Correlating dynamics to conformational properties: An analysis of 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

Atomic displacement parameters (ADPs) obtained by high resolution X-ray diffraction studies on single crystals of proteins represent the mean square displacement of atoms about their mean position. The relationship between the flexibility of the protein molecule and its conformation could be examined by a careful analysis of these ADPs. This communication presents the results of such a statistical analysis. It is shown that the ADPs are related to side chain conformations and are low for energetically favourable rotamers. Examination of the dependence of ADPs on non-planar distortions of the peptide geometry as represented by $\bar{\omega}$ angle reveals that the parameters depend on the direction of non-planar distortion. Those conformations with $\bar{\omega}$ larger than the ideal *trans* geometry (180°–190°) are more flexible when compared to those with $\bar{\omega} < 180°$ (170°–180°). The average ADP of a peptide unit depends weakly on the Ramachandran angles at the corresponding Cα atom. Thus, the flexibility of different segments of the polypeptide appears to be sensitive to the conformation of the side chains as well as the main chain.

ADVANCES in high brilliance X-ray sources as well as computational procedures for the refinement of proteins at high resolution have provided accurate Atomic Displacement Parameters (*B*-values) for several proteins[1,2]. These structures have been analysed with reference to the atomic interactions responsible for protein stability, evolution, function, molecular recognition and principles that govern oligomerization. The information on protein flexibility and mobility available in terms of the ADPs obtained by protein high-resolution structure refinement, in contrast, has not been exploited to the same extent[3–5]. ADPs (or *B*-factors or temperature factors) deposited in the Protein Data Bank correspond to $B = 8\pi^2\langle u^2\rangle$, where $\langle u^2\rangle$ is the average of the mean square atomic displacements (Å^2) along the three coordinate axes and is given by $(u_x^2 + u_y^2 + u_z^2)/3$. These parameters represent the flexibility of protein atoms and are important factors controlling protein function and stability. The atomic displacement parameters determined for different proteins show large variations from one structure to another due to the differences in the static disorder and are also influenced by refinement procedures[6]. However, it has been shown that the frequency distribution of the *B*-factors expressed in units of standard deviation about their mean value at the Cα atoms (B') are similar in protein structures[7]. This distribution could be accurately described as a superposition of two Gaussian functions. The frequency distribution for different amino acids could be used to deduce flexibility indices that are helpful in predicting mobile segments in proteins[3,4,7]. In this communication, we examine the dependence of ADPs on conformation of both the main chain and the side chain: (a) χ^1 representing the side chain rotamer, (b) ω representing the peptide bond geometry, and (c) Ramachandran angles ϕ and φ representing main chain conformation. Generally, energetically unfavourable conformations are associated with high flexibility.

The representative list of protein structures[8] determined by single crystal X-ray diffraction methods and available in the Protein Data Bank[9] (November 1996 release) was chosen for the analysis. None of these structures has sequence identity greater than 25%. The resolution is better than 2.0 Å and the *R* factor is less than 0.2 for all the selected structures. The code numbers of the selected proteins are listed in Table 1.

For each of the selected structures, the *B*-values at the Cα positions were replaced by a normal variate as B'-factor, $B' = (B_i - \langle B\rangle)/\sigma(B)$, where $\langle B\rangle = \Sigma B_i/N$, where B_i is the *B*-value associated with Cα of the *i*th residue and *N* is the total number of residues in the protein, and $\sigma^2(B) = \Sigma(B_i - \langle B\rangle)^2/N$. Normalization of *B*-values is required to compare *B*-values of different protein structures.

