

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/255735265>

Sparsely populated residue conformations in protein structures: Revisiting “experimental” Ramachandran maps

Article in *Proteins Structure Function and Bioinformatics* · July 2014

DOI: 10.1002/prot.24384 · Source: PubMed

CITATIONS

19

READS

99

3 authors, including:



Neha V. Kalmankar
University of Oxford

14 PUBLICATIONS 63 CITATIONS

[SEE PROFILE](#)



Padmanabhan Balaram
Indian Institute of Science

595 PUBLICATIONS 19,813 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Harvard _ post doc at RB Woodward [View project](#)



Design, construction and applications of peptide-based nano-materials [View project](#)

REVIEW

Sparsely populated residue conformations in protein structures: Revisiting “experimental” Ramachandran maps

Neha V. Kalmankar,^{1,2} C. Ramakrishnan,¹ and P. Balaram^{1*}

¹ Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India

² National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore 560065, India

ABSTRACT

The Ramachandran map clearly delineates the regions of accessible conformational (ϕ - ψ) space for amino acid residues in proteins. Experimental distributions of ϕ , ψ values in high-resolution protein structures, reveal sparsely populated zones within fully allowed regions and distinct clusters in apparently disallowed regions. Conformational space has been divided into 14 distinct bins. Residues adopting these relatively rare conformations are presented and amino acid propensities for these regions are estimated. Inspection of specific examples in a completely “arid”, fully allowed region in the top left quadrant establishes that side-chain and backbone interactions may provide the energetic compensation necessary for populating this region of ϕ - ψ space. Asn, Asp, and His residues showed the highest propensities in this region. The two distinct clusters in the bottom right quadrant which are formally disallowed on strict steric considerations correspond to the gamma turn (C7 axial) conformation (Bin 12) and the $i + 1$ position of Type II' β turns (Bin 13). Of the 516 non-Gly residues in Bin 13, 384 occupied the $i + 1$ position of Type II' β turns. Further examination of these turn segments revealed a high propensity to occur at the N-terminus of helices and as a tight turn in β hairpins. The β strand-helix motif with the Type II' β turn as a connecting element was also found in as many as 57 examples.

Proteins 2014; 82:1101–1112.
© 2013 Wiley Periodicals, Inc.

Key words: Ramachandran map; protein conformation; secondary structural motifs; amino acid conformational propensities; Type II' β turn.

INTRODUCTION

The Ramachandran map, which appeared in the literature 50 years ago, delineates the regions of sterically accessible conformational (ϕ - ψ) space for amino acid residues in proteins.^{1–3} The influence of the Ramachandran map in providing insights into the conformations of polypeptide chains and proteins has remained undiminished even after half a century.^{4,5} Although much of the attention has been focused on amino acid propensities to occur in thickly populated regions of sterically

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: DST project at the National Centre for Biological Sciences (to NVK) and DBT-IISc partnership program grant of the Department of Biotechnology.

Dedicated to the memory of Prof. G.N. Ramachandran on the 50th anniversary of the appearance of the seminal paper on the Ramachandran map.

The entire work was performed at Indian Institute of Science, Bangalore, India.

*Correspondence to: P. Balaram, Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India. E-mail: pb@mbu.iisc.ernet.in

Received 7 May 2013; Revised 20 July 2013; Accepted 25 July 2013

Published online 12 August 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24384

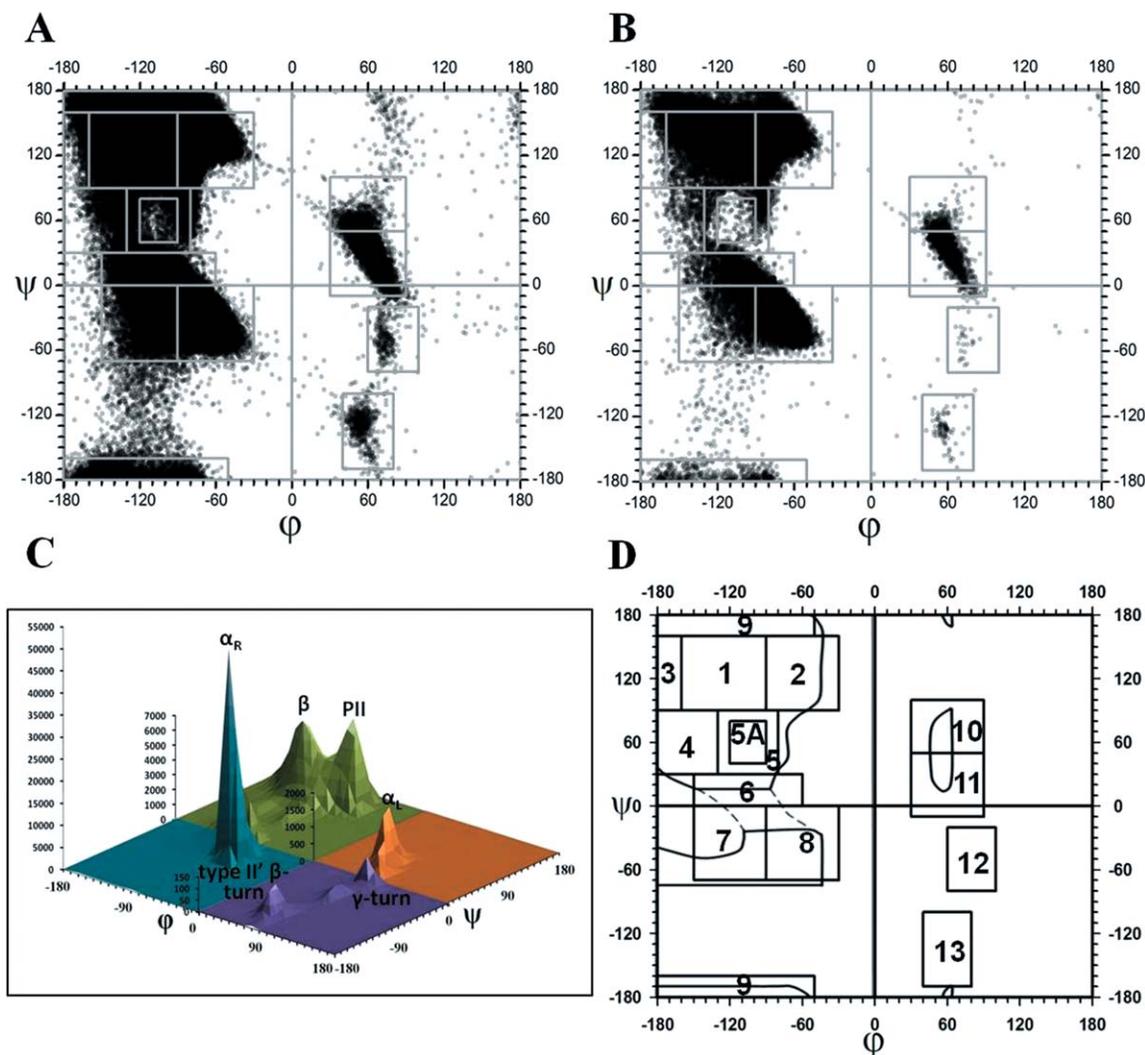
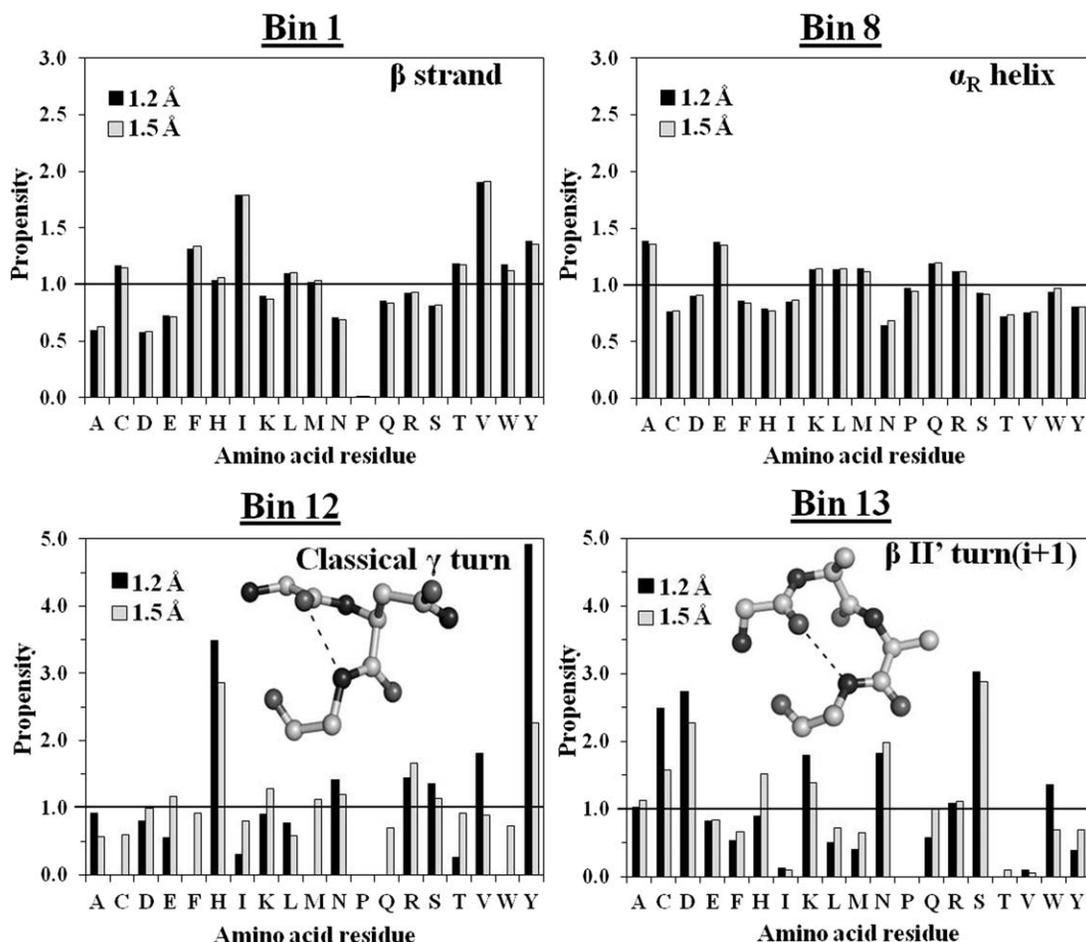


