Journal of Computational Intelligence in Bioinformatics.
ISSN: 0973-385X Volume 4 Number 1 (2011) pp. 23—31

http://www.ripublication.com/jcib.htm

The Cathelicidins: Structural Analysis of Human Peptide Antibiotic

Amit Kumar Singh[1] and Utkarsh Raj[2]

[1]*Division of Bioinformatics, iLife Discoveries P L, India*
[2]*Department of Biotechnology, Amity University, India*

Abstract

Cathelicidin antimicrobial peptides (CAMP) are a recently discovered family of effector molecules in innate immunity. The protein is itself produced in lysosomes of macrophages & neutrophiles, in inactive form & undergoes numerous post translational modifications. Both N & C terminals are cleaved by signal peptidase & peptidylglycine α-amidating monooxygenase respectively, latter also amidates the C terminal. Finally cathelin domain is removed by the action of elastase, in mammalian system, to make the protein active.

Here full structural analysis of Antibacterial Protein FALL-39 of Cathelicidin family is done which is very potent representative of this family. Structure was predicted by using Multiple Mapping Method with Multiple Templates on M4T Server. Further structure verification was done on Anolea, Gromos, Verify3D and Procheck. Statistical Potential Energy calculated, here, Discrete Optimized Protein Energy (DOPE) Score by M4T Server was -8243.072266 & Protein Structure Analysis (ProSA) Program analyzed z score -5.85 along with per residue energy, with 40 window size, is in range of -0.5 to -1.5. PS00946 & PS00947 are Cathelicidins Protein Signatures identified at both the ends. Protein function analysis was done on PDBsum on the other hand ProMotif predicted 1 sheet, 3 β hairpins, 2 β bulges, 4 strands, 2 helices, 9 β turns & 2 disulphide bridges in the structure.

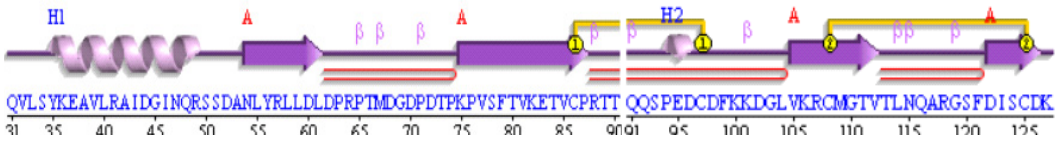
Keywords: Cathelicidin, antimicrobial peptides, Signatures, Homology Modeling, DOPE Score.

Introduction

Small (12/100 amino acids), cationic (net charge of +2 to +9), and amphipathic are some of the properties of Cathelicidins which they share with other antimicrobial

peptides. These proteins are first line of defense and known as “natural antibiotics”. These holoproteins have conserved N-terminal (the cathelin domain) which is cleaved by signal peptidase and the active cationic antimicrobial C-terminal is liberated. This C-terminus is often getting amidated by peptidylglycine α-amidating monooxygenase and cleaves the C-terminal region during degranulation & secretion. Since the peptides are produced in inactive form, so they are stored in granules of circulating immune cells (1). Neutrophil secretory granules are the predominant source of cathelicidins, but they may also be expressed in macrophages, mucosal surfaces in the mouth, lung, and genitourinary tract and in skin keratinocytes in inflammatory disorders. CAMP is positioned at 3p21.3in human genome. This is composed of 4 exons stretched in a length of 2114 nucleotide (2).

Here we have discussed about FALL-39, also known as FA-LL-37, a member of this family. It is a proline/arginine-rich peptide antibiotic, which originates in the bone marrow and in testis (3).



Structural Summary of FALL-39 protein. Diagramdepicts its different secondary & super-secondary structures along with disulfide bridges.

Methods

Protein Structure Prediction

FALL-39 structure was modeled on Multiple Mapping Method with Multiple Templates (M4T) a fully automated modeling server (4, 5).

