

Modeling and structural analysis of human Guanine nucleotide-binding protein-like 3, nucleostemin

Farinaz Nazmi^{1,2}, Mohammad Amin Moosavi^{1,3*}, Marveh Rahmati³, Mohammad Ali Hoessinpour-Feizi²

¹Department of Molecular Medicine, Institute of Medical biotechnology, National Institute for genetic Engineering and biotechnology, Tehran, Iran; ²Department of Zoology, Faculty of Natural Science, The University of Tabriz, Iran; ³Hematology and Oncology Research Center, Tabriz University of medical Science, Tabriz, Iran; Mohammad Amin Moosavi - Email: a-moosavi@nigeb.ac.ir; Phone:+98-44787335; Fax:+98-44787195; *Corresponding author

Received June 02 2015; Revised June 28, 2015; Accepted June 29, 2015; Published July 31, 2015

Abstract:

Human *GNL3* (nucleostemin) is a recently discovered nucleolar protein with pivotal functions in maintaining genomic integrity and determining cell fates of various normal and cancerous stem cells. Recent reports suggest that targeting this GTP-binding protein may have therapeutic value in cancer. Although, sequence analyzing revealed that nucleostemin (NS) comprises 5 permuted GTP-binding motifs, a crystal structure for this protein is missing at Protein Data Bank (PDB). Obviously, any attempt for predicting of NS structure can further our knowledge on its functional sites and subsequently designing molecular inhibitors. Herein, we used bioinformatics tools and could model 262 amino acids of NS (132-393 aa). Initial models were built by MODELLER, refined with Scwrl4 program, and validated with ProsA and Jcsc databases as well as PSVS software. Then, the best quality model was chosen for motif and domain analyzing by Pfam, PROSITE and PRINTS. The final model was visualized by vmd program. This predicted model may pave the way for next studies regarding ligand binding states and interaction sites as well as screening of databases for potential inhibitors.

Background:

Human *GNL3* (nucleostemin) has been viewed as a nucleolar protein with variety of roles, including pre-rRNA processing, cell-cycle control, telomere stability, genomic integrity and self-renewal maintaining of embryonic and tissue stem cells [1, 2]. Since its discovery in 2002, there are accumulating reports that nucleostemin (NS) is also abundantly expressed in many cancerous cells, and contributed directly to formation of cancer stem cells (CSCs), highlighting its importance as diagnostic marker and/or therapeutic target in cancer [1, 2]. In this line, we and other groups evidenced that NS depletion can inhibit tumor growth in *in vitro* and *in vivo* and can lead to inhibition of proliferation and induction of cell death [3, 4]. Although its exact mechanism(s) of action is not clear, this nucleolar protein can interact with some important functional proteins in nucleoplasm, such as p53, mouse double minute protein 2 (MDM2), and telomeric repeat binding factor 1 (TRF1), thereby

modulating different fates of the cells [5]. In fact, GTP status of NS is the key factor in its nucleolar-nucleoplasm recycling and interaction with nucleoplasmic proteins [4-5].

The gene encoding NS is a member of a gene family with an MMR_HSR1 domain [6]. The MMR_HSR1 domain consists of five GTP-binding motifs that have been detected in single-celled microorganisms to high vertebrates [2]. Among them, NS and its homologues, guanine nucleotide binding protein-like 3 (*GNL3L*) constitute a subfamily of GTP-binding proteins with a unique domain of circularly permuted GTP-binding motifs [6]. The chromosomal location of human NS is 3p21.1 with 3 typical transcript variants. The first variant encodes a protein with 549 amino acids while variants 2 and 3 missed N-terminal and contain 537 aa [http://www.ncbi.nlm.nih.gov/gene/26354]. Sequence analysis of isoform 1 demonstrated that the encoded protein has one chain consisting basic (B)

domain (amino acid 2 to 46), intermediate (I) domain (amino acid 282 to 456), acidic domain (amino acid 456 to 543) and a coiled coil region (amino acid 56 to 95) [5-7]. Functionally, the nucleolar entry of NS needs GTP binding (G) domain and I domain whereas its accumulation into nucleolus is dependent to B domain [7]. Despite these data, however, there is no crystal

(three-dimensional) structure for NS in the literature, offering an emerging work for predicting its structure by bioinformatics tools. In this study we represented a predicted model for target sequence of NS, particularly its GTP-binding motifs, which may be helpful for better understanding its functional sites and subsequently designing therapeutic drugs.

