ELSEVIER

Biophysical Chemistry 120 (2006) 24 – 31

Biophysical Chemistry

http://www.elsevier.com/locate/biophyschem

Role of amino acid properties to determine backbone τ(N–Cα–C′) stretching angle in peptides and proteins

S.M. Malathy Sony, K. Saraboji, N. Sukumar [1], M.N. Ponnuswamy *

Department of Crystallography and Biophysics, University of Madras, Guindy campus, Chennai — 600025, India

Received 8 June 2005; received in revised form 27 July 2005; accepted 28 July 2005
Available online 2 November 2005

Abstract

The analysis of the basic geometry of amino acid residues of protein structures has demonstrated the invariability of all the bond lengths and bond angles except for τ, the backbone N–Cα–C′ angle. This angle can be widened or contracted significantly from the tetrahedral geometry to accommodate various other strains in the structure. In order to accurately determine the cause for this deviation, a survey is made for the τ angles using the peptide structures and the ultrahigh resolution protein structures. The average deviation of N–Cα–C′ angles from tetrahedral geometry for each amino acid in all the categories were calculated and then correlated with forty-eight physiochemical, energetic and conformational properties of amino acids. Linear and multiple regression analysis were carried out between the amino acid deviation and the 48 properties. This study confirms the deviation of τ angles in both the peptide and protein structures but similar forces do not influence them. The peptide structures are influenced by physical properties whereas as expected the conformational properties influence the protein structures. And it is not any single property that dominates the deviation but the combination of different factors contributes to the τ angle deviation.

Keywords: Bond angle deviation; Amino acid properties; Correlation analysis; Peptides; Ultrahigh resolution protein structures

1. Introduction

During the past few decades, a number of technical and methodological improvements have taken place in macromolecular crystallography. High-intensity synchrotron radiation sources, efficient two-dimensional detectors and cryogenic techniques are employed in collecting high atomic resolution protein structures of increasing size and complexity. The level of accuracy of these three-dimensional protein structures provides relevant information on biological function and catalysis of the particular enzymes under study. From these unbiased studies, evidences for general and fundamental molecular properties can be derived. Recent studies on the geometry of the polypeptide backbone and on its experimental electron density have demonstrated the potential of crystallography at ultrahigh resolution [1]. In fact, the deviations in backbone bond angle τ(N–Cα–C′) [2], the non-planarity of peptide bond [3], the pyramidalization of carbonyl carbon atom [4] and the lengthening of CO bonds when the CN bond shortens [1] were all identified from the ultrahigh resolution protein structures. These analyses were made keeping in mind the increased need for accurate, efficient, and reliable methods to model protein structures for the structural genomics efforts [5].

Studies on the peptide group were carried out owing to its properties, such as the dipole moment, the planar geometry, and the relatively high rotational barrier around C–N bond, which determine the conformation of polypeptides and proteins [4]. It is the Cα atom, the most important locus for evaluating distortion of covalent geometry in protein structures, that joins sidechain with backbone and respond to both and especially to their compatibility [6]. Database analysis carried out by Karplus [2] showed that the average values of the bond lengths, bond angles, and ω-torsion angles match the ideal values based on small-molecule crystal structures but for well-populated main-chain conformations, these parameters vary over ranges of about 0.015 Å, 4° and 7° (up to 8.8° for the τ angle), respectively. The highly variable interpeptide bond angle τ has relatively large effects on structure because the

* Corresponding author. Tel.: +91 44 2235 1367; fax: +91 44 2230 0122.
E-mail address: mnpsy2004@yahoo.com (M.N. Ponnuswamy).
[1] Presently working in NE-CAT, Building 436E, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439, USA.

0301-4622/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2005.07.012

peptide group magnifies the resulting atomic shifts of ~10° to a 0.31 Å shift in the relative positions of 1–3 related Cα-atoms [2]. This shift may affect any crucial non-covalent interaction that is likely to take place and therefore analysis of τ angle is necessitated.

In the high resolution protein structures this τ(N–Cα–C′) bond angle deviation from tetrahedral geometry was correlated with the change in the conformation of the backbone torsion angles ϕ (N–Cα) and ψ (Cα–C′). It was well established as early as in 1965 by Ramakrishnan and Ramachandran [7] that the allowed domain of the conformation map [(ϕ,ψ) map] of the dipeptide unit increases when τ(N–Cα–C′) angle is increased from 110° to 115°. But in the mid-1990s it was concluded that the widening or contraction of τ angle actually depends on the backbone conformation (ϕ,ψ) of the protein structure [8–11,2]. As one moves from one point in (ϕ,ψ) space to another, bond angle τ vary in a characteristic manner which was consistently found to be contracted in extended forms, intermediate in C_7^{eq}, and relatively large in the helical-bridge regions of peptide (ϕ,ψ) space, with variations far exceeding 10° [8,2]. If large variations in an angle of this kind are neglected in peptide modeling along a chain of hundreds of residues, this may have significant implications for the spatial presence of a protein, or the modeling of an active site [8].

Although deviations from standard geometry are seen in small molecule crystal structures (e.g., [12,13]), and some conformational dependence has been inferred [14], these data are limited and the general relevance of such deviations has remained uncertain. Contradiction was also observed with respect to the role of aminoacid residue type influencing the τ angle variation, but some workers [15] observe only weak dependence whereas Ramachandran et al. [16] and Momany et al. [17] do find differences. Most notable difference is observed for *Gly* and *Pro* having high values [18] and *Val* and *Ile* having low values [2]. Aminoacids are subjected to certain constrained conditions imposed by their physical and chemical properties, which might play an important role in protein structure and function [19]. With these points in mind, the following work was initiated.