Structure Validation

The PROCHECK suite of program provides a check on the stereochemistry of a protein structure (6, 7). This suite takes Brookhaven PDB file format as an input file & finally evaluates different parameters for protein structure. These parameters were evaluated by Morris *et al* (8). A [φ,ψ] plot of amino acid residues in protein, first described by Ramchandran & hence thereafter known on the name of legend as Ramchandran plot/map/diagram (9). The conventional plot is a projection of the torus on the plane, resulting in a distorted view and the presence of discontinuities (10).

Atomic Non-Local Environment Assessment (ANOLEA) is a platform to evaluate the "Non- Local Environment" (NLE) of each heavy atom in the peptide chain (11). Discrete Optimized Protein Energy (DOPE) is a statistical potential calculated for a protein which can be viewed as a conformational energy which measures the relative stability of a conformation with respect to other conformations of the same protein (12, 13). The GROMOS is a molecular dynamics computer simulation package for the study of biomolecular systems and was used for the analysis of conformations (14). The QMEN is a composite scoring function for both the estimation of the global

quality of entire models as well as for the local per-residue analysis of different regions within a model (15).

Verify3D is a platform for structure evaluation. It provides a visual analysis of the quality of a putative crystal structure for a protein. This program analyzes the compatibility of an atomic model (3D) with its own amino acid sequence (1D) (16). Protein Structure Analysis (ProSA) Program is easy-to-use interface for the protein structure validation. ProSA calculates an overall quality score for an input structure on three levels z-score, Plot of Residues & Interactive molecule viewer (17).

The z-score indicates overall model quality and measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations. This plot can be used to check whether the z-score of the protein in question is within the range of scores typically found for proteins of similar size belonging to one of these groups (18). The energy plot shows the local model quality by plotting energies as a function of amino acid sequence position (19).

Results & Discussions

Structure of FALL-39 (Accession Number: CAA86115) was modeled on M4T server by using the templates 1kwi_A and 1n5p_A whose alignment is given in figure 4 & best structure was selected on the basis of DOPE Score. DOPE Score calculated was -8243.072266 which is highly reliable. Protein was visualized (Fig 2 A & B) & secondary structure evaluation was done on Yet Another Scientific Artificial Reality Application Software (YASARA) (20). The secondary structure content predicted was 16.2% Helix, 35.2% Sheet, 8.6% Turn and 40.0% Coil. Ramachandran Plot (fig 1) on RAMPAGE provided 98% residues in favored region, only 2% in allowed region & 0% in outlier region. Complete Statistical Analysis was performed on Procheck & results are summarized in table 1. Consensus patterns identified in the FALL-39 are, Y-x-[ED]-x-V-x-[RQ]-A-[LIVMA]-[DQG]-x-[LIVMFY]-N-[EQ] & F-x-[LIVM]-K-E-T-x-C-x(10)-C-x-F-[KR]-[KE], which are actually Signatures of CAMP family. Empirical force field Energy per amino acid calculated by Gromos lies in the range of +10 to -300 kcal/mol. Where amino acid with negative energy, highlighted in green, shows stable conformation & residues in red shows unstable conformation as they have positive energy. Folding quality assessed by Anolea shows the average folding, which can also be concluded as the protein is produced in the form of precursor & hence cleavable regions are unstable (major red sign at C-terminal which is cleaved at last moment & N-terminal slightly stable as that is cleaved on requirement). Comparison of both the results together shows that improperly packed regions also have higher energy & hence are unstable, fig 6. Similar results were obtained on ProSA server. Per residue energy calculated for protein is shown in fig 7 & local model quality also lies in confidence range. Verfiy3D (Fig 3) assessed the 3D-1D compatibility of modeled protein & 87.4% residues have averaged score > 0.2. Z-Score calculated on QMEN is shown in fig 5. ProMotif result is summarized in fig 8 & table 2 for secondary & super secondary structure prediction. Since secondary structures throw light on their structural & functional characteristics, whereas super secondary structures can act as nucleations in the process of protein folding. Different

secondary structures are summarized in detail like in helix; all properties are provided, including pitch, residues per turn, total sequences participating, length & type. Helices are given in two form net & wheels whereas for β-bulges, hairpins & turns residues are positioned in Ramachandran Plot at their specific φ and ψ angles. Cleft analysis was done PDBSum server & visualized on Jmol (21).

