Bioinformation Volume 5

open access

www.bioinformation.net Issue 3

Hypothesis

Structure modeling of novel DNA glycosylase enzyme from oral pathogen *Streptococcus sanguinis*

Ashwani Sharma*, Anshul Nigam

Dept. of Bioscience & Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai-400076 Maharashtra, India; Ashwani Sharma- Email: ashwansharma@gmail.com, Phone: +91-9819338625, +91-9869856722; *Corresponding author

Received June 16, 2010; accepted August 26, 2010; published September 20, 2010

Abstract

The novel 3-methyladenine DNA glycosylase enzyme from oral pathogen *Streptococcus sanguinisin* involves in DNA repair mechanisms and participates in base excision repair. Its 3D structure is still unknown which may be a potential drug target, therefore here we proposed its putative 3D structure by homology modeling approach. EsyPred3d software produced more précised modeled structure as compare to Swiss model software. The modeled structure was further verified by PROCHECK analysis and subjected to functional site prediction servers for active site residues prediction. The functional site was further validated by molecular docking approach with ligand EDA (3- [2- Deoxyribofuranosyl] - 3H- 1, 3, 4, 5A, 8-Pentaaza- Asindacene-5-monophosphate) from 1F4R. The EDR docked at the cavity of modeled structure of 3-methyladenine DNA glycosylase enzyme with highest Patchdock score of 3966 and lowest Autodock 4 docking energy of -10.30 Kcal/mol. The YA51, LA105, RA107 residues are surrounding the EDA and matching with ligand binding residues predicted by PROFUNC server.

Keyword: DNA glycosylase, Homology Modeling, Procheck, Patchdock, Docking, Autodock

Both authors contributed equally.

Background:

Gram positive bacteria *Streptococcus sanguinis* is member of the viridans group of streptococci **[1].** *S. sanguinis* serves as a tether for the attachment of other oral microorganisms that colonize the tooth surface, form dental plaque, and contribute to development of caries and periodontal disease **[2].** Pathogenicity of this organism may not remain limited to oral infections but extended to cause life-threatening endovascular disease infective endocarditis, a serious infection of the valves or lining of the heart **[3].** Like most oral streptococci, this bacterium produces alpha-hemolysis on blood agar, a characteristic linked to the ability of viridians streptococci to oxidize hemoglobin in erythrocytes by secretion of H_2O_2 **[4]**.

Advances in field of biotechnology and bioinformatics had accelerated the progress of medical research in combating such diseases. Recently genome of *Streptococcus sanguinis* had been sequenced **[5].** Therefore a lot of information regarding *Streptococcus sanguinis* cellular machinery and tools of its pathogenicity can be elucidated using wealth of knowledge available in data bases. Here we report an explicit approach to model *Streptococcus sanguinis* putative DNA repair protein 3-methyladenine DNA glycosylase. A potential drug target. The repair mechanism involves replacement of damaged nitrogenous purine and pyrimidine bases **[6].** Furthermore putative DNA repair proteins possess high degree of sequence conservation with prokaryotic genomes **[7].** Therefore due to important role of 3-methyladenine DNA glycosylase enzyme in DNA repair mechanism in *Streptococcus sanguinis*, but lack of its 3D structure motivated us for proposed investigation. The protein sequence for 3-methyladenine DNA glycosylase is available on website http://www.sanguinis.mic.vcu.edu/. In the present work, we will develop the putative 3D structure model of 3-methyladenine DNA glycosylase protein from *Streptococcus sanguinis* by comparative homology modeling method. In addition, we will subject the modeled structure to functional site prediction servers to find putative active site residues which will validate by molecular docking approach.

Methodology:

The genome sequencing of *Streptococcus sanguinis* has been completed by Virginia Commonwealth University (http://www.sanguinis.mic.vcu.edu). We selected protein sequence coding for putative 3-methyladenine DNA glycosylase enzyme. The structure prediction and analysis was done in following steps: (1) Comparative homology modeling by Swiss model software **[8]** and ESyPred3D server **[9]** for obtaining computational 3D model structure for 3-methyladenine DNA glycosylase protein. (2) The energy minimization of modeled structure by GROMOS96 implemented in Swiss model software. (3) Putative functional sites prediction for modeled structure of 3-methyladenine DNA glycosylase enzyme (4) Validation of functional site by molecular docking studies.