The τ(N–Cα–C′) bond angle for the twenty naturally occurring aminoacid residue types were retrieved from small molecule peptide structures and its average deviation from tetrahedral geometry was calculated. This was then correlated with various physicochemical, energetic and conformational properties of the aminoacids. A careful study of τ(N–Cα–C′) angles on aminoacids and peptides will give an idea about the maximum range to which the angle can stretch in individual aminoacids. Such a study has been undertaken as they are relevant in modeling of polypeptides and protein chains and also in protein refinement.

2. Materials and methods

To enhance our understanding of peptide energetics, it is important to determine the covalent geometry changes as a function of aminoacid properties. These results can, in turn, stimulate theoretical studies to provide explanations for these observations [2]. Calculated geometries have the advantage of providing information on structure–property relations which is difficult to obtain experimentally. Various procedures for deriving and applying molecular geometries have been described in the literature [11]. These geometries are obtained from small-molecular studies, surveys of protein structures, relationships to other proteins, or from experimental observations made from X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy [5]. Herein, a statistical approach is adopted, in which the τ angle average deviation for each of the twenty aminoacid residues are related with various physicochemical, energetic and conformation properties of aminoacids through single and multiple correlation factors [20–23].

In order to access the deviation of backbone τ(N–Cα–C′) bond angle with respect to various aminoacid properties for small-molecular peptide structures, the dataset (referred henceforth as 'peptide dataset') were chosen as follows: Cambridge Structural Database (CSD: [24]), update 5.24 (November 2002), was searched for peptides, excluding data from cyclic, disordered or D-aminoacids containing peptides, as well as structures with the crystallographic error factor $R > 0.08$ and *Cys* containing peptides there should be no disulphide bridge. The τ(N–Cα–C′) bond angles were retrieved and its average deviation from tetrahedral geometry (d) was calculated using the formula:

$$d = \sum \frac{|\tau - 109.5|}{N} \tag{1}$$

where N is the total number of observables and the value 109.5 is the standard angle for the tetrahedral geometry, in radians.

The single correlation between the τ angle average deviation, d (dependent quantity) and aminoacid properties (independent quantities) was calculated using the familiar expression:

$$r = \frac{N\sum XY - \left(\sum X \sum Y\right)}{\left\{\left[N\sum X^2 - \left(\sum X\right)^2\right]\left[N\sum Y^2 - \left(\sum Y\right)^2\right]\right\}^{1/2}} \tag{2}$$

where r is the correlation coefficient, N, X, and Y are the number of data, average deviation and aminoacid properties, respectively. To evaluate the validity of the correlation coefficient, the so-called null hypothesis, i.e., the hypothesis in which the variables are not correlated, was tested by using Student's t distribution. The statistical test yields a p-value, which represents the probability that random sampling would result in a correlation coefficient. Under this hypothesis there is no correlation between the two variables; p-values <0.05 allow one to reject the null hypothesis at the 95% confidence level [1]. The combined effect of these aminoacid properties towards the τ angle deviation are analysed using multiple regression. Multiple correlation coefficients were determined by standard procedures [25,26].

Table 1
Ultrahigh resolution protein structures used in the study[a]

PDB ID	Resolution (Å)	PDB ID	Resolution (Å)	PDB ID	Resolution (Å)	PDB ID	Resolution (Å)
1EJG	0.54	1BYZ	0.90	1HJ9	0.95	1EB6	1.00
3AL1	0.75	1ET1	0.90	1KTH	0.95	1EXR	1.00
1GCI	0.78	1G66	0.90	4LZT	0.95	1G2Y	1.00
1IUA	0.80	1IX9	0.90	7A3H	0.95	1GA6	1.00
1FN8	0.81	1IXB	0.90	1AHO	0.96	1GKM	1.00
1FY4	0.81	2PVB	0.91	1K5C	0.96	1GVW	1.00
1FY5	0.81	1IQZ	0.92	1BYI	0.97	1GVX	1.00
1GDN	0.81	1L9L	0.92	1C75	0.97	1HJ8	1.00
1CBN	0.83	1RB9	0.92	1F94	0.97	1IR0	1.00
1M40	0.85	3LZT	0.92	1G4I	0.97	1JFB	1.00
1MC2	0.85	1B0Y	0.93	1M1Q	0.97	1K2A	1.00
3PYP	0.85	1EA7	0.93	8A3H	0.97	1K4P	1.00
1G6X	0.86	1GDQ	0.93	1GQV	0.98	1K6U	1.00
1MUW	0.86	1GVK	0.94	1GVT	0.98	1KCC	1.00
1DY5	0.87	1GVU	0.94	1IC6	0.98	1LKK	1.00
1GWE	0.88	1IEE	0.94	1IXH	0.98	1LNI	1.00
1AB1	0.89	1KWF	0.94	1JXU	0.99	1O7J	1.00
1JXT	0.89	1NLS	0.94	1MNZ	0.99	2ERL	1.00
1JXW	0.89	2FDN	0.94	1A6M	1.00	5PTI	1.00
1JXX	0.89	1BRF	0.95	1C7K	1.00	8RXN	1.00
1JXY	0.89	1BXO	0.95	1CEX	1.00		

[a] If more than one chain is present in a particular structure, only the first chain is used for the calculation and since the dataset is not too large, no redundancy criteria is imposed.

