Journal of Advanced Bioinformatics Applications and Research
ISSN 0976-2604.Online ISSN 2278 – 6007
Vol 3, Issue 2, 2012, pp 310-316
http://www.bipublication.com

INSILICO 3D STRUCTURE PREDICTION OF E-CADHERIN PROTEIN: A MOLECULE RESPONSIBLE FOR SUPPRESSING MELANOMA METASTASIS

Babita Rathore[1], Madhvi shakya[2]

[1]Department of Bioinformatics, Maulana Azad National Institute of Technology, Bhopal (M.P) India -462051
[2]Department of Mathematics, Maulana Azad National Institute of Technology, Bhopal (M.P) India -462051
*Correspondence-Email: babita.research@gmail.com

Sub Title: 3D STRUCTURE PREDICTION OF E-CADHERIN

[Received-17/04/2012, Accepted-10/05/2012]

ABSTRACT:

Melanoma is one among the most aggressive type of skin tumor. It is malignancy of pigment producing cells called melanocytes, which occur most predominately in epidermis. The cellular contacts greatly limit the ability of epithelial cells to move or migrate. However, this restriction is lost during the last steps of tumorigenesis. The protein that is downregulated at this stage of tumorigenesis is E –Cadherin, a molecule essential for the establishment of homotpic adhesion. Thus, E-cadherin is the central molecule involved in melanoma metastasis. In the present study *Insilico* 3D-strucure of E-cadherin protein, has been elucidated. The 3D-strucure was constructed through homology modeling. The homology modelling is done by using the modeller 9v7 software. The energy of model was minimized by GROMACS. The root mean square deviation (RMSD) for C atom between the template and homology-modelled structure was estimated by GROMACS. The final model was assessed by PROCHECK and WHATCHECK. In modelled structure helix has dominance over other secondary structure i.e. sheet, coil. The E-cadherin protein satisfies stereochemical restraint and passed all criteria carried out in PROCHECK and WHATIF. E-cadherin showed 92.4% residues in most favored region in Ramachandren plot and Z-score of (-4.629). The overall PROCHECK G-factor for E-cadherin homology modelled structure was (-0.16). The RMSD and RMSF of E-cadherin were in acceptable range. The total energy equivalent was -9.2e+0.5kj/mole (approx.) of E-cadherin. This *insilico* study concluded that constructed final E-cadherin structure model was stable and reliable. The data generated could be used for experimental support and approaches like docking can be implemented for identification of inhibitors.

Keywords: Melanoma, GROMACS, RMSD, RMSF, Ramachandren plot.

[I] INTRODUCTION:

After several decades, Cancer is still burning research area due to its diversity in the sites, origin and patho-physio mechanism. Melanoma is one of the popularly known types of skin cancer. Melanoma is a malignancy of pigment producing cells called melanocytes, which are most predominantly found in epidermis [1]. It may begin in a mole (skin melanoma), but can also begin in other pigmented tissues, like eye or intestines. Melanoma causes the majority (75%) of skin cancer related deaths. The incidence of melanoma has rapidly risen in developed countries during the past several decades [2]. According to National Cancer Institute report, in 2011 there will be approximately 70,230 diagnoses of melanoma and almost 8,700 deaths. Both the incidence of and annual mortality from melanoma are rapidly rising. To date, no therapy has been shown for substantially prolong survival in patients with advanced disease [3].

Melanoma is one among most common "metastatic cancer of unknown primary origin" which may reflect either propensity to arise in unexpected site along the neural crest migratory route, or rapid growth of poorly differentiated lesions arising from indolent or unrecognized cutaneous primary lesion. Cell to cell, probably no human cancer is as aggressive as melanoma. It is handful of cancer whose dimension is reported in millimeter. Tumor thickness approaching 4mm represents high risk of metastasis [4].

In recent insight, melanoma indicated that melanoma cells are refractory to the keratinocytes mediated regulation. The loss of regulatory dominance by keratinocytes occurs in concert with down-regulation of E-cadherin expression in melanoma cells [5-7].