Structure Modeling:

The Swiss model is the automated modeling software which develops the 3D structure model of unknown structure protein based on the sequence homology with the known structured protein. It is important to note that for structure prediction, the sequence homology must be higher than 30%. ESyPred3D server predicts the putative 3D modeled structure of 3-methyladenine DNA glycosylase via Modeller (version 6v2) software where it performs the multiple sequence alignments of the query protein sequence (unknown structure) with known structured protein sequences by using different alignment tools such as Matchbox, Clustal W, Dialign and PSI-BLAST etc. In next step, the best alignment subjected to model building. On the other hand, Swiss model software performs the homology modeling and develops the putative 3D model of 3-methyladenine DNA glycosylase. Here we used default parameters for developing the modeled structure. The 3D modeled structure of 3-methyladenine DNA glycosylase so obtained was further analyzed by PROCHECK software.

Energy Minimization via GROMOS96

After obtaining the putative 3D modeled structure of 3-methyladenine DNA glycosylase enzyme, the structure was subjected for energy minimization step in order to get more optimized structure. The

Bioinformation Volume 5

open access

www.bioinformation.net Issue 3

Hypothesis

minimization has been done by GROMOS96 force field implemented in Swiss model software. The GROMOS96 helps in minimization of bond stretch energy of the modeled protein. It incorporates both bonded and non bonded form of energy occupied in the protein molecule.

Functional site prediction

The minimized structure then subjected to different functional site prediction servers PINTS **[10],** PROFUNC **[11],** FIRESTAR **[12]** and Q-SITE FINDER **[13]** in order to obtain putative active site residues. These servers predict the putative active site residues which may act as ligand binding site or reaction site.

Validation of functional site

The functional sites finding was further validated by molecular docking via PATCHDOCK **[14]** and AUTODOCK4 **[15]** softwars (see **Figure 1** for over all methodology). The ligand was extracted from the template proteins, detected in BLAST with lower e value and template pdb matched by Swiss model. The whole modeled protein was taken as centre (without location of any specific amino acid residue) and generated the grid map. The affinity between the ligand and the modeled protein was calculated by measuring the score for PATCHDOCK and docking energy analysis in AUTODOCK4. All the parameters were set to be default in both the docking process.

