Protein Engineering vol.15 no.10 pp.783–798, 2002

A conformational analysis of Walker motif A [GXXXXGKT (S)] in nucleotide-binding and other proteins

C.Ramakrishnan[1,2], V.S.Dani[1] and T.Ramasarma[3]

[1]Molecular Biophysics Unit and [3]Department of Biochemistry, Indian Institute of Science, Bangalore 560012, India

[2]To whom correspondence should be addressed.
E-mail: ramki@crmbu2.mbu.iisc.ernet.in

The sequence GXXXXGKT/S, popularly known as Walker motif A, is widely believed to be the site for binding nucleotides in many proteins. Examination of the crystal structures in the Protein Data Bank showed that about half of the examples having these sequences do not bind or use nucleotides. Data analyses showed 92 different Walker sequences of the variable quartet (XXXX). Ramachandran angles in this segment revealed conformational similarity in the group of 45 proteins, known to bind or utilize nucleotides. The conformations of this segment in other proteins differ widely and it is not known whether they play any role in their functions. A flip of a peptide unit at different locations, with little change in the backbone conformation was noted in nine pairs of these proteins having same Walker sequence. An examination of the immediate neighborhood of the Walker sequence indicates that this region is preceded by a β-strand and followed by an α-helix, resulting in the motif β–W–α, an invariant feature amongst nucleotide-binding proteins.

Keywords: peptide flip/Ramachandran angles/β-turn/Walker motif

Introduction

The motif GXXXXGKT (X, any residue) as a common nucleotide binding fold in the α- and β-subunits of F_1-ATPase, myosin and other ATP-requiring enzymes was first recognized in 1982 by Walker and colleagues (Walker *et al.*, 1982). Since then, this sequence has been found in many proteins that bind nucleotides and thereby gained predictive value for nucleotide binding site in proteins. Crystal structure data of such proteins (Berchtold *et al.*, 1993; Abrahams *et al.*, 1994; Chattopadhyay *et al.*, 2000) indicated that this motif is present in the shape of a loop around nucleotides and utilizes its highly conserved residues of lysine and threonine to bind to their phosphate-oxygen atoms. This consensus sequence of GXXXXGKT (S), with serine substituting threonine in some cases, is more popularly known as Walker loop or P-loop (phosphate binding loop).

In view of growing interest in the proteins containing a segment with Walker sequence, the Brookhaven Protein Data Bank (Berman *et al.*, 2000) was searched and 649 polypeptide chains were found to have such a sequence. Many of these proteins do not bind or use nucleotides in their reactions. Therefore, it appeared that the sequence of the variant quartet and the specific loop structure might have a role in nucleotide binding. To fill the lacunae of information, conformations of the backbone of the peptide fragments of GXXXXGKT (S) were examined using Ramachandran angles. The data analysis in this paper indicates that different foldings are possible for the Walker sequences and only in the nucleotide-binding proteins they have a distinctive loop structure.

Materials and methods

The Ramachandran angles (φ, ψ) (Ramachandran *et al.*, 1963; Ramachandran and Sasisekharan, 1968) were computed from the coordinates of atoms available in the Brookhaven Protein Data Bank (Berman *et al.*, 2000). The segment structure similarity was obtained by evaluating the root mean square (r.m.s.) values of the Ramachandran angles. The package of RASMOL (Sayle and Milner-White, 1995) was used to draw the figures.

Results of data analysis

Proteins containing Walker sequences

Search for the sequence GXXXXGKT (S) in the Protein Data Bank (April 2001 release) revealed 649 entries having this sequence, occurring in 395 protein structures with a resolution of 4 Å or better. Out of the 20^4 combinations of sequence possible for the variable region XXXX, only 92 were found to occur, of which 18 had only one entry. The present analysis is limited to these data

The Ramachandran angles of Walker sequence

Groups having more than one entry were examined from the structural viewpoint. The mean and r.m.s. values of the Ramachandran angles φ and ψ were computed at the eight residues of the segment. Should the same sequence give the same structure, as is widely believed, the r.m.s. values for a group would be small. Using a liberal upper limit of 40°, dissimilar structures were found to be present in 10 of these groups, as revealed by the high r.m.s. values for some of the Ramachandran angles. Using similarity of the Ramachandran angles as the criterion, these were divided into further sub-groups. The various sequences and location of the segment in the protein of the group thus obtained are given in Table I, along with the PDB code, chain identifier, resolution of the structure and r.m.s. for those groupings with more than one entry (the protein names are not included in Table I owing to the large number of examples; however, they are included in Table II, which gives the selected set). The sub-groups with same sequence are indicated by suffixes A, B and C, to the group number. It can be seen that the r.m.s. values are now reasonably small. Some sequences assume more than one conformation: two for six sequences (005 – GAGALGKT, 012 – GLRSDGKT, 016 – GLPAIGKT, 030 – GATGTGKT, 058 – GTAFEGKS and 077 – GLYRTGKS); three for three sequences (006 – GHVDHGKT, 033 – GPTGVGKT and

Table I. Proteins containing the consensus sequence of GXXXXGKT(S): the location of the segment in the chain, PDB code and resolution of the crystal structure are given

No.	Sequence Segment location	PDB code, (resolution in Å) chain identifier		Residue r.m.s. (ϕ,ψ)	Residue r.m.s. (ϕ,ψ)
001	GLSGTGKT 248 --- 255	1AQ2 (1.9)	1AYL (1.8)	G (1,1) L (1,1) S (6,1) G (2,1)	T (1,6) G (6,2) K (1,2) T (2,1)
002	GDRQTGKT 169 --- 176	1BMF (2.8) (A,B,C) 1E1Q (2.6) (A,B,C) 1E79 (2.4) (A,B,C) 1MAB (2.8) (A)	1COW (3.1) (A,B,C) 1E1R (2.5) (A,B,C) 1EFR (3.1) (A,B,C) 1NBM (3.0) (A,B,C)	G (7,8) D (7,6) R (6,5) Q (7,11)	T (9,10) G (13,10) K (8,8) T (10,6)
003	GGAGVGKT 156 --- 163	1BMF (2.8) (D,F) 1E1Q (2.6) (D,E,F) 1E79 (2.4) (D,E,F) 1NBM (3.0) (D,E,F)	1COW (3.1) (D,F) 1E1R (2.5) (D,F) 1EFR (3.1) (D,F)	G (23,11) G (7,19) A (19,18) G (13,17)	V (17,18) G (23,10) K (6,4) T (5,7)
003A	GGAGVGKT 156 --- 163	1BMF (2.8) (E) 1EFR (3.1) (E)	1COW (3.1) (E)	G (1,2) G (2,0) A (1,1) G (2,2)	V (2,2) G (2,5) K (5,0) T (2,10)
003B	GGAGVGKT 156 --- 163	1E1R (2.5) (E)			
003C	GGAGVGKT 156 --- 163	1MAB (2.8) (B)			
004	GAHALGKT 173 --- 180	1A2F (2.1) 1AA4 (2.1) 1AC8 (2.1) 1AED (2.1) 1AEF (2.1) 1AEH (2.1) 1AEK (2.1) 1AEN (2.1) 1AEQ (2.1) 1AET (2.1) 1AEV (2.1) 1BEK (2.2) 1BEP (2.2) 1BES (2.0) 1BVA (1.8) (A) 1CCB (2.1) 1CCI (2.4) 1CCK (2.1) 1CCP (2.2) 1CMQ (2.3) 1CMU (2.1) 1CPE (2.2) 1CPG (2.2) 1DCC (2.2) 1DJ5 (1.9) (A) 2CCP (2.2) 2CYP (1.7) 2PCC (2.3) (A,C) 3CCX (2.3) 4CCX (1.9) 6CCP (2.2)	1A2G (2.1) 1AC4 (2.1) 1AEB (2.1) 1AEE (2.1) 1AEG (2.1) 1AEJ (2.1) 1AEM (2.1) 1AEO (2.1) 1AES (2.1) 1AEU (2.1) 1BEJ (2.4) 1BEM (2.2) 1BEQ (2.1) 1BJ9 (2.2) 1CCA (1.8) 1CCC (2.0) 1CCJ (2.1) 1CCL (2.0) 1CMP (1.9) 1CMT (2.1) 1CPD (2.2) 1CPF (2.2) 1CYF (2.3) 1DJ1 (1.9) (A) 1RYC (1.8) 2CEP (2.2) 2PCB (2.8) (A,C) 3CCP (2.2) 4CCP (2.2) 5CCP (2.2) 7CCP (2.2)	G (5,8) A (5,4) H (5,7) A (8,5)	L (7,4) G (6,5) K (5,5) T (4,4)
005	GAGALGKT 173 --- 180	1CCE (2.3)	1CCG (2.1)	G (5,9) A (5,10) G (11,3) A (1,3)	L (2,1) G (2,1) K (4,2) T (1,3)
005A	GAGALGKT 173 --- 180	1DS4 (2.0) (A) 1DSG (2.5) (A) 1DSP (2.0) (A)	1DSE (2.0) (A) 1DSO (2.0) (A)	G (9,14) A (12,8) G (3,7) A (6,10)	L (13,2) G (4,6) K (3,6) T (5,3)
006	GHVDHGKT 18 --- 25	1B23 (2.6) (P) 1D8T (2.3) (A,B) 1EFC (2.0) (A,B) 1EXM (1.7) A 1G7T (2.0) (A)	1D2E (1.9) (A–D) 1DG1 (2.5) (G,H) 1EFT (2.5) 1G7S (2.0) (A) 1TUI (2.7) (A,B,C)	G (9,4) H (5,6) V (6,5) D (6,7)	H (8,7) G (8,9) K (7,9) T (6,6)

