Atomic displacement parameters of homologous proteins: Conservation of dynamics

V. M. S. Lenin[†], S. Parthasarathy[‡] and M. R. N. Murthy[‡,*]

[†]Industrial Biotechnology, Anna University, Chennai 600 025, India
[‡]Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

Atomic displacement parameters (ADPs) obtained from high-resolution refinement of protein structures represent the mean square displacement of protein atoms from their centroid positions. They contain information regarding the flexibility of the polypeptide. Comparative analysis of the ADPs in homologous proteins shows that the local flexibility of the polypeptide is not correlated to the mutability of the segment. The flexible and rigid regions in the three-dimensional fold of proteins remain largely conserved during the course of evolution. In related proteins, the variation in the flexibility of a given segment is only weakly correlated to the variation of the amino acid sequence at the corresponding positions. These results illustrate that the relationship between sequence and dynamics has degeneracy similar to that of sequence and three-dimensional structure. The observations are consistent with the importance of protein flexibility to protein function.

COMPARATIVE analysis of protein structures has shown that proteins with no detectable sequence similarity could be homologous[1]. It is known that homologous proteins with similar folds can perform very different functions, (e.g. TIM-barrels) while non-homologous proteins with very different folds can be functionally similar, (e.g. serine proteinase inhibitors). The function is determined mainly by the stereo-chemistry and dynamics of the few residues or atoms that constitute the active site. Movement of loops that are located far from the active sites might also play an important role in catalysis. It is of interest, therefore, to study the variation in the dynamics of proteins during the course of evolution. Towards this end an analysis of crystallographic Atomic Displacement Parameters (ADP, *B*-values) of high resolution structures is presented in this article.

In X-ray diffraction studies, intensities of Bragg reflections fall-off with increasing resolution. This intensity fall-off is due both to static disorder and dynamics of the molecule in the crystal. In the structure factor equation, $F = \Sigma f_o \exp(-B \sin 2\theta/\lambda^2)$ the exponential term describes this fall-off in intensity. In protein crystallography, the ADPs are expressed as *B*-values, $B = 8\pi^2\langle u^2\rangle$, where $\langle u^2\rangle$ is the average of the mean square atomic displacements along the three coordinate axes and is given by $(u_x^2 + u_y^2 + u_z^2)/3$ (isotropic approximation). Analyses of the atomic displacement parameters have been carried out to derive flexibility indices of amino acid residues[2,3]. These indices have been used to predict antigenic regions along the polypeptide chain.

Refinement of *B*-values for protein structures is more complicated than that of atomic positions. Different refinement packages apply different restraints on the *B*-values. The final *B*-values are also affected by the weighting schemes followed by individual crystallographers[4]. The frequency distribution of *B*-values, however, in a given protein structure is very characteristic and can be analytically expressed as the summation of two Gaussian functions[5].

Even in the crystalline state, protein atoms are in constant motion about their mean positions. This dynamics or flexibility is essential for activity or function. On the other hand, structural stability requires rigidity. The core of globular proteins is usually densely packed with apolar residues. Surface residues are generally more mobile due to fewer stabilizing interactions. Apart from higher flexibility, they also tend to display larger sequence variation. Further, it is assumed that the regions of the polypeptide that play a decisive role in function need to conserve their dynamics as 'enzyme eye-views' of evolution[6]. It is of interest, therefore, to investigate whether *B*-values can be correlated to the mutability of residues in globular proteins and to make an assessment of the degree of change in the *B*-values of structurally equivalent residues in the course of evolution.

We present here, an analysis of the *B*-values of representative haemoglobin structures and show that the dynamics of these polypeptide chains are conserved in spite of having very low sequence similarity. We have extended the analysis to all high-resolution haemoglobin structures (representing α class), trypsin structures (representing β class) and to triosephosphate isomerase (representing α/β class) to correlate the changes in *B*-values of structurally equivalent resides to amino acid replacements.

Methodology

Selection of structures

Ten representative structures of haemoglobin chains with resolutions of 2.5 Å or better were extracted from the Protein Data Bank (PDB)[7]. Maximum sequence similarity between any pair among these ten structures is 28%. The

*For correspondence. (e-mail: mrn@mbu.iisc.ernet.in)

listing, PDB codes and sources of the ten structures used for the study are given in Table 1. The analysis was extended to all the native structures of haemoglobin currently available in the PDB with resolution 2.0 Å or better (Table 2) and also to structures of trypsin (Table 3) representing β class, and triose phosphate isomerase (TIM; Table 4) representing the α/β class of proteins. The resolution was better than or equal to 2.2 Å for trypsin structures and 2.8 Å for TIM structures, respectively. The maximum sequence identity between any two pairs of structures was 77% and 84% for trypsin and TIM, respectively.

Table 1. List of representative haemoglobin structures used in the analysis

PDB code	Denoted as	Source	Resolution of structure (Å)
1ASH	1ASH	*Ascaris* recombinant haemoglobin expressed in *E. coli*	2.15
1ECA	1ECA	Haemoglobin from *Chironomous thummi thummi*	1.4
1FLP	1FLP	Haemoglobin I from *Lucina pectinata*	1.5
1HLB	1HLB	Sea cucumber haemoglobin	2.5
1ITH	1ITH	Innkeeper worm haemoglobin A chain	2.5
1PBX	1PBXA	Antarctic fish haemoglobin A chain	2.5
1PBX	1PBXB	Antarctic fish haemoglobin B chain	2.5
2HBG	2HBG	Marine bloodworm haemoglobin	1.5
2LHB	2LHB	Sea lamprey haemoglobin	2.0
3SDH	3SDHA	Ark clam haemoglobin A chain	1.4