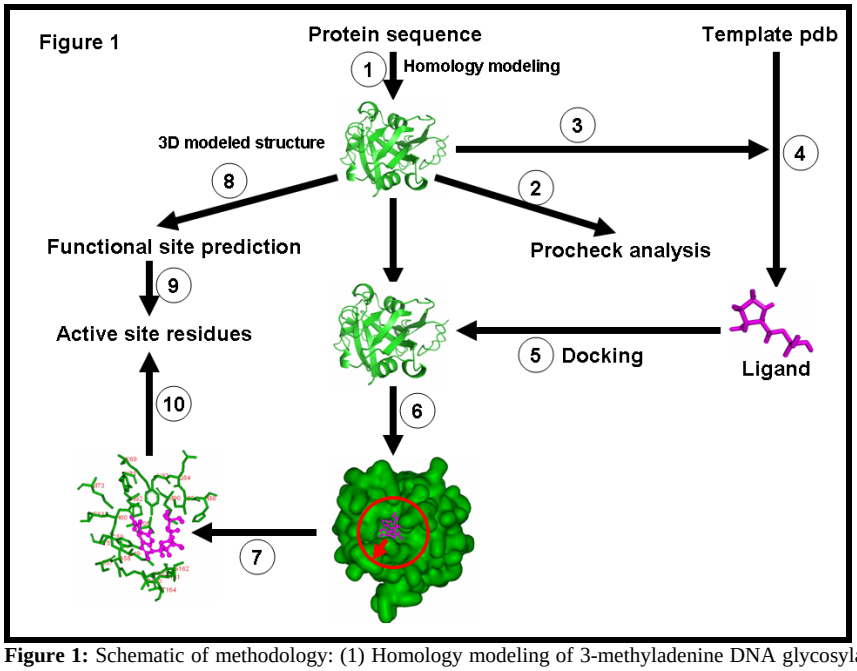

Figure 1: Schematic of methodology: (1) Homology modeling of 3-methyladenine DNA glycosylase protein from *Streptococcus sanguinis*.(2) Procheck analysis of modeled structure (3) Matching with selected template pdb (4)Extraction of ligand from template pdb (5) Screening of ligand against modeled protein (6)Binding analysis of ligand on modeled protein (7) Residues content at $6A^0$ of radius with ligand as center (8)Passed modeled structure through functional site prediction servers (9)Predicted functional sites from servers (10)Matching of putative Active site residues from step 7 and step 9.

Figure 2: Modeled structure of 3-methyladenine DNA glycosylase from *Streptococcus sanguinis* by ESyPred3D.

Bioinformation Volume 5

open access

www.bioinformation.net Issue 3

Hypothesis

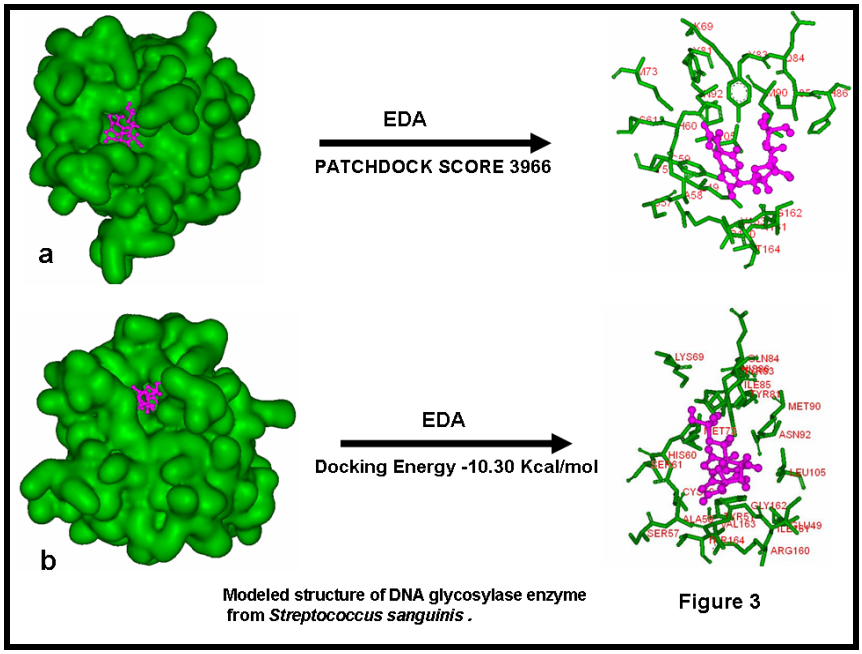

Figure 3: Docking analysis of EDA on modeled 3-methyladenine DNA glycosylase from *Streptococcus sanguinis* : EDA bound at the cavity of modeled 3-methyladenine DNA glycosylase with (a) Patchdock score of 3966 and (b) lowest Autodock4 docking energy of -10.30 Kcal/mol.

Results:

The 3D modeled structure for 3-methyladenine DNA glycosylase protein from *Streptococcus sanguinis* has been generated by Swiss model and ESyPred3D softwares. The Swiss model selected known 3-methyladenine DNA glycosylase protein (PDB ID: 1BNK (A)) as template pdb from *Homo sapiens* for generating *the* putative 3D modeled structure for 3-methyladenine DNA glycosylase protein. . The target-template showed sequence identity of 34.50 % with e-value 0.00e-1. On the other hand ESyPred3D searched known DNA glycosylase protein (PDB ID: 1F4R (A)) as template pdb from *Homo sapiens*. We subjected both the modeled structure for PROCHECK analysis. The Ramachandran plot analysis revealed that 89.0% of amino acid residues from modeled structure generated by ESyPred3D via Modeller (6v2) are incorporated in the favored regions (A, B, and L) of the plot (**Table 1, see supplementary material**). Apart from that 7.1% of residues are in allowed regions (a, b, l, and p) of the plot. On the other hand, modeled structure by Swiss model projected 81.8% of amino acid residues in favored regions (A,B,L) of the Ramachandran plot and 12.3% are in allowed regions (a,b,l,p) of the plot. This analysis concluded that ESyPred3D software more accurately predicted the 3D modeled structure of 3-methyladenine DNA glycosylase (**Figure 2**) protein as compare to Swiss model. Therefore we selected ESyPred3D modeled structure for further analysis. Individual study of all 20 amino acid residues distribution on Ramachandran plot revealed that most of the amino acid residues located in the shaded area (favored regions) of plot. Some residues covered the unflavored regions of plot such as AA97, DA167, CA165, QA40, GA18 and TA164. The Ramachandran plot quality assessment analysis showed that at $2.0A^0$ the most residues are above 90% (favoured+allowed) regions and bad contacts are 5 residues per 100 residues. The modeled structure was further subjected to energy minimization by GROMOS96 program, implemented in swiss model software. The modeled structure was stabilized from initial energy of + 63.012 KJ/mol to final minimized energy of -6509.863 KJ/mol. The BLAST and PSI-BLAST search from pdb database searched 1F4R (A), 1BNK (A) and 1EWN (A) as template proteins with score 88.6-87.4 and e-value of 1e-18 to 3e-18. Dali also produced greater structure homology with 1F4R (Z=33.7, rmsd =0.7), 1EWN (Z=33.1, rmsd=0.7), 1F60 (Z=31.9, rmsd=1.1) and 1BNK (Z=31.6, rmsd=1.0). Functional site prediction servers detected the putative functional site residues in modeled structure of 3-methyladenine DNA glycosylase protein. Q-site finder also found largest cavity on modeled structure with volume of 229 cubic $A^{0.}$ (**Table 2, see supplementary material**).

Functional site finding was further validated by Patchdock and Autodock 4.0 software. Here the ligand molecules were extracted from matched known DNA glycosylase template proteins (1F4R, 1BNK, 1EWN) and their 3D structure generated by CORNIA server. These ligands were screened against Modeled structure of 3-methyladenine DNA glycosylase via Patchdock and Autodock 4.0. Note that the whole modeled structure was taken as docking target (Blind docking). The docking analysis revealed that the ligand EDA (3- [2- Deoxyribofuranosyl]- 3H- 1,3,4,5A,8-Pentaaza- Asindacene-5-monophosphate) bound at the cavity of Modeled structure with highest Patchdock score of 3966 and lowest docking energy of -10.30 Kcal/mol and containing the following residues EA49, YA51, SA57, AA58, CA59, HA60, SA61, KA69, MA73, YA81, YA83, QA84, IA85, HA86, MA90, NA92, LA105, RA107, RA160, IA161, GA162, VA163, TA164 at $6A^0$ of radius (**Figure 3**). The residues YA51, LA105, RA107 are complementary to the residues predicted by PROFUNC functional site prediction server at ligand binding site domain finding. Firestar also confirmed our finding of functional sites residues.

Discussion and Conclusion:

Here we report 3 D model of novel DNA repair protein 3-methyladenine DNA glycosylase from *Streptococcus sanguinis* whose 3D structure is still unknown using homology modeling. The knowledge gained about the structure of DNA repair protein 3-methyladenine DNA glycosylase from *treptococcus sanguinis* may be helpful in discovering drugs against this pathogen. The modeled structure by ESyPred3D (Modeller 6v2) showed high accuracy as compare to structure from Swiss model. The structure was further verified by PROCHECK. The energy minimization via GROMOS96 produced optimized structure for the modeled structure.

The novel DNA repair protein 3-methyladenine DNA glycosylase is playing important/vital role in survival of oral pathogen *Streptococcus sanguinis* in humans. The prediction of modeled structure for novel protein DNA glycosylase from oral pathogen *Streptococcus sanguinis* may provide greater insight for understanding the structure similarity with DNA glycosylase of other organisms. We have also predicted ligand binding sites in modeled structure of DNA glycosylase and also validated by docking method which may be useful for biologist to understand specific role of functional site residues during DNA repair mechanisms. The functional site finding also implicated role in structure based drug

Bioinformation Volume 5

open access

www.bioinformation.net Issue 3 **Hypothesis**

designing against DNA glycosylase protein of oral pathogen Streptococcus *sanguinis.*

Acknowledgement:
AS and AN sincerely acknowledge CSIR and AICTE respectively for fellowship.

