BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net ***Hypothesis***

Volume 8(17)

Molecular interaction analysis of cigarette smoke carcinogens NNK and NNAL with enzymes involved in DNA repair pathways: An *in silico* approach

Qazi Mohd Sajid Jamal[1], Mohtashim Lohani[1], Mohd Haris Siddiqui[1], Mohd Haneef[1], Shailendra Kumar Gupta[2] & Gulshan Wadhwa[3]*

[1]Department of Biotechnology, Microbiology and Bioinformatics, Integral University, Lucknow-226026, India; [2]System Toxicology Group, CSIR-Indian Institute of Toxicology Research, PO Box - 80, MG Marg, Lucknow-226001, India; [3]Department of Biotechnology, Ministry of Science and Technology, CGO complex, Lodhi Road, New Delhi-110 003, India; Gulshan Wadhwa – Email: gulshan.dbt@nic.in; Phone: 011-24369612, +91-9811301820; *Corresponding author

Received August 08, 2012; Accepted August 20, 2012; Published September 11, 2012

Abstract:
DNA damage occurs almost all the times in cells, but is repaired also continuously. Occurrence of all these mutations and their accumulation in one cell which finally becomes tumorigenic/carcinogenic appears possible if the DNA repair mechanism is hampered. We hypothesize that alterations in DNA repair pathways, either all or at least at one i.e. genetic, translational or post-translational level, becomes quite imperative for the initiation and progression of Cancer. Therefore, we investigated the interaction capability of some carcinogens with the enzymes involved in the DNA repair mechanisms. Cigarette smoke's derivatives like NNK and NNAL are well established carcinogens. Hence, we analyzed 72 enzymes involved in the DNA repair Mechanisms for their interactions with ligands (NNK and NNAL). The binding efficiencies with enzymes ranging from +36.96 to -7.47 Kcal/Mol. Crystal Structure of Human Carbonmonoxy-Haemoglobin at 1.25 Å Resolution, PDB ID-1IRD as a +Ve control, showed binding energy -6.31 to -6.68 Kcal/Mol. and Human heat shock factor-binding protein 1, PDB ID- 3CI9 as a -Ve control, showed -3.91 to +2.09 Kcal/Mol. Binding was characterized for the enzymes sharing equivalent or better interaction as compared to +Ve control. Study indicated the loss of functions of these enzymes, which probably could be a reason for fettering of DNA repair pathways resulting in damage accumulation and finally cancer formation.

Key words: Cancer, DNA damage and repair, NNK, NNAL, Molecular docking

Background:
DNA damage occurs almost all the times in cells, but is repaired also continuously. For a cell to become transformed accumulation of all the damage/mutations in cell is mandatory and which appears, to be possible only if mutation escape the repair mechanism, in other words, the repair pathways are hampered. Studies have shown that DNA damage, due to environmental factors and normal metabolic processes inside the cell, occurs at a rate of 1,000 to 1,000,000 molecular lesions per cell per day. While this constitutes only 0.000165% of the human genome's approximately 6 billion bases (3 billion base pairs), unrepaired lesions in critical genes (such as tumor suppressor genes) can impede a cell's ability to carry out its function and appreciably increase the likelihood of cancer formation **[1].**

Occurrence of all these mutations and their accumulation in one cell which finally becomes tumorigenic/carcinogenic appears

possible if the DNA repair mechanism is hampered.A number of chemicals including cigarette smoke carcinogens like NNK and NNAL are well to induce unrepaired / non repairable DNA damage **[2].** 4-(Methylnitrosamino)-1-(3-Pyridyl)-1-butanone (NNK) is one of the nitrosamines derived from tobacco alkaloids and is a proven potent carcinogen **[3].** Presence of substantial amount of NNK in tobacco products plays a significant role as a cause of cancer in population using these products **[4].**

Therefore, we hypothesize that for the initiation and progression of Cancer, alterations in the DNA repair pathways, either all or at least at one i.e. genetic, translational or post-translational level, becomes quite imperative **[5].** That's why we have targeted the lung cancer carcinogens like NNK and NNAL for interaction with enzymes involved in DNA repair pathways **[6].**

We investigated the interaction capability of these carcinogens with enzymes involved in the DNA repair mechanisms. In cell mechanisms normal cells are regularly maintained by a set of instructions that follows whether cell should divide or not. Cancer cell develop an autonomous set of instructions against normal rules, leading to uncontrolled growth and proliferation of abnormal cells [**7].**

Our hypothesis says that if there would be any damage in DNA repair pathways occurs than the main repair machinery and component is caused by carcinogens. The enzymes targeted by foreign particles show the abnormal behavior or failure mechanism to prevent damage. Thus, failure of DNA repair mechanisms leads to cancer progression.

In this study, the docking simulation technique was adopted to preliminarily investigate the binding efficacy of specific carcinogens like NNK and NNAL to genes involved in DNA repair Pathways **[8].** A total of 72 enzymes involved in the DNA repair mechanisms and their interactions with ligands (NNK and NNAL) were analysed. This study was designed to investigate whether the carcinogen like NNK and its metabolite NNAL, apart from directly causing damage to DNA, are capable of affecting the DNA repair pathways.