Table 2. List of all haemoglobin structures used in the analysis

PDB code	Denoted as	Source	Resolution of structure (Å)
1ASH	1ASH	*Ascaris* recombinant haemoglobin expressed in *E. coli*	2.15
1ECA	1ECA	Haemoglobin from *Chironomous thummi thummi*	1.4
1FLP	1FLP	Haemoglobin I from *Lucina pectinata*	1.5
2HBG	2HBG	Marine bloodworm haemoglobin	1.5
2LHB	2LHB	Sea lamprey haemoglobin	2.0
3SDH	3SDHA	Ark clam haemoglobin A chain	1.4
1A3N	1A3NA	Human deoxy haemoglobin A chain	1.8
..	1A3NB	Human deoxy haemoglobin B chain	1.8
1A4F	1A4FA	Haemoglobin A chain from *Anser indicus*	2.0
..	1A4FB	Haemoglobin B chain from *Anser indicus*	2.0
1CG5	1CG5A	Haemoglobin A chain from *Dasyatis akajei*	1.6
..	1CG5B	Haemoglobin B chain from *Dasyatis akajei*	1.6
1HBH	1HBHA	Antarctic fish haemoglobin A chain	2.2
..	1HBHB	Antarctic fish haemoglobin B chain	2.2
1HBR	1HBRA	Chicken haemoglobin A chain	2.3
..	1HBRB	Chicken haemoglobin B chain	2.3
1HDA	1HDAA	Bovine haemoglobin A chain	2.2
..	1HDAB	Bovine haemoglobin B chain	2.2
1HDS	1HDSA	Virginia white-tailed deer A chain	1.98
..	1HDSB	Virginia white-tailed deer B chain	1.98
1IBE	1IBEA	Horse haemoglobin A chain	1.8
..	1IBEB	Horse haemoglobin B chain	1.8
1QPW	1QPWA	Porcine haemoglobin A chain	1.8
..	1QPWB	Porcine haemoglobin B chain	1.8
1SPG	1SPGA	Teleost fish (*Leiostomus xanthurus*) haemoglobin A chain	1.95
..	1SPGB	Teleost fish (*Leiostomus xanthurus*) haemoglobin B chain	1.95
1TIN	1TINA	Fish haemoglobin A chain	2.2
..	1TINB	Fish haemoglobin B chain	2.2

Normalized B-values

For all analyses, B-values of the Cα atoms of the residues alone were considered. Average B-values vary widely between different structures. Therefore, to compare different protein structures B-values at Cα atoms were replaced by normalized B-values (B'-factors) as, $B' = (B_{C\alpha} - \langle B \rangle_{C\alpha})/\sigma(B)$, where $\sigma(B)$ is the standard deviation in B-values of Cα atoms.

Multiple alignment of protein sequences

Relating the B-factors of a family of proteins to a parameter such as mutability requires the identification of sequentially equivalent residues in the same protein family. Multiple sequence alignment was done using the PileUp program present in the Wisconsin Package[8]. For all sequence alignments, the standard scoring matrix provided along with the PileUp program was utilized. The

Table 3. List of trypsin structures used in the analysis

PDB code	Denoted as	Source	Resolution of structure (Å)
1AOJ	1AOJ	North Atlantic Salmon (*Salmo salar*) trypsin	1.7
1MCT	1MCT	Porcine trypsin	1.6
1SGT	1SGT	*Streptomyces griesus*	1.7
1TRN	1TRN	Human trypsin	2.2
1TRY	1TRY	*Fusarium oxysporum* trypsin	1.55
3TGI	3TGI	*Rattus* norvegicus trypsin	1.8
5PTP	5PTP	Bovine trypsin	1.34

Table 4. List of trosephosphate isomerase structures used in the analysis

PDB code	Denoted as	Source	Resolution of structure (Å)
1AMK	1AMK	*Leishmania mexicana*	1.83
1AW2	1AW2	*Vibrio marinus*	2.65
1BTM	1BTM	*Bacillus stearothermophilus*	2.8
1HTL	1HTL	Human TIM	2.8
1TCD	1TCD	*Trypanosoma cruzi*	1.83
1TPF	1TPF	*Trypanosoma brucei brucei*	1.8
1TPH	1TPH	Chicken TIM (*Gallus gallus*)	1.8
1TRE	1TRE	*Escherichia coli*	2.6
1YDV	1YDV	*Plasmodium falciparum*	2.2
1YPI	1YPI	*Saccharomyces cerevisiae*	1.9

correlation coefficients of B-values at aligned positions between these structures were evaluated as,

$$\Sigma_i(B_{1i} - \langle B_1\rangle)(B_{2i} - \langle B_2\rangle) / [\Sigma_i(B_{1i} - \langle B_1\rangle)^2 \Sigma(B_{2i} - \langle B_2\rangle)^2]^{1/2},$$

where B_{1i} is the B-value at position i in protein 1 and B_{2i} is the B-value at the same position in protein 2 and $\langle B_1\rangle$, $\langle B_2\rangle$ are the averages of B-values in structures 1 and 2, respectively.

Correlation between B-values and amino acid replacements

The correlation between B-values and the mutability were studied in terms of the following:

Per cent amino acid replacements. For a given alignment, for each residue number, pair-wise comparison of the ten structures was carried out and the total number of amino acid replacements in all the combinations (${}^nC_2$, where n is the number of sequences) and the percentage of such replacements were determined. Deletions were ignored in that the amino-acid replacements were counted only for those pairs for which neither residue was a gap.

Dayhoff's scores. For a given alignment, the Dayhoff's score[9] of the kth residue number was calculated by the pair-wise comparison of the structures as $D_k = \Sigma D_{ij}$ for all unique pairs i and j among the ten structures, where D_{ij} is the Dayhoff's score between the amino-acid residues at the kth position in structures i and j according to the Dayhoff's similarity score matrix. Deletions were allotted maximum penalty by assigning a score of zero between a deletion and any other amino acid or another deletion.

Smoothening of B-value and other profiles

The smoothened B-value of the nth residue, under a window size, w, is defined as the weighted average of the B-values of the w consecutive residues with the nth residue at the center. Variable weights assigned to the residues increase in arithmetic progression from a certain minimum weight at the residues farthest from the centre to a maximum weight of 1 at the central residue. Therefore, the smoothened B-value of the nth residue under a window size of 5 and a minimum weight of 0.25 would be the weighted average of the $n-2$, $n-1$, n, $n+1$ and $n+2$th residues with weights 0.25, 0.625, 1, 0.625 and 0.25, respectively.