The χ^1 (N–Cα–Cβ–Cγ for most amino acids, Oγ for Ser and Cγ1 for Ile) torsional angle for each side chain was computed as the angle between the planes containing N–Cα–Cβ and Cα–Cβ–Cγ atoms, respectively. Following the usual convention, the angle was taken as the clockwise rotation of the N–Cα vector, looking down the Cα–Cβ direction, required to bring it to overlap with the Cβ–Cγ vector. The rotamer conformation was represented as *gauche*$^-$ for $-120° < \chi^1 < 0°$, *gauche*$^+$ for $0° < \chi^1 < 120°$, and *trans* for $120° < \chi^1 < 240°$. Frequency distribution of

Table 1. PDB codes for the high-resolution structures selected for the analysis

131L	153L	1AMP	1ARB	1ARV	1ATL
1BAM	1BP2	1CCR	1CHD	1CHM	1CNS
1CSE	1CSH	1DAA	1DTS	1DUP	1DYR
1EDE	1FKJ	1FNC	1GOF	1GPR	1HMT
1IAE	1ISC	1KPT	1LCP	1LEN	1LTS
1MOL	1NFP	1NHK	1NAR	1OVA	1PBE
1PDA	1PGS	1PHG	1PTX	1REC	1REG
1RSY	1SAT	1SBP	1SNC	1SRI	1TAG
1TAH	1TCA	1THV	1TRY	1TTB	1TYS
1XNB	2ACQ	2ALP	2AYH	2AZA	2CBA
2CCY	2CDV	2CPL	2CTC	2END	2ER7
2GST	2HMZ	2HTS	2MNR	2NAC	2OLB
2PHY	2POR	2PRK	2SIL	2TGL	3CHY
3CLA	3COX	3DFR	3GRS	3PTE	3SIC
3TGI	4ENL	4GCR	4FGF	5RUB	5TIM
8ABP	8FAB	9RNT			

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

χ^1 was sampled at 10° intervals. Similarly, the ω angle representing the deviation of the peptide geometry from a planar configuration was evaluated as the torsional angle $C\alpha_i$–C'_i–N_{i+1}–$C\alpha_{i+1}$ (where i and $i+1$ represent consecutive residues). 0° and 180° in ω correspond to ideal *cis* and *trans* peptide geometries, respectively. Sampling was at 1° intervals for statistical analysis. Ramchandran angles ϕ and φ representing the polypeptide conformation at each $C\alpha$ atom were calculated as angles C'_{i-1}–N_i–$C\alpha_i$–C'_i and N_i–$C\alpha_i$–C'_i–N_{i+1}, respectively. Frequencies were counted in 60° bins of ϕ and φ.

