

Insilico Comparative Modeling of Lymphoma Causing Nucleophosmin–Anaplastic Lymphoma Kinase Fusion Protein (NPM-ALK) in Homo Sapiens

Santosh Kumar^{1*}, Nem Kumar Jain², Evans Bwalya³, Rozina Dangol⁴, Dhananjay Kashyap⁵

Assistant Professor, Department of Life Sciences, ITM University, Gwalior, M.P., India^{1*}

Assistant Professor, School of Pharmacy, ITM University, Gwalior, M.P., India²

Student, School of Engineering and Technology, Department of Biotechnology, ITM University, Gwalior, M.P., India^{3,4}

Student, Department of Life Sciences, ITM University, Gwalior, M.P., India⁵

ABSTRACT: Nucleophosmin – Anaplastic Lymphoma kinase fusion protein (NPM-ALK) is a chimeric protein. NPM-ALK originates from the fusion of the nucleophosmin (NPM) and the membrane receptor anaplastic lymphoma kinase (ALK) gene. The expression of NPM-ALK oncogene from a chromosomal translocation leads to the fusion of the ALK gene on 2p23 and the NPM gene on 5q35. This fusion yields an 80KD NPM-ALK chimeric protein that shows phototransferase activity and through its interaction with various ALK-adaptor proteins, induces cell transformation and increase cell proliferation. NPM-ALK co-opted several intracellular signal transduction pathways foremost being the STAT3 pathway, normally activated by cytokines from the interleukin-2 (IL-2) family to promote cell proliferation and inhibit apoptosis. In addition NPM-ALK uses epigenetic silencing mechanism to down regulate tumour suppressor genes to maintain its own expression. NPM-ALK protein is involved in the pathogenesis of diverse malignancies including distinct types of lymphoma, lung cancer, carcinoma and neuroblastoma. NPM-ALK serves as the folding member of the ALK fusion protein family and its role in malignant cell transformation is by far the best characterised.

Physiochemical characterization of Nucleophosmin – Anaplastic Lymphoma kinase fusion protein isolated from homo sapiens responsible for lymphoma with 680 amino acids was done to interpret properties like Isoelectric point (pI), EC, AI, GRAVY and instability index. In silico comparative modeling was performed to generate good quality models. The assessment of generated three dimensional structure through Geno3D software against structure verification tools PROCHECK and WHATIF showed that the model generated was stable and acceptable. The predicted model can be used in structure based drug designing.

KEYWORDS: NPM-ALK, Fusion protein, Lymphoma, GENO3D

I. INTRODUCTION

Lymphoma maybe defined as a cancer that begins in the Lymphocytes, which are generally the body's disease fighting cells. These cells are present in organs and tissues such as lymph nodes, thymus, bone marrow and spleen .When lymphocytes grow out of control it becomes known as lymphoma .The protein that is associated with lymphoma is called Nucleophosmin-Anaplastic Lymphoma kinase fusion protein (NPM-ALK).

NPM-ALK is a chimeric protein which belongs to the Anaplastic Large Cells Lymphomas (ALCL). These ALCL are characterised by the fusion of nucleophosmin (NPM) and the membrane receptor Anaplastic lymphoma kinase (ALK) genes [1]. The ALK was first described as a fusion partner in the t (2;5) chromosomal translocation from ALCL from which it takes its name [2]. The NPM-ALK kinase which on dimerization, shows phosphotransferase activity, also through its interaction with various ALK-adapter proteins, enables cell transformation and increases cell proliferation *in vitro* [1]. This cell proliferation when becomes uncontrolled give rise to lymphoma cancer.

Interesting fact, About 50–60% of ALCLs possesses a reciprocal chromosomal translocation, t(2;5)(p23;q35), that fuses the NH₂-terminal portion of the ubiquitously expressed NPM to the entire cytoplasmic portion of the insulin-like receptor ALK tyrosine kinase, thus originating a soluble Mr 80,000 chimeric protein, NPM-ALK [3-6]. The t(2;5)(p23;q35) translocation whose fusion generates NPM-ALK has a transcription under the control of NPM regulatory sites. NPM is a ubiquitously expressed protein that is predominately found in the nucleolus [7]. NPM has many functions and regulates several cellular activities like transcription, ribosome biogenesis, and the shuttling of proteins between the nucleus and cytoplasm. The NPM-ALK fusion gene consists of the first four exons of NPM which encodes for the first 117 amino acids of the NPM protein, and the ALK portion of the fusion includes the exons encoding for the intracellular tail and kinase domain of ALK [2]. Because of an oligomerization motif conserved in the NPM coding sequence, the fusion protein can homo dimerize or associate with wild-type NPM, localizing both in the cytoplasm and nucleus [8-9]. Cytoplasmic, dimerized NPM-ALK was found to interact with and phosphorylate the Shc, IRS-1, and Grb2 ALK-adaptor proteins as well as PLC 1, contributing to the dysregulated growth of the t(2;5) - harboring cell [10-11].

