

Research article

Open Access

Tripeptide analysis of protein structures

Sharmila Anishetty^{1,2}, Gautam Pennathur^{1,2} and Ramesh Anishetty*³

Address: ¹AU-KBC Research Centre, Anna University, Chennai-44, India, ²Centre for Biotechnology, Anna University, Chennai-25, India and ³The Institute of Mathematical Sciences, Chennai-113, India

Email: Sharmila Anishetty - sharmila@au-kbc.org; Gautam Pennathur - pgautam@annauniv.edu; Ramesh Anishetty* - ramesha@imsc.res.in

* Corresponding author

Published: 21 December 2002

Received: 13 July 2002

BMC Structural Biology 2002, 2:9

Accepted: 21 December 2002

This article is available from: <http://www.biomedcentral.com/1472-6807/2/9>

© 2002 Anishetty et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: An efficient building block for protein structure prediction can be tripeptides. 8000 different tripeptides from a dataset of 1220 high resolution ($\leq 2.0^\circ\text{A}$) structures from the Protein Data Bank (PDB) have been looked at, to determine which are structurally rigid and non-rigid. This data has been statistically analyzed, discussed and summarized. The entire data can be utilized for the building of protein structures.

Results: Tripeptides have been classified into three categories: rigid, non-rigid and intermediate, based on the relative structural rigidity between C_α and C_β atoms in a tripeptide. We found that 18% of the tripeptides in the dataset can be classified as rigid, 4% as non-rigid and 78% as intermediate. Many rigid tripeptides are made of hydrophobic residues, however, there are tripeptides with polar side chains forming rigid structures. The bulk of the tripeptides fall in the intermediate class while very small numbers actually fall in the non-rigid class. Structurally all rigid tripeptides essentially form two structural classes while the intermediate and non-rigid tripeptides fall into one structural class. This notion of rigidity and non-rigidity is designed to capture side chain interactions but not secondary structures.

Conclusions: Rigid tripeptides have no correlation with the secondary structures in proteins and hence this work is complementary to such studies. Tripeptide data may be used to predict plausible structures for oligopeptides and for denovo protein design.

Background

The exact conformation of a tripeptide in a protein can be determined from its side chain and hydrogen bond interactions. The backbone conformation is captured in the Ramachandran angles ϕ , ψ [1]. The rotamer degrees of freedom fixes the entire side chain. A fixed tripeptide occurring in different proteins will have different side chain interactions, which in turn determine the range of ϕ , ψ angles. These ranges are typically fairly wide and hence their utility in protein structure prediction is limited. The use of Ramachandran angles comes about from the well known

planarity of the peptide bond. In this work we would like to complement this bias. We expect that C_β locations are more directly constrained by the side chain interactions. The data was therefore analyzed in terms of the C_α and C_β locations, without the bias of the planarity of the peptide bond. For structure prediction, we need the side chain interaction statistics in terms of C_α and C_β locations. The planarity of the peptide bonds can be imposed as a further constraint.

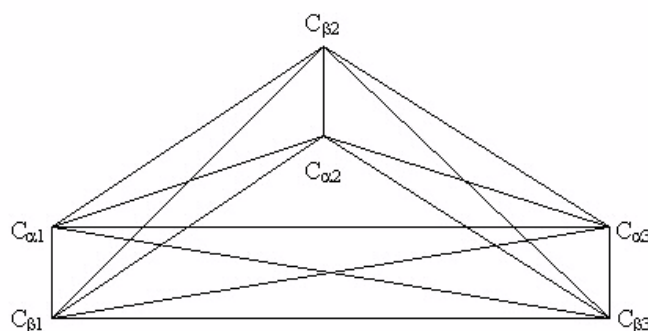


Figure 1
Tripeptide "R₁ R₂ R₃" with C_α and C_β positions

Table 1: Sample of the Mean (M), Standard Deviation (SD) in Angstroms and Frequency (F).