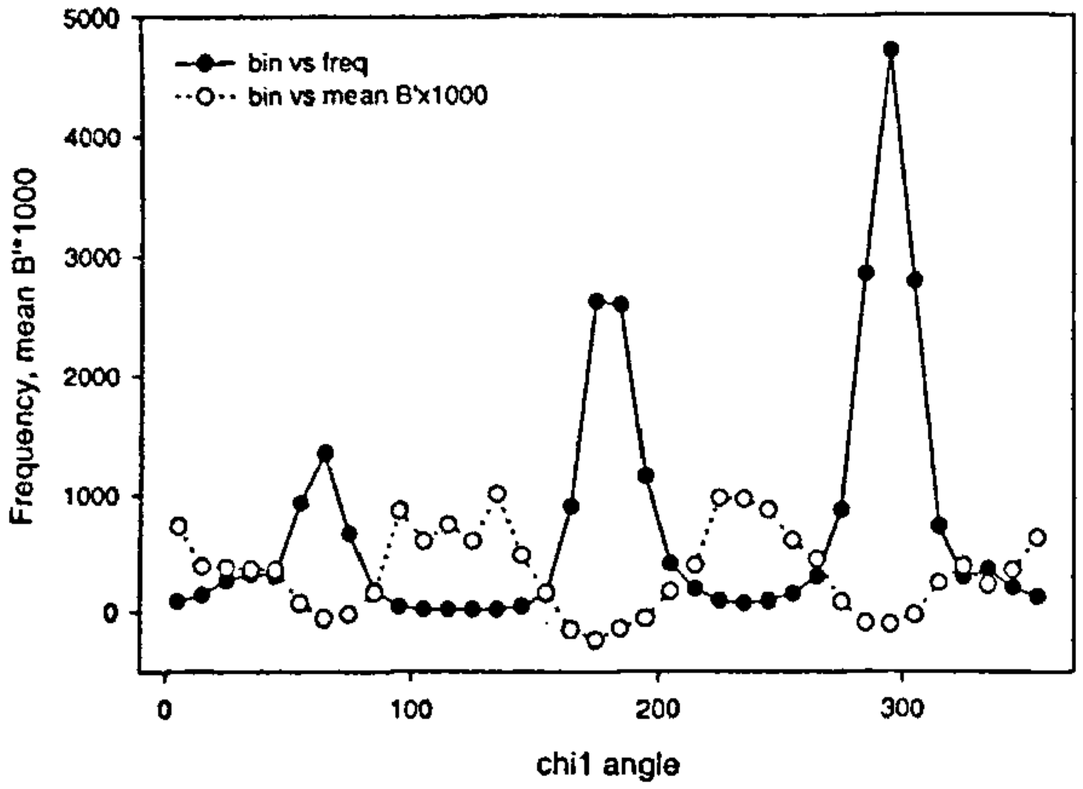
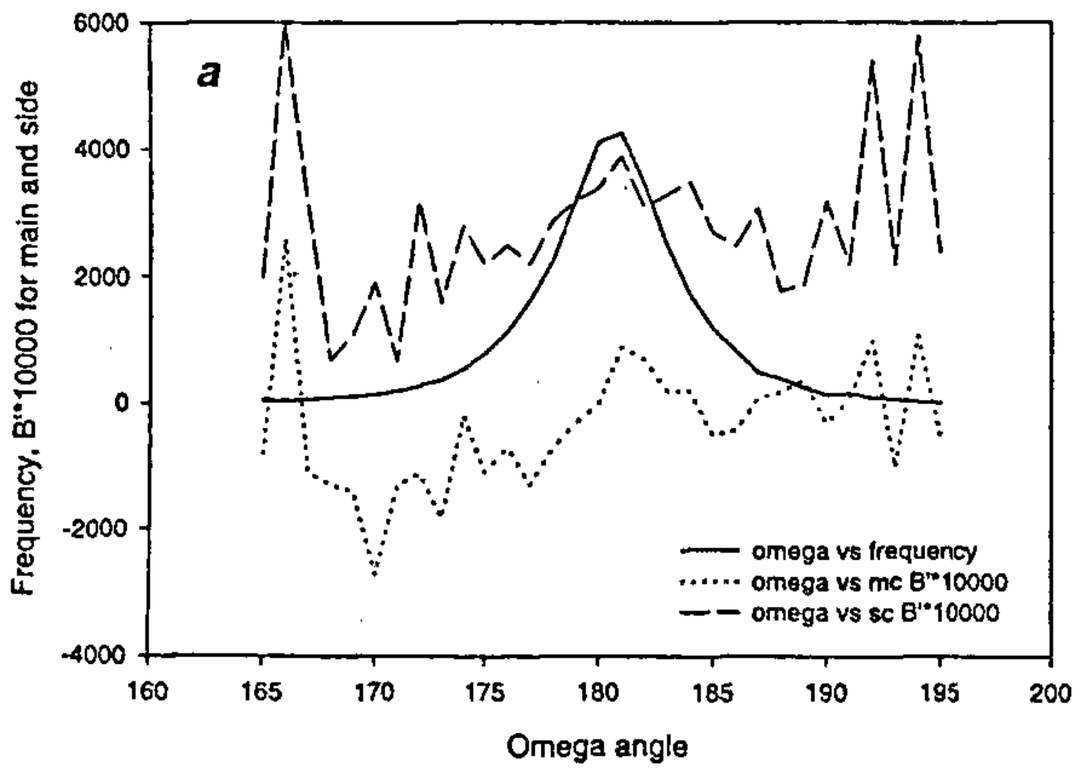


Figure 1. Frequency distribution of χ^1 values and dependence of mean B-factor on χ^1.