In order to cross check our results with the peptide dataset all the above mentioned analyses were also carried out for ultrahigh resolution (≤ 1 Å) protein structures (referred henceforth as 'protein dataset'), which were retrieved from Protein Data Bank (Table 1). The protein dataset were split up into 3 main categories based on resolution and the crystallographic discrepancy factor (R-factor). The first category contains protein structures whose resolution is less than 1 Å and the R-factor is less than 0.19 (named as "<1 and $R_f<0.19$"). The second and third categories do not have any R-factor limitations but the resolution is less than 1 Å (named as "≤ 0.99") and less than or equal to 1 Å (named as "≤ 1"), respectively. This discrimination

Table 2
Average τ angle deviation from tetrahedral geometry (d) for all data sets

Amino acid	Peptide data set		Protein data set													
			<1[a] and $R_f<0.19$		≤ 0.99[a]		≤ 1.00[a]		0–0.80[a]		0.81–0.90[a]		0.91–0.99[a]		0.99–1.00[a]	
			42[b]		60[b]		83[b]		4[b]		22[b]		34[b]		23[b]	
	N	d	N	d	N	d	N	d	N	d	N	d	N	d	N	d
ALA	328	2.125	806	1.491	1059	1.539	1623	1.583	58	1.893	414	1.538	587	1.505	564	1.665
ASP	38	0.235	499	1.348	631	1.416	966	1.418	10	1.749	209	1.529	412	1.351	335	1.423
CYS	3	4.450	174	1.410	280	1.358	369	1.417	9	0.290	105	1.210	166	1.509	89	1.601
GLU	22	0.625	389	1.561	426	1.604	694	1.658	14	0.731	184	1.603	228	1.659	268	1.743
PHE	112	0.764	279	0.770	384	0.913	566	0.793	5	1.936	156	1.257	223	0.650	182	0.539
GLY	295	2.890	733	3.712	1038	3.774	1555	3.842	45	3.832	373	3.693	620	3.818	517	3.98
HIS	21	0.215	172	1.431	201	1.529	305	1.498	8	0.699	95	1.612	98	1.516	104	1.438
ILE	27	1.541	388	0.213	502	0.048	795	0.067	15	0.193	166	0.222	321	0.049	293	0.100
LYS	13	0.784	350	1.227	480	1.261	715	1.271	16	1.630	190	1.278	274	1.227	235	1.293
LEU	293	1.182	594	1.163	714	1.166	1168	1.227	35	1.121	298	1.250	381	1.105	454	1.323
MET	15	0.076	123	1.275	132	1.128	221	1.296	4	0.174	61	1.151	67	1.164	89	1.545
ASN	3	0.622	413	1.705	587	1.692	835	1.591	30	1.703	198	1.809	359	1.626	248	1.353
PRO	154	2.094	397	3.039	458	3.114	710	3.057	24	3.534	185	3.262	249	2.964	252	2.953
GLN	4	0.618	300	1.489	362	1.389	571	1.456	15	1.609	125	1.410	222	1.362	209	1.572
ARG	9	2.015	374	1.344	394	1.320	620	1.370	11	3.017	183	1.163	200	1.370	226	1.458
SER	15	0.092	563	1.685	918	1.635	1348	1.752	38	1.917	328	1.516	522	1.686	430	2.001
THR	7	2.186	504	1.222	723	1.295	1161	1.333	26	1.723	246	1.511	451	1.152	438	1.395
VAL	185	0.899	558	0.038	671	0.140	1081	0.077	31	0.204	233	0.387	407	0.025	410	0.025
TRP	20	0.528	151	0.963	193	1.024	234	1.010	6	0.455	61	1.052	126	1.037	41	0.946
TYR	67	0.769	281	0.962	392	0.991	592	0.966	10	0.553	10	1.146	249	0.926	200	0.918

N represents the total number of each amino acid entries from all the structures.
[a] Resolution (Å).
[b] Number of structures.

Table 3
Single correlation coefficient (r) between the average τ angle deviation (d) and the 48 physicochemical, energetic and conformational properties of the 20 amino acids

