent. Fig. 4(*b*) is a scatter plot of the CC values against the Matthews coefficients for all proteins for which the value could be unambiguously deduced. It is clear that no significant correlation exits.

3.5. Dependence of CC on the mean difference of the B' factors between C^{α} and C^{β} atoms

The correlation between the B values of the main-chain and the side-chain atoms might depend on the size of the difference in the B values between consecutive atoms of the side chain starting from C^{α}. If this value is constrained to be small in the refinement protocol, it will lead to a large CC as well as to long stretches of consecutive C^{α} atoms with large or small B values (see §3.6). Fig. 5 illustrates the relationship between CC values and the mean difference between B' factors of C^{α} and C^{β} atoms for all structures (Fig. 5*a*) and for structures refined using *X-PLOR* (Fig. 5*b*), *PROLSQ* (Fig. 5*c*) and *TNT* (Fig. 5*d*). It is obvious that CCs are correlated to the size of the differences. The correlation coefficient between $\langle B'_{C^{\beta}} - B'_{C^{\alpha}} \rangle$ (mean difference of the B' factors of C^{α} and C^{β} atoms) and the CC values are 0.74, 0.80 and 0.35 for structures refined with *X-PLOR*, *PROLSQ* and *TNT*, respectively, and these average to 0.57 over all the structures. The $\langle B'_{C^{\beta}} - B'_{C^{\alpha}} \rangle$ is not as tightly related to CC in structures refined with *TNT* when compared to those refined with *X-PLOR* or *PROLSQ*. Also, in sets of structures refined with any one of the packages, there are proteins which are clearly outliers in the sense that they have values of $\langle B'_{C^{\beta}} - B'_{C^{\alpha}} \rangle$ which are too large considering the respective CC. These points fall at positions significantly above the regression lines of Fig. 5. These observations probably result from the different constraints used in the three refinement packages currently in extensive use. However, there is a wide variation in the CCs, even for structures refined using any of the individual packages.

3.6. Frequency distribution of high *B*-value stretches

Frequency distribution of high B-value stretches (see §2.4) were calculated for sets of proteins having low CCs (0.5–0.6) and high CCs (0.85–0.95). Table 3 illustrates this distribution. The fraction of residues that occur with high B values is about the same in proteins with low and high correlation coefficients (739 out of 3303 low-CC proteins and 612 out of 2691 high-CC proteins occur with high B value, amounting to 22.4 and 22.7%, respectively). It is obvious that in proteins with low CCs, very long stretches of high B values are not observed when compared to proteins with high CCs. In contrast, short stretches of high B values occur more frequently in proteins with low CCs.

3.7. Possible validation tools

The correlation coefficient between the B'-factor frequency distribution for an individual protein and that of the overall distribution (Fig. 1) averages to 0.97 with an r.m.s. deviation of 0.03. However, for the structures 1CCR and 2TGI the correlation of observed distribution to the overall distribution is unusually low (0.84 and 0.80, respectively). These proteins are cytochrome *c* (1CCR) and transforming growth factor (2TGI), both of which lack the hydrophobic core usually found in the structures of other proteins. The 1CCR structure is loosely wound around the haem group while that of 2TGI has most of its side-chain groups accessible to solvent. The special feature of the latter structure has been noted by the original investigators. Thus, the B'-factor frequency distribution could highlight protein structures with unusual dynamics.

Table 3
Length distribution of stretches with high B values at C^{α} atoms in structures with low CC (0.5–0.6) and high CC (0.85–0.95) values.

The numbers in brackets are the percentage of total residues that belong to these segments.

Length	Low CC	High CC
1	213 (6.4)	37 (1.4)
2	78 (4.7)	16 (1.2)
3	26 (2.4)	12 (1.3)
4	18 (2.2)	15 (2.2)
5	7 (1.1)	6 (1.1)
6	7 (1.3)	9 (2.0)
7	11 (2.3)	3 (0.8)
8	2 (0.5)	10 (3.0)
9	2 (0.5)	3 (1.0)
10	2 (0.6)	3 (1.1)
11	0 (0.0)	4 (1.6)
12	1 (0.4)	3 (1.3)
14	0 (0.0)	1 (0.5)
15	0 (0.0)	1 (0.6)

Choloromuconate cycloisomerase and ferridoxin from *A. vinelandii* are two examples of structures that were initially refined incorrectly and subsequently corrected. The B'-factor frequency distribution for the initial parameters (1CHR and 2FD1; Hoier *et al.*, 1994; Ghosh *et al.*, 1982) and later corrected versions (2CHR and 5FD1; Kleywegt *et al.*, 1996; Stout, 1993) are shown in Fig. 6. It is evident that the corrected structures have the anticipated distribution, while the incorrectly refined structures depart appreciably from the overall distribution. The correlation, between overall distribution and 2FD1 and 5FD1 are 0.79 and 0.95, respectively. Corresponding values for chloromuconate cycloisomerase are 0.91 and 0.99, respectively. Thus, it appears that the B'-factor frequency distribution could be used to qualitatively validate the final reported B values.