Figure 1

(A) Scatter plot in ϕ - ψ space for all non-Gly residues (≤ 1.5 Å), 427766 points; (B) scatter plot in ϕ - ψ space for all non-Gly residues (≤ 1.2 Å) 107599 points; and (C) a 3D plot showing the distribution of ϕ , ψ values for non-Gly residues in the ≤ 1.5 Å data set. Note the clear separation of clusters corresponding to the β and PII regions. Vertical scales for each quadrant are different; (D) Boundaries of the 14 discrete bins are shown in the background of Ramachandran Map.

allowed Ramachandran space, the sparsely populated regions have attracted less attention. In this brief review, written to mark the 50th anniversary of the Ramachandran map, we focus attention on arid regions and apparently disallowed regions which may be of interest in understanding the twists and turns of polypeptide backbones. For 19 of the 20 genetically coded amino acids, with glycine being the sole exception, large regions of conformational space are disallowed, as a consequence of local steric clashes involving the substituent at the C^α -atom. Distributions of the values of backbone torsion angles (ϕ , ψ) for individual amino acids derived from protein crystal structures have provided “experimental Ramchandan maps”, whose outlines largely conform to those predicted theoretically, half a

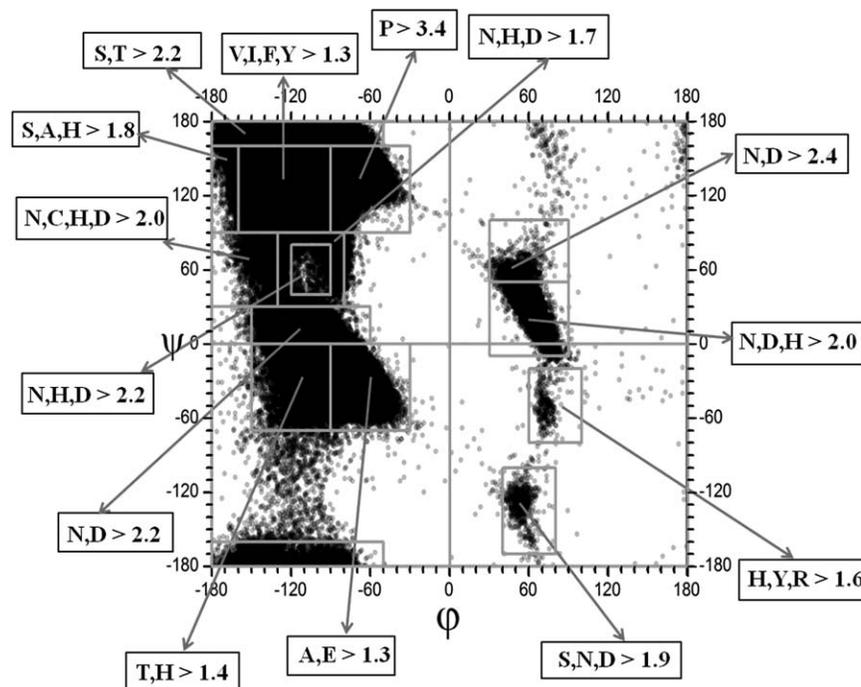
century ago.^{6,7} Indeed, experimental observations from high-resolution protein structures have been the basis for the development of the widely used programs for protein structure validation, PROCHECK,⁸ and Mol-Probity.⁹ A recent overview examines the Ramachandran plot as a powerful device for describing standard structures in proteins.¹⁰ In recent times the Ramachandran plot has been revisited to evaluate the role of hydrogen bonding¹¹ and bond angle distortions¹² in defining the limits of experimental distributions of amino acid conformations. High-resolution protein structures have permitted detailed analysis of the conformation dependence of backbone geometries in proteins.¹³ Careful examination of experimental ϕ - ψ distributions have been used to determine intrinsic

**Figure 2**

Propensity of amino acid residues to occur in Bins 1, 8, 12, and 13 of the Ramachandran map for the 1.2 Å and 1.5 Å data sets. Bin 1 (β strand) and Bin 8 (α_R helix) represent well characterized regions of ϕ - ψ space. Bin 12 (γ turn) and Bin 13 (Type II' β turn) represent the "sterically disallowed" regions in the bottom right quadrant of ϕ - ψ space.

conformational propensities of amino acid residues in proteins.^{14,15} Experimental ϕ - ψ distributions however do reveal that some regions of sterically allowed ϕ - ψ space are sparsely populated, whereas other apparently disallowed regions are populated, albeit to a much lesser extent than the fully allowed regions. These observations are readily explained by the limitations of the original hard-sphere approximation which did not include any consideration of other factors which might compensate for unfavorable van der Waals interactions.^{16,17} Shortle reported the use of the distribution of Ramachandran angles together with rotamer distributions and neighboring residue conformations to examine the effectiveness of composite propensity functions to identify native conformations.¹⁸ Figure 1(A,B,C) shows three experimental distributions for non-Gly amino acid residues using two nonhomologous protein data sets with resolution cutoffs of 1.5 Å and 1.2 Å, respectively. Two features of these scatter plots drew our attention: Firstly, the appearance of a very thinly populated ("arid")

region in the center of the fully allowed region in the top left quadrant. Secondly, the presence of two distinct clusters in the bottom right quadrant which is largely disallowed in the original Ramachandran map. Comparison of the 1.5 Å and 1.2 Å data sets permits delineation of the arid region in the top left quadrant. Rose and coworkers in an analysis of a "protein coil library"¹⁹ identified two basins, termed δ and γ basins, which correspond approximately to the sparsely populated regions in the top left quadrant.²⁰ In a subsequent report Perskie and Rose used the library of coiled segments to examine conformations for basins that lie in the bottom right quadrant.²¹ In this brief review we examine the propensities of individual amino acid residues to adopt these relatively rare Ramachandran conformations in the top left quadrant and also examine the local conformations of residues which populate the right-hand, bottom quadrant of the Ramachandran map, specifically focusing on the gamma turn²² and Type II' β -turn structures.²³

**Figure 3**

Distribution of the highest propensity amino acid residues in specific regions of ϕ - ψ space ($\leq 1.5 \text{ \AA}$ data set).