This study was carried out to generate a three dimensional (3D) structure of NPM-ALK protein *in silico* using different bioinformatics tools, databases, websites and programs.

II. MATERIALS AND METHODS

1. **Retrieval of NPM-ALK protein sequence from NCBI:** In this study the sequence of the NPM-ALK protein was retrieved from NCBI. The URL used in the retrieval is <http://www.ncbi.nlm.nih.gov> which directed us to the ncbi website. A protein database was selected and the accession number AAA58698.1 for the protein was put in the search engine of NCBI. The sequence was then changed into FASTA format and copied onto a notepad or word for easy access when it was needed.

```
>AAA58698.1 nucleophosmin-anaplastic lymphoma kinase fusion protein [Homo sapiens]
MEDSMDMDMSPLRPQNYLFGCELKADKDYHFKVDNDENEHQLSLRTVSLGAGAKDELHIVEAEAMNYEGS
PIKVTLATLKMSVQPTVSLGGFEITPPVVLRLKCGSGPVHISGQHLVVYRRKHQELQAMQMELQSPEYKL
SKLRTSTIMTDYNPNYCFAGKTSSISDLKEVPRKNITLIRGLGHGAFGEVYEGQVSGMPNDPSPLQVAVK
TLPEVCSEQDELDFLMEALIISKFNHQIVRCIGVSLQSLPRFILLELMAGGDLKSFLRETRPRPSQPSS
LAMLDLLHVARDIACGCQYLEENHFIHRDIAARNCLLTCGPGRVAKIGDFGMARDIYRASYYRKGCCAM
LPVKWMPPEAFMEGIFTSKTDTSWFGVLLWEIFSLGYMPYPSKSNQEVLEFVTSGGRMDPPKNCPPGVYR
IMTQCWQHQPEDRPNFAILERIEYCTQDPDVINTALPIEYGPLVEEEEKVPVRPKDPEGVPPLLVSQQA
KREEERSPAAPPPLPTSSGKAAKKPTAAEVSVRVPVPRGPAVEGGHVNMAFSQSNPPSELHKVHGSRNKPT
SLWNPTYGSWFTEKPTKKNPIAKKEPHDRGNLGLGEGSCTVPPNVATGRLPGASLLLEPSSLTANMKEVP
LFRLRHFCGNVNYGYQQQLPLEAATAPGAGHYEDTILKSKNSMNQPGP
```

2. **Homology or Comparative Modelling:** Homology or comparative modelling is based on the concept that when two sequences share high similarity/identity, their respective structures are also similar. It is used in the 3D structure prediction of unknown (Target) protein based on one or more related protein of known structure (Templates). Below is a diagrammatic representation of the steps involved in modelling.

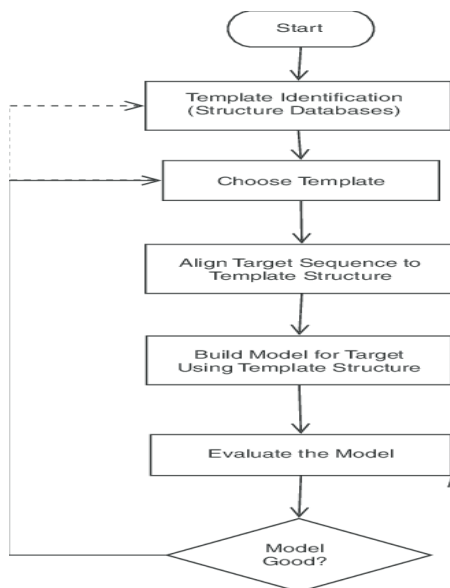


Fig 2.1: Flowchart for the comparative modelling

3. Physico-chemical characterization: The values of theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient [13], instability index [14], aliphatic index [15] and grand average hydropathy (GRAVY) [16] were computed. For physico-chemical characterization, using the Expasy's ProtParam server [12], (<http://us.expasy.org/tools/protparam.html>). The results were shown in Table 1. Atomic Composition of NPM-ALK protein is shown in Table 2.