R ₁ R ₂ R ₃	R ₁ R ₂ distances											
	(α1,α2)			(α1,β2)			(β1,α2)			(β1,β2)		
	M	SD	F	M	SD	F	M	SD	F	M	SD	F
AAA	3.80	0.02	292	4.82	0.10	291	4.50	0.10	291	5.37	0.18	291
AAC	3.80	0.03	37	4.81	0.12	37	4.51	0.12	37	5.39	0.23	37
AAD	3.80	0.02	123	4.79	0.11	122	4.55	0.15	123	5.43	0.25	122
AAE	3.80	0.02	138	4.81	0.11	138	4.50	0.10	138	5.38	0.21	138

Only the R₁R₂ set for four tripeptides is shown here. Similar data for R₁R₂, R₁R₃ and R₂R₃ for 7964 tripeptides is available at http://www.aubc.org/research_areas/bio/projects/protein/tri.html

In theory there are 8000 tripeptides and 160,000 tetrapeptides. With the dataset of about 1220 polypeptide chain structures of high resolution ($\leq 2^\circ\text{A}$), an analysis of tripeptide study is feasible and a tetrapeptide study will be statistically insignificant. In our representative protein data set, for each tripeptide "R₁ R₂ R₃" (Figure 1) we pick the corresponding C_{αi}, C_{βi} i = 1,2,3 positions, and then compute distances (αi,αj), (αi,βj) and (βi,βj) for i not equal to j. These 12 distances along with the covalent bond length (αi,βi) capture the entire solid structure of the tripeptide as embedded in the protein. This information is sufficient to fix a unique φ, ψ value along with the position of the C_β. The same "R₁ R₂ R₃" is picked up from different proteins and from different locations within the same protein. This data is statistically analyzed without

any further physico-chemical bias about the residues. In this way, data about various possible conformations of a tripeptide is gathered and analyzed as explained in the discussion section.

Attempts to understand and capture structural features were first done as inter residue contact statistics [2,3] involving long range interactions. Short range pairwise residue statistics is contained in the Protein Atlas [4]. In the literature [5–12] there have been many attempts to predict new protein or peptide structures using small amino acid subsequences or fragments of length 3 to 10. In the seminal paper [5], Unger *et al.* analyzed hexamers and identified about 100 structural building blocks, in which, each amino acid has a position dependent probability of

Table 2: Column 1 shows the standard deviation ranges in °A. For example the first range is between 0 and 0.1°A and the second range is between 0.1 and 0.2°A and so on. Columns 2 to 12 show for the 12 distances, number of tripeptides, which fall in the range.

The number of tripeptides in different standard deviation ranges												
Std dev °A	($\alpha 1, \alpha 2$)	($\alpha 1, \beta 2$)	($\beta 1, \alpha 2$)	($\beta 1, \beta 2$)	($\alpha 1, \alpha 3$)	($\alpha 1, \beta 3$)	($\beta 1, \alpha 3$)	($\beta 1, \beta 3$)	($\alpha 2, \alpha 3$)	($\alpha 2, \beta 3$)	($\beta 2, \alpha 3$)	($\beta 2, \beta 3$)
0.1	7451	2500	594	192	130	109	151	123	7443	2319	604	188
0.2	153	4829	5862	632	82	50	157	69	190	5025	5885	601
0.3	165	128	873	3540	141	67	255	116	159	118	858	3600
0.4	89	31	112	2380	459	160	615	203	69	23	96	2424
0.5	38	26	57	352	1455	397	1312	434	31	22	52	290
0.6	12	11	19	48	2606	795	1317	604	15	11	15	29
0.7	16	8	12	10	2525	1774	1512	1068	12	13	15	17
0.8	7	4	9	3	424	1746	909	1041	6	3	9	9
0.9	4	1	5	9	62	1469	711	1125	12	9	4	5
1	3	4	0	3	20	624	350	773	2	4	6	6
1.1	2	5	4	2	12	223	162	542	4	2	5	1
1.2	2	2	2	3	7	88	48	427	3	2	2	4
1.3	2	3	4	2	5	27	28	306	1	2	1	2
1.4	4	5	3	1	7	7	6	188	1	2	1	2
1.5	4	1	0	5	4	6	9	83	0	0	2	2
1.6	2	3	0	0	3	3	6	36	2	2	1	0
1.7	2	1	3	2	4	2	2	19	3	4	1	4
1.8	3	1	2	1	1	2	2	11	1	1	4	2
1.9	0	0	2	0	1	2	0	7	1	1	1	0
2	1	0	1	1	2	0	0	3	3	0	2	0