METHODS

Two nonhomologous data sets were used: (i) A set of 2104 protein crystal structures (total of 2152 chains) comprising of 464,060 residues, determined at a resolu-

tion $\leq 1.5 \text{ \AA}$ and (ii) a set of 563 protein crystal structures (total of 578 chains) comprising of 117,255 residues, determined at a resolution $\leq 1.2 \text{ \AA}$. A sequence identity cutoff of $\leq 30\%$ and a length cutoff of ≥ 40 residues have been imposed for both the data sets. All the

Table I

List of Residues With Greatest and Least Propensities to Occur in Specific Bins in ϕ - ψ Space

Bin no.	ϕ	ψ	Total in each bin (1.5 \AA /1.2 \AA data sets)	High propensity residues from the 1.5 \AA (1.2 \AA) data sets	Low propensity residues from the 1.5 \AA (1.2 \AA) data sets
1	-160 to -90	90 to 160	100,014 (23.3%)/25,737 (23.9%)	V,I,F,Y > 1.3 (V,I,F,Y > 1.3)	P < 0.02 (P < 0.02)
2	-90 to -30	90 to 160	60,369 (14.1%)/15,573 (14.4%)	P > 3.4 (P > 3.3)	^a
3	-180 to -160	90 to 160	2,435 (0.57%)/625 (0.58%)	S,A,H > 1.8 (S,A,H > 1.9)	P,V,L,I < 0.4 (P,V,L,I < 0.3)
4	-180 to -130	30 to 90	2853 (0.67%)/750 (0.70%)	N,C,H,D > 2.0 (N,C,H,D > 2.0)	P,I,V,L < 0.4 (P,I,V,L < 0.5)
5	-130 to -80	30 to 90	6372 (1.49%)/1721 (1.60%)	N,H,D > 1.7 (N,H,D > 2.2)	I,V < 0.5 (I,V,T < 0.5)
5A ^b	-120 to -90	40 to 80	1170 (0.27%)/323 (0.30%)	N,H,D > 2.2 (N,H,W > 2.0)	P,V,T < 0.4 (P,V,T < 0.3)
6	-150 to -60	0 to 30	19,069 (4.4%)/4874 (4.5%)	N,D > 2.2 (N,D > 2.2)	P,I,V < 0.5 (P,I,V < 0.5)
7	-150 to -90	-70 to 0	18,624 (4.3%)/4537 (4.2%)	T,H > 1.4 (T,H > 1.3)	P,A < 0.5 (P,A < 0.5)
8	-90 to -30	-70 to 0	184,548 (43.1%)/44,924 (41.7%)	A,E > 1.3 (A,E > 1.3)	^c
9	-180 to -50	-180 to -160 and 160 to 180	28,437 (6.65%)/7532 (7.0%)	S,T > 2.2 (S,T > 2.0)	L,I < 0.5 (L,I < 0.6)
10	30 to 90	50 to 100	1825 (0.43%)/448 (0.42%)	N,D > 2.4 (N,D > 2.7)	I,V,L,T < 0.3 (I,V,L,T < 0.4)
11	30 to 90	-10 to 50	7690 (1.80%)/2064 (1.92%)	N,D,H > 2.0 (N,D,H > 1.8)	I,V,T < 0.2 (I,V,T < 0.1)
12	60 to 100	-80 to -20	246 (0.06%)/56 (0.05%)	H,Y,R > 1.6 (H,Y > 3.4)	A,L,C < 0.6 (Q,C,M,F,W = 0)
13	40 to 80	-170 to -100	516 (0.12%)/131 (0.12%)	S,N,D > 1.9 (S,C,D > 2.4)	I,V,T < 0.1 (I,V,T < 0.2)

^aAll 18 non-Pro residues have propensity values lying between 0.7 and 1.1 for both 1.5 \AA and 1.2 \AA data sets, suggesting that PII conformation may be generally adopted by all the non-Pro residues.

^bBin 5A is a subset of Bin 5.

^cAll 17 residues except Ala and Glu have propensity values lying between 0.6 and 1.2 for both 1.5 \AA and 1.2 \AA data sets, suggesting that α_R helical conformation may be generally adopted by all residues except Ala and Glu.

crystal structures were obtained from the Protein Data Bank (PDB; <http://www.rcsb.org/pdb/home/home.do>)^{24,25} based on the April 2012 release. A total of 427,766 non-Gly residues and 36,294 Gly residues have been found in the 1.5 Å data set, whereas the 1.2 Å data set contains 107,599 non-Gly residues and 9656 Gly residues. The populated regions of ϕ - ψ space have been divided into 14 bins, defined in Figure 1(D). Propensities of 19 non-Gly residues to occur in each bin for the two data sets were computed using the formula:

$$\text{Propensity} = \frac{\left[\frac{\text{Number of individual amino acid (X) in bin 'n'}}{\text{Total number of amino acids in bin 'n'}} \right]}{\left[\frac{\text{Total number of amino acid (X) in data set}}{\text{Total number of amino acids in data set}} \right]}$$

Previously published algorithms were used for the identification of the secondary structure of flanking residues with unusual Ramachandran conformation.²⁶ Subsequently individual structures were examined by Pymol (<http://www.pymol.org>).²⁷

RESULTS AND DISCUSSION

For purposes of this analysis we have divided the populated regions of ϕ - ψ space into 14 square/rectangular bins as described in Figure 1(D). The distribution of amino acid residues in the various bins was determined and propensity values for each residue type computed. Figure 2 summarizes the propensity values determined for Bins 1, 8, 12, and 13. The extremely well characterized regions of the ϕ - ψ space are represented by Bin 1 (β strand), Bin 2 [polyproline II (PII)], Bin 8 (right-handed α helix, α_R), and Bin 11 (left-handed α helix, α_L). Table I lists residues with the greatest and least propensities to occur in a specified region of ϕ - ψ space. Figure 3 schematically illustrates the nature of residues that have the greatest propensity to lie in a specific region of ϕ - ψ space. Residues with the greatest helix propensity Ala and Glu, lie in Bin 8. Val, Ile, Leu, Phe, and Tyr have the highest propensity for the β -strand region, Bin 1, features that have been previously well established in earlier analysis.²⁸⁻³⁰ As anticipated, Pro dominates Bin 2 which encompasses the PII region.^{31,32} Bins 1, 2, and 9 which represent residues in extended and semiextended (PII) conformations account for about 40% of observed residues, whereas Bin 8 which corresponds to the strongly clustered α -helical region accounts for another 40-45%. Approximately about 85% of residues in protein structures thus occur in these limited regions of the Ramachandran map. Bins 10 and 11 which correspond to the small region of allowed Ramachandran space which accommodates *left-handed helical* (α_L) conformations are dominated by Asn, Asp, and His. Notably Asn has the highest propensity in this region (421 and 2149