Normal epithelium shows a strong and well-established network of cell-to-cell contact that maintain proper development and the functionality of epithelial structures. These cellular contacts greatly limit the ability of epithelial cells to move or migrate. However, this restriction is lost during the last steps of tumorigenesis. The protein that is downregulated at this stage of tumorigenesis is E-cadherin, a molecule essential for establishment of homotypic adhesion junctions [8-10].

Cadherin is a family of Ca^{2+} dependent adhesion molecules that mediate cell to cell contacts and are mainly expressed in tissues providing a tight control of morphogenesis [11, 12]. Epithelial (E) cadherin, are found in adherens junctions and forming core protein complexes with β-catenin, α-catenin, and P120 catenin (P120 ctn). Both P120-catenin and β-catenin interact with E-cadherin, where α-catenin associates with the complex through its binding to β-catenin and providing a link with actin cytoskeleton [13]. E-cadherin is often found lost in many human cancers including melanoma. Although there is strong evidence that downregulation of E-cadherin during the process of epithelial de-differentiation occurs at the transcriptional level [14-17]. It is reported that repressions of E-cadherin occur due to the transcription factor SNAIL. Cloning and characterization of the human E-cadherin promoter has revealed that two E-boxes present in a proximal fragment of the promoter are important for the repression of this gene observed in tumor cell line and fibroblast [18]. Similar results were also reported in mouse E-cadherin promoter [19]. The two E-boxes characterized so far in the human and mouse E-cadherin promoters have an identical core, consisting of the sequence 5'-CACCTG. This sequence exactly matches the DNA-binding site of the transcription repressed SNAIL [20, 21].

The 3D structure of the E-cadherin protein is predicted by homology method with the help of various computational tools are available from different sources. The various drawbacks associated with wet-lab based procedures include its high cost and non-amenability. *Insilico* approaches provide a better and viable solution to these problems. In this study an effort has been

made to generate 3D model of E-cadherin protein based on the available template from protein data bank.

[II] METHODOLOGY:

2.1. Template Selection for E-cadherin:

E-cadherin protein sequence was obtained from NCBI data bases [22]. E-cadherin protein sequence is 882 amino acids long. Table 1 depicts the protein sequence considered in the study. The E-cadherin protein sequence was retrieved in FASTA format and used for further analysis. BLAST [23] and PSI BLAST through Protein data Bank (PDB) template was selected. The template, with > 40% sequence identity with the E-cadherin sequence was used as a reference structure to build a 3D model for E-cadherin.

2.2. Homology modeling of E-cadherin:

Modeler 9v7 [24] was used for building model. Out of 5 models generated by modeler for E-cadherin protein, the one with the best G-score of PROECHECK [25] was selected.

2.3. Energy refinement and model evaluation:

The constructed model was energy minimized in *GROMACS* force field using Steepest Descent Minimization Algorithms [26]. The validation of structured model was performed by using WHATCHECK [27] and PROCHECK Table 2. The stereo-chemical properties of the 3D model were assessed by PROCHECK and Ramachandran plot was drawn [28]. The structural analysis was performed to generate figure representation with *Discovery studio 3.1*[29].

[III] RESULTS AND DISCUSSION:

E-cadherin protein has been selected for modeling due to non-availability of its 3D-structure as per literature survey. E-cadherin protein sequence (NCBI- GI – 31073) has been obtained from NCBI (www.ncbi.nlm. nih.gov). The BLAST-P [30] at NCBI has been used to detect similar sequences for E-cadherin. A template (3q2v) of mouse has been obtained from protein data bank which is showing 82% identity with E-cadherin. Multiple Sequence alignment of template and E-cadherin has been performed by ClustalW software for modeling [31]. The 3D structure of E-cadherin has been generated by MODELLER 9v7. Query sequence has found to be 82% homology with PDB entry. In modeled structure (Figure 1), helix has dominance over other secondary structure i.e. sheet, coil. Initially generated model has been refined using energy minimization techniques to optimize stereochemistry, remove bumps and steric clashes among non-bonded interactions using the commands of MODELLER9v7 and final refinement of generated 3D-model has been done through GROMACS (OPLS force field) by using steepest descent and conjugate gradient algorithms [32]. The E-cadherin satisfies stereochemical restraints and passed all criteria carried out in PROCHECK and WHATIF. The generated 3D model of E-cadherin protein has been checked by Ramachandran plot (Figure 2) through PROCHECK program of the SAVS metaserver for analysing and validating protein structure [http://nihserver.mbi.ucla.edu/SAVES/]. In E-cadherin modeled structure, 92.4% residues are in most favored region and 0.5% residues are laying in disallowed region in Ramachandran plot. From WHATIF analysis the Z-score average packing quality of the E-cadherin generated model was (-4.629) which shows well packing of modeled structure. The overall PROCHECK G-factor for E-cadherin homology modeled structure was (-0.16). The score indicates that the modelled structure was acceptable as value is greater than the acceptable value (-0.50) [33]. The comparative study of E-cadherin model by WHATCHECK is shown in (Figure 3) where, H stands for helix, S for strand, T for turn and empty space is for coil [34]. The model obeys completely the folding pattern predicted for protein E-cadherin.