occurrence. Others [10] classified the 20 amino acids into 4 classes (Glycine, Proline, hydrophobic and hydrophilic) and identified structural clusters. Baker *et al* [11,12] used fragments with definite amino acid sequences, and their structural variations in different known proteins, as potential templates for predicting new protein structures. In all the cases they have only used the backbone structure (i.e. C_{α} co-ordinates only). In contrast, we have used the C_{α} and C_{β} co-ordinates, which then captures a significant side chain configuration. Recently, de novo protein design [13–19,22] is gaining importance and our data can also be used for this purpose.

Results

Tripeptides have been classified into three categories: rigid, non-rigid and intermediate based on the relative structural rigidity of various distances between C_{α} and C_{β} atoms in a tripeptide. In our sample of crystallized proteins, of the possible 8000 tripeptides, only 7964 occurred. The 36 tripeptides, which do not occur in our sample, however do occur in the SWISS-PROT [20] sequence data. The tripeptides, which occurred more than 5 times in our sample, were taken into consideration in our analysis. We found that 1294 (18%) tripeptides can be classified as rigid, 302 (4%) as non-rigid and 5731 (78%) as intermediate. We classify rigid tripeptides as those, where at least one (1,3) distance has a fluctuation less

than 0.4°A and non-rigid tripeptides as those with all the (1,3) distance fluctuations greater than 0.7°A. After accounting for rigid and non-rigid, all the intermediates have at least one (1,3) distance fluctuation between 0.4 to 0.7°A. The entire dataset of 7964 tripeptides along with all the 12 relative average distances, standard deviations and frequencies is available at the URL http://www.aukbc.org/research_areas/bio/projects/protein/tri.html. A profile of this large data set can be obtained by studying side chain relative abundance (Table 3) and structural homology (Table 4). To find correlation between the three classes of tripeptides and known secondary structure subsequences, 9986 helices of sequence length ≥ 5 , 4541 helices of sequence length ≥ 12 , 9544 beta strands of length ≥ 3 and 3120 beta strands of length ≥ 7 from DSMP [21] were analyzed and the results are tabulated in Table 5. The distribution of the three classes of tripeptides in these secondary structure elements mirrors their distribution in the entire dataset, occurrence being predominantly intermediate followed by rigid and non-rigid. It can be inferred that the rigid tripeptides have no definite correlation with secondary structure elements.

Table 3 shows the frequency of various amino acid residues occurring in the three categories intermediate, rigid and non-rigid. In any category we have mentioned the relative percentage of occurrence within the same tripeptide

Table 3: Column 1 shows the amino acid. Columns 2 to 4 show percentage of occurrence of the 20 amino acids in Intermediate (I), Rigid (R) And Non-rigid (N) categories respectively.

Amino acid	I%	R%	N%
Glycine G	18	5	18
Alanine A	16	13	18
Valine V	16	12	11
Leucine L	16	15	8
IsoLeucine I	15	16	10
Methionine M	10	23	10
Proline P	13	23	10
Phe Alanine F	15	14	11
Tryptophan W	10	20	12
Serine S	16	6	34
Threonine T	16	7	25
Asparagine N	16	11	14
Glutamine Q	13	20	10
Tyrosine Y	14	15	15
Cysteine C	9	16	16
Lysine K	15	14	13
Arginine R	14	16	14
Histidine H	13	15	11
Asp acid D	16	9	19
Glu acid E	15	17	10

Table 4: Column 1 has the mean ($\alpha 1, \alpha 3$) distance ranges (first bin 5.2 to 5.4°A and so on) Columns 2 to 4 have counts of intermediate, rigid and non-rigid tripeptides respectively, falling in each bin.