Results

Analysis of representative haemoglobin structures

B-value profile along the sequence. Sequence alignment of the ten representative haemoglobin structures using conventional sequence alignment programs resulted in an alignment that did not reflect the actual evolutionary relationship between them because of their very low sequence similarity. To circumvent this, the alignment of these ten sequences was carried out in the presence of a large number of haemoglobin sequences from various other sources so as to form a chain of relationships that might link together the seemingly unrelated sequences of the representative structures. To this end, a total of 657 other haemoglobin sequences were obtained from the PIR library accompanying the Wisconsin Package Version 9.0. The

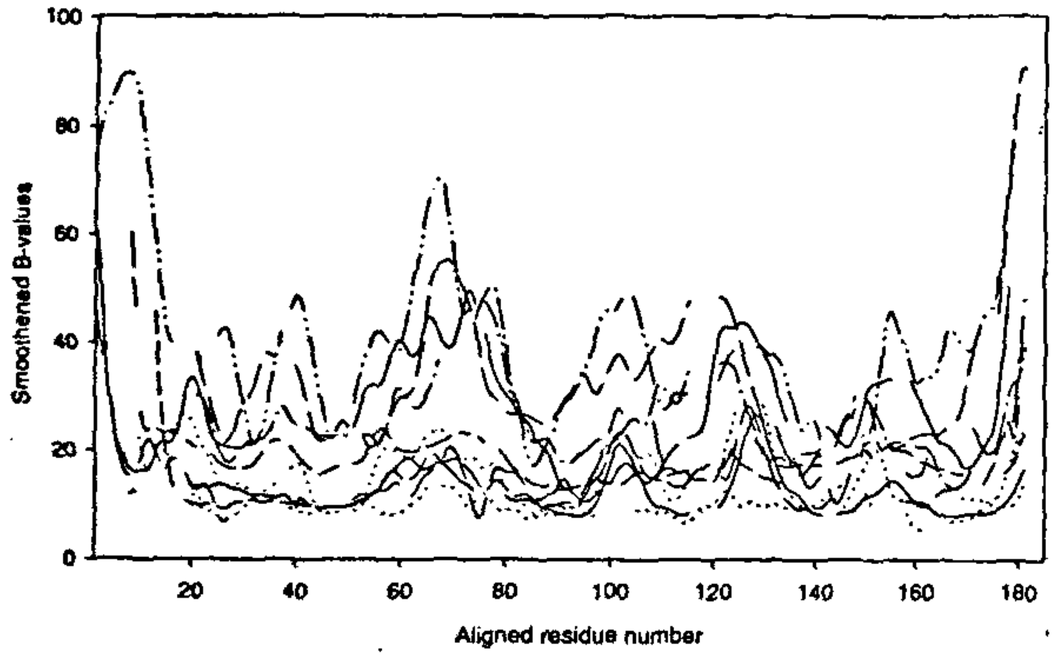


Figure 1. Plot of B-values of representative structures of haemoglobin smoothened with a window size of 5 and variable weightage against aligned residue number.

Table 5. Correlation coefficients between B-values of representative structures of haemoglobin

	1ASH	1ECA	1FLP	1HLB	1ITH	1PBXA	1PBXB	2HBG	1LHB	3SDHA
1ASH	1.00									
1ECA	0.28	1.00								
1FLP	0.33	0.30	1.00							
1HLB	0.21	0.22	0.55	1.00						
1ITH	0.22	0.14	0.41	0.41	1.00					
1PBXA	0.53	0.28	0.42	0.29	0.10	1.00				
1PXAB	0.37	0.30	0.79	0.64	0.53	0.49	1.00			
2HBG	0.47	0.17	0.29	0.39	0.34	0.17	0.16	1.00		
2LHB	0.48	0.19	0.34	0.54	0.36	0.37	0.43	0.77	1.00	
3SDHA	0.40	0.37	0.30	0.50	0.28	0.35	0.23	0.62	0.62	1.00

PileUp program, however, can align atmost 500 sequences at a time. Therefore, alignments were carried out using 490 sequences chosen randomly from the 657 PIR sequences and the sequences of the 10 structures of interest. Ten such random sets were chosen and multiply aligned with gap creation penalty = 12, gap extension penalty = 4 and end-weight = 0.25.

The alignments of the ten sequences were not identical in the ten sets. However, large trends of similarity could be observed in all the ten alignments, especially between residues 51 and 150 of the alignments. All the plots shown here with reference to the representative structures of haemoglobin correspond to one of these ten alignments, which was chosen randomly, viz. alignment number 7. The plot of smoothened B-values (window size 5) versus aligned sequence number is shown in Figure 1. But for a few minor differences, all the ten alignments produce similar profiles for this plot. All the ten structures exhibit large humps between aligned residue numbers 51 and 85 and between 111 and 135, approximately.

Treating gaps in the aligned sequences as deletions, the correlation coefficients between the B-values of the aligned residues were determined for each of the 45 pairs of structures. The correlation coefficients for alignment 7 are shown in Table 5. It can be seen that the correlation coefficients between the B-values of most pairs were significant. This suggests that the dynamics of certain segments were conserved even when the sequence similarity has almost disappeared.

Correlation between B-values and sequence similarity

Plots of smoothened B'-factors and normalized percentage amino acid replacements against residue number suggested that the two parameters do not have significant correlation (data not shown). Similarly, no significant correlation was observed between Dayhoff scores and B'-factors.

The correlation between the B-values will be unity if there is complete sequence identity between two structures. However, as the differences in sequences accumulate, the correlation between the B-values is likely to decrease. To study the profile of B'-factors versus sequence similarity, the correlation coefficients of B-factors of all possible pairs chosen from the ten structures were determined. These were plotted against the sum of the Dayhoff's scores over all the residues and sequence identities of the corresponding pairs of structures (Figure 2). The correlation coefficient was low.