Table1: Ramachandran analytical statistics.

	Number of Residues	%
Number of residues in favoured region		~98
Number of residues in allowed region		~2
Number of non glycine and non proline	91	100
Number of end residues	2	
Number of glycine residues	5	
Number of proline residues	7	
Total no. of residues	105	

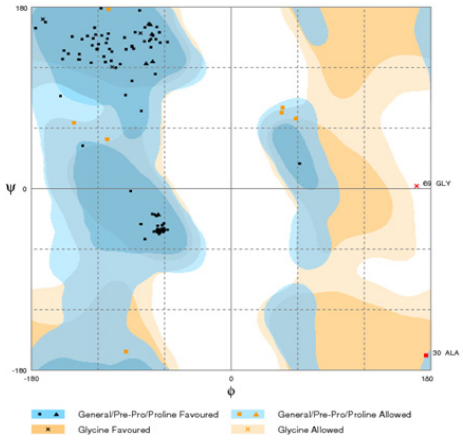


Figure 1: Ramachandra Plot showing the native fold of the protein.

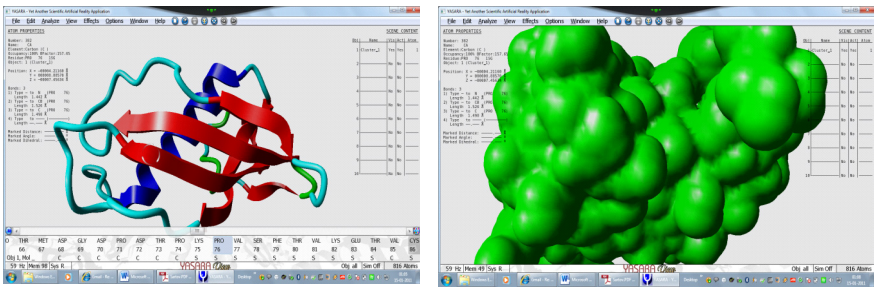


Figure 2:A & B: FALL-39 protein visualized on YASARA Software. Structure shows its helix sheets & coils & turns. Fig 1b shows is solvent accessibility surface of FALL-39 which shows it be very rigid & compact structure.

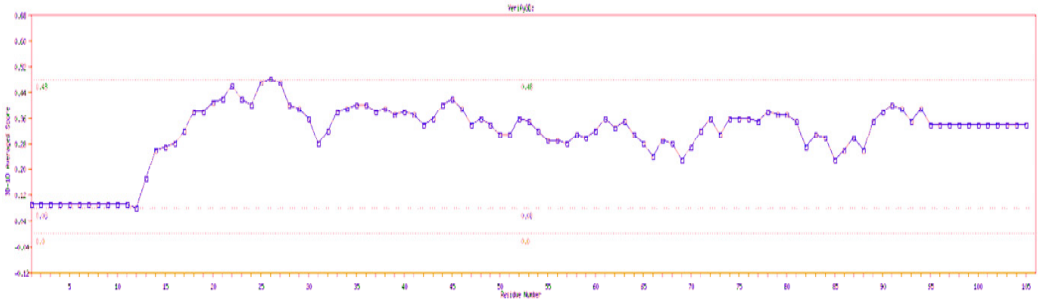


Figure 3: Verify 3D analyzes compatibility of 3D with its own 1D. Since all the atoms are above the threshold value which shows good 3D profile of protein.