The mean value of B'-factors corresponding to each rotamer and residue type was evaluated considering all the proteins. Similarly, the variation of mean B'-factors was monitored as a function of χ^1 angle by grouping side chains into bins of 10°. Residues were also grouped in to different bins in the Ramachandran plot[10]. Even for fairly large bins, the population in certain regions of the Ramachandran map is sparse and hence the estimation of the mean B'-factors for these regions will not be statistically significant. An interval of 60° in ϕ as well as φ was found to provide sufficient number of residues in the bins corresponding to the allowed and partially allowed regions of the map.

Figure 1 shows the frequency distribution of side chain rotamers and their mean B'-factors for structures used in this study. The χ^1 distribution of side chain rotamers in the representative structures follows features that have been noted earlier[11]. For most residues, the preferred rotamers are *gauche*$^-$ and *trans* with somewhat less frequency



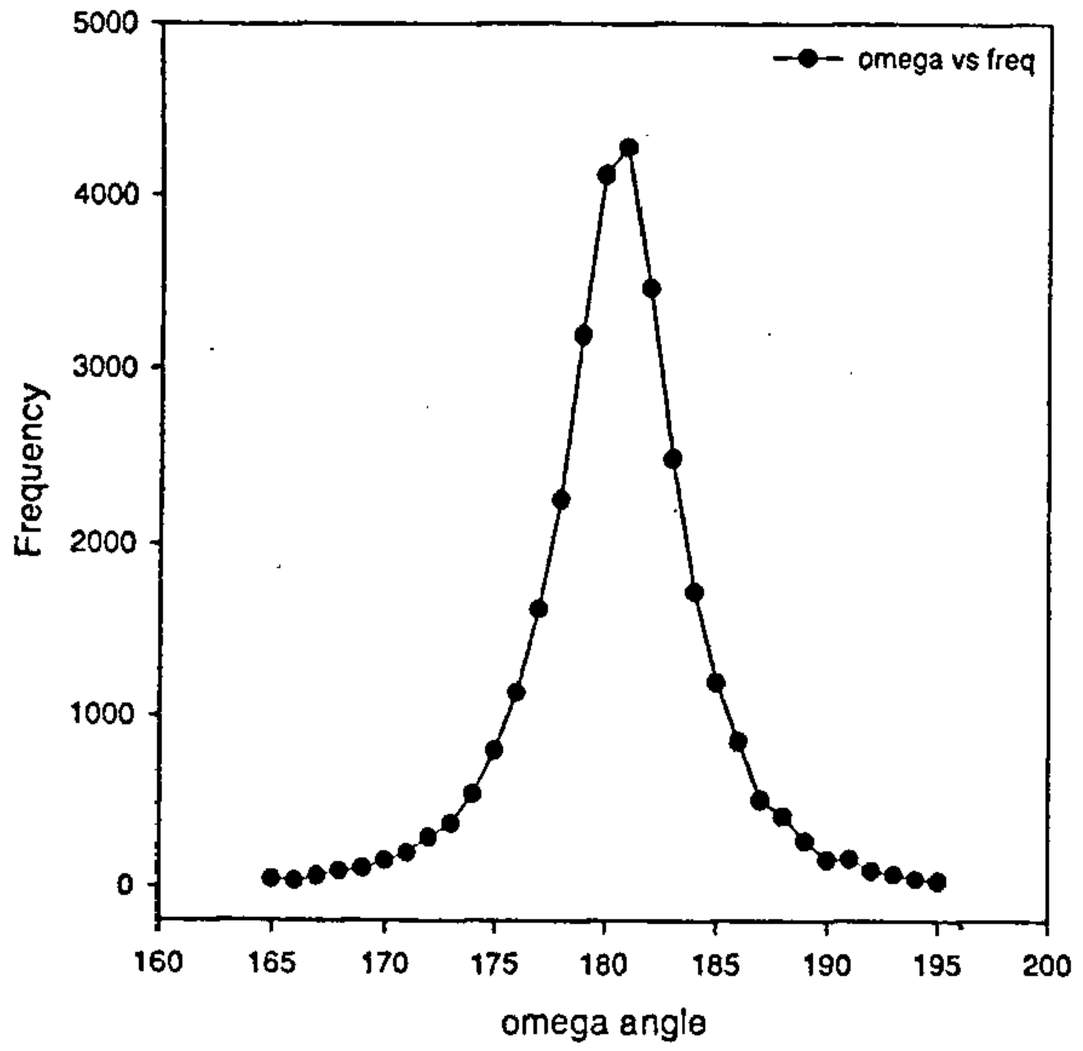


Figure 2. Frequency distribution of peptide non-planar distortion as represented by the dihedral angle ω.

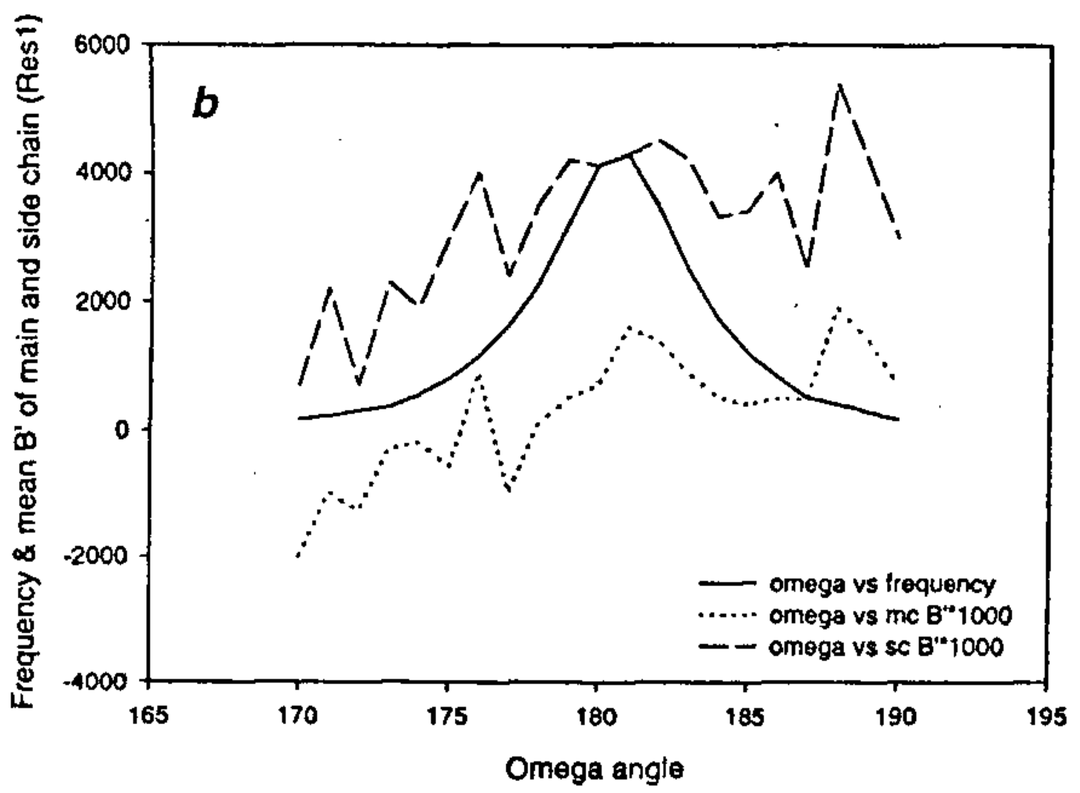


Figure 3. *a*, Variation of mean B'-factors of the main (. . .) and the side chain (– – –) of the $C\alpha$ atom preceding the peptide bond as a function of peptide plane distortion ω. *b*, Variation of mean B'-factors of the main (. . .) and the side chain (– – –) of the $C\alpha$ atom following the peptide bond as a function of peptide plane distortion ω.