Since Dayhoff's score measures the total amino acid invariance at a particular residue, it is likely to be correlated with a parameter that measures the total B'-variation at that particular residue. In order to measure the B'-variation at a particular position, a normalized parameter $\langle|\delta B'|\rangle$ was defined. Here $|\delta B'|$ refers to the absolute value of the difference in B'-values between a pair of residues at a particular position and $\langle|\delta B'|\rangle$ refers to the average of such differences over all possible pairs of residues at that position. Figure 3 *a* shows the profiles of smoothened D′-

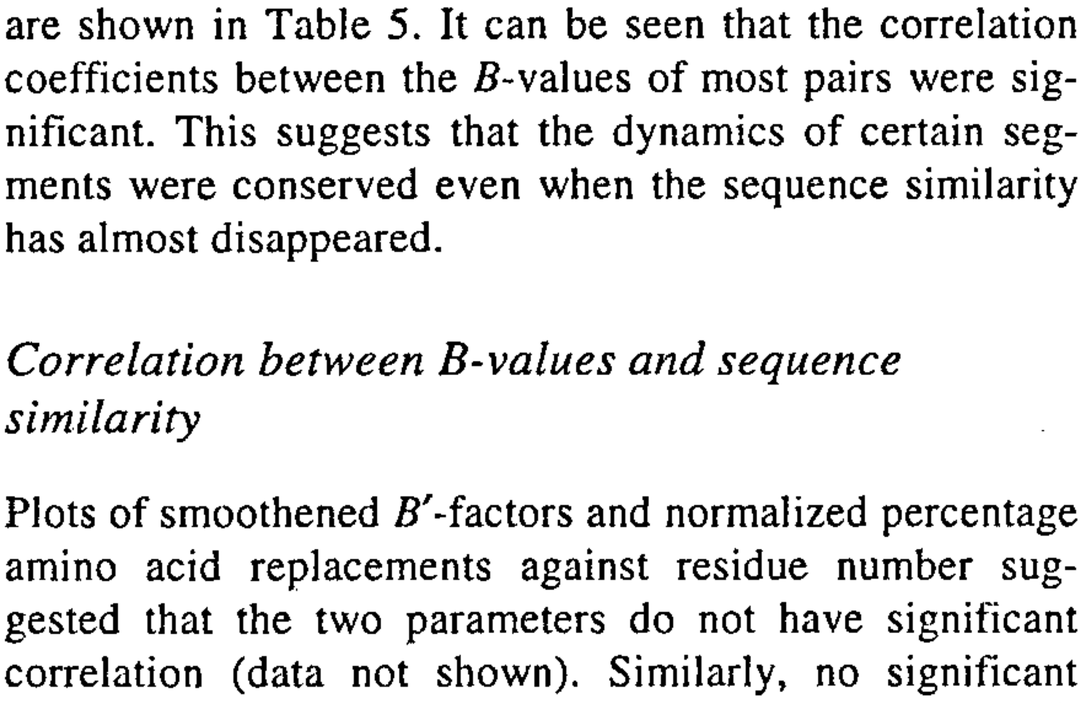


Figure 2. Scatter plot of pair-wise correlation coefficients of B-values and pair-wise Dayhoff's scores.

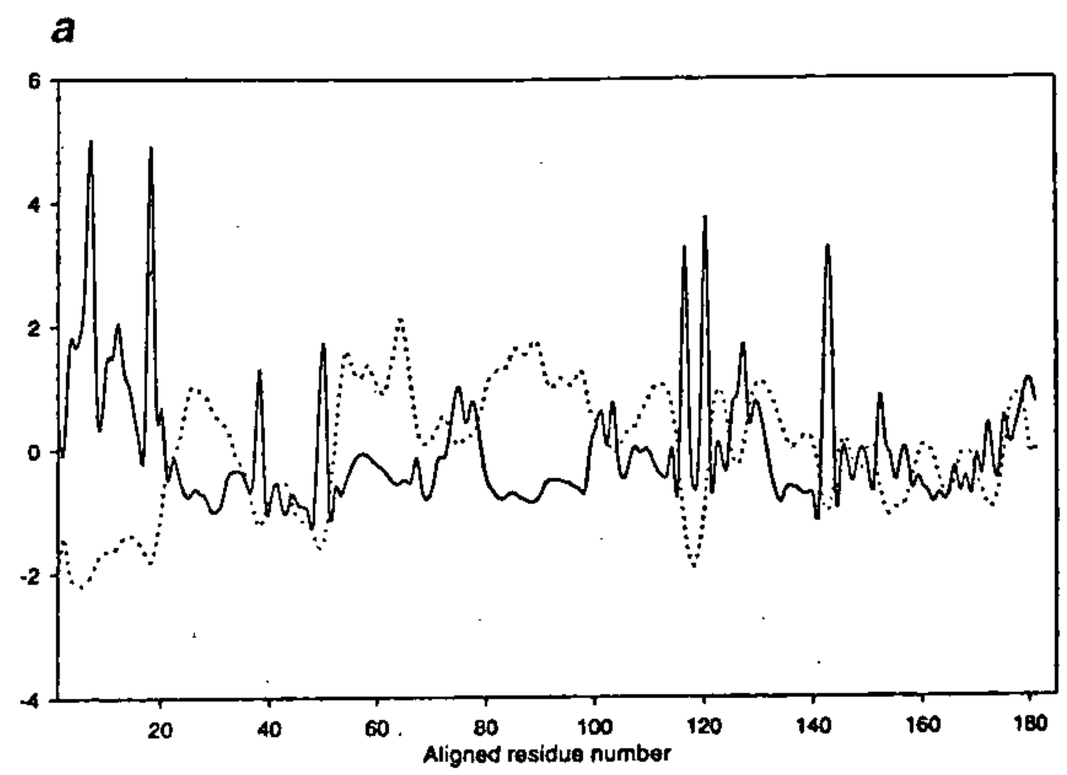


Figure 3. *a*, Profiles of smoothened $\langle|\delta B'|\rangle$ values (dotted) and smoothened D'-values of representative haemoglobin structures (solid), *b*; Correlation between $\langle|\delta B'|\rangle$ values and smoothened Dayhoff's scores of representative haemoglobin structures.

values and $\langle|\delta B'|\rangle$ vs the aligned residue number. The profiles of the two curves indicate that there is significant anti-correlation between the two parameters along the sequence. Figure 3 *b* shows the scatter plot of $\langle|\delta B'|\rangle$ values vs smoothened Dayhoff's scores (cc = – 0.49). The line that best fits this plot has a negative slope indicating that the variation in the thermal parameters is higher in the regions of greater sequence variability. It may also be observed that even in regions of low sequence similarity there are points where $\langle|\delta B'|\rangle$ values are as low as the $\langle|\delta B'|\rangle$ values of regions of high sequence similarity and that the number of points with high $\langle|\delta B'|\rangle$ values are relatively few.