Continued

Table I. *Continued*

No.	Sequence Segment location	PDB code, (resolution in Å) chain identifier		Residue r.m.s. (ϕ,ψ)	Residue r.m.s. (ϕ,ψ)
006A	GHVDHGKT 18 --- 25	1AIP (3.0) (A,B,E,F)	1EFU (2.5) (A,C)	G (18,5) H (5,4) V (9,4) D (11,7)	H (2,4) G (5,14) K (10,5) T (3,5)
006B	GHVDHGKT 18 --- 25	1ETU (2.9)			
007	GYLVNGKT 1264 --- 271	10MH (2.5) (A) HMY (2.5) 2HMY (2.6) (B) 4MHT (2.7) (A) 6MHT (2.0) (A) 8MHT (2.7) (A)	1FJX (2.2) (A) 1MHT (2.8) (A) 3MHT (2.7) (A) 5MHT (2.7) (A) 7MHT (2.8) (A) 9MHT (2.3) (A)	G (7,4) Y (5,4) L (3,7) V (8,6)	N (9,6) G (9,7) K (10,7) T (9,8)
008	GLDAAGKT 24 --- 31	1E0S (2.2) (A) 1HUR (2.0) (A,B) 1RRG (2.4) A,B	1HFV (2.8) (A,B) 1RRF (3.0)	G (4,5) L (9,6) D (7,8) A (3,7)	A (7,9) G (13,9) K (9,9) T (7,6)
009	GPHGMGKT 56 --- 63	1E2H (1.9) (A,B) 1E2J (2.5) (A,B) 1E2L (2.4) (A,B) 1KI3 (2.3) (A,B) 1KI6 (2.3) (A,B) 1KI8 (2.2) (A,B) 1KIN (2.0) (A,B) 1VTK (2.7) 2VTK (2.8)	1E2I (1.9) (A,B) 1E2K (1.7) (A,B) 1KI2 (2.2) (A,B) 1KI4 (2.3) (A,B) 1KI7 (2.2) (A,B) 1KIM (2.1) (A,B) 1QHI (1.9) (A,B) 2KI5 (1.9) (A,B) 3VTK (3.0)	G (14,4) P (5,6) H (5,5) G (7,12)	M (9,14) G (14,10) K (8,9) T (7,8)
010	GVRSDGKT 487 --- 494	1MHY (2.) (D)	1MHZ (2.7) (D)	G (5,9) V (11,2) R (8,3) S (9,8)	D (2,3) G (9,0) K (14,10) T (1,2)
011	GESGAGKT 179 --- 186	1B7T (2.5) (A) 1BR2 (2.9) (A–F) 1D0X (2.0) (A) 1D0Z (2.0) (A) 1D1B (2.0) (A) 1DFK (4.2) (A) 1FMV (2.1) (A) 1G8X (2.8) (A,B) 1MMA (2.1) 1MMG (1.9) 1MND (2.6) 1VOM (1.9)	1BR1 (3.5) (A,C,E,G) 1BR4 (3.6) (A,C,E,G) 1D0Y (2.0) (A) 1D1A (2.0) (A) 1D1C (2.3) (A) 1DFL (4.2) (A,B) 1FMW (2.1) (A) 1LVK (1.9) 1MMD (2.0) 1MMN (2.1) 1MNE (2.7) 2MYS (2.8) (A)	G (26,15) E (7,6) S (10,28) G (28,12)	A (10,17) G (20,18) K (18,6) T (6,12)
012	GLRSDGKT 487 --- 494	1FYZ (2.1) (A,B) 1FZ1 (1.9) (A,B) 1FZ3 (2.0) (A,B) 1FZ5 (2.4) (A,B) 1MMO (2.2) (E)	1FZ0 (2.0) (A,B) 1FZ2 (2.1) (A,B) 1FZ4 (2.3) (A,B) 1FZ7 (1.9) (A,B) 1MTY (1.7) (D,E)	G (2,5) L (4,4) R (4,4) S (8,10)	D (5,4) G (5,5) K (9,8) T (3,2)
012A	GLRSDGKT 487 --- 494	1MMO (2.2) (D)			
013	GLSGSGKT 248 --- 255	1OEN (1.9)			
014	GTAFPGKT 212 --- 219	1QPA (1.8) (A,B)		G (3,3) T (3,1) A (2,2) F (1,1)	P (2,4) G (4,0) K (2,3) T (1,5)
015	GKVTGGKT 102 --- 109	1STE (2.0)			
016	GLPAIGKT 499 --- 506	1BGX (2.3) (T) 1TAQ (2.4)	1CMW (2.6) (A)	G (4,12) L (16,5) P (2,20) A (12,11)	I (8,13) G (19,13) K (22,15) T (2,10)
016A	GLPAIGKT 499 --- 506	1QSS (2.3) (A) 1QTM (2.3) (A) 3KTQ (2.3) (A)	1QSY (2.3) (A) 2KTQ (2.3) (A) 4KTQ (2.5) (A)	G (13,10) L (11,2) P (3,8) A (5,20)	I (15,8) G (4,4) K (2,4) T (5,7)

Continued

Table I. *Continued*

No.	Sequence Segment location	PDB code, (resolution in Å) chain identifier		Residue r.m.s. (ϕ,ψ)	Residue r.m.s. (ϕ,ψ)
017	GSQAGGKT 47 --- 54	1WGT (1.9) (A,B)		G (4,1) S (1,1) Q (2,7) A (3,0)	G (9,4) G (5,7) K (4,1) T (2,0)
018	GPESSGKT 66 --- 73	1G18 (3.8) (A) 2REB (2.3)	1G19 (3.0) (A)	G (5,6) P (6,1) E (4,18) S (20,22)	S (22,12) G (27,28) K (33,30) T (12,9)
019	GDVACGKT 12 --- 19	1A2B (2.4) 1DPF (2.0) (A)	1CXZ (2.2) (A)	G (6,5) D (2,6) V (6,2) A (6,10)	C (6,0) G (1,7) K (9,1) T (1,4)
020	GDGGTGKT 7 --- 24	1A2K (2.5) (C,D,E) 1IBR (2.3) (A,C) 1QG2 (2.5) (A) 1RRP (2.9) (A,C)	1BYU (2.1) (A,B) 1QBK (3.0) (C) 1QG4 (2.5) (A,B) 3RAN (2.1) (A–D)	G (12,8 D (6,10) G (8,5) G (4,13)	T (14,5) G (9,13) K (11,4) T (4,6)
021	GDVAVGKT 210 --- 217	1A4R (2.5) (A,B)		G (0,1) D (1,0) V (1,0) A (1,3)	V (4,1) G (3,3) K (3,3) T (2,0)
022	GDGAVGKT 10 --- 17	1AM4 (2.7) (D,E,F) 1DOA (2.6) (A) 1E96 (2.4) (A) 1G4U (2.3) (R) 1HE1 (2.0) (C,D) 2NGR (1.9) (A)	1AN0 (2.8) (A,B) 1DS6 (2.3) (A) 1FOE (2.8) (B,D,F,H) 1GRN (2.1) (A) 1MH1 (1.3)	G (12,7) D (11,12) G (12,11) A (6,27)	V (26,5) G (6,12) K (11,6) T (4,5)
023	GQTSSGKT 86 --- 93	1BG2 (1.8) 3KIN (3.1) (A,C)	2KIN (1.9) (A)	G (13,11) Q (15,6) T (7,4) S (6,14)	S (14,12) G (12,12) K (8,4) T (2,3)
024	GLPARGKT 45 --- 52	1BIF (2.0) 3BIF (2.3) (A)	2BIF (2.4) (A,B)	G (4,3) L (2,2) P (1,5) A (3,4)	R (2,7) G (3,8) K (4,3) T (6,3)
025	GMDLKGKT 206 --- 213	1BVU (2.5) (A–F)		G (16,25) M (26,9) D (12,6) L (4,12)	K (7,5) G (7,11) K (8,9) T (7,7)
026	GDGACGKT 12 --- 19	1CC0 (5.0) (A,C)	1FTN (2.1) 1TX4 (1.6) (B)	G (1,4) D (3,9) G (3,1) A (7,8)	C (3,1) G (2,1) K (1,1) T (1,0)
027	GLHAMGKT 24 --- 31	1CP2 (1.9) (A,B)		G (3,2) L (1,4) H (2,3) A (3,5)	M (3,7) G (8,1) K (3,3) T (1,1)
028	GAPANGKT 513 --- 520	1CWV (2.3) (A)			
029	GQTGSGKT 474 --- 481	1CZ7 (2.9) (A–D) 3KAR (2.3)	2NCD (2.5) (A)	G (5,4) Q (4,4) T (3,13) G (11,11)	S (11,16) G (19,13) K (11,9) T (6,4)
030	GATGTGKT 39 --- 46	1D2M (1.9) (A)	1D9Z (3.1) (A)	G (8,3) A (3,5) T (5,6) G (14,7)	T (3,2) G (2,15) K (17,13) T (13,1)
030A	GATGTGKT 39 --- 46	1D9X (2.6) (A)			
031	GPPHSGKT 543 --- 550	1D2N (1.7) (A)	1NSF (1.9)	G (1,1) P (0,1) P (1,1) H (1,2)	S (1,1) G (1,2) K (2,0) T (1,2)
0032	GEQAVGKT 18 --- 25	1D5C (2.3) (A)			