in the *gauche*$^+$. However, for amino acids branched at Cβ, the rotamer is either not well-defined or well determined. As these residues form only a small fragment of all the residues, they have not been excluded or treated independently in the present analysis. The peaks corresponding to the preferred rotamers have widths at half maximum of ~ 25°. Although this spread might depend on the resolution at which the structures have been determined as well as the constraints used in refinement, the distribution shown in Figure 1 is typical[11]. The plot also shows the mean B'-factors for these conformational states. It should be noted that the frequency in bins mid way between preferred rotamers are low and hence the estimates of the mean B'-factors have large errors. The frequencies for intervals close to the preferred torsion angles are large and hence the estimated B'-factors are reliable. The B'-factors have clear minima at the preferred conformations of the side chains. It might also be noted that the most favoured torsion angle representing the *gauche*$^-$ form is less than 300°. The minimum in B-factor curve also occurs at this value. Similar observations have been made earlier[11–13].

Figure 2 illustrates the frequency distribution of the ω angle, representing out of plane deviation of the *trans* peptide unit about the C′–N bond and Figures 3 *a* and *b* show the variation of the mean B'-factor as a function of ω for both the preceding and succeeding residues, respectively. It is clear that the highest frequency occurs, not for the ideal *trans* geometry, but for $\omega = 179°$. This deviation from the planarity of the peptide unit in proteins has previously been observed[14]. The distribution is nearly symmetrical about the preferred angle. On the contrary, the mean B'-factor appears to monotonically increase with increasing ω in the range 170°–190°. It implies that the peptide group becomes more flexible when ω increases from the value corresponding to the ideal *trans* geometry. The feature of larger B'-factors associated with $\omega > 180°$ are similar to the trend observed for rotamers – conformations with lower occupancy are associated with larger B'-factors. However, B'-factors continue to decrease when ω changes to values lower than 180°, although the peptide changes to a higher energy state, as reflected by lower frequency of occurrence. Ramakrishnan and Balasubramanian[15] have shown that the area allowed in the

Table 2. Frequency of residues in bins of Ramachandran map and respective average B'-factors in brackets

4378 (– 0.19)	6220 (– 0.06)	733 (0.23)	16 (0.85)	138 (0.25)	91 (0.25)
709 (0.04)	1652 (– 0.06)	55 (0.94)	47 (0.79)	39 (0.46)	11 (1.87)
309 (0.20)	1253 (0.32)	16 (0.65)	453 (0.23)	722 (0.38)	13 (0.00)
167 (0.32)	9611 (– 0.09)	3382 (0.12)	20 (0.33)	348 (0.56)	22 (0.26)
36 (0.35)	110 (0.42)	64 (0.83)	28 (0.64)	34 (1.65)	9 (0.8)
244 (0.00)	243 (0.21)	15 (0.59)	75 (0.17)	177 (0.17)	114 (0.02)