Analysis of all native structures of haemoglobin

Of the 28 native haemoglobin structures with resolution 2.0 Å or better available in the PDB, six were representative structures considered for the earlier analysis. The sequences of these structures were aligned using the PileUp program with the same parameters as for the representative structures, but without any extraneous sequences. The smoothened B'-values of the structures were plotted against the aligned sequence number (Figure 4). It may be observed that the humps seen in the corresponding plot of the representative structures are conserved, but that they are shifted to the left by a few residues, probably due to reduction in the total alignment length resulting from the greater similarity between the structures. Further, the minor humps are amplified in this case.

Analysis of structures of trypsin

The seven structures of trypsin (Table 3) were aligned using the PileUp program (gap creation penalty = 12, gap extension penalty = 4). No end-weight was necessary in this case since the sequences had a high degree of identity. The smoothened B'-factor vs aligned residue number plots were very similar for all the structures (Figure 5 *a*). Table 6 gives the pair-wise correlation coefficients of the B-values of the seven structures, all of which are high. The plot of $\langle|\delta B'|\rangle$ vs Dayhoff's scores (Figure 5 *b*) shows the same trends as observed with the haemoglobin structures. The plots of pair-wise correlation coefficients

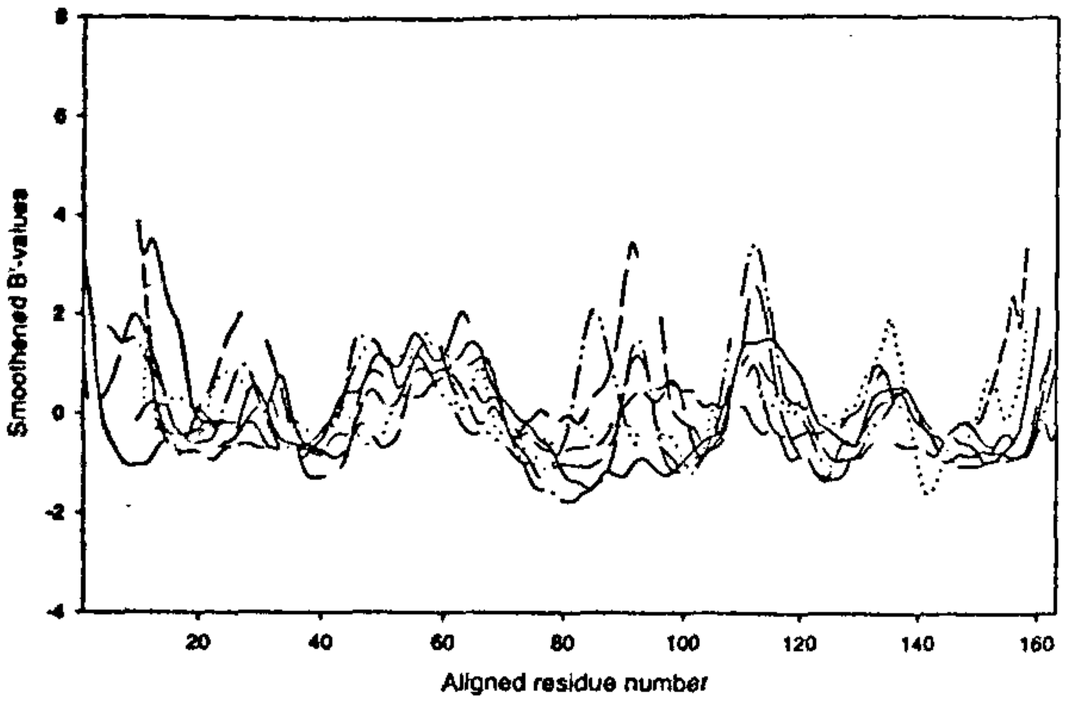
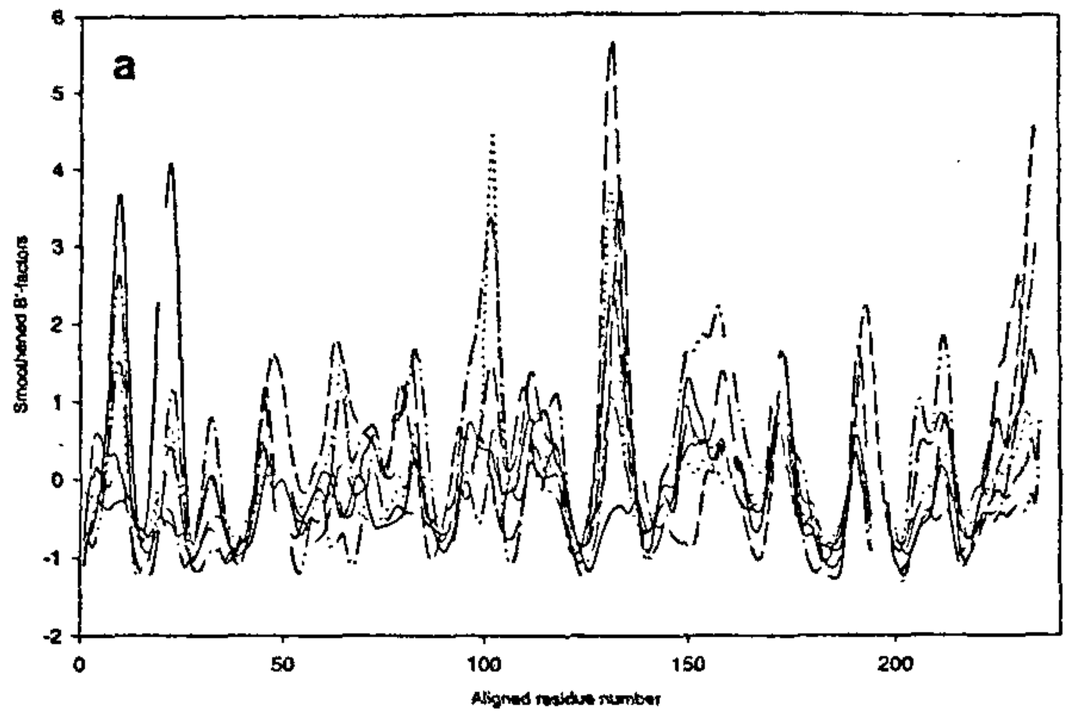
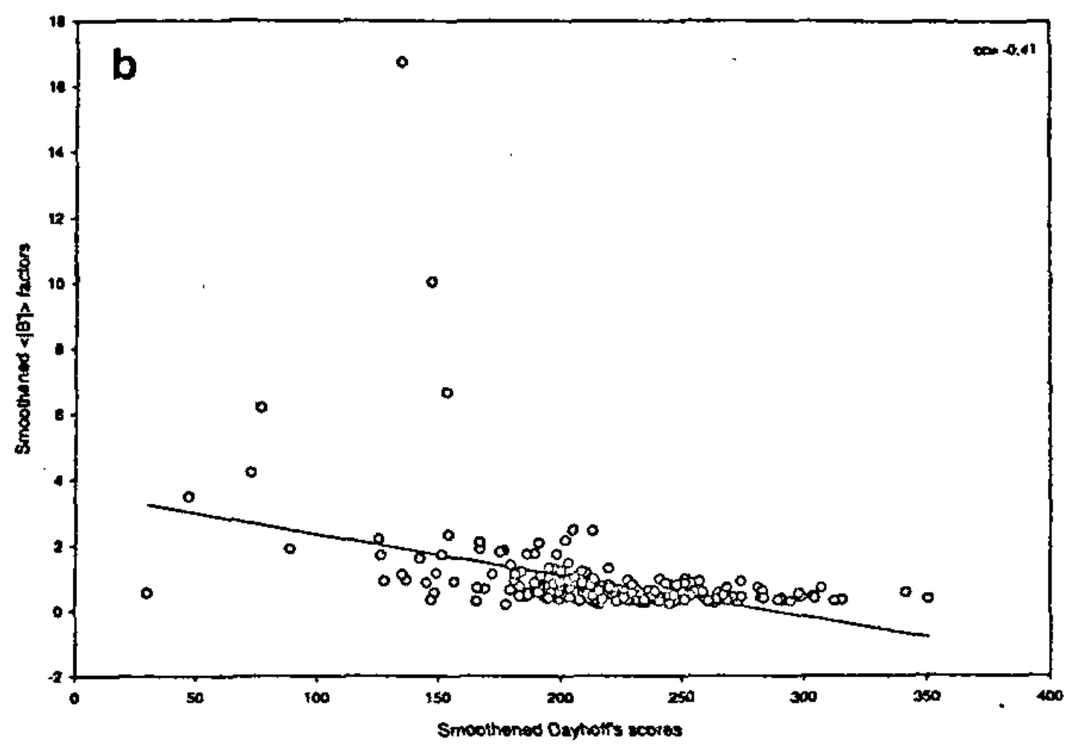