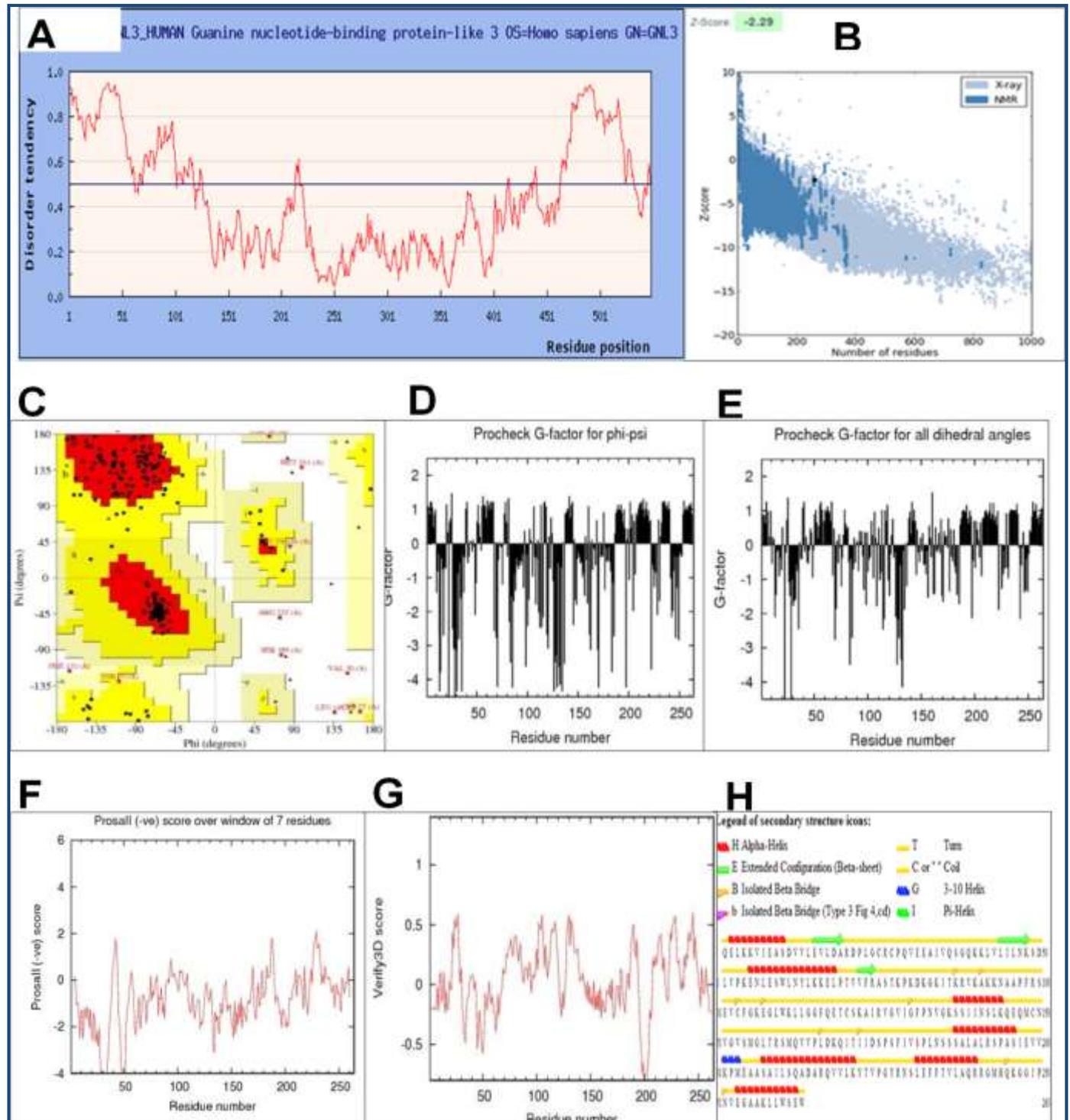


Figure 1: Structural validation of predicted model for NS protein. **A)** Disordered residue prediction. Residues under blue line represent low tendency to be disordered; **(B)** PROSA results for showing Z-score; **(C)** Ramachandran plot of predicted model; **(D-E)** G-factor calculation for phi-psi **(D)** and all dihedral angles **(E)** by PROCHECK. ProsaII **(F)** and Verify3D **(G)** results for modeled residues and secondary structure analysis by STRIDE **(H)**.

Methodology:

Sequence retrieval

The sequence of human *GNL3* in FASTA format was retrieved from Uniprot Knowledge Base (<http://www.uniprot.org/uniprot/Q9BVP2>) with Q9BVP2 accession number [8].

```

sp|Q9BVP2| ----- --ELKKVIE ASDVVLEVD ARDPLGCRCP QVEBAIVQSG
sp|Q9WZM6| MSWYPGHIEK AKRQIKDLLR LVNTVVEVRD ARAPPAT--- --SAYGVDFS
      QKKLVLIINLK SDLVPKENLE SWLNLYLKKEL PTVVFRASFK PKDKGKITKR
      RKETIILLNKK VDIADENTTK NWVEFFKKQG KRVITTHKGE PRK-----
      VKAKKNAAFF RSEVCFGKEG LMKLLGGFQE TCSKAIKRVG VIGFPNVGKSS
      ----- ----VLLKKL SFDRLARVLI VGVENITGKST