3.1. RMSD AND RMSF:

Molecular dynamics simulations have provided significant new information on the nature of protein. RMSD measure the accuracy whereas dynamic fluctuation (RMSF) of proteins of many biological processes such as enzyme activity, macromolecular recognition and complex formation [35].

The changes in structural conformation have been monitored in terms of RMSD and RMSF. Figure 4 shows that protein-protein RMSD and backbone-backbone RMSD has become almost stationary at about 0.13nm in E-cadherin in 200ps simulation time, which is acceptable range. Figure 5 shows RMSF from 0.025-0.075 nm for E-cadherin up to 725 residues only this is because of unfold end of modeled structure. Figure 6 shows the total energy equivalent to be -9.2e+0.5kj/mole (approx.) of E-cadherin.

Solvent accessible surface area (ASA) is a direct measure of interaction of solute and solvent, which in a simple way relate to the hydrophobic energy in the empirical calculations [36].

Figure 7 shows the total ASA (green), hydrophobic (black) and hydrophilic (red). Figure 8 shows radius of gyration (Rg) for E-cadherin protein in the slot up to 175ps the Rg gave the value near 2.15nm but after that time period it stabilized at 2.154nm range.

[IV] CONCLUSION:

In the present study, the structure of E-cadherin, a protein involved in melanoma metastasis has been elucidated. A concise study has been done for identification of sequence through BLAST. Homology modeling has been done for prediction of E-cadherin structure which is shown in Figure 1 and stereochemical checking has been done through Ramachandran plot (Figure 2). The model was also evaluated for comparative study by WHAT CHECK (Figure 3). RMSD, RMSF, total energy, ASA and radius of gyration have been calculated Figure 4, Figure 5, Figure 6, Figure 7 and Figure 8). The 3D structure prediction of protein is important technique in bioinformatics research as it has wider application in fields like drug design and disease prediction. The 3D protein model determination of target protein can be used for rational drug design and their occurrence in other disease could be served for the discovery of broad spectrum of drugs. Computational approaches like docking can be used for identification of inhibitors, which can bind to the targets with experimental or modeled structure. For docking approaches, ligand and receptor protein are necessary. Two approaches are most commonly used in docking. One approach uses matching technique that describes ligand and receptor as complimentary surface. Another approach simulates the actual docking process in which ligand protein pairwise energies are calculated. In present study 3Dmodelling data has been generated for experimental support and could provide a good foundation for functional analysis of experimentally derived crystal structure.

ACKNOWLEDGEMENT:

Authors are highly thankful the Director of institute for providing facilities and humble support during this work

REFERENCES:

1. Owens, D.M., Watt, F.M.[2003] Contribution of stem cells and differentiated cells to epidermal tumours. Nat. Rev. Cancer; 3: 444–451.
2. Berwick, M., Erdei, E., Hay, J.[2009] Melanoma epidemiology and public health. Dermatol. Clin ; 27: 205–214.
3. http://www.cancer.gov/cancertopics/types/melanoma
4. Lynda Chin., Levi A. Garraway., and David E. Fisher [2006] “Malignant melanoma: genetics and therapeutics in the genomic era” Genes Dev; 20: 2149-2182.
5. Tang A, Eller MS, Hara M, Yaar M, Hirohashi S, Gilchrest BA.[1994] “E-cadherin is the major mediator of human melanocyte adhesion to keratinocytes in vitro. *J Cell Sci*; 107:983-992.
6. Cowley GP, Smith MEE. [1996] “Cadherin expression in melanocytic naevi and malignant melanomas. *J Pathol*;179:183-187.
7. Jouneau A, Yu YQ, Pasdar M, Larue L.[2000]“Plasticity of cadherin-catenin expression in