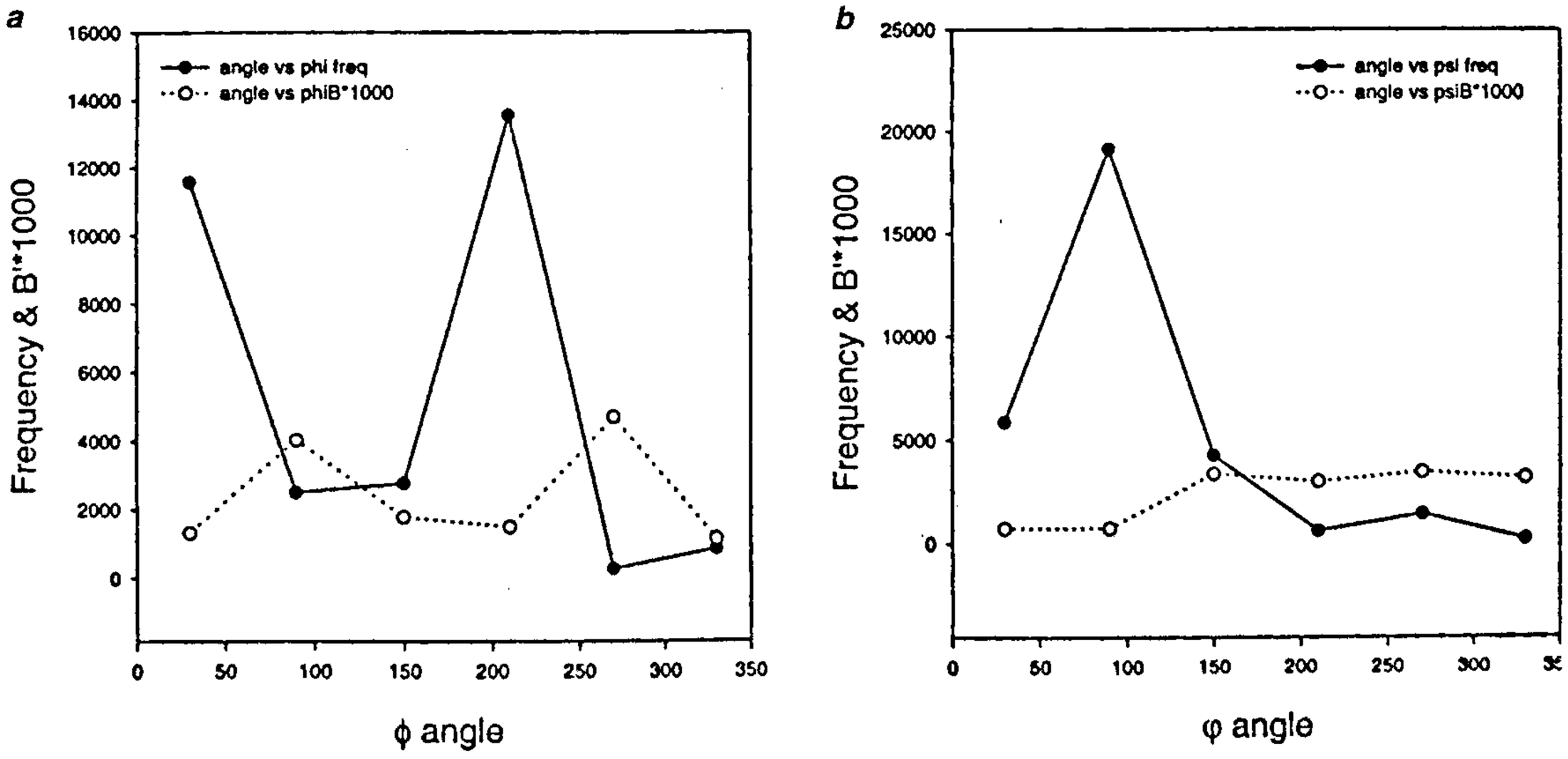


Figure 4. Variation of the B'-factor as a function of (*a*) ϕ and (*b*) φ.

Ramachandran diagram for non-glycyl residues reduces with increasing ω, in the range 170°– 190°. Therefore, the reduction of B'-factor for non-planar distortions with $\omega < 180°$ might be due to the fact that the energy corresponding to the peptide geometry is determined by the partial double bond character of the C′–N bond, while the flexibility might be related to the allowed area (lack of close contacts) of the Ramachandran diagram for the corresponding ω. Similar trends are observed for the side chain and both the preceding and succeeding residues of the peptide bond.