Mean ($\alpha 1, \alpha 3$)	I	R	N
5.4	2	5	0
5.6	22	121	0
5.8	772	335	6
6	1744	276	40
6.2	1755	200	128
6.4	1052	209	89
6.6	330	113	23
6.8	52	27	9
7	2	8	5
7.2	0	0	1
7.4	0	0	0
7.6	0	0	1

category. The amino acids Methionine, Proline, Tryptophan and Glutamine occur predominantly in the rigid tripeptides and not as much in the other two categories, suggesting that Methionine, Proline, Tryptophan and Glutamine can cause structural rigidity in a tripeptide. Similarly, it can be inferred that Serine and Threonine can cause structural non-rigidity in a tripeptide. The amino acid Cysteine tends to be either in the rigid or in the non-rigid category, but rarely in the intermediate category. The sum of all the percentage numbers within a category falls

a little short of 300, since the residues occur at any of the three possible positions in a tripeptide. Multiple occurrences of a residue within a tripeptide are counted only once.

Table 3 does not reflect the absolute frequency of occurrence in our dataset. For example Methionine and Tryptophan actually occur rarely. Therefore, we looked at those tripeptides that occur with more than the average frequency which is greater than 40 in our sample. This

Table 5: Intermediate, rigid and non-rigid tripeptide occurrence in relative percentages \pm Standard deviation (in %), in secondary structures and in the entire dataset.

Secondary Structure Type	Sequence length	I %	R %	N %
Helix	≥ 5	82 ± 16	14 ± 14	3 ± 7
Helix	≥ 12	82 ± 11	15 ± 11	3 ± 5
β Strands	≥ 3	84 ± 24	13 ± 22	3 ± 11
β Strands	≥ 7	85 ± 15	11 ± 13	3 ± 7
Entire dataset	> 0	12	4	84

high frequency sample dataset again shows that Serine and Threonine are predominant amongst non-rigids, while Proline, Alanine, Isoleucine, and Leucine occur often amongst the rigids.

The structural homology of most tripeptides is shown in Table 4. The fluctuating nature of tripeptides is quite often captured by (α_1, α_3) distances. So, we chose to classify various structural similarities by looking at the (α_1, α_3) distances alone. This was done by taking the (α_1, α_3) mean distances in bins of 0.2°A (range being 5.0 to 7.8°A) and counting the number of tripeptides falling into each bin. Table 4 shows that almost all the intermediate tripeptides have an (α_1, α_3) distance of $(6.0 \pm 0.7)^\circ\text{A}$. We take the range to be 1.4 as the allowed fluctuation in any single tripeptide is 0.7°A . The tabulated results show that all intermediates are broadly similar in structure. Among the rigid tripeptides where the maximum allowed fluctuation in any individual tripeptide is 0.4°A only, we find that there are essentially two structural categories namely $(5.8 \pm 0.4)^\circ\text{A}$ and $(6.4 \pm 0.4)^\circ\text{A}$. Finally the non-rigid tripeptides, whose fluctuations are certainly larger than 0.7°A , can be thought of having essentially one structure.

The frequency with which each of the tripeptides occurs in our sample set of 0.27 million tripeptides from 1220 polypeptide chains is shown in Figure 2.