      IINSLKQEQM CNVGVSMGLT RSMQVVFLOD QITTIIDSPSF IVSPLNSSSA
      IINKLKGKRA SSVGAQFGIT KGIQWFSLEN GVKILDTPGI LYKNIFSEDL
      LALRSPASIE VVKPNEAASA ILSQADARQV VLKTYVPGYR NSLEFFTULA
      AAKLLLVGSL EVERIEDQRI FERAFELFAR SIGIESS--- -FSEFFEDFA
      QRRGMHQKGG IPNVEGAAKL LKSEW-----
      RKRGLLKKGG VPDIEBALML FFTVEVAQGA GRVSRFERPED ITPVQKQTR
      --
      GV
  
```

Figure 2: Sequence alignment using clustalW.

Disordered residues

Disordered proteins are a kind of protein that lacks a fixed or ordered three dimensional structures and therefore cannot be predicted, so we firstly tried to find which residues of NS can be potentially disordered through a disorder prediction server (<http://iupred.enzim.hu/>).

Sequence alignment and homology modeling

Different blast algorithms (blastp, PSI-blast and PHI-blast) were used against Protein Data Bank (PDB) to choose a suitable template [9]. The template and target sequence were aligned subsequently using ClustalW with the default parameters. Finally, the aligned sequence was used as the input for modeler 9.14 to generate a model.