AA Prop.	Peptide	Protein							AA Prop.	Peptide	Protein						
		<1 and R_f <0.19	≤0.99	≤1.00	0–0.80	0.81–0.90	0.91–0.99	0.99–1.00			<1 and R_f <0.19	≤0.99	≤1.00	0–0.80	0.81–0.90	0.91–0.99	0.99–1.00
K^0	**0.366**	**0.529**	**0.525**	**0.523**	**0.644**	**0.506**	**0.516**	**0.515**	α_χ	−0.087	0.257	0.281	0.265	0.191	0.324	0.250	0.226
H_t	−0.036	−0.390	−0.377	−0.401	−0.348	−0.328	−0.397	−0.442	α_μ	−0.140	−0.314	−0.327	−0.307	−0.344	−0.390	−0.287	−0.264
Hp	0.190	**−0.597**	**−0.607**	**−0.603**	**−0.595**	**−0.596**	**−0.593**	**−0.579**	V^0	**−0.382**	**−0.564**	**−0.557**	**−0.571**	−0.370	**−0.538**	**−0.568**	**−0.601**
P	−0.254	−0.010	0.010	0.010	0.066	−0.030	0.032	0.007	N_m	−0.269	−0.357	−0.378	−0.348	−0.433	−0.425	−0.347	−0.306
pH_i	0.077	−0.003	−0.005	−0.003	0.242	−0.056	0.009	0.002	N_l	0.393	−0.388	−0.394	−0.398	−0.420	−0.379	−0.383	−0.387
pK'	−0.295	0.214	0.240	0.254	0.102	0.239	0.226	0.269	H_{gm}	0.262	−0.391	−0.405	−0.402	−0.477	−0.418	−0.379	−0.384
Mw	**−0.438**	−0.472	−0.462	−0.477	−0.395	−0.460	−0.461	**−0.510**	ASA_D	**−0.398**	**−0.561**	**−0.566**	**−0.569**	−0.423	**−0.570**	**−0.562**	**−0.580**
B_l	−0.228	**−0.720**	**−0.708**	**−0.729**	**−0.545**	**−0.643**	**−0.740**	**−0.759**	ASA_N	**−0.328**	0.111	0.112	0.110	0.282	0.082	0.113	0.097
R_I	−0.208	−0.365	−0.357	−0.373	−0.263	−0.272	−0.406	−0.405	ΔASA	−0.194	**−0.661**	**−0.668**	**−0.670**	**−0.630**	**−0.652**	**−0.664**	**−0.671**
μ	0.019	−0.499	−0.494	**−0.502**	**−0.526**	**−0.519**	−0.471	**−0.512**	ΔG_h	0.294	0.048	0.048	0.047	0.001	0.109	0.018	0.049
H_{nc}	0.010	−0.224	−0.211	−0.218	−0.337	−0.137	−0.238	−0.225	G_{hD}	0.144	0.035	0.072	0.076	−0.198	0.112	0.065	0.085
E_{sm}	**0.372**	0.302	0.290	0.314	0.005	0.290	0.299	0.360	G_{hN}	0.234	0.006	0.013	0.008	−0.190	0.072	−0.012	0.004
E_l	0.155	−0.457	−0.454	−0.457	−0.454	−0.433	−0.451	−0.452	ΔH_h	**0.383**	0.327	0.325	0.329	0.202	0.362	0.306	0.340
E_t	**0.353**	−0.153	−0.158	−0.145	−0.341	−0.143	−0.149	−0.111	$-T\Delta S_h$	**−0.324**	**−0.605**	**−0.600**	**−0.611**	−0.421	**−0.580**	**−0.608**	**−0.631**
P_α	**−0.397**	−0.463	−0.468	−0.437	−0.435	−0.483	−0.454	−0.389	ΔC_{ph}	−0.143	**−0.537**	**−0.526**	**−0.539**	−0.398	−0.463	**−0.557**	**−0.562**
P_β	0.050	**−0.679**	**−0.680**	**−0.689**	**−0.531**	**−0.639**	**−0.692**	**−0.690**	ΔG_c	−0.281	−0.121	−0.109	−0.117	−0.119	−0.152	−0.087	−0.136
P_t	0.152	**0.712**	**0.712**	**0.695**	**0.569**	**0.692**	**0.714**	**0.659**	ΔH_c	−0.215	−0.399	−0.381	−0.392	−0.454	−0.377	−0.376	−0.412
P_c	**0.301**	**0.768**	**0.772**	**0.752**	**0.663**	**0.761**	**0.768**	**0.713**	$-T\Delta S_c$	0.080	0.474	0.456	0.466	**0.554**	0.416	0.468	0.480
C_a	**−0.369**	**−0.566**	**−0.562**	**−0.568**	−0.382	**−0.568**	**−0.558**	**−0.584**	ΔG	−0.002	−0.209	−0.177	−0.202	−0.330	−0.135	−0.195	−0.249
F	0.077	**0.651**	**0.643**	**0.648**	**0.648**	**0.604**	**0.649**	**0.651**	ΔH	0.073	−0.237	−0.213	−0.225	−0.432	−0.172	−0.225	−0.243
B_r	**0.309**	−0.306	−0.309	−0.306	−0.419	−0.286	−0.304	−0.288	$-T\Delta S$	−0.100	0.238	0.218	0.225	0.454	0.179	0.228	0.230
R_a	−0.146	**−0.597**	**−0.592**	**−0.597**	**−0.518**	**−0.544**	**−0.603**	**−0.597**	*v*	**−0.469**	−0.446	−0.425	−0.448	−0.318	−0.408	−0.435	−0.499
N_s	**0.376**	−0.418	−0.425	−0.420	−0.488	−0.418	−0.409	−0.392	*s*	−0.265	−0.252	−0.243	−0.255	0.004	−0.255	−0.242	−0.283
α_n	**−0.337**	**−0.533**	**−0.527**	**−0.506**	−0.474	**−0.526**	**−0.522**	−0.469	*f*	**−0.384**	−0.363	−0.387	−0.371	−0.184	−0.437	−0.360	−0.349

Abbreviations: K^0 = compressibility; H_t = thermodynamic transfer hydrophobicity; H_p = surrounding hydrophobicity; *P* = polarity; pH_i = isoelectric point; pK' = equilibrium constant with reference to the ionization property of COOH group; Mw = molecular weight; B_l = bulkiness; R_I = chromatographic index; μ = refractive index; H_{nc} = normalized consensus hydrophobicity; E_{sm} = short and medium range non-bonded energy; E_l = long range non-bonded energy; E_t = total non-bonded energy ($E_{sm}+E_l$); P_α, P_β, P_t and P_c = respectively, α-helical, β-structure, turn and coil tendencies; C_a = helical contact area; *F* = mean rms fluctuational displacement; B_r = buriedness; R_a = solvent accessible reduction ratio; N_s = average number of surrounding residues; α_n, α_c and α_m = respectively, power to be at the N-terminal, C-terminal and middle of α-helix.; V^0 = partial-specific volume; N_m and N_l = respectively, average medium and long-range contacts; H_{gm} = combined surrounding hydrophobicity (globular and membrane); ASA_D, ASA_N and ΔASA = respectively, solvent accessible surface area for denatured, native and unfolding; ΔG_h, G_{hD} and G_{hN} = Respectively, Gibbs free energy change of hydration for unfolding, denatured and native protein; ΔH_h = unfolding enthalpy change of hydration; $-T\Delta S_h$ = unfolding entropy change of hydration; ΔC_{ph} = Unfolding hydration heat capacity change; ΔG_c, ΔH_c and $-T\Delta S_c$ = respectively, unfolding Gibbs free energy, unfolding enthalpy and unfolding entropy changes of chain; ΔG, ΔH, and $-T\Delta S$ = respectively, unfolding Gibbs free energy change, unfolding enthalpy change and unfolding entropy change; *v* = volume (number of non-hydrogen side chain atoms); *s* = shape (position of branch point in a side-chain); f = flexibility (number of side-chain dihedral angles). K^0 in m^3/mol/Pa ($\times 10^{-15}$); H_t, H_p, H_{nc}, H_{gm}, ΔG_h, G_{hD}, G_{hN}, ΔH_h, $-T\Delta S_h$, ΔG_c, ΔH_c, $-T\Delta S_c$, ΔG, ΔH and $-T\Delta S$ in kcal/mol; *P* in Debye; pH_i and pk' in pH units; E_{sm}, E_l and E_t in kcal/mol/atom; B_1, C_a, ASA_D, ASA_N and ΔASA in $Å^2$; *F* in Å; V^0 in m^3/mol ($\times 10^{-6}$); ΔC_{ph} in cal/mol/K and the rest are dimensionless quantities. Brief description of these properties can be found in articles, Oobatake and Ooi, 1993 [21]; Gromiha and Ponnuswamy, 1993 [22]; and Gromiha et al., 1999 [23].
For peptide dataset $r \geq 0.3$ and for protein dataset $r \geq 0.5$ are highlighted in bold font.