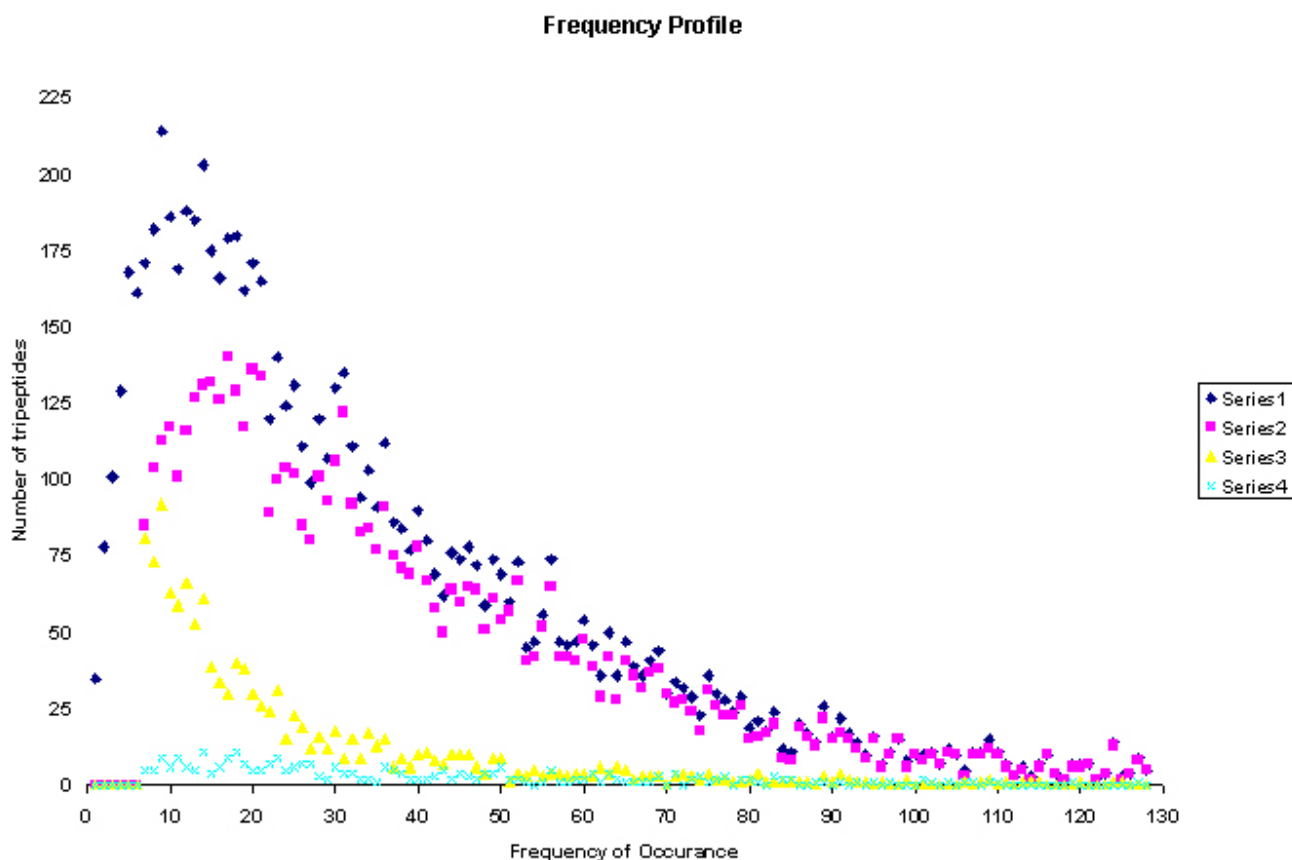
In the graph (Figure 2), Series 1 indicates the frequency profile of all the tripeptides, Series 2 is for intermediate, Series 3 is for rigid and Series 4 is for non-rigid. About 10% of the tripeptides occur with a frequency of less than 6. All the categories except the non-rigid do show that tripeptides that occur very frequently are fewer. They obey Poisson like statistics. Most tripeptides occur with a frequency of 15. Non-rigids are rare and almost each non-rigid tripeptide occurs with a different frequency. We find in our sample size of 0.27 million tripeptides, 12.6% fall into the rigid category, 4% in non-rigids and the remaining in intermediate.

Discussion

Every tripeptide, which occurs in a particular protein and in a particular position, may yield one possible conformation. Examining the entire crystallized data will give the other possible conformations of the tripeptide structures in proteins. When this is statistically analyzed, it will give a clue to the magnitude of conformational fluctuations. It should be borne in mind that in every event, there are particular chemical bonds or steric hindrance, which makes the conformation possible.