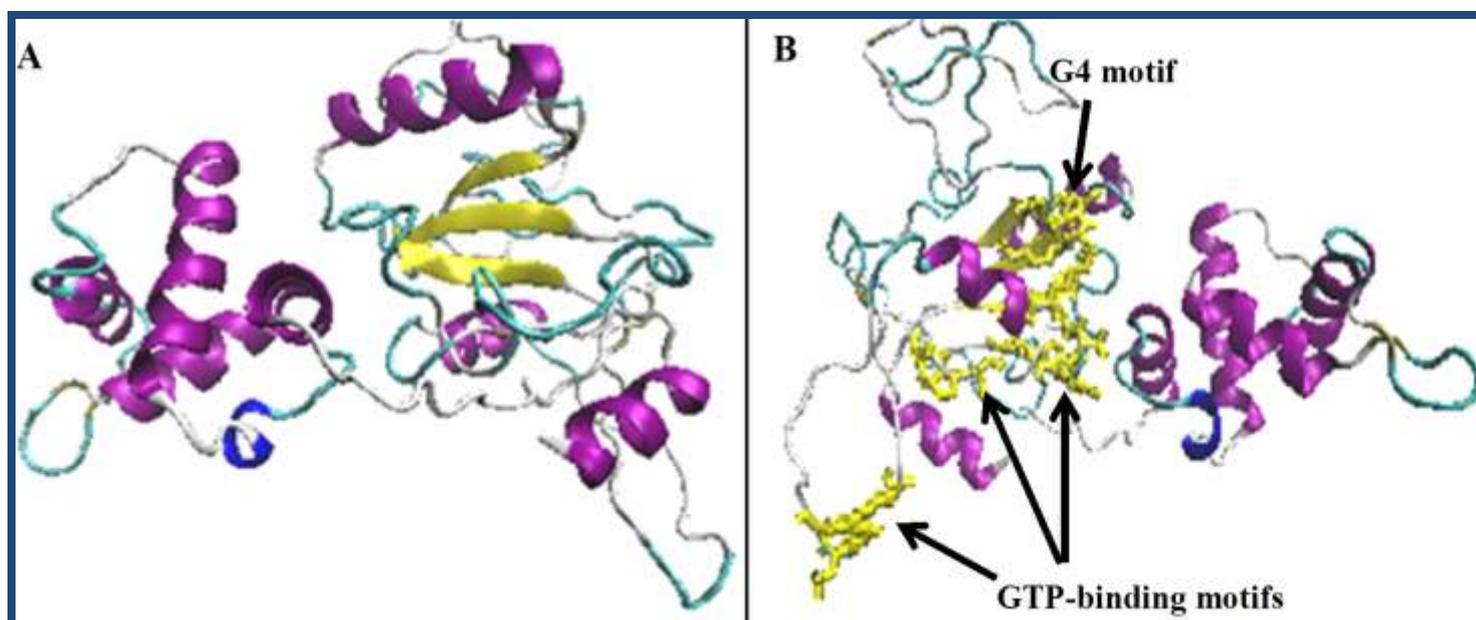


Figure 3: Three dimensional (3D) structure predicted for target sequence of human *GNL3* (A) and GTP-binding motifs (B) by MODELLER and visualized VMD

Model refinement

Predicted models were then refined using Scwrl4 program for prediction of protein side chain conformations. This program is based on a new algorithm and new potential function that result in improved accuracy [10].

Structural validation

The resulted structure was subjected to structural quality assessment. The ProSA program was used to assess the energy of residue-residue interaction using a distance based pair potential and the energy was transformed to a score called Z-score. Also, predicted models were assessed with Jcsg [11] server indexes and the best one was chosen. Moreover, we used PSVS program for model validation that give results in PROCHECK [12], MolProbity [13], Verify3D [14] and ProsaII plots and graphs.

Secondary structure

Stride database was used for secondary structure prediction and computation of α - helical, β - strand and coiled regions [15].

Domain and motif analysis

The Pfam database is a large collection of protein families [16], each represented by multiple sequence alignments and hidden Markov models (HMMs). Pfam results showed one conserved MMR-HSR1 domain in our query sequence. Amino acid residues within a domain that occur consistently and are responsible for specific function are called motifs. They can be used as fingerprints in evolutionary studies and in assigning a recently sequenced protein to a particular family. Fingerprint in our query sequence were found by motif search server PROSITE [17] and PRINTS [18].

Structure visualization

The predicted model is visualized by vmd software [19].