is intended because the number of protein structures, which has 1 Å resolution contributes 28% to the whole data. To verify whether the resolution of the protein structure plays any role in the bond angle deviation, the dataset were further divided into four groups containing structures in the resolution range 0–0.8, 0.8–0.9, 0.9–0.99 and 0.99–1 Å. On the whole, the protein dataset has 7 subgroups whose bond angle τ was calculated using the *bond angle* program of WHATIF server (http://swift.cmbi.kun.nl/ WIWWWI/) and its average deviation (d) was correlated with the 48 aminoacid properties. In addition the three main chain torsion angles (ϕ,ψ,ω) were also calculated using the program *Torsions* for all the structures. The conformational (ϕ,ψ) space was subdivided into regions determined by a 90° grid and the parameter sets ϕ and ψ from all residues were then collected in accordance with the 4 regions to which they belong, namely (ϕ,ψ), ($\phi,-\psi$), ($-\phi,\psi$) and ($-\phi,-\psi$). The average values of ϕ, ψ and ω in each of the four regions for all the 20 aminoacid residues were calculated.

3. Results and discussion

The average τ angle deviation from tetrahedral geometry (d) was calculated for both the peptide and protein datasets for each aminoacid and are reported in Table 2. In the peptide dataset the number of data for *Cys*, *Asn* and *Gln* residues are very low when compared with the residues *Ala*, *Phe*, *Gly*, *Leu*, *Pro* and *Val*. The residues *Cys*, *Asn* and *Gln* are classified under polar groups whereas the larger data category residues are hydrophobic except *Gly*. Maximum deviation of 4.5° is found for *Cys* but the number of observation is only three and is not statistically significant. As expected *Gly* and *Pro* show large deviations but *Ala*, *Arg* and *Thr* residues also show similar trend. Surprisingly *Met* and *Ser* residues show least deviation and aromatic residues *Phe*, *Trp* and *Tyr* and the negatively charged residues, *Asp* and *Glu*, show consistent lower values. In the protein dataset *Cys*, *His*, *Met* and *Trp* has fewest data, which typically represents the proportions of soluble proteins [2]. The *Gly* and *Pro* residues show largest deviation whereas the Cβ branched residues *Ile* and *Val* show least deviation. As observed in the peptide dataset, the aromatic residues show lower values and except the aliphatic residues (*Ala*, *Gly*, *Ile*, *Leu* and *Val*), the charged (both positive and negative) and polar residues show consistent deviations.

3.1. Individual contribution of aminoacid properties

The linear single correlation analysis keeping the τ angle deviation (d) as the dependent quantity and the 48 aminoacid properties as independent quantities were performed (Table 3). The p-value of the Student's t test stood in the range 10^{-3}–10^{-9}, which indicates the reliability of r-value at 95% confidence level. In the peptide dataset the maximum correlation was observed for volume (v) property, i.e., the number of non-hydrogen sidechain atoms of the aminoacid residues has better relationship with d. In addition to that other 17 aminoacid properties, K^0, Mw, E_{sm}, E_t, P_α, P_c, C_a, B_r, N_s, α_n, V^0, N_l, ASA_D, ASA_N, ΔH_h, $T\Delta S_h$, and f were showing significant correlation ($r \geq 0.3$). The correlation of quantities K^0, Mw, E_{sm}, E_t, B_r, N_s, α_n, V^0, ASA_D, ASA_N, ΔH_h, $-T\Delta S_h$, v and f indicate the importance of physical properties of the aminoacids that also include short-range interaction. All the protein datasets show similar tendency and the conformational properties dominate the relationship with d as expected and show relatively high correlation coefficient [8–11,2]. This again confirms our statistical approach to be significant. The maximum correlation is observed for the coil propensity (P_c) value in all the sub-groups. In addition to this, good relationship was observed for the properties that are influenced by the backbone conformational angles, i.e., B_1, P_t, P_c, H_p, P_β, F, ΔASA, $-T\Delta S_h$, K^0, C_a, R_a, α_n, V^0, ASA_D, ΔG_{ph}. Here P_t, P_c, P_β, C_a, R_a and α_n properties represent the secondary interactions and Hp, F, ΔASA, $-T\Delta S_h$, ASA_D and ΔG_{ph} depends on tertiary interactions. Surprisingly the helical tendency (P_α) did not seem to play major role in the protein dataset in this analysis, even though the r-value is moderate but when compared with all other conformational properties it is fairly low. From this study it was observed that the physical, compressibility (K^0), and the conformational property, coil tendency (P_c) has significant influence in the peptide and protein datasets.