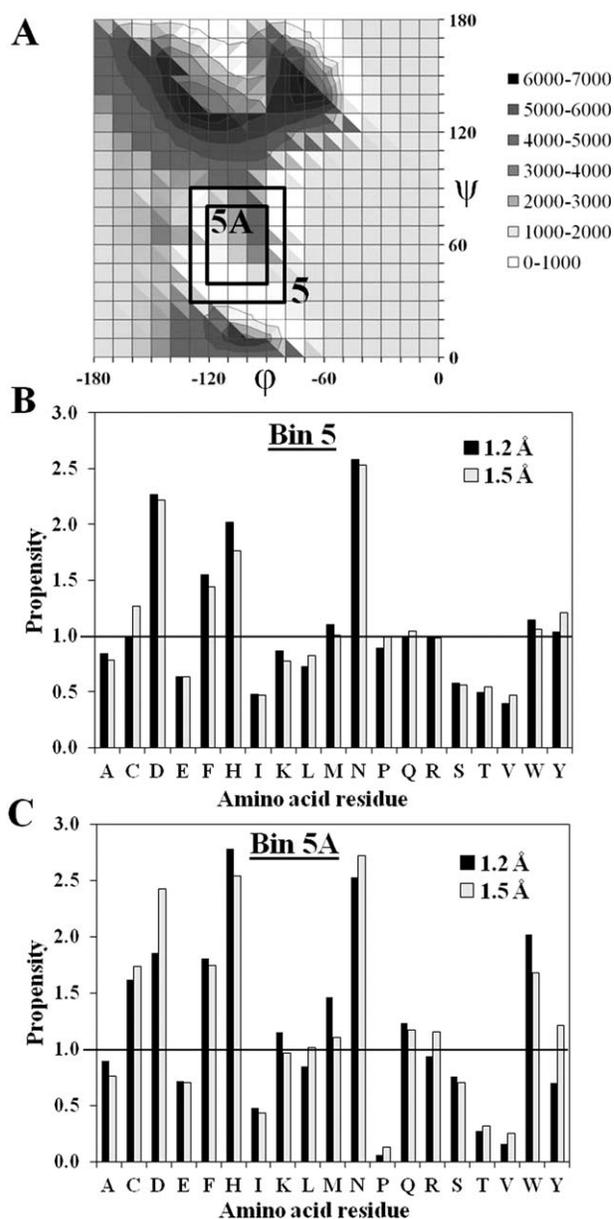


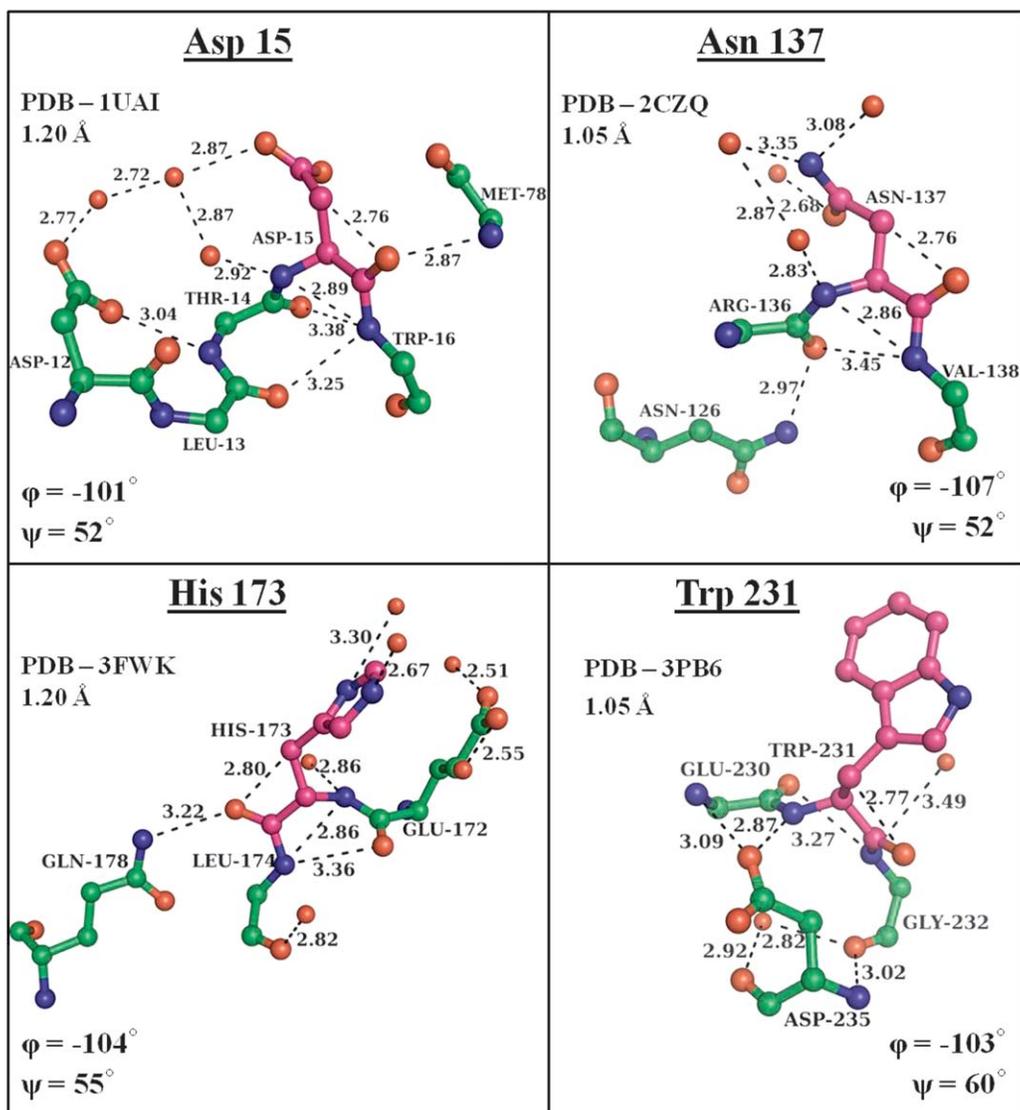
Figure 4

(A) Top view of a 3D ϕ - ψ plot (top left quadrant) indicating the arid region corresponding to Bins 5 and 5A (data set ≤ 1.5 Å, 212125 points); (B) distribution of amino acid residue propensity in Bin 5; and (C) distribution of amino acid residue propensity in Bin 5A.

points in Bins 10 and 11, respectively). The β branched residues Ile and Val have the lowest propensities to adopt α_L conformations. These observations on residue propensities are completely consistent with earlier analysis reported in the literature.^{33,34}

Top left quadrant arid region

The arid region in the top left quadrant is represented by Bin 5. Figure 4(B) shows the propensities for