The data set that we studied typically has about 0.2°A (R factor) uncertainty in the position of any co-ordinates of the atom. Consequently, most of the nearest neighbour distance data show a standard deviation ranging from 0 to 0.4°A . Next nearest neighbour (R_1, R_3) data will typically have a standard deviation of 0.4°A to 0.7°A . Let us say, we start with a rigid structure in three dimensions given by mutual distances between $C_{\alpha 1}, C_{\beta 1}, C_{\alpha 2}, C_{\beta 2}$ points. We have to find the $C_{\alpha 3}$ and $C_{\beta 3}$ co-ordinates. Eight additional distances with corresponding standard deviations (fluctuations) are given in Table 1 for each of the tripeptides. Many of the distances are actually redundant. Therefore, the best distances are picked to achieve our goal. It can be noted that the nearest neighbour distances have smaller fluctuations, and to fix a point we need the distance of a said point from three other known points. This translates to saying that we should know, at least one next to nearest neighbour distance. We can choose from amongst all the (1,3) distances, the one, which has the lowest fluctuation and thus fix $C_{\alpha 3}$ and then $C_{\beta 3}$ or vice versa. This yields the optimum procedure for fixing the three co-ordinates accurately. In Table 2 we have boldfaced the significant range around the medians within the category. This helps us in demarcating those with fewer fluctuations than the median.

In Table 2 there are occasions where R_1, R_2 or R_2, R_3 distances show a large standard deviation, greater than 0.2°A . These abnormalities are artificial fluctuations in our crystallized data sample. For example, in certain PDB data, the authors have given more than one possible coor-

**Figure 2**

Legend: Series 1 (All tripeptides) Series 2 (Intermediate tripeptides) Series 3 (Rigid tripeptides) Series 4 (Non-rigid tripeptides).

dinate information due to experimental uncertainties in data interpretation. In our analysis, the first choice with its uncertainties has been taken. This is reflected in Table 2. These occasions are rare, and therefore, we have taken them as random errors causing fluctuations.

We discuss some possibilities for the rigidity of certain tripeptides. At the outset, we cannot make any strict criterion for rigidity. However, statistically more often than not the following observations hold. Rigidity due to Proline is well understood because of the side chain interacting covalently with the backbone. Consequently, C_{β} is held rigidly upto *trans* and *cis* ambiguity. This amounts to the fact that essentially ϕ is frozen to $-60^{\circ} \pm 20^{\circ}$ [22]. The tripeptide therefore tends to be rigid. On closer examination, we find Proline in position 3 in the tripeptide, makes the tripeptide rigid. This is in agreement with the expectations from the covalent bond structure.

Methionine and Tryptophan are fairly bulky; perhaps, the good space filling is the cause for rigidity. Rigid tripeptides with Glutamine invariably also have another polar side chain residue; consequently they form a weak ionic bond within the tripeptide. Usually residues with long side chains have more rotameric fluctuations [23]. Occasionally we find they may bind with another residue within the tripeptide and end up being rigid. Lastly, non-rigids tend to have Serine and Threonine residues, which is consistent with their high polarity. Cysteine, which is well known for its tendency to form di-sulfide bridge, can fall into either rigid or non-rigid category but rarely in the intermediate category.

A single tripeptide does not have a unique structure, indeed it varies as the position of the tripeptide changes within a protein and across proteins. This fluctuation is captured by the standard deviations in Table 1. The classification of rigid, non-rigid and intermediate is most often

determined by ($\alpha 1, \alpha 3$) distance alone. Therefore we can assess that two different tripeptides have similar structure, when their ($\alpha 1, \alpha 3$) distance along with their standard deviations overlap. This criterion implies that structurally homologous tripeptides have similar backbones. We have also looked at various other cross co-relations such as ($\alpha 1, \alpha 3$) vs ($\beta 1, \beta 3$) distances, ($\alpha 1, \alpha 3$) vs ($\alpha 1, \beta 3$) distances etc. The structural homology conclusion, as presented in the results remains unaltered.