4. Discussion

The diffraction data from protein crystals do not provide a sufficient number of reflections to refine the positional and individual thermal parameters of all the atoms independently in most cases. The geometry of the chemical groups of proteins are known in sufficient detail from small-molecule structures to deduce very reasonable constraints on the bond lengths, angles and torsion angles, which in turn help to reduce the effective parameters that are refined independently. However, the constraints that are suitable for the refinement of thermal parameters are less certain. Reasonable constraints for the

thermal parameters of bonded atoms and atoms bonded to the same central atom have been deduced and incorporated into refinement protocols (Konnert & Hendrickson, 1980; Tronrud, 1996).

The thermal parameters associated with the main-chain and the side-chain atoms exhibit a high degree of correlation in a large fraction of protein structures determined at high resolution. However, CCs vary between 0.47 and 0.99. This variation was found not to be correlated to the size of proteins or to the packing density in the crystals. The size of the differences in the B' factors of bonded atoms (in particular, C^{α} and C^{β} atoms) was found to be related to the observed correlation. The observed correlation was also related to the frequency of occurrence of longer stretches of consecutive C^{α} atoms with high B values. These properties may, in fact, vary from protein to protein. On the other hand, the observations could be related to the method of refinement used in the determination of the protein structures. For example, the correlation might be related to the relative weights assigned to the energy and crystallographic terms in the refinement protocols. If this is the reason for the observed variations in correlation, it has important implications for protein refinement and necessitates a closer scrutiny of the weights assigned to various terms in protein refinement (Smith, 1997).

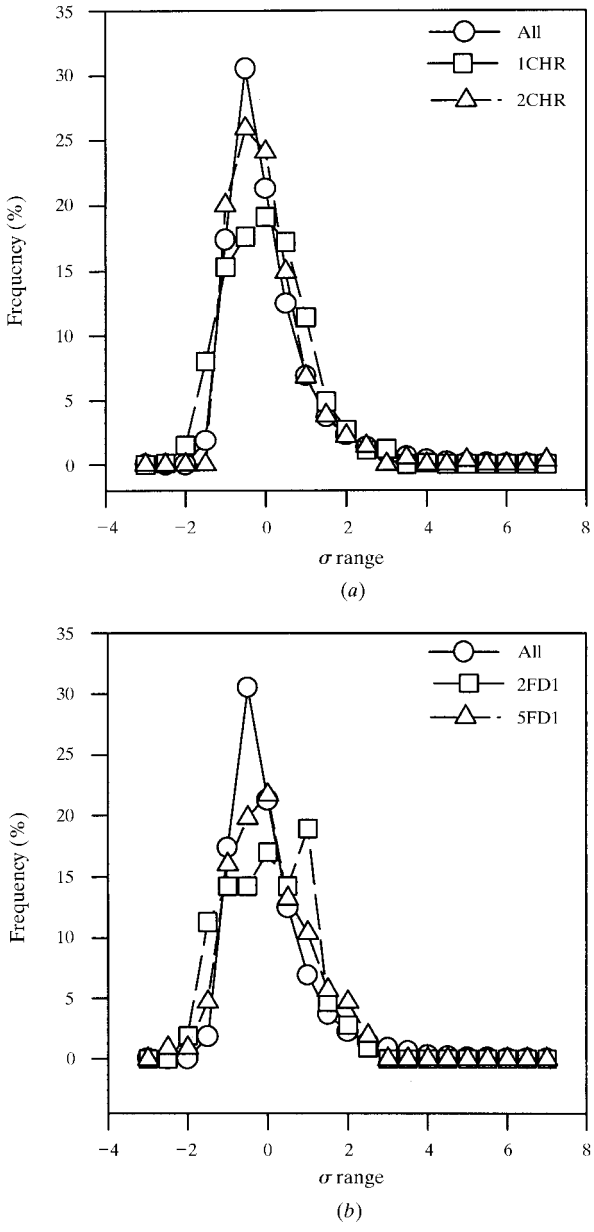

Figure 6
Observed B'-factor frequency distribution over all the structures (○), initial erroneous structures (□) and subsequently corrected structures (△). (*a*) Chloromuconate cycloisomerase (1CHR and 2CHR), (*b*) ferridoxin (2FD1 and 5FD1).