Figure 4. Profiles of B'-factors of all haemoglobin structures.





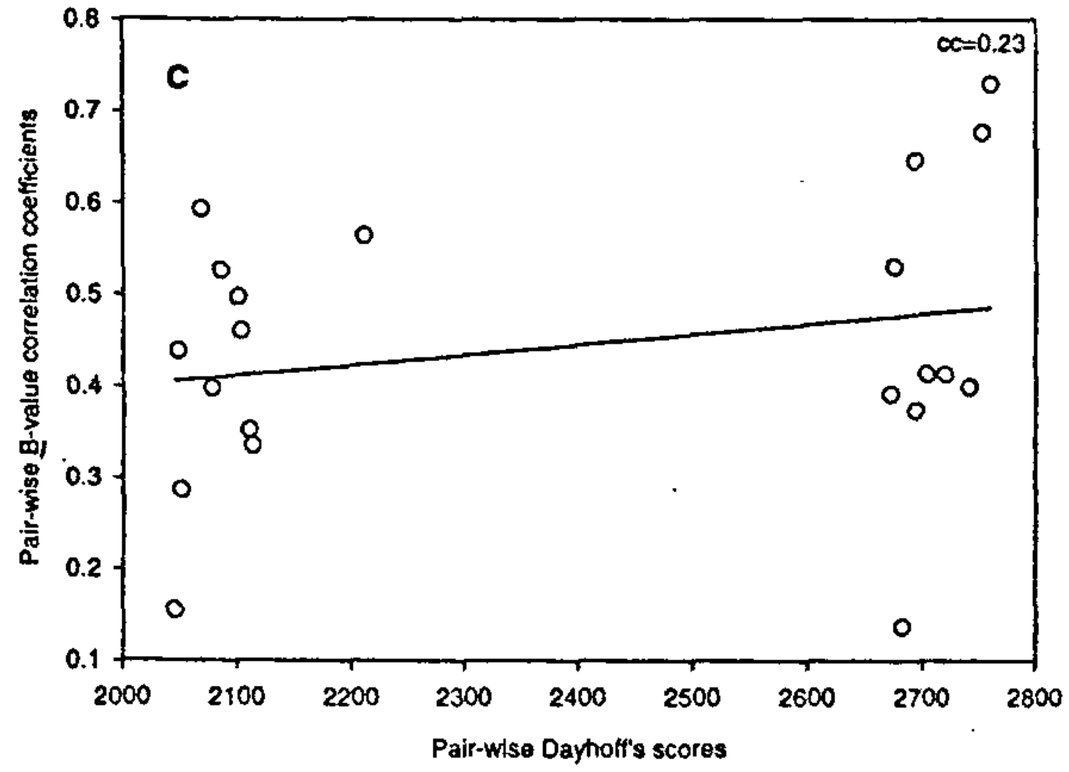


Figure 5. *a*, Profile of smoothened B'-values of trypsin structures; *b*, Scatter plot of $\langle|\delta B'|\rangle$ and Dayhoff's scores of trypsin structures; *c*, Scatter plot of pair-wise correlation coefficients of B-values and pair-wise Dayhoff's scores.

versus Dayhoff's scores (Figure 5 *c*) showed little correlation. It may be noted that, in this case, both these plots are divided into two distinct regions of high and low similarity and that the average of the correlation coefficients in both these regions taken separately would be comparable.