the melanocyte lineage". *Pigment Cell Res*;13:260-272.
8. Vleminckx, K., Vakaet, L. Jr, Mareel, M., Fiers, W. & van Roy, F. [1991] Genetic manipulation of E-cadherin expression by epithelial tumor cells reveals an invasion suppressor role. *Cell*;66: 107–119.
9. Perl, A. K., Wilgenbus, P., Dahl, U., Semb, H. & Christofori, G. [1998] "A causal role for E-cadherin in the transition from adenoma to carcinoma". *Nature*; 392**:** 190–193.
10. Shiozaki, H., Oka, H., Inoue, M., Tamura, S. & Monden, M. [1996] "E-cadherin mediated adhesion system in cancer cells". *Cancer*; 77: 1605–1613.
11. Yagi, T. and Takeichi, M. [2000] Cadherin superfamily genes: functions,genomic organization, and neurologic diversity. Genes Dev; 14: 1169-1180.
12. Gumbiner, B. M. [2005] Regulation of cadherin-mediated adhesion in morphogenesis.*Nat. Rev. Mol. Cell Biol*; 6:622-634.
13. Isabel Molina-Ortiz, Rube´n A. Bartolome´, Pablo Herna´ndez-Varas, Georgina P. Colo, and Joaquin Teixido[2009]Overexpression of E-cadherin on Melanoma Cells InhibitsChemokine-promoted Invasion Involvingp190RhoGAP/p120ctn-ependent Inactivation of RhoA. *the journal of biological chemistry* ; 284: 22:15147–15157.
14. Vleminckx, K., Vakaet, L. Jr, Mareel, M., Fiers, W. and Van Roy, F. [1991] Genetic manipulation of E-cadherin expression by epithelial tumor cells reveals an invasion suppressor role. *Cell*;66: 107-119.
15. Frixen, U. H., Behrens, J., Sacha, M., Eberle, G., Voss, B., Warda, A., Lochner, D., and Birchmeier, W. [1991] E-cadherin-mediated cell-cell adhesion prevents invasiveness of human carcinoma cells. *J. Cell Biol*; 113: 173-185.
16. Navarro, P., Gómez, M., Pizarro, A., Gamallo, C., Quintanilla, M. and Cano, A. [1991] A role for the E-cadherin cell-cell adhesion molecule during tumor progression of mouse epidermal carcinogenesis. *J. Cell Biol*; 115: 517-533.
17. Thiery, J. P. [2002] Epithelial–mesenchymal transitions in tumour progression *Nat. Rew. Cancer*; 2: 442-454.
18. Giroldi, L. A. et al. [1997] Role of E boxes in the repression of E-cadherin expression. Biochem. Biophys. Res. Commun; 241: 453–458.
19. Hennig, G., Lowrick, O., Birchmeier, W. & Behrens, J. [1996] Mechanisms identified in the transcriptional control of epithelial gene expression. J. Biol. Chem; 271: 595–602.
20. Mauhin, V., Lutz, Y., Dennefeld, C. & Alberga, A. [1993] Definition of the DNA-binding site repertoire for the Drosophila transcription factor SNAIL. Nucleic Acids Res; 21: 3951–3957.
21. Nakayama, H, Scott, I. C. & Cross, J. C. [1998] The transition to endoreduplication in trophoblast giant cells is regulated by the mSNA zinc finger transcription factor. Dev. Biol;199: 150–163.
22. http://www.ncbi.nlm.nih.gov/
23. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. [1990] Basic local alignment search tool.J. Mol. Biol;215: 403 [PMID: 2231712]
24. A Sali & TL Blundelll. [1993] Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol.; 2349(3): 779-815 [PMID: 8254673].
25. Laskoswki RA, MacArthur MW, Moss DS, Thorton JM. [1993] PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst;26:283–291. doi: 10.1107/S0021889892009944.
26. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark A, Berendsen H. [2005] GROMACS: fast, flexible, and free. J Comput Chem; 26:1701-18. [PMID: 16211538]
27. Hooft RWW, Vriend G, Sander C, Abola EE. [1963] Errors in protein structures. Nature.1996;381:272–27
28. GN Ramachandran *et al. J. Mol. Biol*;**7**: 95 [PMID:13990617].
29. http://accelrys.com/products/discovery-studio/visualization-download.php
30. Altschul, S.F., Thomas, L.M., Alejandro, A.S., Jinghui, Z., Zheng, Z., Webb, M., David, J.L.[1997] Gapped BLAST and PSI BLAST: a new generation of protein database search programs. Nucleic Acids Res.;25: 17: 3389–3402.
31. JD Thompson *et al. [1994] Nucleic Acids Res*;22: 4673 [PMID:7984417].
32. Jorgensen, W.L., Tirado, R.J., [1988] The OPLS force field for proteins. Energy minimizations for crystals of cyclic peptides and crambin. J. Am. Chem. Soc.;110:1657–1666.
33. Archna Sahay1 & Madhvi Shakya. [2010] Structure Prediction of Dihydroflavonol 4-Reductase and Anthocyanidin Synthase from Spinach. Bioinformation.,;5:6: 001-006.
34. Pallavi Chauhan and Madhvi Shakya., Comparative homology modeling and simulation studies of protein c-jun., IJIB., 2009; 8:25.
35. Bhattacharya, D.K., Clementi, E., Xue, W. [1992] Stochastic dynamic simulation of a protein. Int. J. Quantum Chem.;42: 1397–1408.
36. Lu, B.Z., Chen, W.Z., Wang, C.X., Xu3, X.J. [2002] Protein molecular dynamics with electrostatic force entirely determined by a single