Results & discussion:

Although NS plays an important role in physiological and pathological conditions of stem and cancerous cells, the data on its structure is insufficient [3-7]. In a try to predict structure of NS we engaged bioinformatics tools in this study. The first issue that should be consider before starting the modeling process is sequence evaluation of target protein regarding to its tendency for being disordered. The results in **Figure 1A** showed a low disorder tendency in our target sequence while high tendency was seen in amino acids of 1-131 and 393-549. After evaluation of disordered residues we started modeling process. Different blast algorithms were also performed. Among blast results, the best score belonged to 3CNL chain A which contains 263 amino acids and had 99% coverage with amino acids of 132-393 of NS; there was no structural template with acceptable scores for whole protein at PDB. Although, a secondary structure prediction for whole NS were performed via <http://cho-fas.sourceforge.net/index.php> (data not shown), we decided to predict our model based on amino acids 132-393 largely because any models predicted from disorder regions (1-131 and 393-549 aa) may not reliable. In addition, the functional GTP-binding motifs were found among amino acids of 132-393. Blast results for target sequence showed 86.7 total score, 8e-20 E value and 26% identity, meaning that it is proper for choosing as a template. In the next step, the template and target sequence were aligned using ClustalW with the default parameters (**Figure 2**) and further, the aligned sequence was used as the input for modeler 9.14 to generate ten (10) models. After model refinement, different indexes of models were compared by jcsg database to choose the best one and then validated with Prosa and PSVS software (**Figure 1**). The constructed model is monomer with 263 amino acids; its molecular weight estimated as 28875 Dalton. The Prosa analysis showed that Z-score in the predicted model is -2.29, indicating reasonable side chain interactions (**Figure 1B**). Ramachandran Plot from Procheck evaluation for refined model showed 82.6% residues in allowed regions, 13.0% in additionally allowed regions and 2.6% in disallowed regions (**Figure 1C**). Global quality scores are given in supplementary **Tables 1 & 2** (see **supplementary material**). In addition, PROCHECK G-factor for phi-psi and all dihedral angles was performed (**Figure 1D & E**). With respect to mean and standard deviation for a set of 252 X-ray structures < 500 residues, of resolution <= 1.80 Å, R-factor <= 0.25 and R-free <= 0.28; a positive value indicates a 'better' score. Results showed no close contacts within 2.2 Å, 2.4 ° RMS deviation for bond angles and 0.018 Å for RMS deviation for bond lengths (**Figure 1D & E**). ProsaII and Verify3D results also demonstrated that approximately 60% of residues represent a score over 0.2 which indicate an acceptable model (**Figure 1F & G**). Secondary structure was predicted with Stride database and analyzed with PSVS (**Figure 1H**). As depicted from Figure 1H, the predicted secondary structure contains 7 α - helices, 3 β -strands, a 3- helix and a lot of Turn and Coil structures. In fact, amino acids 2A-10A, 55A-67A, 137A-144A, 187A-195A, 207A-221A, 232A-240A, 253A-262A forms α - helices and 15A-19A, 44A-48A, 72A-74A forms β - strands. By contrast, much less β -sheets were observed in 1-132 and 393-549 regions (data not shown). Finally, the predicted three-dimensional (3D) structure was shown in Figure 3A. Motif and domain analysis showed the conserved MMR-HSR1 domain is located between amino acids 256-370 where order of GTP-binding motifs is G4 (amino

acids 47-50), G5 (amino acids 75-77), G1 (amino acids 131-138), G2 (amino acids 157-161) and G3 (amino acids 175-178) in our model (**Figure 3B**).

GTP binding sites are good therapeutic targets for drug designing against GTP-binding proteins. Therefore, the sites of 5 GTP-binding motifs in the predicted model are highly important and can be considered as valuable targets. Indeed, GTP-binding is the main mechanism that can control the functions of NS by its shuttling between nucleolus and nucleoplasm [5-7]. For instance, it has been reported that NS can enter into nucleoplasm to interact with MDM2 and as a results induce p53-dependent cell-cycle arrest and/or apoptosis [4-7]. To get more insights on predicted model, we served the target sequence to corresponding database (<http://iupred.enzim.hu/>) and predicted disorder and anchor tendency in all residues. Then, we compared given scores for each GTP-binding motif and calculated the average number of score for being disorder and ANCHOR (data not shown). The results presented in Figure S1 showed that G1 and G4 orderly present the lowest scores (0.11 and 0.18) for being disorder while the best ANCHOR score belonged to G4 (0.25). Since a higher ANCHOR score represents a more fixed structure, we proposed that G4 motif may be more reliable target for drug design.