Analysis of structures of TIM

The TIM structures in Table 4 were aligned using the PileUp program with gap creation penalty = 12 and gap extension penalty = 4. Here too, no end-weight was necessary for alignment. The correlation coefficients between the *B*-values were high for all the pairs (Table 7). The smoothened *B'*-value vs aligned residue number plots for all the TIM structures were almost identical (Figure 6 *a*). The plot of $\langle|\delta B'|\rangle$ versus Dayhoff's scores (Figure 6 *b*) shows the same trends as observed with the haemoglobin and trypsin structures. The plots of pair-wise correlation coefficients versus Dayhoff's scores and identity scores (Figure 6 *c*) showed little correlation.

Discussion

Protein function depends on both its structure and dynamics. In the early days of X-ray diffraction studies on protein crystals, only a 'static' image of the protein structure was represented. This was essentially due to limitations on the resolution of data collection set by the X-ray intensities available and lack of reliable protein structure refinement protocols. However, with the advent of powerful X-ray sources such as rotating anode X-ray generator and synchrotron radiation it has been possible to collect near-atomic resolution data on crystals of a large number of proteins. Also advances in computer technology have provided the resources required for refinement of protein structures. The information on the dynamics of a large number of proteins is therefore now available in terms of the atomic displacement parameters. The dynamics of three protein families were investigated in this analysis, viz. haemoglobin, trypsin and triose phosphate isomerase, representing α, β, α/β class of structures.

It was not possible to obtain a unique alignment of representative haemoglobin sequences using standard programs such as PileUp. The uncertainties in alignment are probably due to the low sequence similarity between the sequences used. This problem was circumvented by using a large number of other sequences (a total of 500 at a time) in the alignment that provided links between the original low-similarity sequences. The validity of this approach was justified by the observation that different random selection for the 490 sequences used for linking the original sequences led to closely similar alignments.

In spite of the very low sequence identity of the representative structures of haemoglobin, the stretches of high flexibility are conserved to a large extent as is evident from Figure 1. The similarity in the flexibilities of different segments of haemoglobin chains is also reflected in the correlation coefficients between the *B*-values of aligned residues in these structures (Table 5). It might be anticipated that the physical attributes of polypeptide segments in proteins including flexibility would change with mutations. Since the exposed loops are both likely to be more flexible and also mutate at a faster rate when

Table 6. Correlation coefficients between *B*-values of trypsin structures

	1AOJ	1MCT	1SCT	1TRN	1TRY	3TGI	5PTP
1AOJ	1.00						
1MCT	0.37	1.00					
1SCT	0.29	0.40	1.00				
1TRN	0.53	0.41	0.15	1.00			
1TRY	0.35	0.34	0.56	0.46	1.00		
3TGI	0.39	0.73	0.59	0.40	0.53	1.00	
5PTP	0.14	0.68	0.44	0.41	0.50	0.65	1.00

Table 7. Correlation coefficients between triosephosphate isomerase structures

	1AMK	1AW2	1BTM	1HTL	1TCD	1TPF	1TPH	1TRE	1YDV	1YPI
1AMK	1.00									
1AW2	0.23	1.00								
1BTM	0.47	0.48	1.00							
1HTL	0.48	0.27	0.58	1.00						
1TCD	0.37	0.59	0.56	0.58	1.00					
1TBF	0.47	0.44	0.48	0.54	0.57	1.00				
1TPH	0.46	0.25	0.56	0.71	0.64	0.54	1.00			
1TRE	0.26	0.65	0.49	0.43	0.57	0.47	0.40	1.00		
1YDV	0.38	0.66	0.52	0.49	0.64	0.61	0.58	0.59	1.00	
1YPI	0.35	0.46	0.46	0.69	0.69	0.47	0.73	0.45	0.61	1.00