3.2. Combined effect of aminoacid properties

Multiple regression analyses were carried out by varying the number of independent quantities in order to identify the combined influence of properties towards τ angle deviation (Table 4). For this only the selected 18 properties for peptide

Table 4
Multiple regression analysis of average τ angle deviation (d) with the selected physicochemical, energetic and conformational properties of the 20 amino acids

Regression	Peptide		Protein													
	AA prop.	r	<1 and R_f <0.19		≤0.99		≤1.00		0–0.80		0.81–0.90		0.91–0.99		0.99–1.00	
			AA prop.	r	AA prop.	r	AA prop.	r	AA prop.	r	AA prop.	r	AA prop.	r	AA prop.	r
Single	v	0.469	P_c	0.768	P_c	0.772	P_c	0.752	P_c	0.663	P_c	0.761	P_c	0.713	B_1	0.759
2-term	$N_l+(-T\Delta S_h)$	0.675	B_1+P_c	0.827	P_c+P_β	0.829	B_1+P_c	0.822	K^0+H_p	0.813	$P_c+\alpha_n$	0.810	B_1+P_c	0.837	K^0+B_1	0.831
3-term	$+ASA_N$	0.777	$+K^0$	0.865	$+K^0$	0.863	$+K^0$	0.859	$+P_c$	0.833	$+K^0$	0.832	$+K^0$	0.87	$+P_c$	0.858
4-term	$+K^0$	0.810	$+P_\beta$	0.879	$+\alpha_n$	0.877	$+\alpha_n$	0.878	$+R_a$	0.843	$+P_\beta$	0.849	$+\alpha_n$	0.887	$+P_\beta$	0.873
5-term	$+V^0$	0.848	$+\Delta G_{ph}$	0.900	$+B_1$	0.887	$+P_\beta$	0.887	$+P_t$	0.851	$+R_a$	0.861	+ΔASA	0.899	$+\Delta G_{ph}$	0.892
6-term	$+\Delta H_h$	0.859	$+(-T\Delta S_h)$	0.907	$+\Delta G_{ph}$	0.906	$+\Delta G_{ph}$	0.906	$+B_1$	0.860	$+B_1$	0.867	$+T\Delta S_h$	0.921	$+(-T\Delta S_h)$	0.900

Correlation coefficient (r) for the respective amino acid properties are given for different combinations. The plus sign (+) here represents that, in addition to the previous row properties of that particular column this row property is added to give the multiple correlation coefficient.

Table 5
Cross correlation coefficient for the peptide and the protein datasets

	<1 and R_f <0.19	≤0.99	≤1	0–0.8	0.81–0.9	0.91–0.99	0.99–1
Peptide	0.331	0.323	0.327	0.277	0.286	0.346	0.349
<1 and R_f <0.19		0.995	0.997	0.765	0.974	0.996	0.982
≤0.99			0.996	0.780	0.988	0.994	0.972
≤1				0.765	0.975	0.997	0.988
0–0.8					0.786	0.744	0.723
0.81–0.9						0.966	0.934
0.91–0.99							0.984
0.99–1							

dataset and the 15 properties of the protein dataset was used. Even though the single property correlation is comparatively low for peptide dataset, the combined aminoacid properties show drastic rise in the correlation. In addition to the influence of the physical properties [K^0 (compressibility) and V^0 (partial-specific volume)] the tertiary interaction, such as N_1 (average long range contacts) and $-T\Delta S_h$ (unfolding entropy change of hydration), also play a role in varying the τ angle, probably due to the various crystal contacts. The multiple regression analysis of the protein datasets reaffirms that in addition to the conformational properties [P_c and P_β (β-sheet and coil tendencies)] the physical properties, especially B_1 (bulkiness) and K^0 (compressibility) of the aminoacid influence the τ angle deviation. All the classifications in the protein dataset show similar tendency, as observed for single correlation coefficient. In this multiple regression analysis, it is the compressibility (K^0) that stands out for both the peptide and the protein datasets.

Cross correlation analysis between the protein datasets showed similar tendency, irrespective of the resolution cut-off in single and multiple correlation analysis. Table 5 presents the correlation coefficient (r) for the datasets from which it can be inferred that all the protein datasets are correlated nicely with $r \geq 0.9$ except for the group (0–0.8) which can be ignored owing to paucity of data (i.e. it has only 4 structures). The "<1 and R_f<0.19" category can be chosen as a representative dataset for proteins since it contains structures, which were solved at ultrahigh resolution and are well refined. Comparatively the peptide dataset show poor correlation with all the protein dataset, the correlation being around 0.3.

3.3. Role of backbone torsion angles

In order to obtain a good picture to see the factors which influence 'd', instead of using the theoretical aminoacid property values which were derived from unrelated data, the

Table 6
Average torsion angles for the "<1 and R_f <0.19" protein dataset which is correlated with the τ angle deviation (d) for the four regions in the (ϕ,ψ) conformational space

	No. of entries	d	$(-\phi,\psi)$			$(-\phi,-\psi)$			$(\phi,-\psi)$[a]			(ϕ,ψ)[a]		
			ϕ	ψ	ω[b]	ϕ	ψ	ω[b]	ϕ	ψ	ω[b]	ϕ	ψ	ω[b]
ALA	806	1.491	−100.302	130.744	6.233	−66.140	−36.230	3.626				51.564	44.689	4.348
ASP	499	1.348	−103.376	94.105	6.129	−75.792	−47.778	4.008	57.173	−75.502	7.134	56.614	46.845	3.631
CYS	174	1.410	−110.979	130.965	5.489	−75.764	−37.687	4.262						
GLU	389	1.561	−104.192	124.974	6.168	−69.662	−33.489	3.191						
PHE	279	0.770	−112.529	125.326	6.019	−70.049	−40.947	4.531						
GLY	733	3.712	−103.673	134.196	5.339	−81.607	−74.877	4.021	92.412	−83.107	4.578	91.838	66.098	3.868
HIS	172	1.431	−108.748	116.076	5.688	−72.293	−38.251	3.627				57.945	42.800	3.464
ILE	388	0.213	−108.655	125.155	5.200	−69.564	−40.303	4.127						
LYS	350	1.227	−104.695	122.812	4.788	−70.745	−36.564	3.621				57.396	35.074	4.258
LEU	594	1.163	−98.139	121.930	5.488	−69.756	−36.786	4.075				58.141	51.634	5.765
MET	123	1.275	−110.808	126.785	5.435	−68.583	−37.446	2.918						
ASN	413	1.705	−104.199	99.602	5.232	−79.172	−41.757	4.044				59.037	39.237	3.681
PRO	397	3.039	−66.787	138.492	4.954	−64.576	−25.709	3.260						
GLN	300	1.489	−98.286	118.890	4.715	−70.401	−33.069	3.924				38.393	76.264	4.361
ARG	374	1.344	−103.164	119.818	6.506	−69.159	−40.107	3.804				61.915	35.232	6.425
SER	563	1.685	−105.600	133.460	6.727	−72.269	−30.659	5.083	50.738	−101.247	18.638	53.289	55.782	5.816
THR	504	1.222	−108.836	125.361	5.081	−79.251	−39.291	4.083				53.528	54.722	9.251
VAL	558	0.038	−112.180	127.611	5.844	−71.186	−40.124	3.388						
TRP	151	0.963	−105.788	121.257	5.453	−72.257	−39.790	3.499						
TYR	281	0.962	−109.653	126.918	5.906	−76.833	−36.008	4.049				58.778	31.745	4.251
r*			0.584	0.230	−0.159	−0.205	−0.377	−0.038						