Conclusions

Tripeptide analysis is shown to be feasible and statistically significant results have been obtained. Some regular features of side chain interactions suggest that there are essentially three classes of tripeptides: rigid, non-rigid and intermediate. These have no correlation with the secondary structures in proteins and hence are complementary to such studies. This data may be used to predict plausible structures for oligopeptides and for denovo protein design. Attempts are being made by us to develop useful pseudo energy functions to realize the above aims.

Methods

A representative data set of 1220 protein structures was obtained from Protein Data Bank PDB [24] with a sequence identity ranging from 0–90% and $\leq 2^\circ \text{A}$ resolution. This PDB list was generated from CulledPDB (R.L.Dunbrack Jr., <http://www.fccc.edu/research/labs/dunbrack/pisces/>), which uses an algorithm similar to the remove-until-done algorithm of Hobohm and Sander [25]. A statistical analysis of 0.27 million tripeptides from this set was done. If the distribution of tripeptides across proteins were random, we would expect each of these tripeptides to occur on an average about 34 times. For each of these 0.27 million tripeptides 12 distances ($\alpha 1, \alpha 2$), ($\alpha 1, \beta 2$), ($\beta 1, \alpha 2$), ($\beta 1, \beta 2$), ($\alpha 1, \alpha 3$), ($\alpha 1, \beta 3$), ($\beta 1, \alpha 3$), ($\beta 1, \beta 3$), ($\alpha 2, \alpha 3$), ($\alpha 2, \beta 3$), ($\beta 2, \alpha 3$), ($\beta 2, \beta 3$) were calculated. The mean and standard deviation was then computed for each of these distances. The frequency of occurrence of each of the 12 distances was also calculated. The objective is to find if there is a conservation of distances in these tripeptides irrespective of where they occur in a protein sequence, or which protein they came from.

As a countercheck for the statistical reliability of our analysis, another representative dataset of 0.15 million tripeptides in 700 proteins with sequence identity ranging from 0–20% and $\leq 2^\circ \text{A}$ resolution was taken and similar computational analysis was done. The quantitative results as reflected in Table 2 remained the same in both the sets. Statistical significance became better in the 0–90% sequence identity set. The standard deviation obtained in the 0–90% set is therefore much more reliable.

In order to determine the structural homologs in the tripeptides, ($\alpha 1, \alpha 3$) mean distances were taken in bins of 0.2°A in a range of 5.0°A to 7.8°A . The number of tripeptides from each of the three categories rigid, non-rigid and intermediate which fall into these 0.2°A bins was counted. The results are tabulated in Table 4. Helix and beta strand subsequences from DSMP [21] were extracted and the frequency of the three classes of tripeptides was computed. The results are tabulated in Table 5.

Supplementary material

The tripeptide distance data is available on the World Wide Web at the following URL: http://www.au-kbc.org/research_areas/bio/projects/protein/tri.html

Authors' contributions

SA did the data analysis, programming and webcasting. PG introduced the subject and background chemistry to RA and SA. The concept was conceived by RA. Development of the concept was done by all the three authors.

Acknowledgements

We thank S.V.Ramanan, G.Baskaran and N.Gautham for their valuable discussions and comments. P.Gautam is grateful to DBT India for support through Department of Biotechnology grant number BT/PR1737/BID/18/015/99. We thank the anonymous referees' for their suggestions, which helped us make the manuscript more contextual.