Continued

Table I. *Continued*

No.	Sequence Segment location	PDB code, (resolution in Å) chain identifier		Residue r.m.s. (ϕ,ψ)	Residue r.m.s. (ϕ,ψ)
033	GPTGVGKT 57 --- 64	1DO0 (3.0) (A–F) 1E94 (2.8) (E,F) 1G41 (2.3) (A) 1G4B (7.0) (E,F,K,L)	1DO2 (4.0) (A,C) 1G3I (3.4) (A–F) 1G4A (3.0) (E,F)	G (8,8) P (9,10) T (10,11) G (17,24)	V (19,19) G (13,20) K (18,11) T (13,10)
033A	GPTGVGKT 57 --- 64	1DO2 (4.0) (B,D)		G (2,1) P (1,3) T (3,1) G (2,1)	V (1,1) G (1,2) K (1,2) T (1,1)
033B	GPTGVGKT 57 --- 64	1G3I (3.4) (S,T,U,V,W)		G (1,1) P (1,1) T (0,1) G (0,0)	V (0,1) G (1,1) K (0,0) T (0,0)
034	GTEFEGKT 44 --- 51	1DT0 (2.1) (A,B,C)		G (0,1) T (2,1) E (1,1) F (1,2)	E (2,1) G (1,2) K (1,2) T (2,2)
035	GKGGVGKT 15 --- 22	1F48 (2.3) (A)			
036	GLQGSGKT 105 --- 112	1FFH (2.0) 2FFH (3.2) (A,B,C) 3NG1 (2.3) A,B	1NG1 (2.0) 2NG1 (2.0)	G (5,7) L (6,4) Q (4,5) G (10,19)	S (19,7) G (9,7) K (3,4) T (4,5)
037	GRPGTGKT 50 --- 57	1FNN (2.0) (A,B)		G (2,1) R (2,4) P (2,2) G (1,7)	T (4,4) G (2,2) K (1,2) T (2,1)
038	GAPVDGKT 116 --- 123	1FS7 (1.6) (A) 1FS9 (2.0) (A)	1FS8 (1.6) (A)	G (2,2) A (3,1) P (2,1) V (1,3)	D (1,2) G (1,0) K (2,0) T (2,2)
039	GVNGVGKT 300 --- 307	1FTS (2.2)			
040	GLDNAGKT 24 --- 31	1FZQ (1.7) (A)			
041	GPSGCGKT 36 --- 43	1G29 (1.9) (1,2)		G (3,2) P (2,1) S (1,3) G (1,2)	C (1,1) G (5,5) K (1,1) T (2,0)
042	GGTGSGKT 178 --- 185	1G6O (2.5) (A,B)		G (1,4) G (8,3) T (2,5) G (4,1)	S (2,2) G (1,3) K (3,1) T (0,3)
043	GPPGLGKT 45 --- 52	1HQC (3.2) (A,B)		G (1,9) P (9,6) P (2,15) G (25,13)	L (5,8) G (8,11) K (16,4) T (0,2)
044	GKGGTGKT 10 --- 17	1HYQ (2.6) (A)			
045	GKVTSGKT 102 --- 109	1JCK (3.5) (B,D)		G (0,0) K (0,0) V (0,0) T (0,0)	S (0,0) G (0,0) K (0,0) T (0,0)
046	GARGCGKT 9 --- 16	1SHK (1.9) (A,B)	2SHK (2.6) (A,B)	G (5,5) A (4,1) R (4,2) G (3,4)	C (7,2) G (2,8) K (7,3) T (4,1)
047	GLDRTGKT 12 --- 19	1TMK (2.1) (A,B) 3TMK (2.0) (A–H)	2TMK (2.4) (A,B)	G (5,1) L (4,9) D (5,14) R (9,8)	T (9,9) G (13,10) K (7,4) T (2,5)

Continued

Table I. *Continued*

No.	Sequence Segment location	PDB code, (resolution in Å) chain identifier		Residue r.m.s. (ϕ,ψ)	Residue r.m.s. (ϕ,ψ)
048	GNSSVGKT 29 --- 36	1ZBD (2.6) (A)	3RAB (2.0) (A)	G (5,2) N (1,2) S (3,3) S (2,3)	V (3,2) G (1,5) K (1,5) T (4,2)
049	GLEGAGKT 10 --- 17	4TMK (1.9) (A)	5TMP (1.9) (A)	G (3,1) L (2,2) E (3,3) G (2,9)	A (1,9) G (15,7) K (9,9) T (2,3)
050	GAGGVGKS 10 --- 17	121P (1.5) 1CTQ (1.2) (A) 1GNQ (2.5) 1LFD (2.1) (B,D) 1QRA (1.6) (A) 221P (2.3) 5P21 (1.3) 6Q21 (1.9) (A–D)	1BKD (2.8) (R) 1GNP (2.7) 1GNR (1.8) 1Q21 (2.2) 1WQ1 (2.5) (R) 4Q21 (2.0) 621P (2.4) 721P (2.0)	G (6,10) A (13,7) G (9,11) G (10,10)	V (14,13) G (15,9) K (7,11) S (7,7)
051	GADGVGKS 10 --- 17	1AGP (2.3)			
052	GAGESGKS 36 --- 43	1AGR (2.8) (A,D) 1AZT (2.3) (A,B) 1BOF (2.2) 1CJK (3.0) (C) 1CJU (2.8) (C) 1CS4 (2.5) (C) 1FQJ (2.0) (A,D) 1GDD (2.2) 1GG2 (2.4) (A) 1GIL (2.3) 1GOT (2.0) (A) 1TAD (1.7) (A,B,C) 1TND (2.2) (A,B,C)	1AZS (2.3) (C) 1BH2 (2.1) 1CIP (1.5)(A) 1CJT (2.8) (C) 1CJV (3.0) (C) 1CUL (2.4) (C) 1FQK (2.3) (A,C) 1GFI (2.2) 1GIA (2.0) 1GIT (2.6) 1GP2 (2.3) (A) 1TAG (1.8)	G (14,5) A (7,5) G (7,7) E (8,8)	S (9,10) G (10,9) K (7,5) S (6,5)
053	GIVSYGKS 211 --- 218	1AU8 (1.9) (A)	1CGH (1.8) (A)	G (3,1) I (0,0) V (1,3) S (5,2)	Y (1,1) G (1,3) K (6,1) S (2,3)
054	GDGTGGKS 78 --- 85	1CYN (1.8) (A)			
055	GPSGTGKS 8 --- 15	1EX6 (2.3) (A,B) 1GKY (2.0)	1EX7 (1.9) (A)	G (6,5) P (3,5) S (10,6) G (4,6)	T (7,19) G (27,7) K (5,3) S (4,9)
056	GSGGVGKS 10 --- 17	1C1Y (1.9) (A) 1KAO (1.7) 3RAP (2.2) R,S	1GUA (2.0) (A) 2RAP (2.6)	G (5,6) S (5,4) G (5,7) G (7,10)	V (10,3) G (4,4) K (3,4) S (6,6)
057	GDTSDGKS 183 --- 189	1HYL (1.8) (A,B)		G (1,5) D (3,2) T (2,2) S (5,6)	D (5,3) G (2,3) K (4,7) S (6,2)
058	GTAFEGKS 44 --- 51	1ISA (1.8) (A) 1ISC (1.8) (A)	1ISB (1.8) (A)	G (1,2) T (0,0) A (1,1) F (1,0)	E (1,0) G (1,4) K (2,0) S (0,1)
058A	GTAFEGKS 44 --- 51	1ISA (1.8) (B) 1ISC (1.8) (B)	1ISB (1.8) (B)	G (1,2) T (1,1) A (2,2) F (1,2)	E (2,1) G (2,2) K (3,2) S (3,1)
059	GKGGIGKS 9 --- 16	1CP2 (1.9) (A,B) 1FP6 (2.1) (A–D) 1G21 (3.0) (E–H) 1N2C (3.0) (E–H) 2NIP (2.2) (A,B)	1DE0 (2.4) (A,B) 1G1M (2.2) (A,B) 1G5P (2.2) (A,B) 1NIP (2.9) (A,B)	G (17,11) K (14,17) G (16,19) G (12,7)	I (24,21) G (23,13) K (13,11) S (10,12)
059A	GKGGIGKS 9 --- 16	1G20 (2.2) (E,F)		G (8,9) K (13,6) G (8,1) G (2,7)	I (6,14) G (16,2) K (1,4) S (3,5)

Continued

Table I. *Continued*

No.	Sequence Segment location	PDB code, (resolution in Å) chain identifier		Residue r.m.s. (ϕ,ψ)	Residue r.m.s. (ϕ,ψ)
059B	GKGGIGKS 9 --- 16	1G20 (2.2) (G,H)		G (12,7) K (9,4) G (19,30) G (1,22)	I (6,9) G (18,9) K (2,12) S (4,5)
060	GAVGVGKS 10 --- 17	1HE8 (3.0) (B) 2Q21 (2.2)	1RVD (1.9) (A) 521P (2.6)	G (16,2) A (4,6) V (13,3) G (9,15)	V (13,9) G (8,10) K (10,10) S (8,9)
061	GMVVEGKS 190 --- 197	1DJO (2.0) (A,B) 3PGA (2.) (1 – 4)	1DJP (1.9) (A,B) 4PGA (1.7) (A,B)	G (5,3) M (3,4) V (5,2) V (2,5)	E (9,6) G (10,12) K (6,4) S (5,3)
062	GARGVGKS 10 --- 17	421P (2.2)			
063	GIIAPGKS 539 --- 546	1AHU (2.7) (A,B) 1AHZ (3.3) (A,B) 1E0Y (2.7) (A,B) 1E8G (2.1) (A,B) 1QLT (2.2) (A,B) 1VAO (2.5) (A,B)	1AHV (3.1) (A,B) 1DZN (2.8) (A,B) 1E8F (2.9) (A,B) 1E8H (2.6) (A,B) 1QLU (2.4) (A,B) 2VAO (2.8) (A,B)	G (8,13) I (14,12) I (6,5) A (7,11)	P (8,6) G (6,7) K (5,6) S (5,6)
064	GAVESGKS 40 --- 47	1AS0 (2.0) 1AS3 (2.4)	1AS2 (2.8)	G (3,1) A (1,4) V (7,3) E (2,3)	S (4,3) G (6,3) K (2,3) S (4,3)
065	GSSGSGKS 39 --- 46	1B0U (1.5) (A)			
066	GVRFPGKS 313 --- 320	1BAG (2.5)			
067	GGARSGKS 406 --- 413	1C9K (2.2) (A,B,C)	1CBU (2.3) (A,B,C)	G (7,9) G (7,10) A (8,7) R (6,5)	S (5,6) G (5,7) K (7,3) S (6,10)
068	GFAKTGKS 195 --- 202	1CA1 (1.9) 1QMD (2.2) (A,B)	1QM6 (2.5) (A,B)	G (1,3) F (2,1) A (2,3) K (1,1)	T (2,1) G (2,3) K (2,1) S (1,2)
069	GLLEAGKS 174 --- 181	1CD1 (2.6) (A,C)		G (9,4) L (6,3) L (5,5) E (6,12)	A (13,4) G (10,21) K (16,3) S (3,6)
070	GAPGVGKS 10 --- 17	1CLU (1.7) (A) 1JAI (1.8)	1JAH (1.8) 821P (1.5)	G (5,2) A (4,3) P (5,2) G (8,3)	V (2,2) G (4,3) K (3,2) S (3,3)
071	GGSCTGKS 434 --- 441	1CLV (2.0) (A) 1TMQ (2.5) (A)	1JAE (1.6) 1VIW (3.0) (A)	G (5,13) G (7,8) S (5,7) C (11,2)	T (5,8) G (8,28) K (21,16) S (14,32)
072	GRSGRGKS 138 --- 145	1HJB (3.0) (C,F) 1IO4 (3.0) (C)	1HJC (2.6) (A,D)	G (8,4) R (11,9) S (12,12) G (14,5)	R (3,6) G (5,11) K (11,3) S (4,3)
073	GPLYLGKS 587 --- 594	1CMX (2.2) (A,C)		G (0,1) P (1,1) L (0,2) Y (2,2)	L (2,3) G (3,1) K (1,2) S (3,0)
074	GLSASGKS 32 --- 39	1D6J (2.0) (A,B)		G (5,2) L (1,2) S (2,1) A (4,7)	S (5,4) G (1,1) K (2,3) S (4,2)
075	GQAMPGKS 105 --- 112	1DDZ (2.2) (A,B)		G (1,1) Q (0,0) A (1,0) M (1,0)	P (1,1) G (1,1) K (0,1) S (1,1)