[a] The missing values implies that these entries are not observed for that particular amino acid.
[b] The average ω values reported here is derived by calculating the deviation of the ω torsion angle from 180°.
* r = single correlation coefficient.

Table 7
Multiple regression analyses of the backbone torsion angles and the selected 15 aminoacid properties with the τ angle deviation

Correlation	Only torsion angle		4 torsion angle+15 AA prop	
	Properties	r	Properties	r
single	ϕ of $(-\phi,\psi)$	0.584		
2-term	ϕ of $(-\phi,\psi)$+ψ of $(-\phi,-\psi)$	0.817	B_1+ϕ of $(-\phi,\psi)$	0.910
3-term	+ϕ of $(-\phi,-\psi)$	0.840	.+ψ of $(-\phi,-\psi)$	0.940
4-term	+ψ of $(-\phi,\psi)$	0.866	+ΔASA	0.954
5-term	+ω of $(-\phi,\psi)$	0.876	+ψ of $(-\phi,\psi)$	0.962
6-term	+ω of $(-\phi,-\psi)$	0.878	+ϕ of $(-\phi,-\psi)$	0.975

The plus sign (+) in the beginning of the properties column represents that, in addition to the previous row properties of that particular column, this row property is added to give the multiple correlation coefficient.

analysis was extended by using the real experimental values that are related to the dataset. As the backbone torsion angle is a suitable candidate, multiple regression analysis between τ angle deviation (d) and the three backbone torsion angles (ϕ,ψ,ω) for the protein dataset "<1 and R_f<0.19" were done. This analysis was not carried out for the peptide dataset as they are not influenced by any of the conformational properties, even though single correlation showed influence of P_c but not in multiple regression analysis. We are aware of the fact that the torsion angles calculated were subdivided into four regions with 90° grid [(ϕ,ψ), $(\phi,-\psi)$, $(-\phi,\psi)$ and $(-\phi,-\psi)$] in the conformation space (ϕ,ψ). For the ω torsion angle only the deviation from planarity, that is $180-\omega$, are calculated. The average values of ϕ, ψ and ω in each of the four regions for each of the 20 aminoacid residues are reported in Table 6. *Asp*, *Gly* and *Ser* are found in all the four regions of the (ϕ,ψ) space. It has been reported by Karplus [2] that *Asp* and *Ser* adopt a conformation which forms intradipeptide hydrogen bonded rings with adjacent peptide that stabilize the strained conformations that is found in $(\phi,-\psi)$ region. This explains the reason for the occurrence of *Asp* and *Ser* residues in the conformationally constrained area.

The peptide bond is non-planar as can been seen from the average ω torsion angle [3], but this deviation also shows conformation dependence. The extended region $(-\phi,\psi)$ shows relatively much deviation (around 5.5°) than the helical region $(-\phi,-\psi)$ (around 4°) presumably due to the cooperative effects of the secondary structures *helix compression* and *β-expansion* [10]. The single correlation analysis shows only moderate correlation for the ϕ torsion angle of the extended region $(-\phi,\psi)$ all other torsion angles relate very poorly. Combinations with multiple regression analysis were tried whose correlation coefficient with the combined properties is presented in Table 7. Data in the regions $(-\phi,\psi)$ and $(-\phi,-\psi)$, representing the extended and helical conformation were only considered for the regression analysis, and so the total number of properties is six [$(-\phi,\psi,\omega)$ and $(-\phi,-\psi,\omega)$]. In addition to the torsion angle combinations, the 15 selected aminoacid properties were also combined and reported (Table 7). Here the ω torsion angle was not considered since its contribution is not much to the correlation.

The combined effect raises the correlation coefficient proportionally with the number of independent parameters in multiple regression and the highest correlation coefficient value (r=0.878) is obtained when all the six conformational angles were considered. When the 15 aminoacid properties were combined with the torsion angles, an excellent correlation of 0.975 is obtained. The physical property, bulkiness (B_1) with solvent accessible surface area for unfolding (ΔASA) contribute much to the τ angle deviation in addition to the backbone torsion angles.

4. Conclusion

From this study we conclude that the peptides and the proteins show deviations in their backbone τ bond angle but are not influenced by similar forces. The reason for the difference between them may be that, compared with small systems, large subunits of polymers (proteins) have the advantage of displaying some of the long-range interactions, which are among the effective factors that determine polymer properties [8]. And also small molecule structures are influenced by crystal-field effects that are only marginal in macromolecular crystal structures which are dominated by intramolecular interactions [1,4]. The τ bond angle deviation in peptides is influenced by the physical properties of the aminoacids involved whereas in proteins it is the backbone conformations that influence it. This study also reports the influence of two more properties bulkiness and solvent accessible surface area for unfolding that contribute to the τ bond angle deviation in the proteins. And we also confirm that not any single properties of the aminoacids contribute to the deviation but the combined effect of various related properties that influence it.