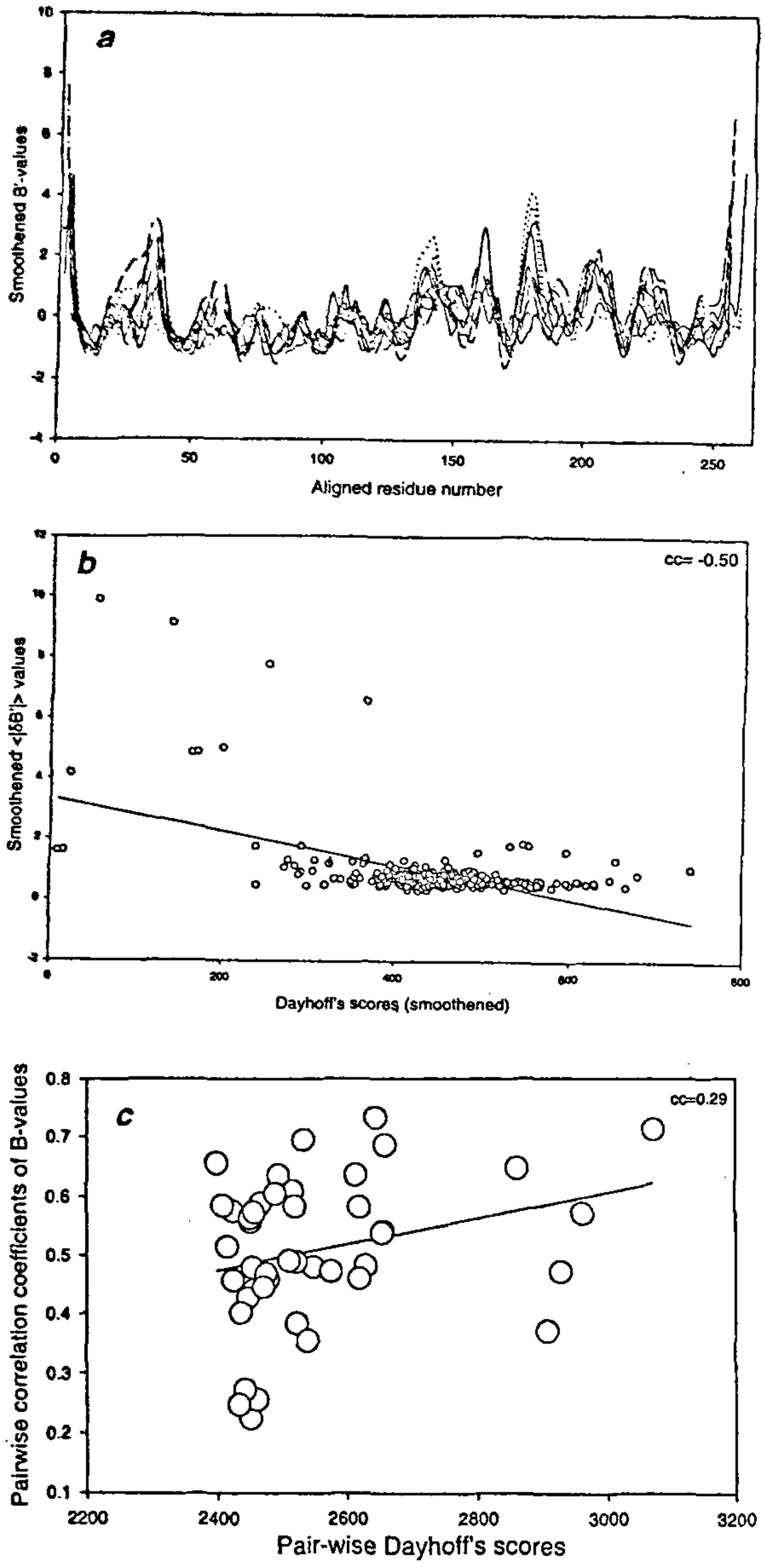


Figure 6. *a*, Profile of smoothened B'-values of triosephosphate isomerase structures; *b*, Scatter plot of $\langle|\delta B'|\rangle$ and Dayhoff's scores of TIM structures; *c*, Scatter plot of pair-wise correlation coefficients of B-values and pair-wise Dayhoff's scores.

compared to buried segments, a correlation might exist between mutability and flexibility. However, Figure 2 clearly shows that such a correlation is limited, if any.

Figure 3 *a* illustrates the variation of D′ and $\langle|\delta B'|\rangle$ values against aligned residue number for representative haemoglobin structures. The Dayhoff's scores and the variation in B'-values are represented as scatter plots in Figure 3 *b*. These plots indicated a negative correlation. These plots were made for alignment number 7; however, when similar examination was made for other alignments, the correlation was found to be weaker (data not shown). These observations suggest that the changes in flexibility are only weakly linked to the changes in the sequence. Further, although points of high B'-value variation occur only in the regions of low sequence similarity, there still exist, even in such regions, many points where the B'-value variation is as low as that observed in regions of high sequence similarity. These analyses were extended to all haemoglobin structures determined at high resolution and consistent results were obtained (Figure 4). This implies that different sequences are compatible with similar flexibility. Thus the relationship between sequence and flexibility has a similar degeneracy as the relationship between sequence and structure. It is also possible that the amino acid sequence distribution determines the precise three-dimensional structure, which in turn dictates the flexibility.

Globins are α-helical proteins. In order to examine the validity of the observations made in this class of proteins for other classes, a set of structures representing β-sheet proteins (trypsin; Table 3) and a representative set for α/β proteins (TIM; Table 4) were selected and the analyses were repeated.

Trypsin sequences had a relatively higher degree of sequence identity and correspondingly the B'-profiles of all the members selected for analysis were closely similar (Figure 5 *a*). There was a small negative correlation between $\langle|\delta B'|\rangle$ and Dayhoff's scores (Figure 5 *b*) suggesting only a weak link between sequence changes and the corresponding changes in flexibility. Interestingly, the sequences of trypsin chosen clustered into two distinct classes with low Dayhoff's scores between members of one class and the other. However, for sequences of high and low sequence similarities, the correlation coefficients between the B-values of corresponding residues were similarly scattered (Figure 5 *c*) suggesting that the changes in sequence do not result in substantial changes in flexibility. Similar observations were also made on the α/β TIM structures (Figure 6).

The closely similar results in the three distinct classes of proteins examined here suggest that the broad conclusions on the retention of polypeptide flexibility in the course of amino acid replacements resulting from mutation is of general validity. This conservation of protein flexibility during the course of evolution is consistent with the generally accepted importance of flexibility to protein function.

1. Branden, C. and Tooze, J., *Introduction to Protein Structure*, New York, Garland Publishing Inc., 1991.
2. Karplus, P. A. and Schulz, G. E., *Naturwissenschaften*, 1985, **72**, 212–213.
3. Vihinen, M., Torkkila, E. and Riikonen, P., *Proteins: Struct. Funct. Genet.*, 1994, **19**, 141–149.
4. Parthasarathy, S. and Murthy, M. R. N., *Acta Crystallogr. D*, 1999, **55**, 173–180.

5. Parthasarathy, S. and Murthy, M. R. N., *Protein Sci.*, 1997, **6**, 2561–2567; 1998, **7**, 525.
6. Hasson, M. S., Schlichting, I., Moulai, J., Taylor, K., Barrett, W., Kenyon, G. L., Babbit, P. C., Gerlt, J. A., Petsko, G. A. and Ringe, D., *Proc. Natl. Acad. Sci. USA*, 1999, **95**, 10396–10401.
7. Bernstein, F. C., Koettzde, T. F., Williams, G. J. B., Meyer, E. F. Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchui, T. and Tasumi, M., *J. Mol. Biol.*, 1997, **112**, 535–542.
8. Wisconsin Package, Version 9.0, Genetics Computer Group (GCG), Madison, Wisconsin.
9. Dayhoff, M. O., Schwartz, R. M. and Orcutt, B. C., in *Atlas of Protein Sequence and Structure* (ed. Dayhoff, M. O.), National Biomedical Research Foundation, Washington DC, 1978, vol. 5, pp. 353–362.

Received 14 October 1999; revised accepted 9 February 2000