Continued

Table I. *Continued*

No.	Sequence Segment location	PDB code, (resolution in Å) chain identifier		Residue r.m.s. (ϕ,ψ)	Residue r.m.s. (ϕ,ψ)
076	GVVQPGKS 510 --- 517	1DEE (2.7) (B,D,F)		G (1,1) V (1,1) V (0,2) Q (5,1)	P (2,1) G (1,2) K (1,0) S (2,3)
077	GLYRTGKS 45 --- 52	1DG3 (1.8) (A)			
077A	GLYRTGKS 45 --- 52	1F5N (1.7) (A)			
078	GENGIGKS 42 --- 49	1DYW (1.8) (A)	1E3B (1.8) (A)	G (2,2) E (2,1) N (3,2) G (3,3)	I (1,1) G (0,2) K (2,2) S (4,2)
079	GRPNVGKS 15 --- 22	1EGA (2.4) (A,B)		G (2,0) R (2,3) P (2,1) N (1,3)	V (4,3) G (4,1) K (2,1) S (1,1)
080	GAAAAGKS 893 --- 900	1EJ6 (3.6) (A)			
081	GEAAVGKS 14 --- 21	1EK0 (1.4) (A)			
082	GSVAVGKS 95 --- 102	1ESM (2.5) (A–D)	1ESN (2.6) (A–D)	G (3,5) S (13,14) V (4,2) A (13,10)	V (3,2) G (8,3) K (6,17) S (8,10)
083	GAEAAGKS 188 --- 195	1F07 (2.0) (A–D)		G (1,3) A (4,2) E (3,4) A (4,2)	A (4,4) G (4,2) K (2,4) S (5,3)
084	GQNGSGKS 30 --- 37	1F2T (1.6) (A)	1F2U (1.6) (A,C)	G (13,3) Q (3,12) N (8,4) G (16,14)	S (13,8) G (7,3) K (7,13) S (2,6)
085	GHVDSGKS 14 --- 21	1F60 (1.6) (A)	1G7C (2.0) (A)	G (0,1) H (5,4) V (3,4) D (0,2)	S (4,6) G (5,3) K (1,1) S (1,3)
086	GDSGVGKS 27 --- 34	1G16 (1.8) (A–D)	1G17 (2.0) (A,B)	G (3,3) D (3,3) S (1,5) G (4,4)	V (4,2) G (2,3) K (3,2) S (2,2)
087	GLVSPGKS 951 --- 958	1HQM (3.3) (C)			
088	GESAVGKS 28 --- 35	1HUQ (1.8) (A)			
089	GIPGVGKS 8 --- 15	1NKS (2.5) (A–F)		G (7,2) I (4,3) P (4,7) G (12,11)	V (10,8) G (12,9) K (6,3) S (5,8)
090	GDVSPGKS 457 --- 464	1QHB (2.3) (A–F)		G (2,2) D (2,2) V (1,2) S (1,1)	P (2,1) G (3,2) K (3,3) S (4,3)
091	GGNGAGKS 34 --- 41	1QHL (2.2) (A)			
092	GGSSAGKS 10 --- 17	1QHN (2.7) (A) 1QHX (2.5) (A)	1QHS (2.8) (A) 1QHY (2.6) (A)	G (3,2) G (4,3) S (3,2) S (2,1)	A (1,2) G (4,3) K (3,1) S (4,5)

For those groups which have more than one entry, structural similarity is brought out by the small r.m.s. values of the Ramachandran angles (ϕ,ψ), given in the last column.

059 – GKGGIGKS); and four for one sequence (003 – GGAGVGKT). These data implied that highly localized conformational variants are possible in these segments retaining overall structural similarity.

Conformational variants of segments of Walker sequences

The next step was the grouping of the conformations irrespective of the sequence of the variable region of the Walker segment. This was done as follows: (1) for those

Table II. The entries selected from Table I regrouped based on their structural similarity [the examples in set VII do not possess any structural similarity, as analyzed using the Ramachandran angles (ϕ,ψ)]

Sr. No.	Sequence of Walker motif	Segment location	PDB code (chain)	Name of the protein
Set I				
1	G L S G T G K T	248–255	1AYL	Phosphoenol pyruvate kinase
2	G D R Q T G K T	169–176	1E79 (A)	F_1-ATPase α subunit
3	G P E S S G K T	66–73	2REB	REC A protein.
4	G L P A R G K T	45–52	1BIF	6-Phosphofructo-2-kinase/fructose-2,6-bisphosphatase
5	G Q T G S G K T	474–481	3KAR	Kinesin-like protein KAR3
6	G A T G T G K T	36–43	1D2M (A)	UvrB protein
7	G P P H S G K T	551–558	1D2N (A)	*N*-Ethylmaleimide-sensitive fusion protein
8	G P T G V G K T	57–64	1G41 (A)	Heat shock protein HslU
9	G L Q G S G K T	105–112	1FFH	Signal recognition protein FFH
10	G R P G T G K T	50–57	1FNN (A)	CDC6p
11	G L D N A G K T	24–31	1FZQ (A)	ADP-ribosylation-like factor
12	G P S G C G K T	36–43	1G29 (1)	Mal K
13	G L D R T G K T	12–19	3TMK (A)	Thymidylate kinase (*S.cerevisiae*)
14	G L E G A G K T	10–17	4TMK (A)	Thymidylate kinase (*E.coli*)
15	G P S G T G K S	8–15	1EX7 (A)	Guanylate kinase
16	G S G G V G K S	10–17	1KAO	Small G-protein rap2a + GDP
17	G K G G I G K S	8–15	1CP2 (A)	Nitrogenase iron protein
18	G S S G S G K S	39–46	1B0U (A)	Histidine permaease HisP–ATP binding subunit
19	G G A R S G K S	6–13	1C9K (A)	Adenosylcobinamide kinase
20	G L Y R T G K S	45–52	1F5N (A)	Guanylate binding protein
21	G E A A V G K S	14–21	1EK0 (A)	YPT51
22	G Q N G S G K S	30–37	1F2T (A)	Rad50 ABC ATPase
23	G D S G V G K S	27–34	1G16 (A)	Sec4
24	G E S A V G K S	28–35	1HUQ (A)	Rab5c
25	G I P G V G K S	8–15	1NKS (A)	Adenylate kinase
26	G G S S A G K S	10–17	1QHX (A)	Chloramphenicol phosphotransferase
27	G H V D H G K T	18–25	1EXM (A)	EF-Tu
28	G P H G M G K T	56–63	1E2K (A)	Thymidine kinase
29	G E S G A G K T	179–186	1MMG	Myosin motor domain
30	G D G G T G K T	17–24	1BYU (A)	RAN-GTPase (+ GDP)
31	G D G A V G K T	10–17	1MH1	RAC1
32	G D G A C G K T	12–19	1TX4 (B)	Rho A
33	G E Q A V G K T	18–25	1D5C (A)	Rab6 + GDP
34	G K G G V G K T	15–22	1F48 (A)	Arsenite transporting ATPase
35	G V N G V G K T	300–307	1FTS	Signal recognition particle receptor
36	G G T G S G K T	178–185	1G6O (A)	Traffic ATPase (+ ADP)
37	G A R G C G K T	9–16	1SHK (A)	Shikimate kinase
38	G N S S V G K T	29–36	3RAB (A)	Rab3a
39	G A G G V G K S	10–17	1CTQ (A)	P21 Ras
40	G A G E S G K S	40–47	1CIP (A)	Guanine nucleotide binding protein alpha-I
41	G R P N V G K S	15–22	1EGA (A)	GTP binding protein ERA
42	G S V A V G K S	95–102	1ESM (A)	Pantothenate kinase
43	G G A G V G K T	156–163	1E79 (D)	F_1-ATPase β subunit
44	G L D A A G K T	24–31	1HUR (A)	ADP-ribosylation factor-1
45	G Q T S S G K T	85–92	1BG2	Kinesin motor domain
Set II				
46	G H V D H G K T	18–25	1EFU (A)	EF-Tu
47	G L S A S G K S	32–39	1D6J (A)	APS kinase
48	G H V D S G K S	14–21	1F60 (A)	Elongation factor EEF1A
49	G K G G I G K S	9–16	1G20 (G)	Nitrogenase iron protein
50	G L Y R T G K S	45–52	1DG3 (A)	Guanylate binding protein-1
Set III				
51	G Y L V N G K T	264–271	6MHT (A)	HhaI methyltransferase
52	G M V V E G K S	190–197	4PGA (A)	Glutaminase–asparaginase
Set IV				
53	G L H A M G K T	24–31	1CP2 (A)	Nitrogenase iron protein
54	G A E A A G K S	188–195	1F07 (A)	Tetrahydromethanopterin reductase
Set V				
55	G F A K T G K S	195–202	1CA1	Alpha-toxin
56	G L L E A G K S	174–181	1CD1 (A)	CD1
Set VI				
57	G A T G T G K T	39–46	1D9X (A)	UvrB
58	G P P G L G K T	45–52	1HQC (A)	RuvB
Set VII				
59	G G A G V G K T	156–163	1BMF (E)	F_1-ATPase β subunit (bovine)
60	G G A G V G K T	156–163	1E1R (E)	F_1-ATPase β subunit (bovine)