Poisson–Boltzmann calculation. Proteins: Struc. Funct. Genet.; 48: 497–504.

Fig: 1. 3D visualization of Modelled Structure of E-cadherin protein

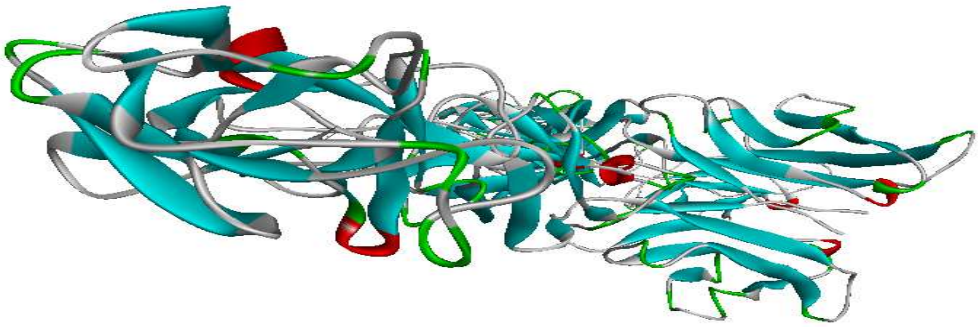

Fig: 2. Torsion angles of φ and ψ in the generated models through Ramachandran plot.