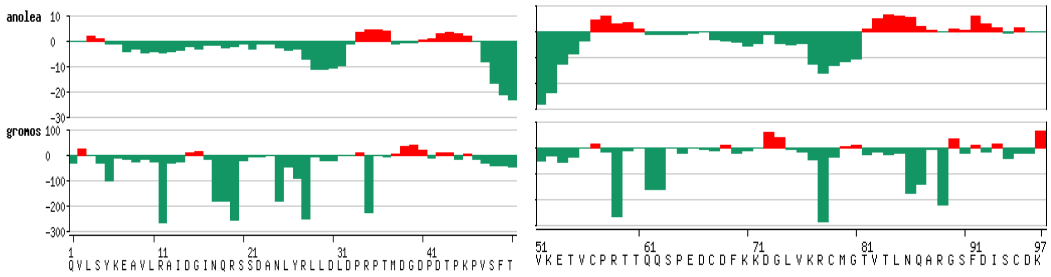


Figure 6: Anolea gives the packing quality of protein. Red sign shows protein is not properly packed at certain positions. Gromos is empirical force field energy per amino acid. Negative energy values (in green) represent favourable energy environment whereas positive values (in red) unfavourable energy environment for a given amino acid.

```
>P1;1kwiA
structureX:1kwi:FIRST :A:LAST :A:undefined:undefined:-1.00:-1.00
-----ALSYREAVLRAVDRLNEQSSEANLYRLLELD---------
GTPKPVSFTVKETVCPRPTRQPPELCDFKE
NGRVKQCVGTVTLD----PLDITCNEVQ--
*
>P1;1n5pA
structureX:1n5p:FIRST :A:LAST :A:undefined:undefined:-1.00:-1.00
GSHMQALSYREAVLRAVDRLNEQSSEANLYRLLELDQPPKADEDPGTPKPVSFTVKETVCPRPTR
QPPELCDFKE
NGRVKQCVGTVTLDQIKDPLDITCNEVQGV
*
>P1;target
sequence:target:     : :     : :::-1.00:-1.00
AIIAQVLSYKEAVLRAIDGINQRSSDANLYRLLDLDPRPTMDGDPDTPKPVSFTVKETVCPRTTQ
QSPEDCDFKK
DGLVKRCMGTVTLNQARGSFDISCDKDNKR
*
```

Figure 4: Alignment of target with multiple templates.

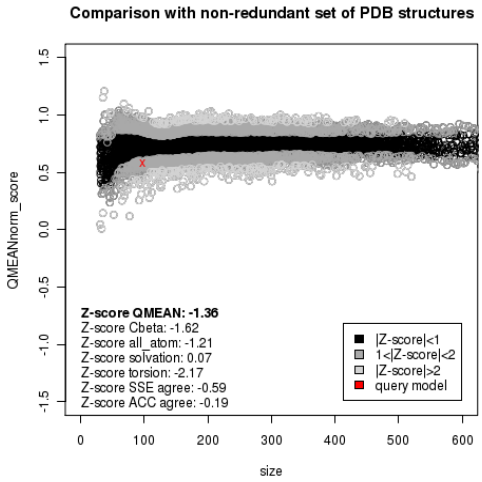


Figure 5: QMEN result, which shows the overall quality of protein. *Nearer the confidence point better is the protein quality*. Here, FALL-39 marked by Red Cross is towards confidence point showing good quality of protein.