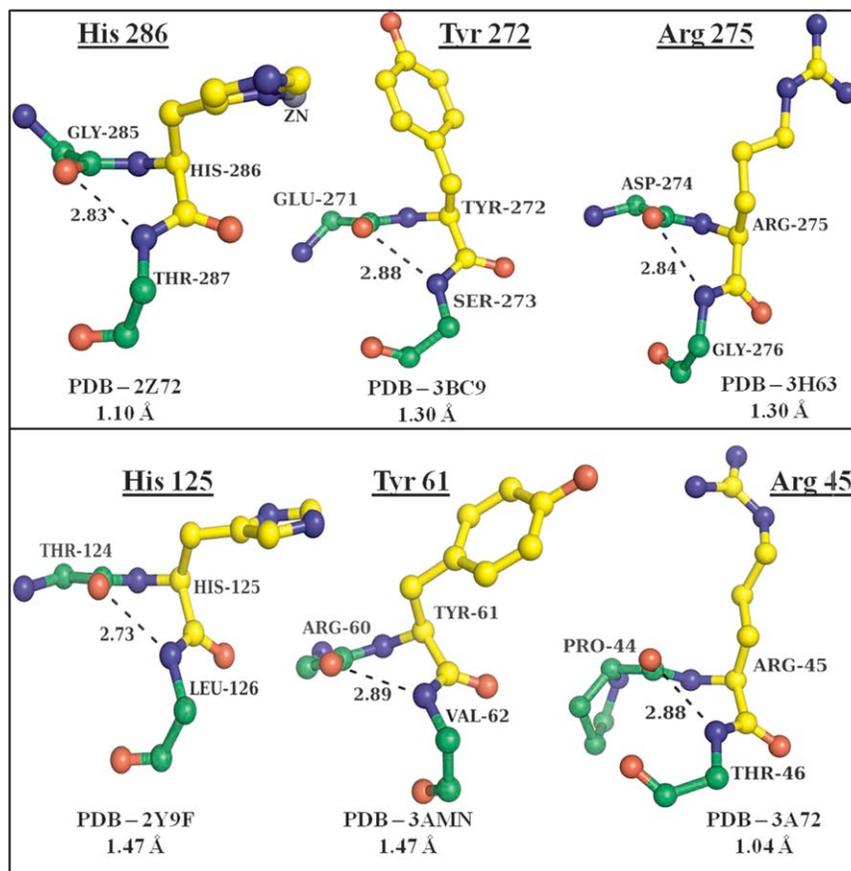
**Figure 5**

Illustrative examples of residues (pink) with high propensity to occur in Bin 5A [Asp 15, PDB:1UAI ($\phi, \psi = -101^\circ, 52^\circ$); Asn 137, PDB:2CZQ ($\phi, \psi = -107^\circ, 52^\circ$); His 173, PDB:3FWK ($\phi, \psi = -104^\circ, 55^\circ$); Trp 231, PDB:3PB6 ($\phi, \psi = -103^\circ, 60^\circ$)]. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the 19 non-Gly residues to occur in this region computed for both 1.5 Å and 1.2 Å data sets. The highest propensities are observed for Asn, Asp and His residues. The aromatic amino acids also show a significantly greater tendency to populate this region, as compared to the aliphatic residues. We further narrowed down our focus to Bin 5A which yielded a total of 1170 points in the 1.5 Å data set and 323 points in 1.2 Å data set. Calculated propensities of occurrence are shown in Figure 4(C). Once again, His, Asn, and Asp show the greatest frequency of occurrence, whereas the aromatic amino acids Phe, Trp, and Cys show a stronger preference to occur in this region as compared to other amino acids. It should be noted that the area rep-

resented by Bin 5A also corresponds to a relatively high energy region in computed energy surfaces or “Ramachandran energy maps.”^{35–37} We therefore turned our attention to an examination of specific examples that occur in this arid region by superimposing a $10^\circ \times 10^\circ$ grid on Bin 5A and examining the region corresponding to the least populated box ($\phi = -110^\circ$ to -100° , $\psi = 50^\circ$ to 60°). This yielded a total of 33 examples from the 1.5 Å data set. Figure 5 provides four illustrative examples.

In the four examples shown the values of the backbone torsion angles (ϕ, ψ) are: PDB: 1UAI, Asp15, $-101^\circ, 52^\circ$; PDB: 2CZQ, Asn137, $-107^\circ, 52^\circ$; PDB: 3FWK, His173, $-104^\circ, 55^\circ$; and PDB: 3PB6, Trp231,

**Figure 6**

Illustrative examples of high propensity residues (yellow) of Bin 12 [classical γ turn; His 286, PDB:2Z72 ($\phi, \psi, \chi^1 = 72^\circ, -46^\circ, -174^\circ$); Tyr 272, PDB:3BC9 ($\phi, \psi, \chi^1 = 75^\circ, -45^\circ, -51^\circ$); Arg 275, PDB:3H63 ($\phi, \psi, \chi^1 = 74^\circ, -64^\circ, -48^\circ$); His 125, PDB:2Y9F ($\phi, \psi, \chi^1 = 78^\circ, -65^\circ, -166^\circ$); Tyr 61, PDB:3AMN, ($\phi, \psi, \chi^1 = 73^\circ, -68^\circ, -160^\circ$); Arg 45, PDB:3A72 ($\phi, \psi, \chi^1 = 79^\circ, -54^\circ, -56^\circ$)]. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

$-103^\circ, 60^\circ$. All four examples may be considered as distorted *inverse* gamma turns^{22,38} in which the stabilizing $3 \rightarrow 1$ (C7) hydrogen bond has been significantly lengthened (N---O 3.27–3.45 Å). All four examples lie on the protein surface and a hydration network is evident, in which key hydrogen bonds hold the backbone of the distorted residue. Listings of backbone dihedral angles for all the 33 examples in the central grid of Bin 5A are listed in Supporting Information Table S1. The possibility that the sterically allowed regions may depend on the N–C $^\alpha$ –C bond angles was first considered by Ramakrishnan & Ramachandran in 1965.² An examination of the bond angles for the 33 examples in the arid region (Bin 5A) yielded an average value of $112 \pm 2^\circ$, establishing that no specific distortion at the C $^\alpha$ was present. The arid region clearly corresponds to a situation in which the loss of $3 \rightarrow 1$ hydrogen bond stabilization in an inverse gamma turn is compensated by local hydrogen bond interactions involving either neighboring residues or waters of hydration.

The proximity to a stable minimum in the energy surface represented by the inverse gamma turn rationalizes the rarity of conformations represented by Bin 5A. Although the absence of unfavorable steric clashes is a necessary condition for a conformation to be significantly populated, it is clearly not a sufficient condition. Hydrogen bonds and local electrostatic interactions determine the population distributions within the Ramachandran allowed regions.

Bottom right quadrant

In the classical representation of the Ramachandran map, (see Figure 1 of Porter and Rose¹¹) for a recent illustration, this quadrant is largely devoid of a sterically allowed region. Inspection of the population distributions shown in Figure 3 reveal that there are significant number of residues observed in the two distinct clusters represented by Bins 12 and 13. Bin 12 contains 246

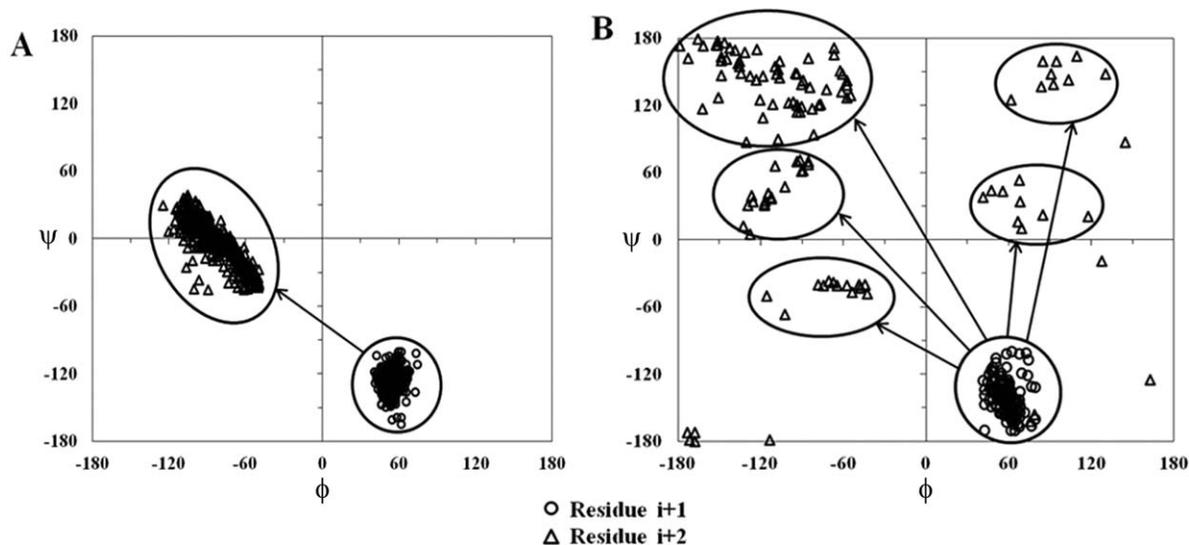


Figure 7

(A) Scatter plot in ϕ - ψ space for all non-Gly, Type II' β turn ($i + 1$, $i + 2$ residues); (B) scatter plot in ϕ - ψ space for all non-Gly, non-Type II' β turn ($i + 1$, $i + 2$ residues). $i + 1$ residues are represented as "o" and $i + 2$ residues are represented as " Δ ". Arrows represent the local conformation of the two-residue segment ($i + 1/i + 2$).