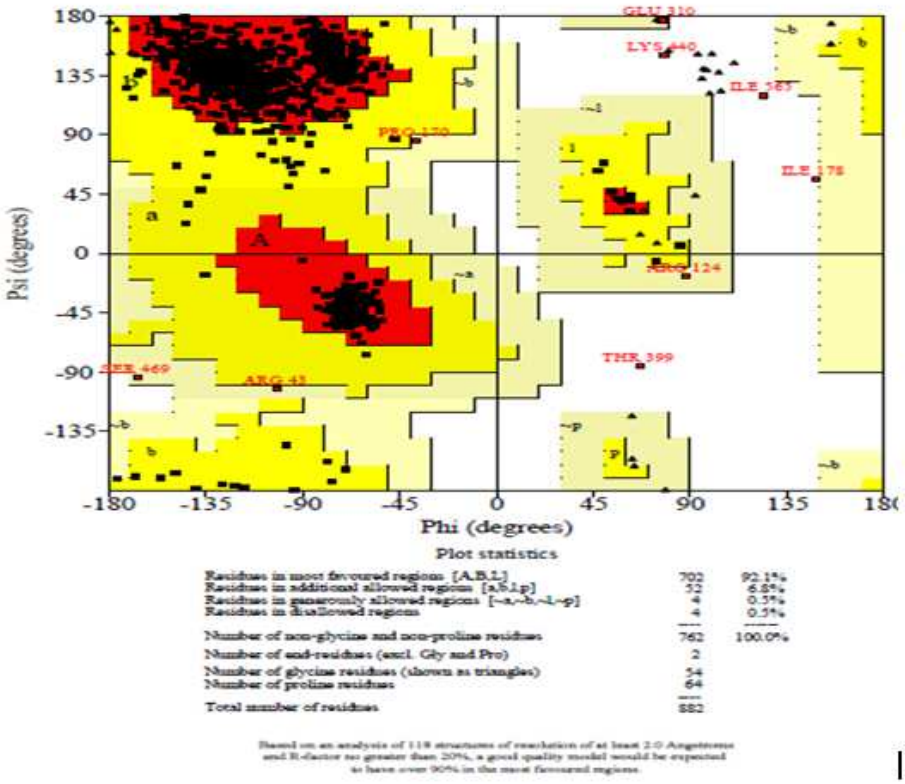

Fig: 3. 2D structure of E-cadherin protein by WHAT CHECK

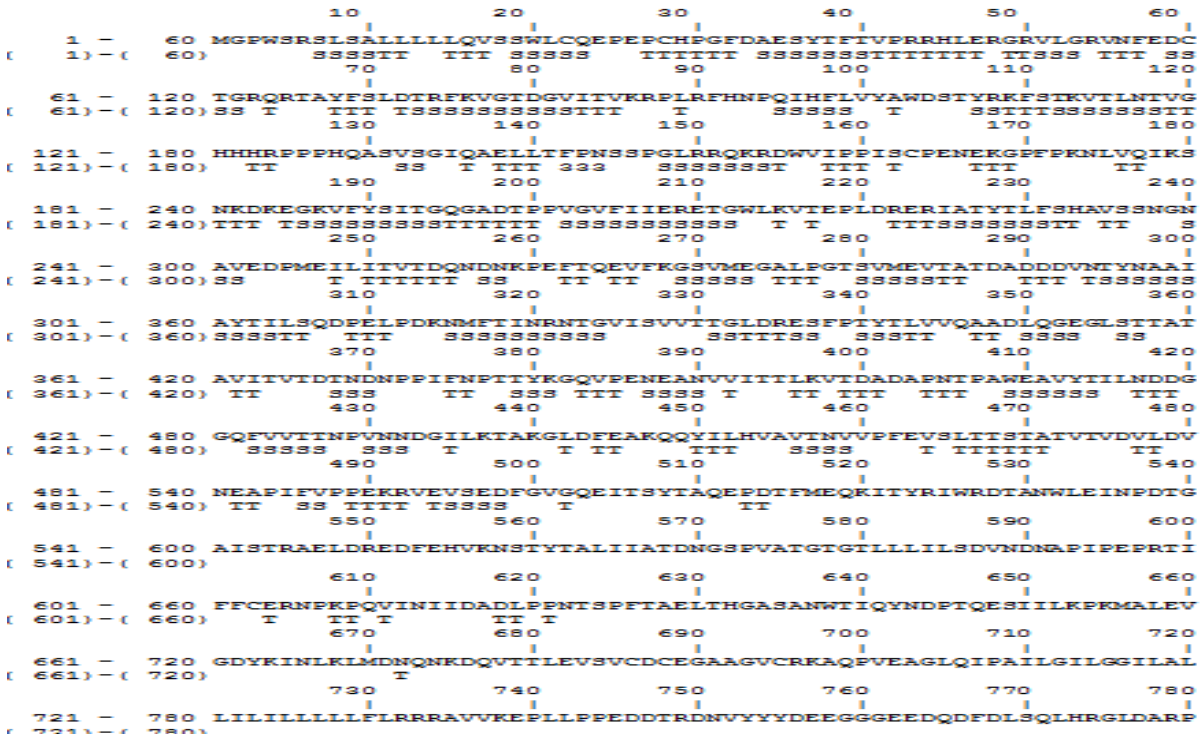

Fig: 4. RMSD of protein after 200ps stimulation

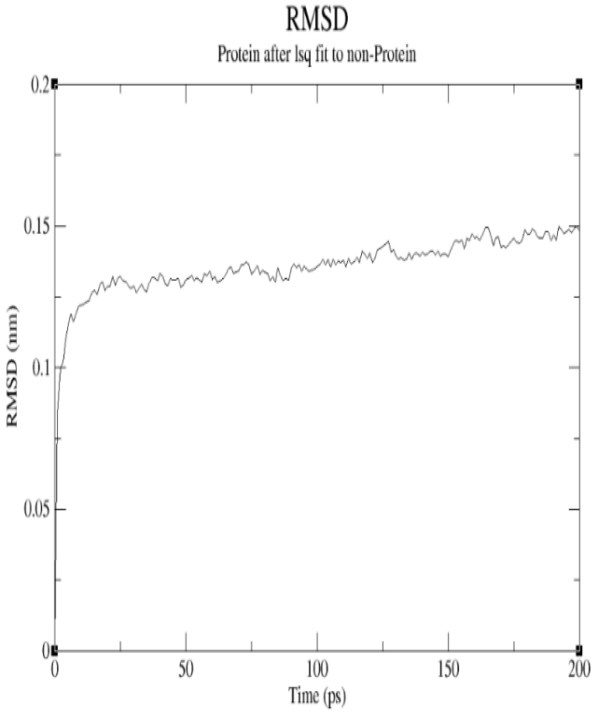

Fig: 5. RMSF of protein after 200ps simulation

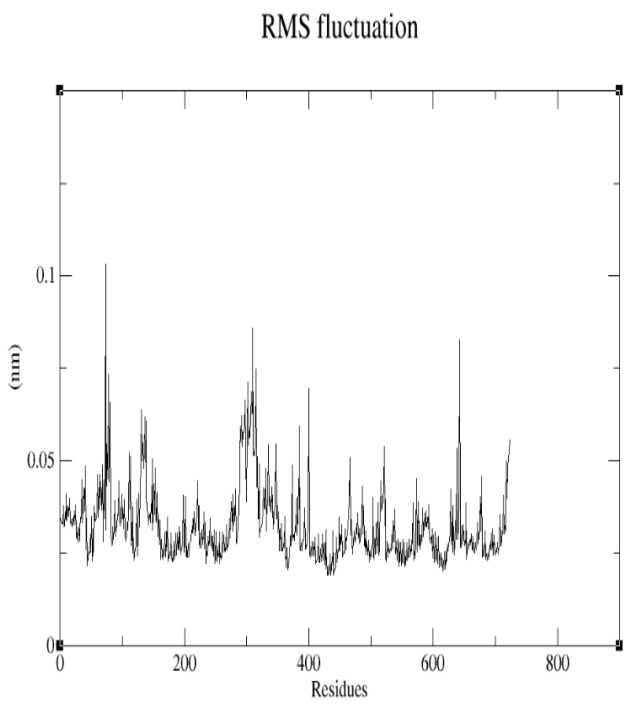

Fig: 6. total energy of protein after 300ps simulation