4. Secondary structure prediction: Secondary structure has been predicted using PHYRE 2 (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>) software where the FASTA format of the sequence was given as input. SOPMA [17] was employed for calculating the secondary structural features of the protein sequence considered for this study. The results were presented in Table 3.

5. Tertiary structure prediction using GENO3D: In order to perform comparative modelling we used a web server known as Geno3D. The server geno3d (<http://geno3d-pbil.ibcp.fr>) is an automatic web server for protein molecular modelling [18]. The strategy which Geno3D uses is comparative protein structure modelling by spatial restraints (distances and dihedral) satisfaction.

III. RESULTS AND DISCUSSION

In the present study, physico-chemical parameters computed using Expasy's ProtParam tool was represented in Table 1. A protein having instability index smaller than 40 is predicted as stable, on the other hand a value above 40 predicts that the protein may be unstable [15]. The Instability index of 51.55 indicates the unstable nature of NPM-ALK protein. The high extinction coefficient (65165) indicates presence of high concentration of Cys, Trp and Tyr. Aliphatic index (76.71) of query protein suggests that the protein may be stable for a wide temperature range. The Atomic composition of Nucleophosmin–Anaplastic Lymphoma kinase fusion protein (NPM-ALK) is Carbon (2103), Hydrogen (3371), Nitrogen (671), Oxygen (677) and Sulphur (4) which is computed by Expasy's ProtParam shown in Table 2. The low value (-0.415) of Grand Average hydropathicity (GRAVY) indicates the possibility of better interaction with water. SOPMA predicts amino acids correctly to describe secondary structure prediction [23]. The results of SOPMA (shown in Table 3) show higher number of random coils in comparison to other secondary structure elements (alpha helix, extended strand and beta turns) and graphs are plotted in Figure 1. Secondary structure and disorder prediction made using phyre 2 software shown in Figure 2 represents that 21.0% of residues were found in the core right handed alpha helices (A), beta sheets (B) and left-handed alpha helix (L) region. 16.0% of residues were found in the allowed right-handed alpha helix (A), beta sheets (B) and left-handed alpha helices regions. The 47.0% of the residues were found in the disordered region. The 4.0% of the residues was found to be TM Helix (Figure 3). Above results indicate to a good quality of predicted model.

The comparative modelling of nucleophosmin-anaplastic lymphoma kinase fusion protein using Geno3d software resulted into three models. Each model has different parameters which include information about whole molecule i.e. Ramachandran plot, deviation, restraints and comparison to templates. The stereo chemical quality of Model two suggests that it is the most stable model having 71.4% in the core region, 25.2% in the allowed region rest in generously allowed and disallowed region. The lower energy value of -11523.50Kcal/mol also indicates about its stability (Fig 4). Finally chosen model which was most stable was visualized by Rasmol software (Figure 5).

Physicochemical Properties:

S. No.	Property	Value
1.	Number of Amino Acid	680
2.	Molecular Weight	75314.22
3.	Theoretical pI	6.44
4.	Total Number of Negatively Charged Residue (Asp + Glu)	77
5.	Total Number of Positively Charged Residue (Arg + Lys)	72
6.	Extinction coefficient	65165
7.	Instability Index	51.55
8.	Aliphatic Index	76.71
9.	Grand Average of Hydropathicity	-0.415

Table 1: Physicochemical parameters computed using Expasy's ProtParam tool

Atomic Composition:

S. No.	Atoms	Number of atoms
1.	Carbon [C]	3335
2.	Hydrogen [H]	5252
3.	Nitrogen [N]	918
4.	Oxygen [O]	993
5.	Sulfur [S]	38

Table 2: Atomic composition computed by ExPASy Protparam tool

SOPMA RESULTS:

S. No.	Property	Value %
1.	Alpha Helix	23.97%
2.	₃ 10 Helix	0.00%
3.	Pi Helix	0.00%
4.	Beta Bridge	0.00%
5.	Extended Strand	16.91%
6.	Beta Turn	5.29%
7.	Bend Region	0.00%
8.	Random Coil	53.82%
9.	Ambiguous States	0.00%
10.	Other States	0.00%