Collectively, this study represented a general view of NS structure. Predicted structural model can fit a variety of applications in future studies including, ligand binding states and interaction sites as well as screening of databases for potential lead compounds or inhibitors. It should be mentioned that our model for target sequence of NS is only a predictive model and need to be confirmed in further studies.

Acknowledgement:

The authors appreciate the financial support of National Institute of Genetic Engineering and Biotechnology and Hematology and Oncology Research Center of Shahid Ghazi Tabatabai Hospital of Tabriz for this study.

References:

- [1] Tsai RY & Meng L, *Int J Biochem Cell Biol.* 2009 **41**: 2122 [PMID: 19501670]
- [2] Tsai RY, *J Cell Sci.* 2014 **127**: 3885 [PMID: 25128565]
- [3] Sijin L *et al. J Exp Clin Cancer Res.* 2004 **23**: 529 [PMID: 15595646]
- [4] Rahmati M *et al. Hematology* 2014 **19**: 455 [PMID: 24628257]
- [5] Yang HX *et al. World J Gastroenterol.* 2005 **11**: 4812 [PMID: 16097049]
- [6] Meng L *et al. Mol Cell Biol.* 2007 **27**: 8670 [PMID: 17923687]
- [7] Leipe DD *et al. J mol Biol.* 2002 **317**: 41 [PMID: 11916378]
- [8] Westbrook J *et al. Nucleic Acids Res.* 2002 **30**: 245 [PMID: 11752306]
- [9] Altschul SF *et al. Nucleic Acids Res.* 1997 **25**: 3389 [PMID: 9254694]
- [10] Krivov GG *et al. Proteins* 2009 **77**: 778 [PMID: 19603484]
- [11] Elsliger MA *et al. Acta Crystallogr Sect F Struct Biol Cryst Commun.* 2010 **66**: 1137 [PMID: 20944202]
- [12] Davis IW *et al. Nucleic Acids Res.* 2004 **32**: W615 [PMID: 15215462]
- [13] Laskowski RA *et al. J Biomol NMR.* 1996 **8**: 477 [PMID: 9008363]

- [14] Eisenberg D *et al.* *Methods Enzymol.* 1997 **277**: 396 [PMID: 9379925]
- [15] Heinig M & Frishman D, *Nucleic Acids Res.* 2004 **32**: W500 [PMID: 15215436]
- [16] Finn RD *et al.* *Nucleic Acids Res.* 2014 **42**: D222 [PMID: 24288371]
- [17] Sigrist CJ *et al.* *Brief Bioinform.* 2002 **3**: 256 [PMID: 12230035]
- [18] Attwood TK, *Brief Bioinform.* 2002 **3**: 252 [PMID: 12230034]
- [19] Humphrey W *et al.* *J Mol Graph.* 1996 **14**: 33 [PMID: 8744570]

Edited by P Kanguane

Citation: Nazmi *et al.* Bioinformation 11(7): 353-358 (2015)

License statement: This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License.

Supplementary material:

Table 1: Ramachandran Plot Summary from Procheck

Most favored regions	Additionally allowed regions	Generously allowed regions	Disallowed regions
82.6%	13.0%	1.7%	2.6%

Table 2: Structure Quality Factors - overall statistics

	Mean score	SD	Z-score ^g
Procheck G-factor ^e (phi / psi only)	-0.47	N/A	-1.53
Procheck G-factor ^e (all dihedral angles)	-0.09	N/A	-0.53
Verify3D	-0.05	0.0000	-6.58
ProsaII (-ve)	-0.43	0.0000	-4.47
MolProbity clashscore	159.65	0.0000	-25.87