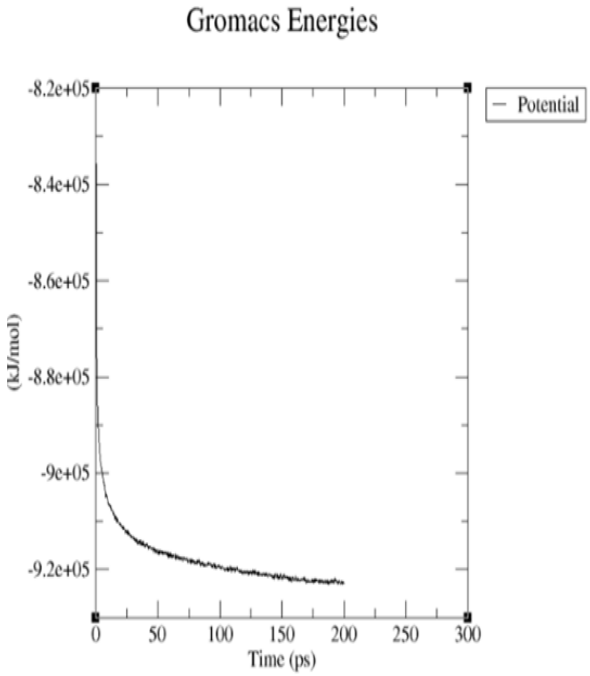

Fig: 7. solvent accessible surface area of modelled protein

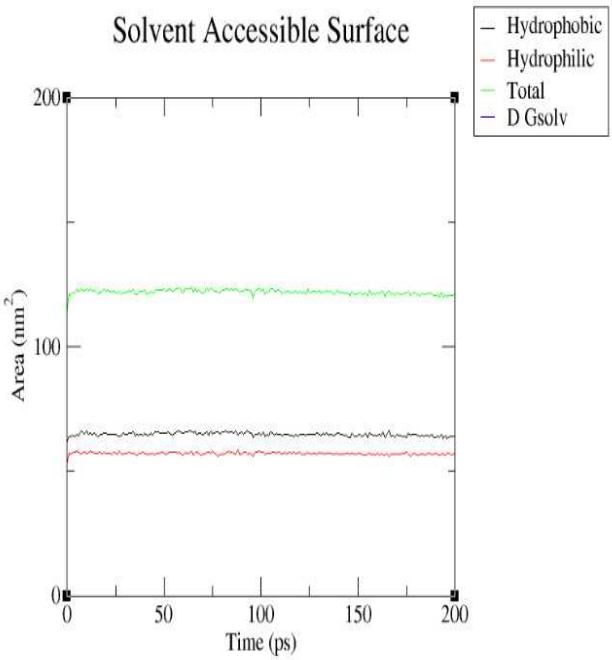

Fig: 8. Radius gyration of E-cadherin protein

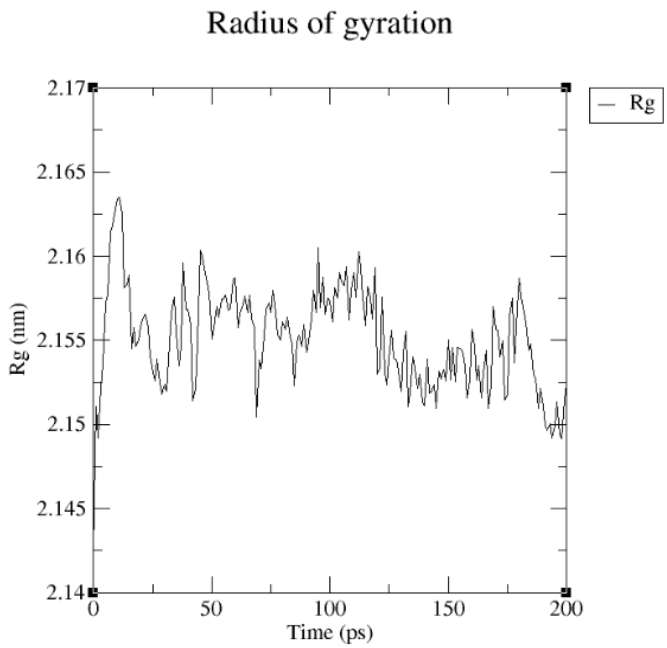

Table 1: Discription of E-cadherin protein

Protein Name	Accession No	Length	Description
E-cadherin	CAA78353	882	Cell adhesion protein

Table 2: analysis of E-cadherin protein by PROCHECK and WHATCHECK

Server	Property	E-cadherin
PROCHECK	Residues in most favored Region	92.1%
	Residues in additionally allowed region	6.8%
	Residues in generously allowed region	0.5%
	Residues in disallowed region	0.5%
WHATCHECK	RMS Z-score for bond angle	1.285