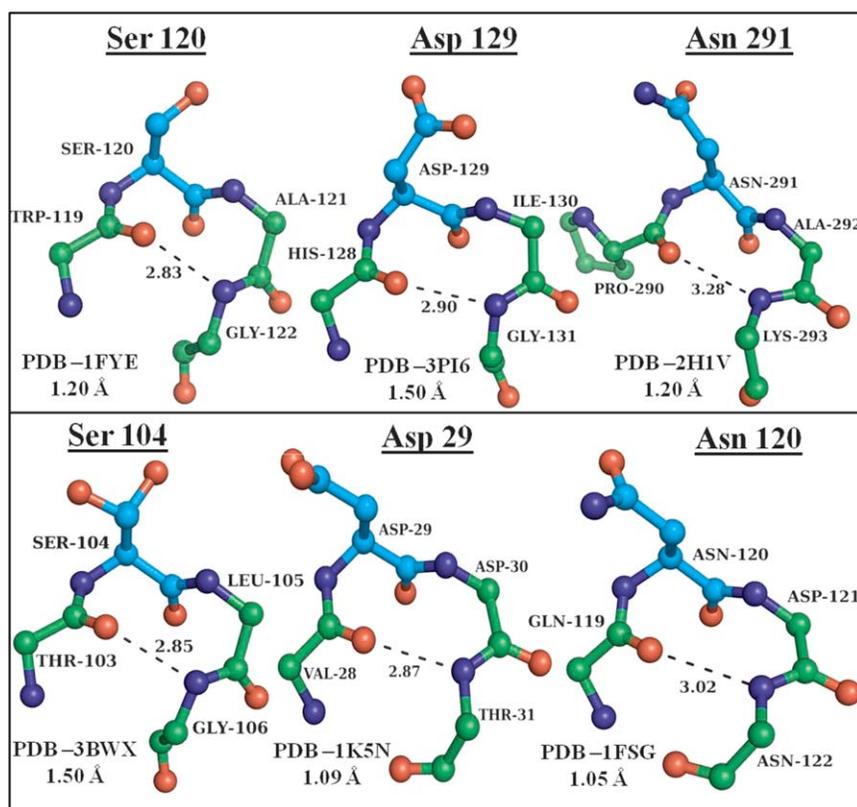
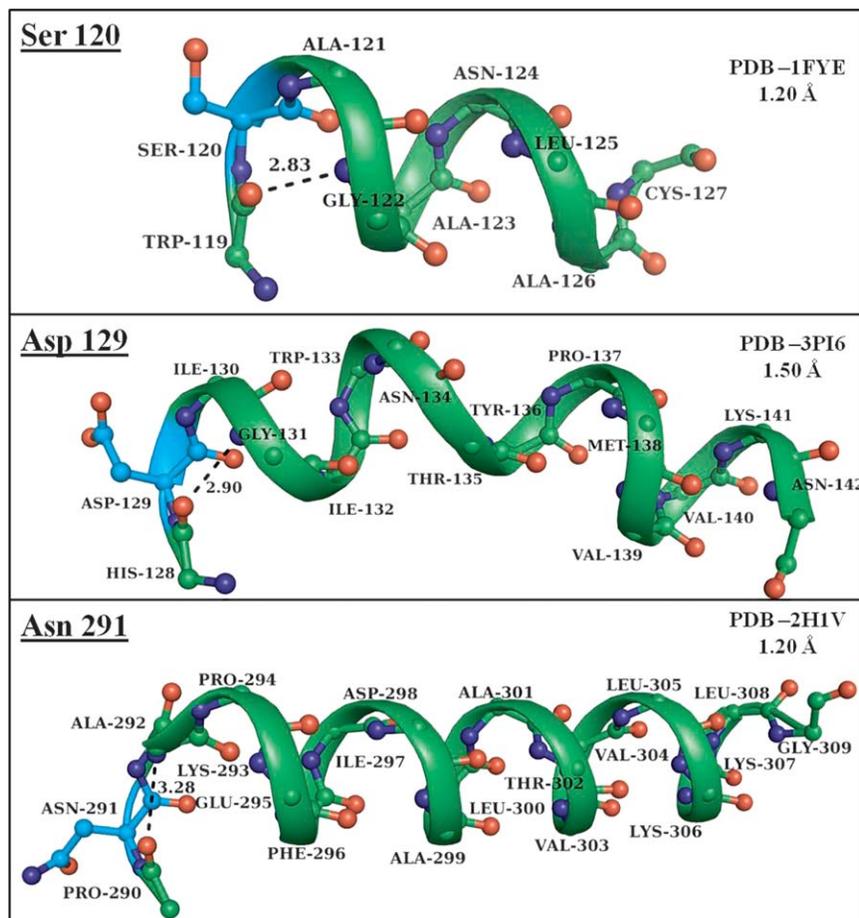


Figure 8

Illustrative examples of high propensity residues (cyan) of Bin 13 [Type II' β turn; Ser 120, PDB:1FYE (ϕ, ψ) $_{i+1} = (57^\circ, -122^\circ)$, (χ^1) $_{i+1} = (-175^\circ)$, and (ϕ, ψ) $_{i+2} = (-53^\circ, -34^\circ)$; Asp 129, PDB:3PI6 (ϕ, ψ) $_{i+1} = (61^\circ, -131^\circ)$, (χ^1) $_{i+1} = (-169^\circ)$, and (ϕ, ψ) $_{i+2} = (-68^\circ, -10^\circ)$; Asn 291, PDB:2H1V (ϕ, ψ) $_{i+1} = (49^\circ, -134^\circ)$, (χ^1) $_{i+1} = (-65^\circ)$, and (ϕ, ψ) $_{i+2} = (-110^\circ, 22^\circ)$; Ser 104, PDB:3BWV (ϕ, ψ) $_{i+1} = (58^\circ, -122^\circ)$, (χ^1) $_{i+1} = (-48^\circ)$, and (ϕ, ψ) $_{i+2} = (-53^\circ, -36^\circ)$; Asp 29, PDB:1K5N (ϕ, ψ) $_{i+1} = (54^\circ, -127^\circ)$, (χ^1) $_{i+1} = (-55^\circ)$, and (ϕ, ψ) $_{i+2} = (-103^\circ, 24^\circ)$; Asn 120, PDB:1FSG (ϕ, ψ) $_{i+1} = (50^\circ, -129^\circ)$, (χ^1) $_{i+1} = (-59^\circ)$, and (ϕ, ψ) $_{i+2} = (-113^\circ, 25^\circ)$]. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

**Figure 9**

Illustrative examples of helices in proteins with Type II' β turn at the N-terminus [$i + 1$ residues are shown in cyan; Ser 120, PDB:1FYE (ϕ, ψ)_{*i*} = $(-127^\circ, 150^\circ)$, (ϕ, ψ)_{*i*+1} = $(57^\circ, -122^\circ)$, (χ^1)_{*i*+1} = (-175°) , (ϕ, ψ)_{*i*+2} = $(-53^\circ, -34^\circ)$, and (ϕ, ψ)_{*i*+3} = $(-72^\circ, -25^\circ)$; Asp 129, PDB:3PI6 (ϕ, ψ)_{*i*} = $(-126^\circ, 134^\circ)$, (ϕ, ψ)_{*i*+1} = $(61^\circ, -131^\circ)$, (χ^1)_{*i*+1} = (-169°) , (ϕ, ψ)_{*i*+2} = $(-68^\circ, -10^\circ)$, and (ϕ, ψ)_{*i*+3} = $(-57^\circ, -30^\circ)$; Asn 291, PDB:2H1V (ϕ, ψ)_{*i*} = $(-53^\circ, -45^\circ)$, (ϕ, ψ)_{*i*+1} = $(49^\circ, -134^\circ)$, (χ^1)_{*i*+1} = (-65°) , (ϕ, ψ)_{*i*+2} = $(-110^\circ, 22^\circ)$, and (ϕ, ψ)_{*i*+3} = $(-67^\circ, 146^\circ)$]. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

points, (0.057%), whereas Bin 13 contains 516 points, (0.12%). Bin 12 corresponds to the *classical* gamma turn conformation in which the substituent at the C $^\alpha$ atoms lies in a quasixial position. Figure 6 provides illustrative examples. His, Tyr, and Arg residues show the highest propensity to adopt these conformations.