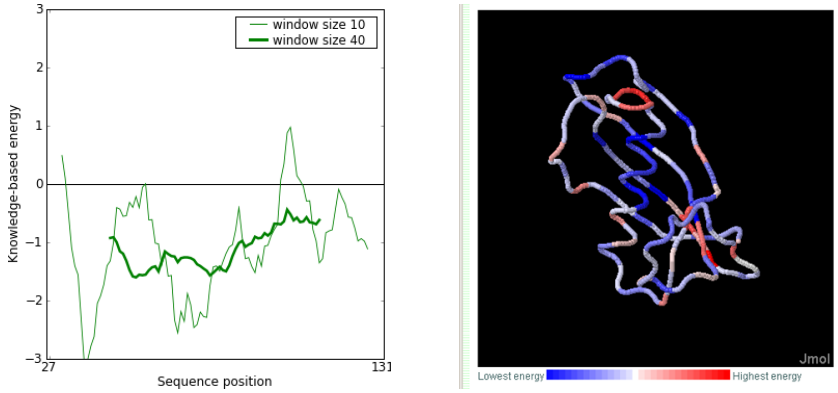
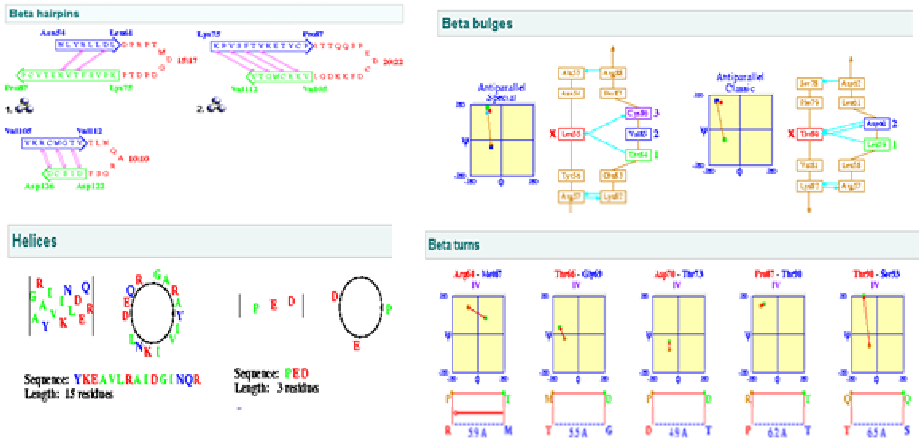


Figure 7: ProSA result. 7 A shows the local model quality & 7 B shows the per residue energy for protein. More towards blue more the stable structure & higher the intensity of red more is the instability.



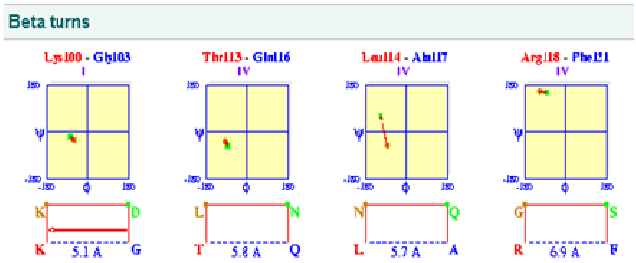


Figure 8: Color postscript diagram generated by PROMOTIF server. Different Super Secondary structures present in FALL-39 protein are plotted on Ramachandran graph & each indicates φ,ψ values for respective residues. Helical nets and helical wheels drawn for helices, here. In both cases, the residues are indicated by their one- letter amino acid codes and color-coded for hydrophobic (green), polar (blue), and charged (red) amino acid types. The N-terminal residue of each helix is at the bottom left-hand corner of the helical net. In the helical wheel, the first residue is indicated by an asterisk and subsequent residues are plotted at 100' intervals, going around the circle in a clockwise fashion.

Acknowledgements

Authors acknowledges the public availability of various servers & softwares for research work with special thanks to YASARA team to provide academic license, Mr. Harjinder Singh Bajwa, for his excellent technical assistance & BCS-insilico team to provide us a platform to present this work.

Table 2: Tabular summary of Secondary & Super secondary structures showing their positions, type, angles & participating residues.

Table of beta hairpins

No.	Strand 1 Start	Strand 1 End	Strand 1 Length	Strand 2 Start	Strand 2 End	Strand 2 Length	Hairpin class
1.	Asn54	Leu61	8	Lys75	Pro87	13	15:17
2.	Lys75	Pro87	13	Val105	Val112	8	20:22
3.	Val105	Val112	8	Asp122	Asp126	5	10:10

Number of beta hairpins in chain : 3

Beta sheets

Sheet	View	No. of strands	Type	Barrel	Topology
A		4	Antiparallel	No	1 1 1

Number of sheets in chain : 1

Beta strands

No.	Start	End	Sheet	No. resid	Edge	Sequence
1.	Asn54	Leu61	A	8	Yes	NLYRLLDL
2.	Lys75	Pro87	A	13	No	KPVSFTVKETVCP
3.	Val105	Val112	A	8	No	VKRCMGTV
4.	Asp122	Asp126	A	5	Yes	DISCD