Table: Calculated Secondary Structure elements by using SOPMA

Model	Core	Allowed	Generously	Disallowed	Energy kcal/mol
1	69.3%	24.4%	5.0%	1.3%	-11048.90
2	71.4%	25.2%	2.1%	1.3%	-11523.50
3	67.6%	25.6%	4.6%	2.1%	-11171.20

Table 4: Stereochemical quality of models with procheck

(The table shows the various parameters which associated with the stereochemical quality of the models i.e. core value, allowed and disallowed and energy)

Figures:

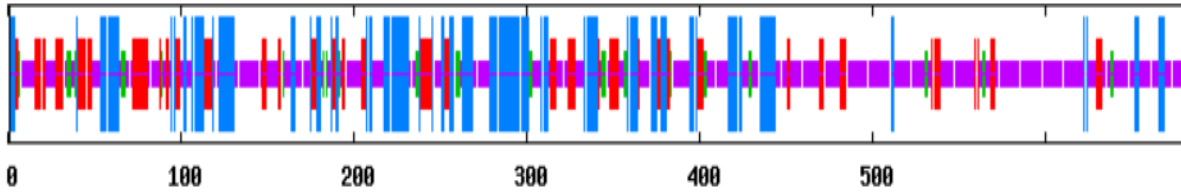


Figure 1: Secondary structure prediction by consensus prediction from multiple alignments

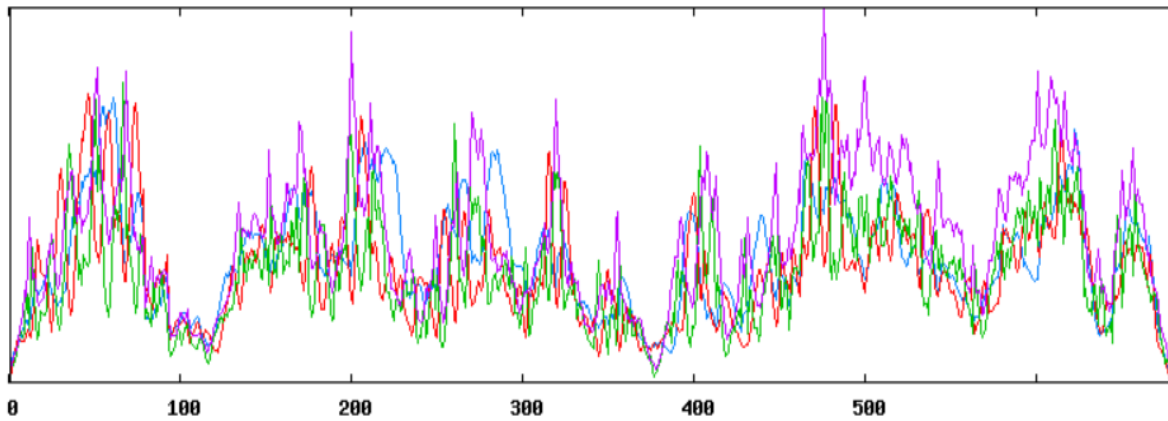


Figure 2: Graphical representation of consensus prediction of secondary structure