Bin 13 corresponds to the inverse of the PII conformation (PII'), which is most often found at the $i + 1$ residue in Type II' β turns.^{23,39} Of the 516 examples of non-Gly residues occurring in Bin 13, 384 occupied the $i + 1$ position of the Type II' β turn. Figure 7 shows the scatter plot in ϕ - ψ space for the Bin 13 residue $i + 1$ and the following residue $i + 2$. The tight clustering for the residues $i + 1/i + 2$ of Type II' β turns is evident. For the non-Type II' β -turn conformations, there is significantly greater diversity of conformations at the $i + 2$ position. The prime turns (I' and II') have a high propensity to facilitate β -hairpin formation.^{40,41} Ser,

Asn, and Asp have the highest propensity to occur in Bin 13. Interestingly, there is a distinct difference in the nature of the residues with the highest propensity to occur in Bins 12 and 13, although both these clusters are in a region of conformational space where unfavorable van der Waals contacts must necessarily be compensated by other favorable interactions. Figure 8 provides illustrative examples of residues occurring in Bin 13 and adopting the Type II' β -turn conformations. To further examine the role of Type II' β turns in the nucleation of local structures, we examined an expanded data set in Bin 13 which also included Gly residues (non-Gly residues = 516, Gly residues = 1387, and total residues = 1903). A total of 1395 residues occupied the $i + 1$ position of Type II' β turns. These examples were further examined to identify secondary structural elements at both the N- and C-terminal sides. In principle, the Type II' β turn can occur at the N-terminus of a helical

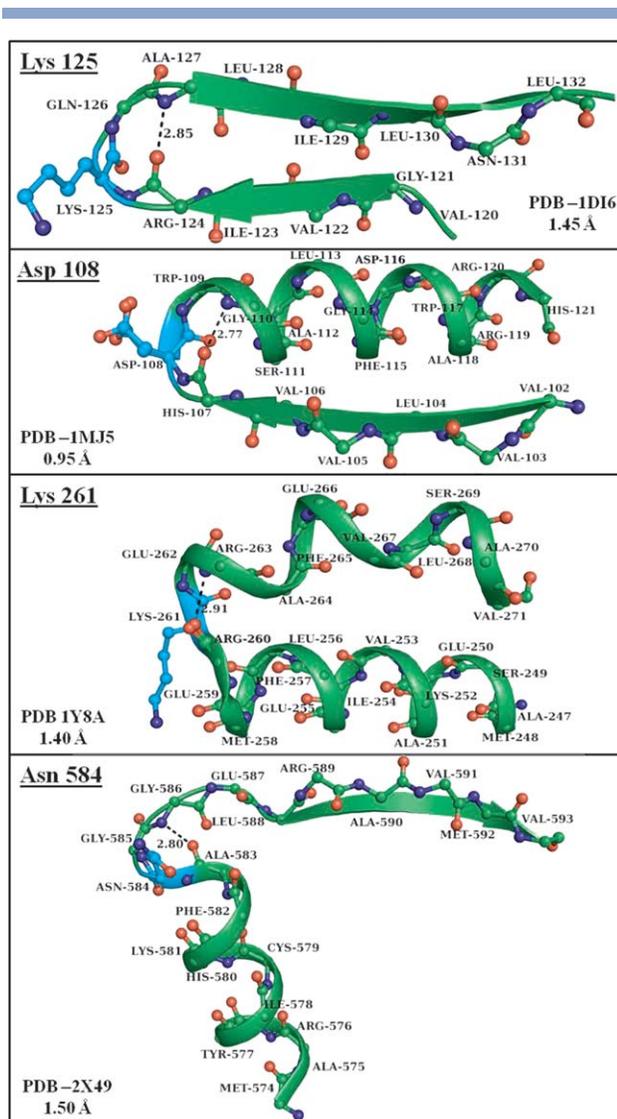


Figure 10

Illustrative examples of all four motifs wherein Type II' β turns serve as connecting segments [$i + 1$ residues are shown in cyan; Lys125, PDB:1DI6 ($(\phi, \psi)_i = (-122^\circ, 120^\circ)$, $(\phi, \psi)_{i+1} = (59^\circ, -118^\circ)$, $(\chi^1)_{i+1} = (-70^\circ)$, $(\phi, \psi)_{i+2} = (-113^\circ, 14^\circ)$, and $(\phi, \psi)_{i+3} = (-99^\circ, 142^\circ)$; Asp 108, PDB:1MJ5 ($(\phi, \psi)_i = (-135^\circ, 143^\circ)$, $(\phi, \psi)_{i+1} = (56^\circ, -133^\circ)$, $(\chi^1)_{i+1} = (-175^\circ)$, $(\phi, \psi)_{i+2} = (-63^\circ, -20^\circ)$, and $(\phi, \psi)_{i+3} = (-64^\circ, -31^\circ)$; Lys 261, PDB:1Y8A ($(\phi, \psi)_i = (-113^\circ, -7^\circ)$, $(\phi, \psi)_{i+1} = (57^\circ, -122^\circ)$, $(\chi^1)_{i+1} = (-71^\circ)$, $(\phi, \psi)_{i+2} = (-55^\circ, -26^\circ)$, and $(\phi, \psi)_{i+3} = (-88^\circ, 1^\circ)$; Asn 584, PDB:2X49, $(\phi, \psi)_i = (-90^\circ, -10^\circ)$, $(\phi, \psi)_{i+1} = (53^\circ, -121^\circ)$, $(\chi^1)_{i+1} = (-70^\circ)$, $(\phi, \psi)_{i+2} = (-117^\circ, 9^\circ)$, and $(\phi, \psi)_{i+3} = (-115^\circ, 167^\circ)$]. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

segment since the $i + 2$ residues occur in the α_R region. Interestingly, 148 examples were recovered in which a well developed right-handed α helix followed a Type II' β turn. Such Type II' β turn initiated helical segments have also been characterized in the crystal structures of short synthetic peptides.⁴² Figure 9 shows three illustrative

examples of helical segments in proteins, containing Type II' β turn at the N-terminus. We further examined the Type II' β turns for the development of flanking secondary structures, helices and β strands, in the limited set of 384 examples with a non-Gly residue at the $i + 1$ position. Of these, 126 correspond to examples where secondary structural elements were connected by a two residue, tight Type II' β turn. Of these 59 correspond to antiparallel strand arrangements resulting in β -hairpin structures, whereas 57 resulted in a β strand-helix motif. Only eight examples could be classified as antiparallel helix motif with the central connecting Type II' β turn. While only two examples of a helix-strand motif were found. Figure 10 provides illustrative examples of all four motifs wherein Type II' β turns serve as connecting segments. Clearly, the β hairpin and the antiparallel strand-helix arrangements are the favored super secondary structural motifs facilitated by the Type II' β turn.

CONCLUSIONS

The analysis of amino acid residue conformations presented above has focused on the sparsely populated regions of conformational space. Within the fully allowed region of L-amino acid residues, the central arid region in the top left quadrant of the Ramachandran map is clearly disfavored since minor changes in dihedral angles allow the residue to form a C7 hydrogen bond resulting in an inverse gamma turn conformation. The relatively few examples of high fidelity protein structures, which lie in this arid region, reveal that the energetic penalty for distortion from the favored gamma turn region is compensated by hydrogen bonds to solvents. Residues with these unusual Ramachandran conformations almost invariably occur at significantly hydrated sites.