Continued

Table II. *Continued*

Sr. No.	Sequence of Walker motif	Segment location	PDB code (chain)	Name of the protein
61	G G A G V G K T	156–163	1MAB (B)	F_1-ATPase β subunit (rat)
62	G A H A L G K T	173–180	2CYP	Cytochrome *c* peroxidase
63	G A G A L G K T	173–180	1DS4 (A)	Cytochrome *c* peroxidase
64	G H V D H G K T	18–25	1ETU	EF-Tu domain 1
65	G T A F P G K T	212–219	1QPA (A)	Lignin peroxidase
66	G L R S D G K T	487–494	1MTY (D)	Methane monoxygenase (Mc)[a]
67	G L R S D G K T	487–494	1MMO (D)	Methane monoxygenase (Mt)[a]
68	G K V T G G K T	102–109	1STE	Sec2 superantigen
69	G L P A I G K T	499–506	1BGX (T)	Taq polymerase
70	G L P A I G K T	499–506	1QSS (A)	Taq Klenow fragment
71	G S Q A G G K T	47–54	1WGT (A)	Wheat germ agglutinin
72	G M D L K G K T	206–213	1BVU (A)	Glutamate dehydrogenase
73	G A P A N G K T	513–520	1CWV (A)	Invasin
74	G P T G V G K T	57–64	1DO2 (B)	HslU
75	G P T G V G K T	57–64	1G3I (S)	HslU protease
76	G A P V D G K T	116–123	1FS7 (A)	Cytochrome *c* nitrite reductase
77	G K G G T G K T	10–17	1HYQ (A)	MinD-1
78	G K V T S G K T	102–109	1JCK (B)	Sec3 superantigen
79	G I V S Y G K S	211–218	1CGH (A)	Cathepsin G
80	G D G T G G K S	78–85	1CYN (A)	Cyclophilin B
81	G D T S D G K S	183–189	1HYL (A)	Collagenase
82	G T A F E G K S	44–51	1ISA (A)	Superoxide dismutase
83	G T A F E G K S	44–51	1ISA (B)	Superoxide dismutase
84	G K G G I G K S	9–16	1G20 (E)	Nitrogenase iron protein
85	G I I A P G K S	539–546	1E8G (A)	Vanillyl-alchohol oxidase
86	G V R F P G K S	313–320	1BAG	α-Amylase (*B.subtilis*)
87	G G S C T G K S	434–441	1JAE	α-Amylase (yellow mealworm)
88	G R S G R G K S	138–145	1HJC (A)	RUNT-related transcription factor-1
89	G P L Y L G K S	187–194	1CMX (A)	Ubiquitin
90	G Q A M P G K S	105–112	1DDZ (A)	Carbonic anhydrase
91	G V V Q P G K S	510–517	1DEE (B)	IgM heavy chain
92	G E N G I G K S	42–49	1DYW (A)	Cylophilin 3
93	G A A A A G K S	893–900	1EJ6 (A)	Reovirus core protein λ2
94	G L V S P G K S	951–958	1HQM (C)	Bacterial RNA polymerase β subunit
95	G D V S P G K S	457–464	1QHB (A)	Vanadium bromo-peroxidase
96	G G N G A G K S	34–41	1QHL (A)	MukB N-terminal domain

[a]Mc, *Methylococcus capsulatus*; Mt, *Methylosinus trichosporum*

Table III. Mean and r.m.s. (ϕ,ψ) values (°) for the first six sets given in Table II

	Position No.							
	1	2	3	4	5	6	7	8
Set I [45]								
Mean (ϕ,ψ)	(159,166)	(–67,162)	(–63,141)	(79,12)	(–87,–11)	(104,20)	(–60,–46)	(–66,–40)
R.m.s.(ϕ,ψ)	(18,11)	(10,10)	(8,11)	(14,15)	(17,14)	(16,11)	(7,10)	(5,5)
Set II [5]								
Mean (ϕ,ψ)	(158,171)	(–66,157)	(–60,–36)	(–82,6)	(–86,–5)	(73,29)	(–63,–32)	(–66,–43)
R.m.s.(ϕ,ψ)	(11,7)	(7,8)	(13,7)	(19,15)	(20,14)	(12,13)	(5,9)	(4,8)
Set III [2]								
Mean (ϕ,ψ)	(–138,176)	(–138,154)	(–106,114)	(–127,125)	(41,54)	(84,–11)	(–115,149)	(–104,134)
R.m.s.(ϕ,ψ)	(35,25)	(2,6)	(17,8)	(8,4)	(1,3)	(3,5)	(10,13)	(7,7)
Set IV [2]								
Mean (ϕ,ψ)	(–63,–45)	(–62,–44)	(–65,–38)	(–62,–21)	(–97,7)	(82,16)	(–99,163)	(–114,141)
R.m.s.(ϕ,ψ)	(4,4)	(9,7)	(1,3)	(3,9)	(7,8)	(2,2)	(12,19)	(13,16)
Set V [2]								
Mean (ϕ,ψ)	(–52,–47)	(–67,–45)	(–61,–28)	(–76,–44)	(–73,–36)	(–71,–34)	(–44,–38)	(–63,–49)
R.m.s.(ϕ,ψ)	(8,2)	(4,6)	(5,7)	(7,5)	(8,2)	(4,12)	(17,3)	(1,1)
Set VI [2]								
Mean (ϕ,ψ)	(–141,138)	(–65,151)	(–62,147)	(89,–2)	(–95,144)	(–70,85)	(–78,–40)	(–59,–52)
R.m.s.(ϕ,ψ)	(8,2)	(11,3)	(6,14)	(3,32)	(21,29)	(23,18)	(26,9)	(11,4)

The corresponding values do not have any meaning for set VII, which has structurally dissimilar conformations. The number of examples in each set is given in parentheses along with the set number in the first column.