Figure 3: PHYRE 2 Secondary Structure prediction

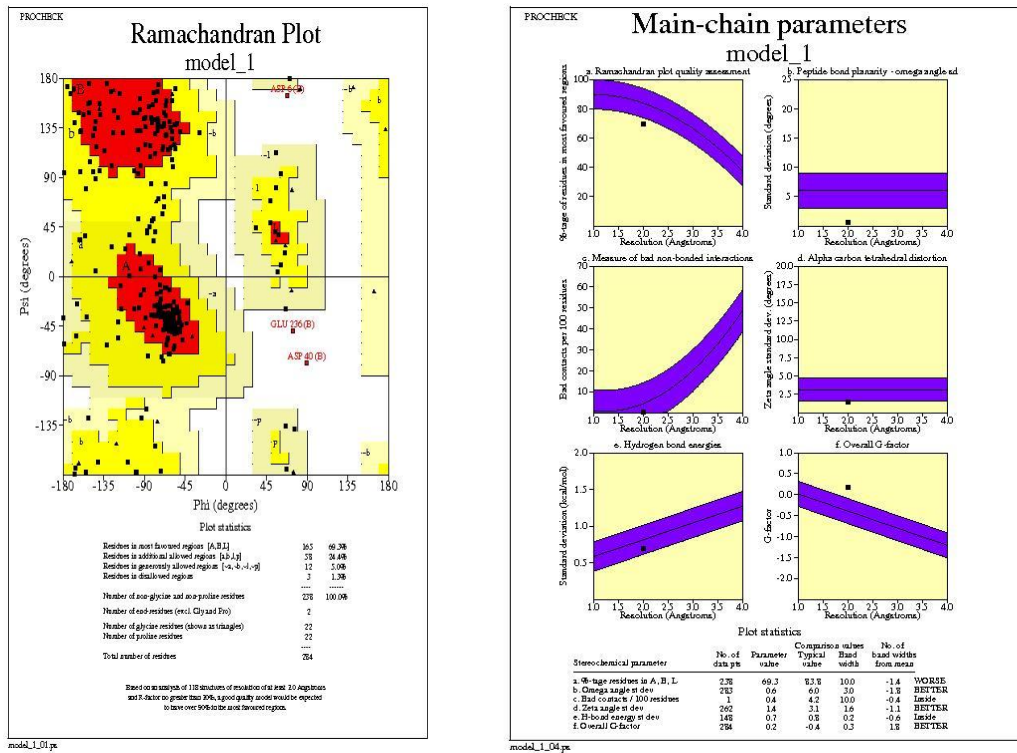


Figure 4: Ramchandran Plot for residue Types

STRUCTURE OF NPM-ALK PROTEIN

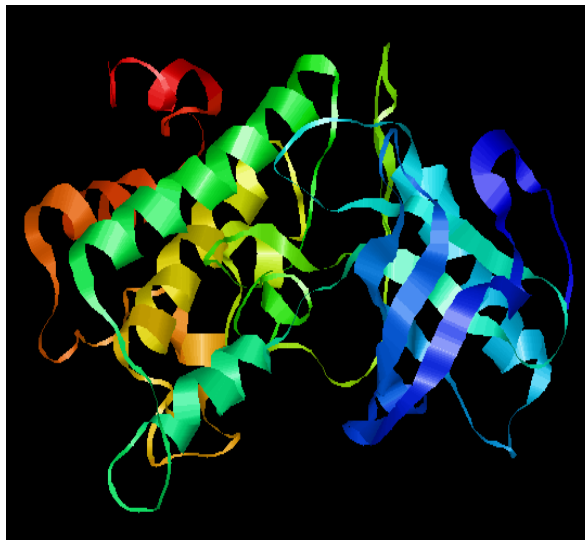


Figure 5: Structure of NPM-ALK protein generated by Geno3D software

IV. CONCLUSION

On the basis of various structural and physiochemical parameters assessment, it can be concluded that the predicted three dimensional structure of protein of NPM-ALK is stable. The Ramchandran plot data showed that nearly 7% of the residues for each protein lies in the core region. The maximum percentage of the residues lying in the core region

represents the best stereo chemical quality of the protein structure which is further supported by its nearly 95% residues lying in the most favoured and allowed regions. Only about 2 % residues lie in the disallowed regions which generally involve steric hindrance between the side chain C-beta methylene group and main chain atoms. Since the percentage residues lying in the disallowed region is much lower than its presence in core and allowed regions the structures generated show best stereochemical quality and stability. The three dimensional structures generated using Geno3D software and validated by PROCHECK software showed the best identity, lowest energy in Kcal/mol and root mean square deviation (RMSD) were considered as the best optimized and superimposed model. One ought to bear in mind that the structure is just a model and thus not experimentally elucidated. Since no effective therapeutic or vaccine is available for lymphoma. Structural information of this model can be effectively used and can be further implemented in future drug designing

V. ACKNOWLEDGEMENTS

We would like to acknowledge ITM University, Gwalior for its timely support and also the groups who supported us directly or indirectly in making this research a success. We would also like to thank the softwares without whose support this work would have not been completed.