Table 2 lists the frequencies of residues occurring in different bins of the Ramchandran plot sampled at 60° intervals in the structures selected for analysis. The numbers are consistent with the well-known features of the Ramachandran map[16]. The largest occupancy is found for the helical regions in the third quadrant and for strand regions in the second quadrant of the map. The B'-factors associated with the atoms N_i, $C_i\alpha$, C'_i, N_{i+1} of all residues falling in the bins were averaged. The mean B'-factor of all Cα atoms was subtracted from this average. The values shown in Table 2 correspond to these. It is evident that the mean B'-factor is low for all cells where the occupancy is high. Since the frequency of occurrence depends on the energy corresponding to the particular cell, the flexibility is linked to the energy as observed for χ^1. Similar observations were made for the B'-factors associated with the Cα atom alone. Thus the occurrence of a particular pair of Ramachandran angle has the effect of influencing the flexibility of the local segment of the polypeptide. Figure 4 *a* illustrates the frequency and the mean B'-factor as a function of ϕ (including all φ), while Figure 4 *b* shows the same parameters as functions of φ (including all ϕ). The correlation between high frequency and low flexibility is evident in these plots also.

Availability of a large number of well-determined three-dimensional structures of proteins makes possible studies on protein structure, function, evolution and dynamics. Preliminary examination of the ADPs determined during the course of high resolution X-ray structure determination of a set of unique proteins carried out in this communication clearly demonstrates that flexibility of the protein chain is dependent on conformation. Both main chain and side chain conformations appear to affect the mobility of residues. Residue mobility is minimum when the side chain confirms to the most preferred rotamer. On the contrary, the mobility of the backbone depends asymmetrically on the departure of the peptide plane from ideal *trans* geometry. For non-planar distortions resulting from increasing ω, the flexibility increases. Peptides with $\omega < 180°$ tend to be less flexible although the distortions are not in the direction of reducing energy. This might be a consequence of a larger area of Ramachandran diagram allowed for peptides with $\omega < 180°$. Flexibility of the polypeptide is also found to depend on the Ramachandran angles, partially allowed and disallowed regions correspond to larger B'-factors. Thus, the local flexibility of polypeptides appears to be influenced by main chain as well as side chain conformations.

1. Dauter, Z., Lamzin, V. S. and Wilson, K. S., *Curr. Opin. Struct. Biol.*, 1997, **7**, 681–688.
2. Merrit, E. A., *Acta Crystallogr. D*, 1999, **55**, 1109–1117.
3. Vihinen, M., Torkkila, E. and Riikonen, P., *Proteins: Struct. Funct. Genet.*, 1994, **19**, 141–149.
4. Karplus, P. A. and Schulz, G. E., *Naturwissenschaften*, 1985, **72**, 212–213.
5. Carugo, O. and Argos, P., *Acta Crystallogr. D*, 1999, **55**, 473–478.
6. Parthasarathy, S. and Murthy, M. R. N., *Acta Crystallogr. D*, 1999, **55**, 173–180.
7. Parthasarathy, S. and Murthy, M. R. N., *Protein Sci.*, 1997, **6**, 2561–2567; 1998, **7**, 525.
8. Hobohm U. and Sander C., *Protein Sci.*, 1994, **3**, 522–524.
9. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr. and Brice, M. D., *J. Mol. Biol.*, 1977, **112**, 535–542.
10. Ramachandran, G. N. and Sasisekharan, V., *Adv. Prot. Chem.*, 1968, **23**, 283–437.
11. Thanki, N., Umrania, Y., Thronton, J. M. and Goodfellow, J. M., *J. Mol. Biol.*, 1991, **221**, 669–691.
12. Dunbrack, R. L. Jr. and Karplus, M., *J. Mol. Biol.*, 1993, **230**, 543–574.
13. Carugo, O. and Argos, P., *Protein Eng.*, 1997, **10**, 777–787.
14. MacArthur, M. W. and Thornton, J. M., *J. Mol. Biol.*, 1996, **264**, 1180–1195.
15. Ramakrishnan, C. and Balasubramanian, R., *Int. J. Pept. Protein Res.*, 1972, **4**, 79–90.
16. Ramachandran, G. N., Ramakrishnan C. and Sasisekharan, V., *J. Mol. Biol.*, 1963, **7**, 95–99.

Received 14 October 1999; revised accepted 9 February 2000