Reference List

[1] J Kreth *et al.* Journal *of Bacteriolog.* **190:**4632 (2008) [PMID: 18441055].
[2] M Yamaguchi *et al. Microbes and Infectio.* **8**:2791 (2006) [PMID : 17045503].
[3] LS Turner *et al. Infection and Immunity.* **77:**4966 (2009) [PMID : 19703977].
[4] JP Barnard *et al. Infection and Immunity.* **64:**3853 (1996) [PMID : 8751938].
[5] P Xu *et al. Journal of Bacteriology*. **189:**3166 (2007) [PMID : 17277061].
[6] EJ O'Rourke *et al. Journal of Biological Chemistry*. **275:**20077 (2000) [PMID : 10777493].
[7] KG Berdal *et al. EMBO Journal*. **9:**4563 (1990) [PMID : 2265619].
[8] F Kiefer *et al. Nucleic Acids Research.* **37:**D387 (2009) [PMID : 18931379].
[9] C Lambert *et al. Bioinformatics.* **18:**1250 (2002) [PMID : 12217917].
[10] A Stark *et al. Nucleic Acids Research.* **31:**3341(2003) [PMID : 12824322].
[11] RA Laskowski *et al. Nucleic Acids Research.* **33:**W89 (2005) [PMID : 15980588].
[12] G Loüpez *et al. Nucleic Acids Research.* **35:**W573 (2007) [PMID : 17584799].
[13] ATR Laurie & RM Jackson, *Bioinformatics.* **21:**1908 (2005) [PMID : 15701681] .
[14] D Schneidman-Duhovny *et al. Nucleic Acids Research.* **33:**W363 (2005) [PMID : 15980490].
[15] GM Morris *et al. Journal of Computational Chemistry*. **19:**1639 (1998).

Edited by P. Kangueane
Citation: Sharma *et al.* Bioinformation 5(3): 136-140 (2010)

Bioinformation Volume 5

open access

www.bioinformation.net Issue 3

Hypothesis

Supplementary material:

Table 1: Comparison of Ramachandran plot statistics for modeled structure of 3-methyladenine DNA glycosylase from ESyPred3D (modeler 6v2) and Swiss model software.

Properties	EsyPred3D		Swiss model	
Residues in most favoured regions [A,B,L]	138	89.0%	126	81.8%
Residues in additional allowed regions [a,b,l,p]	11	7.1%	19	12.3%
Residues in generously allowed regions [~a,~b,~l,~p]	2	1.3%	7	4.5%
Residues in disallowed regions	4	2.6%	2	1.3%
Number of non-glycine and non-proline residues	155	100.0%	154	100.0%
Number of end-residues (excl. Gly and Pro)	2		1	
Number of glycine residues (shown as triangles)	18		18	
Number of proline residues	8		8	
Total number of residues	183		181	

Table 2: Functional site prediction for modeled 3-methyladenine DNA glycosylase from *Streptococcus sanguinis*

FUNCTIONAL SITE SERVERS	PUTATIVE RESIDUES
PINTS	NO HIT
CSA	NO HIT
PROFUNC	YA51, LA105, RA107 putative ligand binding residues, LA31, EA47, TA48 by reverse template search RA107, LA105, GA120 as DNA binding site
Q-SITE FINDER	DA14, FA15,HA16,LA14,IL19,TA21,AA26,AA29,LA30 ,MA33,AA77,GA78,LA93,LA139,AA140,LA146
FIRE STAR	YA127,AA134,AA135,HA136,MA149,YA157,YA159,IA161, CA167,NA169,CA178,LA180,RA182, VA262,GA263,VA264l
DOCKING ANALYSIS	YA51, SA57, AA58, CA59,HA60,SA61,KA69, MA73,YA81,YA83,QA84,IA85,HA86,MA90, NA92,LA105,RA107,RA160,IA161,GA162,VA163,TA164