Acknowledgements

SMMS and KS acknowledge the financial assistance from Council of Scientific and Industrial Research, Government of India and the authors also thank Dr. M. Michael Gromiha, CBRC, AIST, Japan, for the suggestions.

References

[1] L. Esposito, L. Vitagliano, A. Zagari, L. Mazzarella, Experimental evidence for the correlation of bond distances in peptide groups detected in ultrahigh-resolution protein structures, Protein Eng. 13 (2000) 825–828.
[2] P.A. Karplus, Experimentally observed conformation-dependent geometry and hidden strain in proteins, Protein Sci. 5 (1996) 1406–1420.
[3] M.W. MacArthur, J.M. Thornton, Deviations from planarity of the peptide bond in peptides and proteins, J. Mol. Biol. 264 (1996) 1180–1195.
[4] L. Esposito, L. Vitagliano, A. Zagari, L. Mazzarella, Pyramidalization of backbone carbonyl carbon atoms in proteins, Protein Sci. 9 (2000) 2038–2042.
[5] M.A. Depristo, P.I.W. De Bakker, R.P. Shetty, T.L. Blundell, Discrete restraint-based protein modeling and the C{alpha}-trace problem, Protein Sci. 12 (2003) 2032–2046.
[6] S.C. Lovell, I.W. Davis, W.B. Arendall III, P.I.W. De Bakker, J.M. Word, M.G. Prisant, J.S. Richardson, D.C. Richardson, Structure validation by C^{α} geometry: ϕ, ψ and Cβ deviation, Proteins: Structure, Function and Genetics 50 (2003) 437–450.

[7] C. Ramakrishnan, G.N. Ramachandran, Stereochemical criteria for polypeptide and protein chain conformations: II. Allowed conformations for a pair of peptide units, Biophys. J. 5 (1965) 909–933.
[8] L. Schafer, S.Q. Newton, F.A. Momany, V.J. Klimkowski, On the use of conformationally dependent geometry trends from ab initio dipeptide and oligopeptide studies in empirical peptide modeling, J. Mol. Struct., Theochem 232 (1991) 275–289.
[9] L. Schafer, M. Cao, Predictions of protein backbone bond distances and angles from first principles, J. Mol. Struct., Theochem 333 (1995) 201–208.
[10] X. Jiang, M. Cao, B. Teppen, S.Q. Newton, L.J. Schafer, Predictions of protein backbone structural parameters from first principles: systematic comparisons of calculated N–C(α)–C′ angles with high-resolution protein crystallographic results, J. Phys. Chem. 99 (1995) 10521–10525.
[11] X. Jiang, C.-H. Yu, M. Cao, S.Q. Newton, E.F. Paulus, L. Schafer, ϕ/ψ-Torsional dependence of peptide backbone bond-lengths and bond-angles: comparison of crystallographic and calculated parameters, J. Mol. Struct. 403 (1997) 83–93.
[12] V. Sasisekharan, in: N. Ramanathan (Ed.), Stereochemical Criteria for Polypeptide and Protein Structures, Wiley and Sons, India, 1962, pp. 39–78.
[13] F. Winkler, J. Dunitz, The non-planar amide group, J. Mol. Biol. 5 (1971) 169–182.
[14] P. Chakrabarti, J.D. Dunitz, Structural characteristics of the carboxylic amide group, Helv. Chim. Acta 65 (1982) 1555–1562.
[15] K. Siam, S.Q. Kulp, J.D. Ewbank, L. Schafer, C. Van Alsenoy, Ab initio studies of structural features not easily amenable to experiment: Part 64. Conformational analysis and local geometry maps of the model dipeptide N-acetyl N′-methyl serine amide, J. Mol. Struct. 184 (1989) 143–157.
[16] G.N. Ramachandran, A.S. Kolaskar, C. Ramakrishnan, V. Sasisekharan, The mean geometry of the peptide unit from crystal structure data, Biochim. Biophys. Acta 359 (1974) 298–302.
[17] F.A. Momany, R.F. McGuire, A.W. Burgess, H.A. Scheraga, Energy parameters in polypeptides: VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids, J. Phys. Chem. 79 (1975) 2361–2381.
[18] R.A. Engh, R. Huber, Accurate bond and angle parameters for X-ray protein structure refinement, Acta Crystallogr., A 47 (1991) 392–400.
[19] Q. Du, D. Wei, K.-C. Chou, Correlations of amino acids in proteins, Peptides 24 (2003) 1863–1869.
[20] K. Tomii, M. Kanehisa, Analysis of amino acid indices and mutation matrices for sequence comparison and structure prediction of proteins, Protein Eng. 9 (1996) 27–36.
[21] M. Oobatake, T. Ooi, Hydration and heat stability effects on protein unfolding, Prog. Biophys. Mol. Biol. 59 (1993) 237–284.
[22] M.M. Gromiha, P.K. Ponnuswamy, Relationship between amino acid properties and protein compressibility, J. Theor. Biol. 165 (1993) 87–100.
[23] M.M. Gromiha, M. Oobatake, A. Sarai, Important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins, Biophys. Chem. 82 (1999) 51–67.
[24] F.H. Allen, O. Kennard, 3D search and research using the Cambridge structural database, Chem. Des. Autom. News 8 (1993) 31–37.
[25] D.H. Sanders, R.J. Eng, A.F. Murph, Statistics — A Fresh Approach, Mcgraw-Hill, New York, 1985.
[26] P.S. Grewal, Numerical Methods Of Statistical Analysis, Sterling Publishers, New Delhi, 1997.