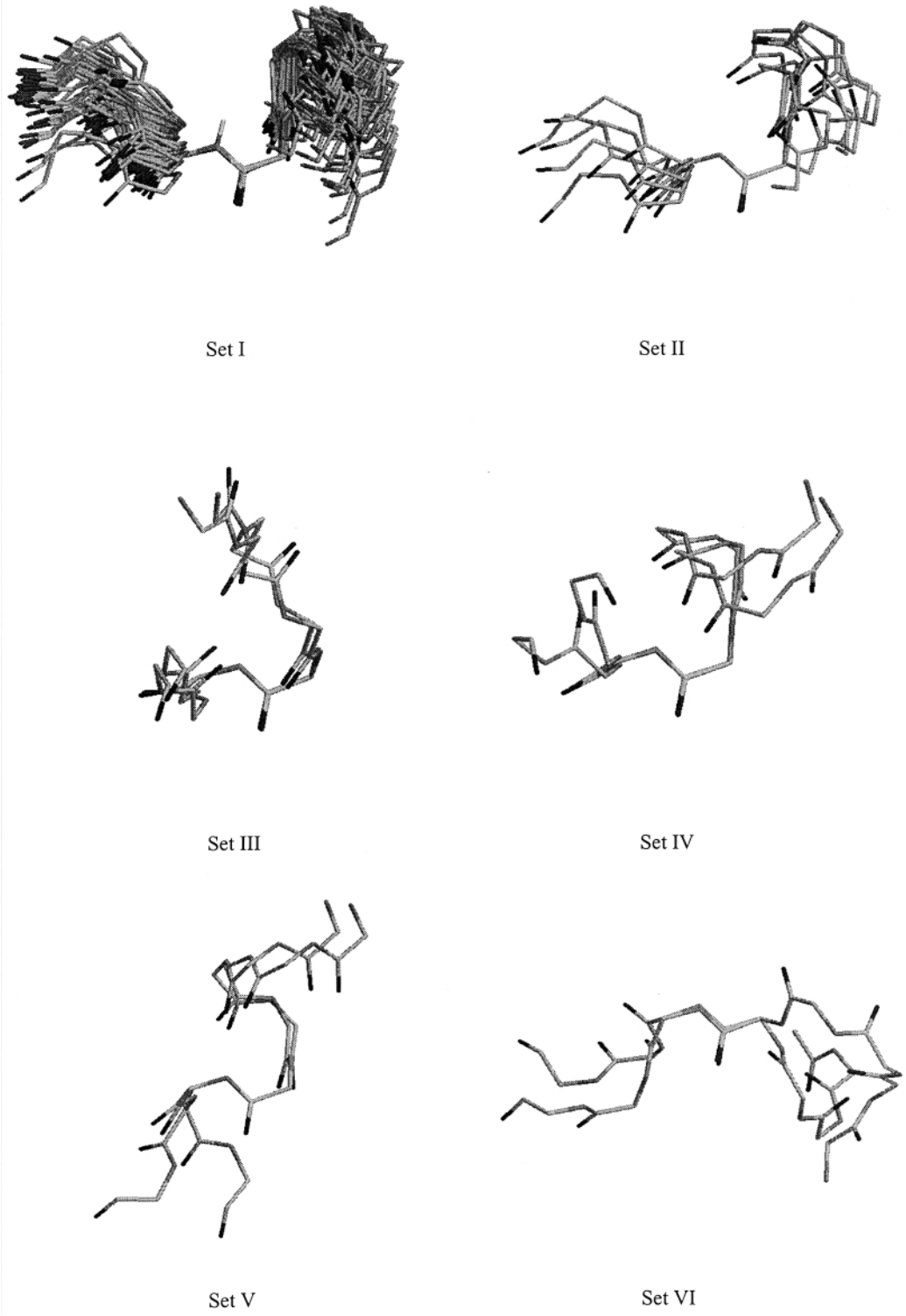
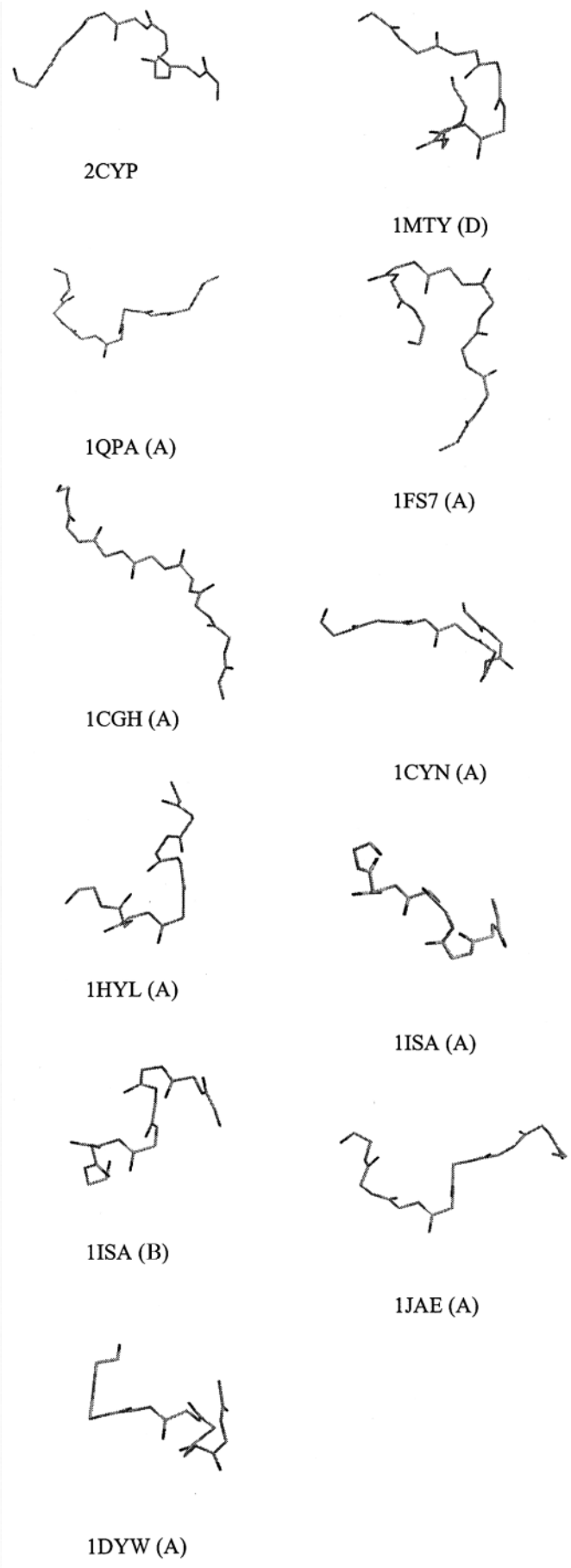


Fig. 1. Wire-frame diagrams of the backbone atoms in the segment GXXXXGKT (S) in the proteins given in sets I–VI of Table II.

groups in Table I which had only one entry, the choice was unambiguous; (2) for those groups having more than one entry, one with the best resolution shown in bold face in Table I had been picked up as the representative of the group/sub-group. These collectively gave 107 examples which were regrouped solely on the basis of similarity of the Ramachandran angles (ϕ, ψ). Of the new sets thus obtained, 53 (out of a total of 107) entries constituted the major set. Another set had five entries, while seven others had two entries each. The last set comprised 35 entries, without any structural similarity among them. In any particular set, proteins with high overall sequence homology could be found, although the sequences of the variable region were different. These are as follows: (1AYL, 1OEN);



(1A4R, 1MH1); (1DPF, 1TX4); (1CIP, 1AS0), (1AGP, 1CTQ, 1RVD, 421P, 821P); pairs [(2CYP, 1CCG); (1MHY (D), 1MTY (D)], as well as [1DT0 (A), 1ISA (B)]. Since the structures in such cases are expected to be similar, the entries that had the best resolution were retained. These were 1AYL, 1MH1, 1TX4, 1CIP, 1CTQ, 2CYP, 1MTY (D) and 1ISA (B). The final grouping thus obtained is given in Table II, which has 45 proteins in set I, five in set II, two each in sets III, IV, V and VI and the remaining 38 in set VII. The mean and r.m.s. (ϕ, ψ) values of sets I–VI are given in Table III and these are small enough to warrant structural similarity among the members. The r.m.s. has no relevance for the last set (set VII).

For easy comprehension of the structural grouping, the line diagrams of the backbone of GXXXXGKT (S) segments of the proteins in sets I–VI, drawn with the peptide unit spanning residues 5–(X) and 6–(G) as the common internal frame of reference, are shown in Figure 1. The sickle-like folding with overlap of the atoms of the backbone is seen with members of the set I (Figure 1). Nearly the same structure of segment 5–8 is found in set II, but that of segment 1–4 is different (Table III). Set VII is comprised of structures of differing conformations indicating the flexibility of Walker sequences to acquire random folding. Out of the 38 examples in this set, only those which have resolution of 1.8 Å or better are shown in Figure 2.

Differing structures with same Walker sequence

Structural differences between segments having the same Walker sequence are also perceivable from the foregoing data. There are nine such examples in the present data set. The PDB codes along with the Ramachandran angles at the eight positions of Walker sequence of these nine pairs are given in Table IV. Large differences in Ramachandran angles are observed at four different locations within the segment (shown in bold face in the table). These are as follows: (i) 1VOM–2MYS (A), (ii) 1F5N (A)–1DG3 (A), (iii) 1FP6 (A)–1G20 (E) and (iv) 1EFT–1EFU (A) all at locations 3/4; (v) 1MMO (D)–1MMO (E) at locations 4/5; (vi) 1ISA (A)–1ISA (B), (vii) 1D9X (A)–1D9Z (A) and (viii) 1BMF (D)–1BMF (E) at locations 5/6; and (ix) 1G3I (A)–1G3I (S) at locations 6/7. These large changes arise owing to a flip of the peptide unit spanning the two residues.

There are more than two examples with differing conformations for two of the sequences. The first example consists of 1EFT, 1EFU (A) and 1ETU, which have the same sequence, GHVDHGKT (iv in Table IV), but the peptide unit between residues 3 and 4 in 1EFT and 1EFU (A) is flipped. The conformation of the third member of this group, 1ETU, does not match in entirety with the other two examples. A close examination reveals that the conformations of 1EFT and 1ETU differ only in the segment 18–20. The second example is of 1BMF (D, E), 1MAB (B) and 1E1R (E) having the same sequence GGAGVGKT. The peptide unit between locations 5 and 6 in 1BMF (E) and 1BMF (D) is flipped, as also is the one between locations 6 and 7 in 1BMF (D) and 1MAB (B) (viii in

Fig. 2. Wire-frame diagrams of the backbone atoms in the segment GXXXXGKT (S) in the proteins belonging to set VII of Table II. Only those examples occurring in protein structures which have a resolution of 1.8 Å or better are shown.

Table IV Ramachandran angles (ϕ,ψ) (°) in the segment GXXXXGKT for those examples with same sequence for XXXX but with different conformations