References

1. Ramachandran GN, Ramakrishnan C and Sasisekharan V **Stereochemistry of polypeptide chain configurations.** *J Mol Biol* 1963, **87**:95-99
2. Miyazawa S and Jernigan RL **Estimation of Effective Inter-Residue Contact Energies in Protein Crystal Structures: Quasi-Chemical Approximation.** *Macromolecules* 1985, **18**:534-552
3. Miyazawa S and Jernigan RL **Residue-Residue Potentials with a Favorable Contact Pair Term and an Unfavorable High Packing Density Term for Simulation and Threading.** *J Mol Biol* 1996, **256**:623-644
4. Singh J and Thornton JM **Atlas of Protein Side-Chain Interactions IRL press 1992, I & II:**
5. Unger T, Harel D, Wherland S and Sussman JL **A 3D building blocks approach to analyzing and predicting structures of proteins.** *Proteins* 1989, **5**:355-373
6. Rooman MJ, Rodriguez J and Wodak SJ **Automatic definition of recurrent local structure motifs in proteins.** *J Mol Biol* 1990, **213**:327-336
7. Rooman MJ, Rodriguez J and Wodak SJ **Relations between protein sequence and structure and their significance.** *J Mol Biol* 1990, **213**:337-350
8. Levitt M **Accurate Modeling of Protein Conformation by Automatic Segment Matching.** *J Mol Biol* 1992, **226**:507-533
9. Fetrow JS, Palumbo MJ and Berg G **Patterns, Structures and Amino Acid Frequencies in Structural Building Blocks, a Protein Secondary Structure Classification Scheme.** *PROTEINS* 1997, **27**:249-271
10. Micheletti C, Seno F and Maritan A **Recurrent Oligomers in Proteins: An Optimal Scheme Reconciling Accurate and Concise Backbone Representations in Automated Folding and Design Studies.** *PROTEINS* 2000, **40**:662-674
11. Simons KT, Bonneau R, Ruczinski I and Baker D **Ab Initio Protein Structure Prediction of CASP III Targets Using ROSETTA.** *PROTEINS* 1999, **Suppl 3**:171-176
12. Bonneau R, Tsai J, Ruczinski I, Chivian D, Strauss CE and Baker D **Rosetta in CASP4: Progress in ab initio protein structure prediction.** *PROTEINS* 2001, **Suppl 5**:119-126

13. Pabo C **Molecular technology: designing proteins and peptides.** *Nature* 1983, **301**:200
14. Quinn TP, Tweedy NB, Williams RW, Richardson JS and Richardson DC **De-novo design, synthesis and characterization of a beta sandwich protein.** *Proc Natl Acad Sci USA* 1994, **91**:8747-8751
15. Johnson MS, Srinivasan N, Sowdhamini R and Blundell TL **Knowledge based protein modeling.** *Crit Rev Biol Mol Biol* 1994, **29**:1-68
16. Fechteler T, Dengler U and Schomburg D **Prediction of protein 3-dimensional structures in insertion and deletion regions—a procedure for searching databases of representative protein fragments using geometric scoring criteria.** *J Mol Biol* 1995, **253**:114-131
17. Bassil Dahiyat I and Stephen Mayo L **De Novo Protein Design: Fully Automated Sequence Selection.** *Science* 1997, **278**:82-87
18. Rohl C and Baker D **Denovo determination of protein backbone structure from residual dipolar couplings using Rosetta.** *J Am Chem Soc* 2002, **124**:2723-2729
19. Bonneau R, Strauss CE, Rohl CA, Chivian D, Bradley P, Malmstrom L, Robertson T and Baker D **De Novo Prediction of Three dimensional Structures for Major Protein Families.** *J Mol Biol* 2002, **322**(1):65-78
20. Bairoch A and Apweiler R **The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000.** *Nucleic Acids Res* 2000, **28**:45-48
21. Guruprasad K, Shivaprasad M and Ranjit Kumar G **Database of Structural Motifs in Proteins (DSMP).** *Bioinformatics* 2000, **16**:372-375
22. Gunasekharan K, Ramakrishnan C and Balaram P **Beta hairpins in proteins revisited: lessons from denovo design.** *Protein Eng* 1997, **10**:1131-1141
23. Dunbrack RL Jr and Karplus M **Backbone dependent Rotamer library for proteins: Application to Side-chain prediction.** *J Mol Biol* 1993, **230**:543-574
24. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN and Bourne PE **The Protein Data Bank.** *Nucleic Acids Res* 2000, **28**:235-242
25. Hobohm U, Scharf M, Schneider R and Sander C **Selection of representative protein data sets.** *Protein Science* 1992, **1**:409-417

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