The analysis presented here illustrates that there are differences in the distribution of CCs for structures refined using different programs (Fig. 3). The degree of dependence of the CC on the mean difference in the B' factors at C^{α} and C^{β} atoms is also related to the refinement package used (Fig. 5). These results could be attributed to the differences in the way B-value refinement is handled in these packages. However, it is

unlikely that the choice of the package is the sole source of the observed variations. More frequent observation of longer stretches of consecutive high B values (Table 3), smaller difference between the B' factors of C^{α} and C^{β} atoms and high CCs are almost always found simultaneously. These structures might have been refined with larger weighting of the X-ray term relative to weighting of stereochemical constraints when compared to structures with lower CCs. Significantly, the outliers in Fig. 5 with large values for mean difference between B' factors of C^{α} and C^{β} atoms and low CCs might represent over-refined structures. The proteins for which B'-factor distribution departed appreciably from the overall distribution were found to have structures devoid of a hydrophobic core and hence were likely to possess genuinely novel dynamical properties. Similar departure from overall distribution was also observed in those cases where the initial models were wrong, suggesting that the B'-factor distribution could be used as a qualitative validation tool. It is of obvious importance to separate real protein attributes from the results obtained as a consequence of the constraints used in refinement and the weights associated with various terms in the quantity minimized. The analysis presented in this communication brings to focus some of the plausible artifacts of protein thermal-parameter refinement and it is hoped that this analysis will stimulate further investigations leading to better refinement procedures.

References

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchui, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.
Brünger, A. T., Kuriyan, K. & Karplus, M. (1987). *Science*, **235**, 458–460.
Daopin, S., Piez, K. A., Ogawa, Y. & Davies, D. R. (1992). *Science*, **257**, 369–372.
Dauter, Z., Lamzin, V. S. & Wilson, K. S (1997). *Curr. Opin. Struct. Biol.* **7**, 681–688.

Ghosh, D., Donnell, S. O., Furey, W., Robbins, A. H. Jr & Stout, C. D. (1982). *J. Mol. Biol.* **158**, 73–109.
Hendrickson, W. A. & Konnert, J. H. (1980). *Computing in Crystallography*, edited by R. Diamond & K. Venkatesan, ch. 13, pp. 13.01–13.26. Bangalore: Indian Academy of Sciences.
Hobohm, U. & Sander, C. (1994). *Protein Sci.* **3**, 522–524.
Hoier, H., Schlomann, M., Hammer, A., Glusker, J. P., Carrell, H. L., Goldman, A., Stezowski, J. J. & Heinemann, U. (1994). *Acta Cryst.* D**50**, 75–84.
Karplus, P. A. & Schulz, G. E. (1985). *Naturwissenschaften*, **72**, 212–213.
Kleywegt, G. J., Hoier, H. & Jones, A. T. (1996). *Acta Cryst.* D**52**, 858–863.
Konnert, J. H. & Hendrickson, W. A. (1980). *Acta Cryst.* A**36**, 344–350.
Kurinov, I. V. & Harrison, R. W. (1995). *Acta Cryst.* D**51**, 98–109.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Ochi, H., Hata, Y., Tanaka, N., Kakudo, M., Sakurai, T., Aihara, S. & Morita, Y. (1983). *J. Mol. Biol.* **166**, 407–418.
Parthasarathy, S. & Murthy, M. R. N. (1997). *Protein Sci.* **6**, 2561–2567.
Parthasarathy, S. & Murthy, M. R. N. (1998). *Protein Sci.* **7**, 525.
Peters-Libeu, C. & Adman, E. T. (1997). *Acta Cryst.* D**53**, 56–77.
Ragone, R., Facchinao, F., Facchinao, A., Facchiano, A. M. & Colonna, G. (1989). *Protein Eng.* **2**, 497–504.
Ringe, D. & Petsko, G. A. (1986). *Methods Enzymol.* **131**, 389–433.
Smith, G. D. (1997) *Acta Cryst.* D**53**, 41–48.
Stout, C. D. (1993). *J. Biol. Chem.*, **268**, 25920–25927.
Tronrud, D. E. (1996). *J. Appl. Cryst.* **29**, 100–104.
Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). *Acta Cryst.* A**43**, 489–501.
Vihinen, M., Torkkila, E. & Riikonen, P. (1994). *Proteins Struct. Funct. Genet.* **19**, 141–149.