Number of strands in chain : 4

Disulphides

1st cysteine	2nd cysteine	Type	Chi1	Chi2	Chi3	Chi2p	Chi1p
86	97		-67.3	155.0	-79.8	-134.9	-72.0
108	125	SRH	-65.0	-64.1	101.0	-101.1	-53.8

Number of disulphides in chain : 2

Table of beta bulges

No.	Bulge type	Res X	Res 1	Res 2	Res 3	Res 4
1.	Antiparallel special	Leu55	Thr84	Val85	Cys86	
2.	Antiparallel classic	Thr80	Leu59	Asp60		

Number of beta bulges in chain : 2

Table of helices

No.	Start	End	Type	No. resid	Length	Unit rise	Residues per turn	Pitch	Deviation from ideal	Sequence
1.	Tyr35	Arg49	H	15	22.37	1.46	3.65	5.34	11.1	YKEAVLRAIDGINQR
2.	Pro94	Asp96	G	3	-	-	-	-	-	PED

Number of helices in chain : 2

Table of beta turns

No.	Turn	Sequence*	Turn type	Residue i+1 Phi	Residue i+1 Psi	Residue i+1 Chi1	Residue i+2 Phi	Residue i+2 Psi	Residue i+2 Chi1	i to i+3 CA-dist	H-bond
1.	Arg64-Met67	RPTM	IV	-65.6	126.0	29.2	55.7	73.4	64.1	5.9	Yes
2.	Thr66-Gly69	TMDG	IV	-96.4	-21.4	63.8	-130.2	31.8	-65.6	5.5	
3.	Asp70-Thr73	DPDT	IV	-63.2	-73.3	28.9	-64.7	-36.9	-168.0	4.9	
4.	Pro87-Thr90	PRTT	IV	-115.9	132.3	177.8	-96.4	141.4	65.9	6.2	
5.	Thr90-Ser93	TGQS	IV	-53.5	-55.8	-175.2	-92.6	179.4	-95.5	6.5	
6. *	Lys100-Gly103	KKDG	I	-56.9	-32.2	-172.3	-71.2	-18.7	-161.2	5.1	Yes
7. *	Thr113-Gln116	TLNQ	IV	-92.0	-34.4	-176.9	-80.6	-52.2	-68.1	5.8	
8. *	Leu114-Ala117	LNQA	IV	-80.6	-52.2	-68.1	-112.5	61.0	-177.9	5.7	
9. *	Arg118-Phe121	RGSF	IV	-115.4	155.7	-	-80.3	151.3	80.2	6.9	

Number of beta turns in chain : 9

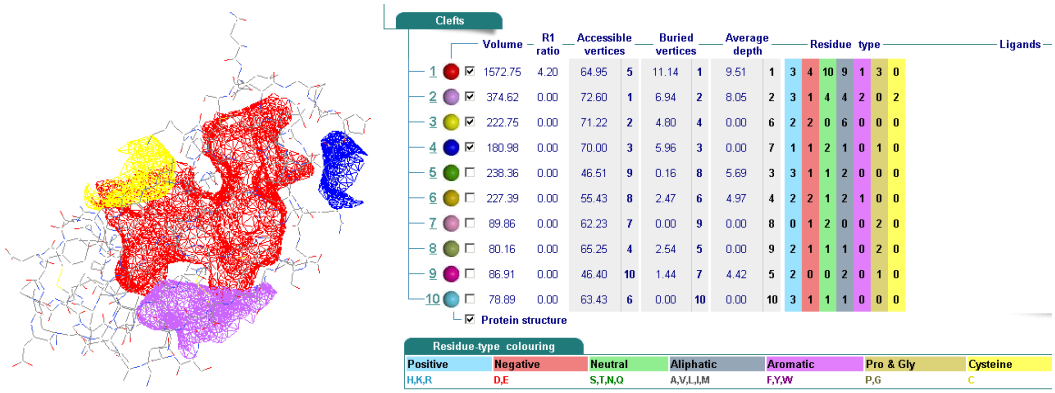


Figure 9: Cleft analysis done on PDBSum. Showing major 4 clefts in the left diagram visualised on Jmol.