No.	Sequence, PDB code (chain) and segment locations	(ϕ,ψ) at position							
		1	2	3	4	5	6	7	8
(i)	GESGAGKT, 1VOM, 179–186	(150,155)	(–73,171)	**(–59,131)**	**(86,4)**	(–67,–26)	(117,18)	(–61,–47)	(–67,–37)
	GESGAGKT, 2MYS(A), 179–186	(178,174)	(–72,166)	**(–35,–42)**	**(–101,2)**	(–61,–25)	(82,57)	(–82,–47)	(–44,–48)
(ii)	GLYRTGKS, 1F5N(A), 45–52	(155,173)	(–53,147)	**(–63,157)**	**(60,34)**	(–111,13)	(79,28)	(–66,–52)	(–60,–41)
	GLYRTGKS, 1DG3(A), 45–52	(149,178)	(–75,161)	**(–60,–42)**	**(–54,–22)**	(–71,–14)	(65,31)	(–67,–17)	(–60,–32)
(iii)	GKGGIGKS, 1FP6(A), 9–16	(155,176)	(–64,163)	**(–73,131)**	**(74,13)**	(–79,–24)	(130,20)	(–75,–25)	(–76,–35)
	GKGGIGKS, 1G20(E), 9–16	(137,156)	(–53,138)	**(–80,–35)**	**(–62,10)**	(–97,–49)	(134,39)	(–61,–46)	(–70,–32)
(iv)	GHVDHGKT, 1EFT, 18–25	(147,165)	(–65,164)	**(–66,133)**	**(75,18)**	(–91,–20)	(125,31)	(–67,–46)	(–73,–34)
	GHVDHGKT, 1EFU(A), 18–25	(175,176)	(–63,141)	**(–45,–41)**	**(–102,24)**	(–101,4)	(84,40)	(–68,–44)	(–67,–47)
	GHVDHGKT, 1ETU, 18–25	(–126, –166)	(–120,–41)	(143,82)	(133,–27)	(–47,–37)	(119,10)	(–49,–47)	(–57,–46)
(v)	GLRSDGKT, 1MMO(D), 487–494	(–80,42)	(–125,158)	(–79,161)	**(–62,137)**	**(125,–29)**	(74,23)	(–135,–59)	(–69,136)
	GLRSDGKT, 1MMO(E)[a], 487–494	(–84,50)	(–129,148)	(–70,170)	**(–70,–1)**	**(–86,–2)**	(65,27)	(–145,–53)	(–80,138)
(vi)	GTAFEGKT, 1ISA(A), 44–51	(81,8)	(–112,173)	(–70,–13)	(–74,–23)	**(–69,134)**	**(92,–12)**	(–87,162)	(–79,166)
	GTAFEGKT, 1ISA(B), 44–51	(83,9)	(–107,174)	(–67,–19)	(–68,–21)	**(–64,–25)**	**(–100,31)**	(–126,159)	(–76,161)
(vii)	GATGTGKT, 1D9X(A), 39–46	(–122,133)	(–75,149)	(–68,161)	(92,–34)	**(–73,173)**	**(–92,68)**	(–52,–48)	(–70,–48)
	GATGTGKT, 1D9Z(A), 39–46	(–110,129)	(–68, –169)	(–68,156)	(48,26)	**(–99,2)**	**(76,84)**	(–97,–53)	(–41,–41)
(viii)	GGAGVGKT, 1BMF(E), 156–163	(93,153)	(–125,–107)	(–86,159)	(78,–37)	**(–86,133)**	**(–48,30)**	(–41,–58)	(–69,22)
	GGAGVGKT, 1BMF(D), 156–163	(–174,145)	(–75,176)	(–65,127)	(68,41)	**(–112,–8)**	**(128,5)**	**(–52,–60)**	(–64,–43)
	GGAGVGKT, 1MAB(B), 156–163	(–153,160)	(–76,116)	(–100,105)	(69,89)	(–160,–25)	**(179, –158)**	**(88, –106)**	(–45,–52)
	GGAGVGKT, 1E1R(E), 156–163	(–170,–87)	(111, –176)	(–74,–71)	(66,17)	(–131,23)	(79,17)	(–51,–54)	(–63,–35)
(ix)	GPTGVGKT, 1G3I(A), 57–64	(–161,179)	(–82, –172)	(–80,136)	(43,65)	(–124,8)	**(126,–23)**	**(–34,–51)**	(–80,–40)
	GPTGVGKT, 1G3I(S), 57–64	(–174,158)	(–47, –165)	(–118,92)	(141,–19)	(73,–4)	**(82,98)**	**(–120,–23)**	(–89,–33)

[a]The entry 1MMO (E) has been taken in the place of 1MTY (D) of Table II since both of these are structurally and sequentially highly homologous.

Table V. Distribution of the different types secondary structures flanking Walker sequence GXXXXGKT (S) in the examples given in Tables I and II

Secondary structure flanking Walker sequence		Examples in Table I	Examples in Table II
Preceding	Following		
α	α	62	9
α	β	19	4
α	X	91	9
β	α	418	60
β	β	21	8
β	X	4	3
X	α	9	0
X	β	23	3
X	X	2	1

α = α-Helix; β = β-strand; X = neither α nor β.

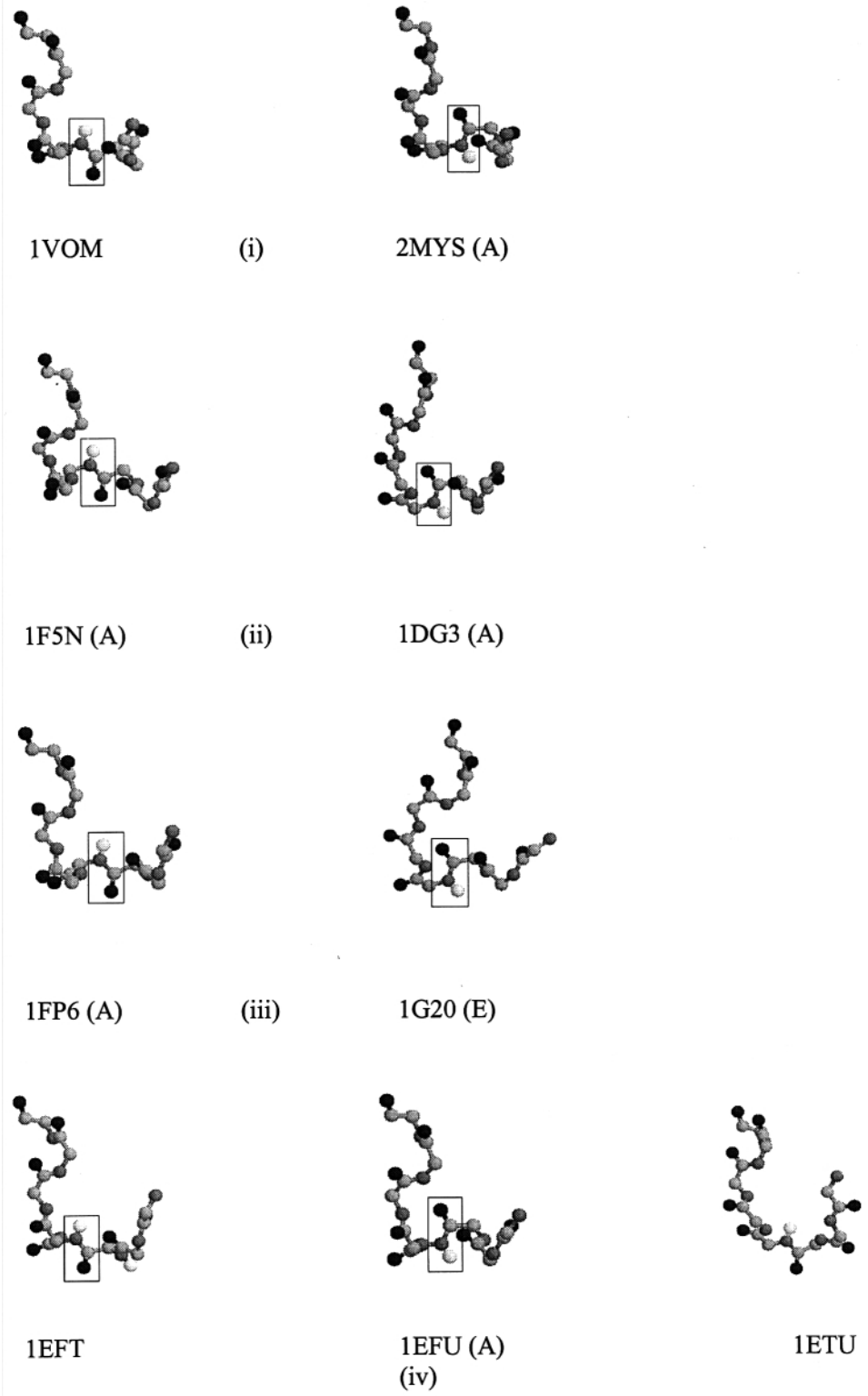
Table IV, shown as overlapping boxes). The fourth entry, 1EIR (E), has an altogether different conformation.

The last entry in Table IV corresponds to the sequence GPTGVGKT occurring in the two chains A and S of the protein 1G3I and the conformations are different. In this case the peptide unit between locations 6 and 7 show a rotation of ≈90° about the virtual C^{α}–C^{α} bond, instead of a flip, as is found in the other examples.

The ball and stick diagrams of these nine examples with a flipped peptide unit shown within a box are given in Figure 3. The overlap of the polypeptide backbones appears good. The examples of pairs i–vi correspond to the flip occurring at the middle peptide unit of the well-known 4→1 hydrogen-bonded β-turns of types I and II (Venkatachalam, 1968; Gunasekaran *et al.*, 1998). However, the flip of the peptide unit observable in pairs vii–ix does not correspond to the β-turn flip as the values of (ϕ, ψ) are far different from those characteristic of β-turn ranges. Further, the 4→1 hydrogen bond is also absent. Notwithstanding the flip, the same overall backbone structure is retained.

The examples of nucleotide-binding proteins are arranged in Figure 3 with the nucleotide bound forms on the left and the free forms on the right. These are as follows: myosin ATPase [1VOM–2MYS (A)], guanylate binding protein [1F5N (A)–1DG3(A)], nitrogenase, [1FP6(A)–1G20(E)] elongation factor Tu [1EFT–1EFU(A), uvrB protein (A)–DNA helicase [1D9Z(A)–1D9X(A)], F_1-ATPase [1BMF(D)–1BMF(E) and 1MAB(B)] and the HSLUV protease chaperone complex [1G3I(A)–1G3I(S)]. Wherever the nucleotide is bound the N–H of the flipped peptide unit projects inwards of the loop. In the case of F_1-ATPase (Abrahams *et al.*, 1994), this N–H forms a hydrogen bond with P=O of the β-phosphate. It appears that the presence or absence of the nucleotide makes the difference between the two structural forms. The residues of Walker sequence in such proteins not only bind to the nucleotide phosphates but also show consequent localized structural changes. This feature has important implications in the biochemical events that occur at this site.

In the case of proteins with oxygen-related reactions, the difference appears to be present in the polypeptides as isolated. The two proteins of methane monooxygenase, showing a flip at position 4/5, are derived from two organisms. The Walker sequence is present only in the Fe-form of superoxide dismutase and the two identical subunits of this enzyme protein exhibit this flip of a peptide unit. It is possible that the O=O group may act as the P=O in nucleotide phosphate in protein–substrate interactions. No relationship has so far been found between the peptide flips in Walker sequences and the activities of these proteins.