The bottom right quadrant was a largely disallowed region of conformational space for L-amino acid residues in the original formulation of Ramachandran map. Nevertheless, two distinct clusters of experimental points may be observed corresponding to the gamma turn conformation and the inverse of the PII structure (PII'). The PII' conformation well followed by a residue in the helical α_R region results in the Type II' β turn. Indeed as many as 384 examples of the 516 residues occurring in Bin 13 (PII') occupy the $i + 1$ position of Type II' β turn. The formation of the $4 \rightarrow 1$ hydrogen bond in the turn presumably compensates for the relatively high van der Waals energy in this region of conformational space. The Type II' β turn is an important element in promoting helix folding in both peptides and proteins. A survey of the examples examined in the present study revealed that the Type II' β turns serve as a connecting element in β hairpins, a feature recognized nearly three decades ago by Sibanda and Thornton.⁴⁰ In addition to the β hairpin, antiparallel strand-helix motifs are also favored.

Relatively few examples of antiparallel helices are observed. Two-dimensional (2D) ϕ - ψ conformational space provides a powerful means for analyzing a local conformation of polypeptide chain. Although considerable attention has been devoted to regions of ϕ - ψ space corresponding to regular secondary structures, further insights may follow from a closer examination of sparsely populated regions.

The Ramachandran map anticipated many features of amino acid residue conformations in proteins in the years before X-ray diffraction provided detailed insights into the folded structures of proteins. Backbone torsion angles have come to be widely used as descriptors of polypeptide conformations and parameters that help in classifying structural motifs in proteins. Torsion angle space, bounded by the constraints of the Ramachandran map, defines the range of structures that may be explored by individual amino acid residues as proteins fold and unfold. This brief review of amino acid propensities for the sparsely populated allowed regions is intended to mark an anniversary and also to suggest that a great deal may still be learnt by examining the growing database of high-resolution protein structures.

ACKNOWLEDGMENT

NVK thanks Prof. K.S. Krishnan for the encouragement and support.

REFERENCES

- Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. *J Mol Biol* 1963;7:95–99.
- Ramakrishnan C, Ramachandran GN. Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophys J* 1965;55:909–933.
- Ramachandran GN, Sasisekharan V. Conformation of polypeptides and proteins. *Adv Protein Chem* 1968;23:283–438.
- Ting D, Wang G, Shapovalov M, Mitra R, Jordan MI, Dunbrack RL, Jr. Neighbor-dependent Ramachandran probability distributions of amino acids developed from a hierarchical Dirichlet process model. *PLoS Comput Biol* 2010;6:e1000763.
- Hermans J. The amino acid dipeptide: small but still influential after 50 years. *Proc Natl Acad Sci USA* 2011;108:3095–3096.
- Herzberg O, Moult J. Analysis of the steric strain in the polypeptide backbone of protein molecules. *Proteins* 1991;11:223–229.
- Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. Stereochemical quality of protein structure coordinates. *Proteins* 1992;12:345–364.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK—a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 1993;26:283–291.
- Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 2010;66:12–21.
- Hollingsworth SA, Karplus PA. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomol Concepts* 2010;1:271–283.
- Porter LL, Rose GD. Redrawing the Ramachandran plot after inclusion of hydrogen-bonding constraints. *Proc Natl Acad Sci USA* 2011;108:109–113.
- Zhou AQ, O'Hern CS, Regan L. Revisiting the Ramachandran plot from a new angle. *Protein Sci* 2011;20:1166–1171.
- Berkholz DS, Shapovalov MV, Dunbrack RL Jr, Karplus PA. Conformation dependence of backbone geometry in proteins. *Structure* 2009;17:1316–1325.
- Swindells MB, MacArthur MW, Thornton JM. Intrinsic ϕ , ψ propensities of amino acids, derived from the coil regions of known structures. *Nat Struct Biol* 1995;2:596–603.
- Beck DAC, Alonso DOV, Inoyama D, Daggett V. The intrinsic conformational propensities of the 20 naturally occurring amino acids and reflection of these propensities in proteins. *Proc Natl Acad Sci USA* 2008;105:12259–12264.
- Mandel N, Mandel G, Trus BL, Rosenberg J, Carlson G, Dickerson RE. Tuna cytochrome c at 2.0 Å resolution. *J Biol Chem* 1977;252:4619–4635.
- Ho BK, Thomas A, Brasseur R. Revisiting the Ramachandran plot: hard-sphere repulsion, electrostatics, and H-bonding in the alpha-helix. *Protein Sci* 2003;12:2508–2522.
- Shortle D. Composites of local structure propensities: evidence for local encoding of long-range structure. *Protein Sci* 2002;11:18–26.
- Fitzkee NC, Fleming PJ, Rose GD. The Protein Coil Library: a structural database of nonhelix, nonstrand fragments derived from the PDB. *Proteins* 2005;58:852–854.
- Perskie LL, Street TO, Rose GD. Structures, basins, and energies: a deconstruction of the Protein Coil Library. *Protein Sci* 2008;17:1151–1161.
- Perskie LL, Rose GD. Physical-chemical determinants of coil conformations in globular proteins. *Protein Sci* 2010;19:1127–1136.
- Milner-White EJ. Situations of gamma-turns in proteins: their relation to alpha-helices, beta-sheets and ligand binding sites. *J Mol Biol* 1990;216:385–397.
- Wilmot CM, Thornton JM. Analysis and prediction of the different types of beta-turn in proteins. *J Mol Biol* 1988;203:221–232.
- Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol* 1977;112:535–542.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. *Nucleic Acids Res* 2000;28:235–242.
- Ramakrishnan C, Srinivasan N. Glycyl residues in proteins and peptides: an analysis. *Curr Sci* 1990;59:851–861.
- DeLano WL. The PyMOL molecular graphics system. San Carlos, CA: DeLano Scientific; 2002. Available at: <http://www.pymol.org>
- Chou PY, Fasman GD. Conformational parameters for amino acids in helical, beta-sheet and random coils calculated from proteins. *Biochemistry* 1974;13:211–222.
- Chou PY, Fasman GD. Empirical predictions on protein conformation. *Annu Rev Biochem* 1978;47:251–276.
- Creighton TE. *Proteins: structure and molecular properties*. New York, NY: W.H. Freeman and Company; 1993.
- Han WG, Jalkanen KJ, Elstner M, Suhai S. Theoretical study of aqueous N-acetyl-L-alanine N'-methylamide: structures and Raman, VCD, and ROA spectra. *J Phys Chem B* 1998;102:2587–2602.
- Shi Z, Woody RW, Kallenbach NR. Is polyproline II a major backbone conformation in unfolded proteins? *Adv Protein Chem* 2002;62:163–240.
- Richardson JS, Richardson DC. Principles and patterns in protein conformation. In: Fasman GD, editor. *Prediction of protein structure and the principles of protein conformation*. New York: Plenum Press; 1989. pp 1–98.
- Srinivasan N, Anuradha VS, Ramakrishnan C, Sowdhamini R, Balam P. Conformational characteristics of asparaginyl residues in proteins. *Int J Pept Protein Res* 1994;44:112–122.

35. Scott RA, Scheraga HA. Method for calculating internal rotation barriers. *J Chem Phys* 1965;42:2209–2215.
36. Gibson KD, Scheraga HA. Influence of flexibility on the energy contours of dipeptide maps. *Biopolymers* 1966;4:709–712.
37. Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M. Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem Sci* 2000;25:331–339.
38. Rose GD, Gierasch LM, Smith JA. Turns in peptides and proteins. *Adv Protein Chem* 1985;37:1–109.
39. Venkatachalam CM. Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. *Biopolymers* 1968;6:1425–1436.
40. Sibanda BL, Thornton JM. Beta-hairpin families in globular proteins. *Nature* 1985;316:170–174.
41. Gunasekaran K, Ramakrishnan C, Balam P. Beta-hairpins in proteins revisited: lessons for de novo design. *Protein Eng* 1997;10:1131–1141.
42. Raghavender US, Kantharaju, Aravinda S, Shamala N, Balam P. Hydrophobic peptide channels and encapsulated water wires. *J Am Chem Soc* 2010;132:1075–1086.