REFERENCES

1. Nucleophosmin-Anaplastic Lymphoma Kinase (NPM-ALK), a Novel Hsp90-Client Tyrosine Kinase: Down-Regulation of NPM-ALK Expression and Tyrosine Phosphorylation in ALK CD30 Lymphoma Cells by the Hsp90 Antagonist 17-Allylamino,17-demethoxygeldanamycin1z ..Paolo Bonvini,² Tamara Gastaldi, Brunangelo Falini, and Angelo Rosolen
2. Morris, S. W., Kirstein, M. N., Valentine, M. B., Dittmer, K. G., Shapiro, D. N., Saltman, D. L., and Look, A. T. Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science (Wash. DC)*, 263: 1281–1284, 1994.
3. Morris, S. W., Naeve, C., Mathew, P., James, P. L., Kirstein, M. N., Cui, X., and Witte, D. P. ALK, the chromosome 2 gene locus altered by the t(2;5) in nonHodgkin's lymphoma, encodes a novel neural receptor tyrosine kinase that is highly related to leukocyte tyrosine kinase (LTK). *Oncogene*, 14: 2175–2188, 1997.
4. Wellmann, A., Otsuki, T., Vogelbruch, M., Clark, H. M., Jaffe, E. S., and Raffeld, M. Analysis of the t(2;5)(p23;q35) translocation by reverse transcription-polymerase chain reaction in CD30_____ anaplastic large-cell lymphomas, in other non-Hodgkin's lymphomas of T-cell phenotype, and in Hodgkin's disease. *Blood*, 86: 2321–2328, 1995.
5. Falini, B., Bigerna, B., Fizzotti, M., Pulford, K., Pileri, S. A., Delsol, G., Carbone, A., Paulli, M., Magrini, U., Menestrina, F., Giardini, R., Pilotti, S., Mezzelani, A., Ugolini, B., Billi, M., Pucciarini, A., Pacini, R., Pelicci, P. G., and Flenghi, L. ALK expression defines a distinct group of T/null lymphomas ("ALK lymphomas") with a wide morphological spectrum. *Am. J. Pathol.*, 153: 875–886, 1998
6. Morris, S. W., Xue, L., Ma, Z., and Kinney, M. C. Alk_____ CD30_____ lymphomas: a distinct molecular genetic subtype of non-Hodgkin's lymphoma. *Br. J. Haematol.*, 113: 275–295, 2001
7. H. Umekawa, and M. O. J. Olson, "Expression and subcellular locations of two forms of nucleolar protein B23 in rat tissues and cells," *Cellular and Molecular Biology Research*, vol. 39, no. 1, pp. 33–42, 1993.
8. Bischof, D., Pulford, K., Mason, D. Y., and Morris, S. W. Role of the nucleophosmin (NPM) portion of the non-Hodgkin's lymphoma-associated NPM-anaplastic lymphoma kinase fusion protein in oncogenesis. *Mol. Cell. Biol.*, 17: 2312–2325, 1997. 8. Cordell, J. L.,
9. Pulford, K. A., Bigerna, B., Roncador, G., Banham, A., Colombo, E., Pelicci, P. G., Mason, D. Y., and Falini, B. Detection of normal and chimeric nucleophosmin in human cells. *Blood*, 93: 632–642, 1999
10. Pulford, K., Lamant, L., Morris, S. W., Butler, L. H., Wood, K. M., Stroud, D., Delsol, G., and Mason, D. Y. Detection of anaplastic lymphoma kinase (ALK) and nucleolar protein nucleophosmin (NPM)-ALK proteins in normal and neoplastic cells with the monoclonal antibody ALK1. *Blood*, 89: 1394–404, 1997
11. Mason, D. Y., Pulford, K. A., Bischof, D., Kuefer, M. U., Butler, L. H., Lamant, L., Delsol, G., and Morris, S. W. Nucleolar localization of the nucleophosmin-anaplastic lymphoma kinase is not required for malignant transformation. *Cancer Res.*, 58: 1057–1062, 1998.
12. Gasteiger E (2005) Protein Identification and Analysis Tools on the ExPASy Server. In: John M. Walker, *The Proteomics Protocols Handbook*, Humana Press 571-607.
13. Gill SC, Von Hippel PH (1989) Extinction coefficient. *Anal Biochem* 182: 319-328
14. Guruprasad K, Reddy BVP, Pandit MW (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Prot Eng* 4: 155-16



15. Ikai AJ (1980) Thermo stability and aliphatic index of globular proteins. J Biochem 88: 1895-1898
16. Kyte J, Doolittle RF (1982) A simple method for displaying the hydrophobic character of a protein. J Mol Biol 157: 105- 132
17. Geourjon C, Deléage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 11: 681-684.
18. Combet C, Jambon M, Deleage G, Geourjon C (2002) Geno3D: Automatic comparative molecular modelling of protein. Bioinformatics 18: 213-214