The secondary structures flanking the Walker sequence

The foregoing analysis indicated that the variable region is unlikely to determine the conformation of the Walker sequence A found in many nucleotide-binding proteins. The characteristic loop structure of the Walker sequence in these proteins is known to be preceded by a β-strand and followed by an α-helix (see, for an example, Abrahams *et al.*, 1994). It was therefore of interest to examine the occurrence of the flanking secondary structure of Walker sequences in proteins listed in Tables I and II. For this purpose, segments of eight residues on either side of Walker sequences were examined for the presence of secondary structures (α = α-helix; β = β-strand; X = neither α nor β; W = Walker sequence A). All nine possible combinations do occur and their distribution is given in Table V. The majority of the examples fall into the category of β–W–α. This structural motif is present in all cases in the sets 1, 2 and 6 and some in set 7 of Table II. Interestingly, each of these proteins can bind to nucleotides leading to

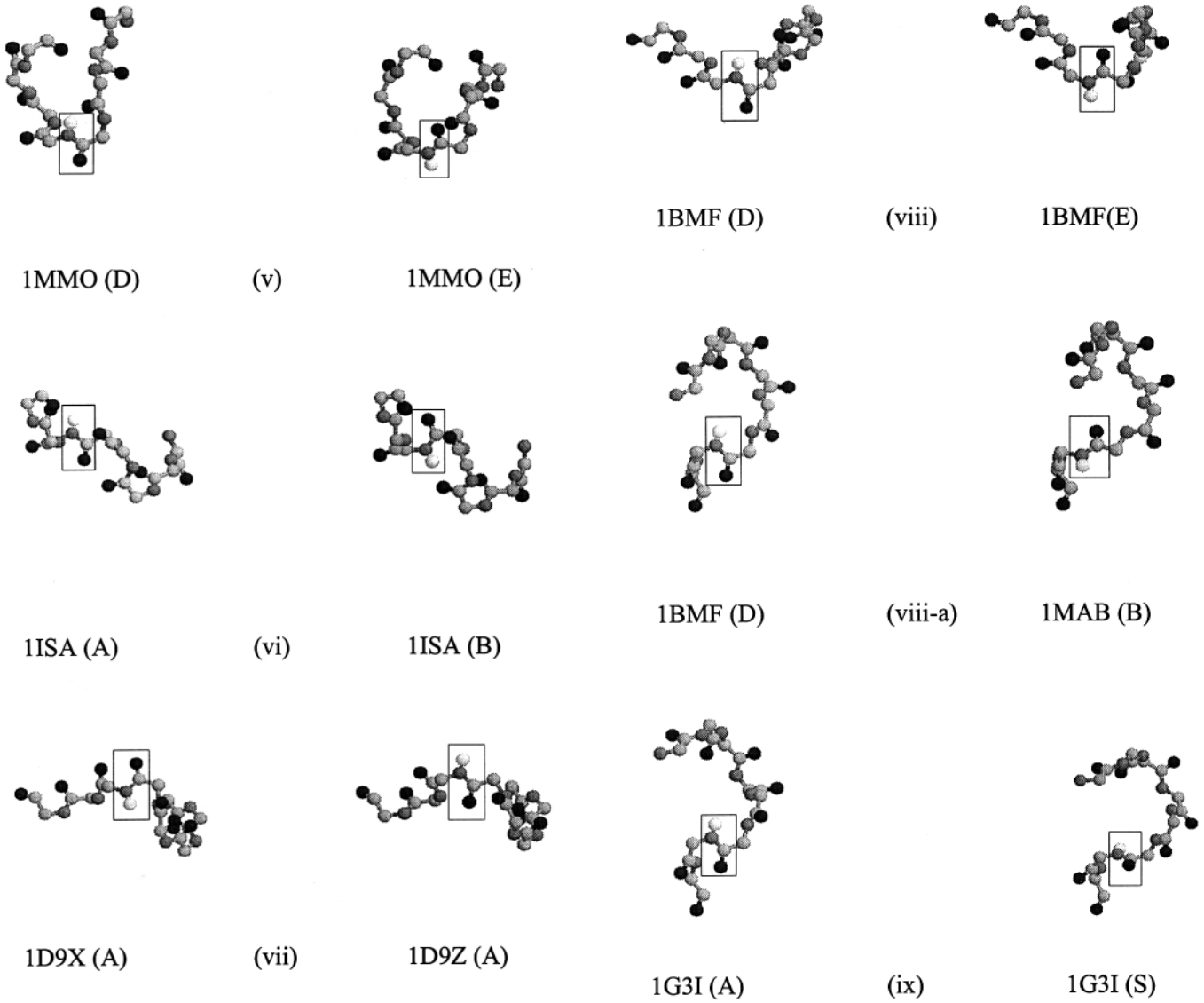


Fig. 3. Ball and stick diagrams of the backbone atoms in the segment GXXXXGKT (S) in the examples given in Table II. The flip of the peptide unit can be seen in the box, shown at the stated positions; 3 and 4 for i–iv; 4 and 5 for v; 5 and 6 for vi–viii; 6 and 7 for viii-a and ix. Shown as a white ball in this peptide unit, the hydrogen atom has been geometrically fixed.

hydrolysis of the terminal phosphate to provide energy for accompanying reactions (e.g. ATPases) in a large number of cases and in some cases transfer the phosphate to acceptors (kinases). This is true of the examples of proteins in the miscellaneous set 7. Hence it appears that the structural motif β–W–α, but not W alone, is the determining factor for nucleotide binding. The examples in sets 3, 4 and 5 of Table II, although small in number (only two each), show distinctive motifs of X–W–β, α–W–β and α–W–α, respectively.

Discussion

The noteworthy observation in this study is that the Walker sequence is present in many proteins and is not limited to those that bind and/or use nucleotides in their actions. Because of the belief that it provides the loop for phosphate binding, the so-called P-loop, this sequence was looked for only in such proteins and was invariably found. A search in the PDB files for its general occurrence, undertaken in this study, revealed its broad distribution (Tables I and II). The diversity of these proteins is truly amazing. These include peroxidases (of cytochrome, lignin), proteases (cathepsin, collagenase, serine protease), enzymes (methane monooxygenase, superoxide dismutase, α-amylase, glutamate dehydrogenase, carbonic anhydrase, Taq polymerase, etc.), binding proteins (lectin, trypsin inhibitor) and miscellaneous proteins (α-toxin, cyclophilin B, enterotoxin). An examination of the structures of cytochrome peroxidase and superoxide dismutase indicated that these sequences are present at some distance from the active metal centers. It is to be ascertained whether Walker sequences in these proteins are utilized in their actions or their presence is incidental. It becomes obvious that the Walker sequence is more widely distributed and presence of the P-loop seems to be restricted to the nucleotide binding proteins.

The second observation in this study is the sharing of a common loop structure in proteins of the major group which use and bind nucleotide phosphates. These include kinases, phosphatases, ATPases, heat shock proteins, transfer/transport

ATPases, permeases, myosin motor domain and elongation factor. The variable quartet (XXXX) has little influence on the bend as seen from the minor variation of overlap in this region (Figure 1). Indeed, the variable quartet is so highly random in sequence that it gives no clue on the looping. Of these, G (13.3%), A (11.9%), S (9.8%), V (8.4%) and T (5.9%) occur more commonly than other amino acids, but no sequence can be identified with a set or a sub-set of proteins. Thus the formation of the β-turn loop seems to depend less on this sequence and more on the polypeptide chains on either side of the P loop, characteristically a β-sheet at the N-terminus and an α-helix at the C-terminus. The absence of the classical 4→1 hydrogen bond in these loop structures appears to provide more room to surround and manipulate the phosphate chain of nucleotides for exchanging terminal phosphate.

Finally, the minor, local differences in the structures with the same Walker sequence, in our opinion, are of importance as they offer possibilities of participation in the functions of these proteins. These relate to the flip of peptide units in four positions (3–4, 4–5, 5–6, 6–7 in Table IV) in these sequences. The large differences in Ramachandran angles indeed brings to light these structural variants. Three examples are noted in the pairs that show these flips: the same enzyme protein from two different organisms (methyl monooxygenase), the two subunits of a homodimer protein (Fe-superoxide dismutase) and the binding of nucleotide to one of the two subunits (F_1-ATPase, β-subunit). The last example is a case with possible interaction of the substrate and the backbone structure of the enzyme active site and offers interesting mechanistic possibilities. Details of this have been reported elsewhere (Ramasarma and Ramakrishnan, 2002).

Acknowledgements

T.R. is a senior scientist of the Indian National Science Academy, New Delhi. C.R. and V.S.D. acknowledge financial assistance from the Council of Scientific and Industrial Research, New Delhi, India.

References

Abrahams,J.P., Leslie,A. G., Lutter,R. and Walker,J. E. (1994) *Nature*, **370**, 621–628.

Berchtold,H., Reshetnikova,L., Reiser,C.O., Schirmer,N.K., Sprinzl,M. and Hilgenfeld,R. (1993) *Nature*, **365**, 126–132.

Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) *Nucleic Acids Res.*, **28**, 235–242.

Chattopadhyay,D., Langsley,G., Carson,M., Recacha,R., DeLucas,L, and Smith,C. (2000) *Acta Crystallogr.*, **D56**, 937–944.

Gunasekaran,K., Gomathi,L., Ramakrishnan,C. Chandrasekhar,J. and Balaram,P. (1998) *J. Mol. Biol*., **284**, 1505–1516.

Ramachandran,G.N., Ramakrishnan,C. and Sasisekharan,V. (1963) *J. Mol. Biol.*, **7**, 95–99.

Ramachandran,G.N. and Sasisekharan,V. (1968) *Adv. Protein Chem*., **23**, 283–437.

Ramasarma,T. and Ramakrishnan,C. (2002) *Indian J. Biochem. Biophys*., **39**, 5–15.

Sayle,R.A. and Milner-White,E.J. (1995) *Trends Biochem. Sci*., **20**, 374–376.

Venkatachalam,C.M. (1968) *Biopolymers*, **6**, 1425–1436.

Walker,J.E., Saraste,M., Runswick,M.J. and Gay,N.J. (1982) *EMBO J.*, **1**, 945–951.

Received August 31, 2001; revised June 11, 2002; accepted July 11, 2002