References

[1] Håvard Jenssen, Pamela Hamill, and Robert E. W. Hancock. Peptide Antimicrobial Agents. Clinical Microbiology Reviews, Vol. 19, No. 3, p. 491–511, July 2006.

[2] Entrez Gene, NCBI.

[3] Birgitta Agerberth, Hans Gunne, Jakob Odeberg, Per Kogner, Hans G. Boman, and Gudmundur H. Gudmundsson. FALL-39, a putative human peptide antibiotic, is cysteine-free and expressed in bone marrow and testis. Proc. Natl. Acad. Sci. USA, Vol. 92, pp. 195-199, January 1995.

[4] Rykunov D, Steinberger E, Madrid-Aliste CJ, Fiser A., J Struct Funct Genomics. 2009 Mar;10(1):95-9. Epub 2008 Nov 5. Improved scoring function for comparative modeling using the M4T method.

[5] Narcis Fernandez-Fuentes, Carlos J. Madrid-Aliste, Brajesh Kumar Rai, J. Eduardo Fajardo and Andra´ s Fiser, Nucleic Acids Research, 2007, Vol. 35. M4T: a comparative protein structure modeling server.

[6] Laskowski R A, MacArthur M W, Moss D S, Thornton J M (1993). PROCHECK - a program to check the stereochemical quality of protein structures. J. App. Cryst., 26, 283-291.

[7] Laskowski R A, Rullmannn J A, MacArthur M W, Kaptein R, Thornton J M (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR, 8, 477-486.

[8] Morris A L, MacArthur M W, Hutchinson E G & Thornton J M (1992). Stereochemical quality of protein structure coordinates. Proteins, 12, 345-364.

[9] Ramachandran GN, Ramakrishnan C, Sasisekharan V (July 1963). "Stereochemistry of polypeptide chain configurations". J. Mol. Biol. 7: 95–9.

[10] S. Hovmöller, T. Zhou & T. Ohlson (2002) Conformations of amino acids in proteins. In: Acta Cryst. vol. D58, p. 768-776.

[11] Melo, F. and Feytmans, E. (1998) Assessing Protein Structures with a Non-local Atomic Interaction Energy. Journal of Molecular Biology 277, 1141-1152.
[12] M.-y. Shen and A. Sali. Statistical potential for assessment and prediction of protein structures. Protein Science 15, 2507–2524, 2006.
[13] D. Eramian, M.-y. Shen, D. Devos, F. Melo, A. Sali, and M.A. Marti-Renom. A composite score for predicting errors in protein structure models. Protein Science 15, 1653–1666, 2006.
[14] van Gunsteren, W. (1996) Biomolecular Simulations: The GROMOS96 Manual and User Guide. VdF Hochschulverlag ETHZ.
[15] Benkert, P., Tosatto, S.C.E. and Schomburg, D. (2008) QMEAN: A comprehensive scoring function for model quality assessment. Proteins: Structure, Function, and Bioinformatics, 71(1):261-277.
[16] Eisenberg D, Luthy R, Bowie JU. (1997), VERIFY3D: assessment of protein models with three-dimensional profiles, Methods Enzymol., 277, 396-404.
[17] Markus Wiederstein and Manfred J. Sippl, Nucleic Acids Research, 2007, Vol. 35. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins.
[18] Sippl,M.J. (1995) Knowledge-based potentials for proteins. Curr. Opin. Struct. Biol., 5, 229–235.
[19] Sippl, M.J. (1993). Proteins 17, 355-362. Recognition of Errors in Three-Dimensional Structures of Proteins.
[20] Proteins 47, 393-402.
[21] Laskowski R A (2009). PDBsum new things. Nucleic Acids Res., 37, D355-D359.