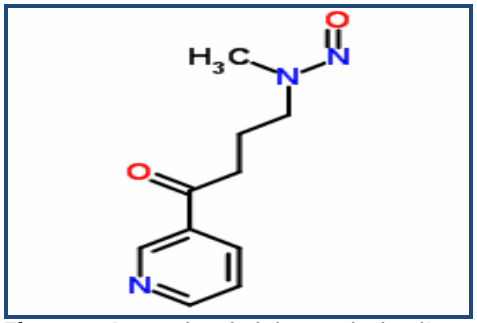

Figure 1: 4-(Methylnitrosamino)-1-(3-pyridyl)-1-butanone PubChem Compound ID- 47289, ChemSpider ID-43038

Methodology:

Preparation of ligand structures

Ligand file of NNK (4-(Methylnitrosamino)-1-(3-pyridyl)-1-butanone) and NNAL (4-(methylnitrosamino)-1-(3-pyridyl)-1-butan-1-ol) were downloaded in .mol format **(Table 1, Figure 1 & Figure 2)** from ChemSpider Chemical Database. These files could not directly used by Autodock 4.0 tools **[9]** thus they were converted it into .pdb files using Discovery Studio Visualizer version 2.5.5. Discovery Studio is a software package of biological molecular design solutions for computational chemists and computational biologists. Discovery Studio makes it easier to examine the properties of large and small molecules. Further the ligands were submitted for minimization using Chimera version 1.5.3 using with Genetic Algorithm Steps 2000 and 0.5 grid units Optimized **[10]**.

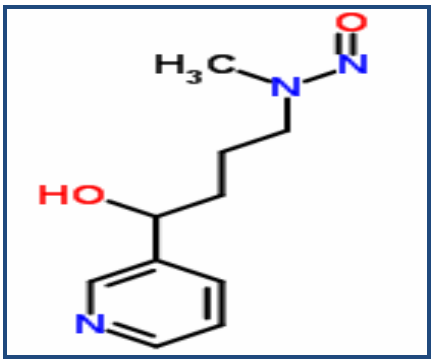

Figure 2: 4-(methylnitrosamino)-1-(3-pyridyl)-1-butan-1-ol PubChem Compound ID- 104856, ChemSpider ID- 94646

Preparation of protein structures

The structures of enzymes involved in the DNA Repair mechanisms were obtained from Protein Data Bank and some structure were downloaded from MODBASE server Table 2 (see supplementary material). MODBASE is a queryable database of annotated protein structure models, contains theoretically calculated models, which may contain significant errors, not experimentally determined structures **[11].** Published structures were edited to remove HETATM using Discovery Studio Visualizer (Version 2.5.5). Chimera was used for energy minimization, removal of steric collision with the steepest descent steps 1000, steepest descent size 0.02 Å, Conjugated gradient steps 1000 and the conjugate gradient step size 0.02 Å for the conjugate gradient minimization **[12, 13].**

Docking Studies

Docking studies were performed by Autodock version 4.0 suit **[14, 15]** and Cygwin interface was used in the Microsoft Windows XP professional Version 2002, Service pack 3 operating System on Intel (R) Core (TM) 2 Duo, CPU T6500 @ 2.10 GHz, 1.19 GHz, and 2.96 GB of RAM of Dell Machine. We implemented molecular docking methods followed by the searching the best conformation of enzymes and carcinogens complex on the basis of binding energy. Water molecules were removed from the protein structures before docking and hydrogen atoms were added to all target proteins. Kollman united charges and salvation parameters were added to the proteins. Gasteiger charge was added to the ligands. Grid box was set to cover the maximum part of proteins and ligand. The values were set to 60×60×60 Å in X, Y and Z axis of grid point. The default grid points spacing was 0.375 Å. Lamarckian Genetic Algorithm (LGA) **[16]** was used for proteins ligands flexible docking calculations. The LGA parameters like population size (ga_pop_size), energy evaluations (ga_num_generation), mutation rate, crossover rate and step

size were set to 150, 2500000, 27000, 0.02, 0.8 and 0.2 Å, respectively. The LGA runs were set to 40 runs. All obtained 40 conformations of proteins and ligand complex were analysed for the interactions and binding energy of the docked structure using Discovery Studio Visualizer version 2.5.5.

Protein-Protein Interaction analysis

The interacting proteins of selected enzymes were found using STRING 9.0 database that predict, interacting proteins against your query. We found interface residues using PDBe PISA, an interactive tool for the exploration of macromolecular (protein, DNA/RNA and ligand) interfaces **[17]**. The discovery studio 2.5 was used for Zdock (Dock Proteins). Zdock scores obtained from both Protein-Protein interactions as well as from Protein-Complex (ligand protein+NNK/NNAL) interaction.

Z dock calculations

Discovery studio 2.5 was used to perform protein-protein docking. ZDOCK is an initial stage rigid body molecular docking algorithm that uses a fast Fourier transform (FFT) algorithm to improve performance for searching in translational space **[18]**. All of the available structures from NMR were used to calculate the docking poses and the structures obtained were subjected to energy minimization using the smart minimize algorithm (Max steps 200, RMS gradient 0.01) in the program Discovery studio 2.5. The resulting Zdock scores with the highest value were used as appropriate conformational pose **Table 5 (see supplementary material).**

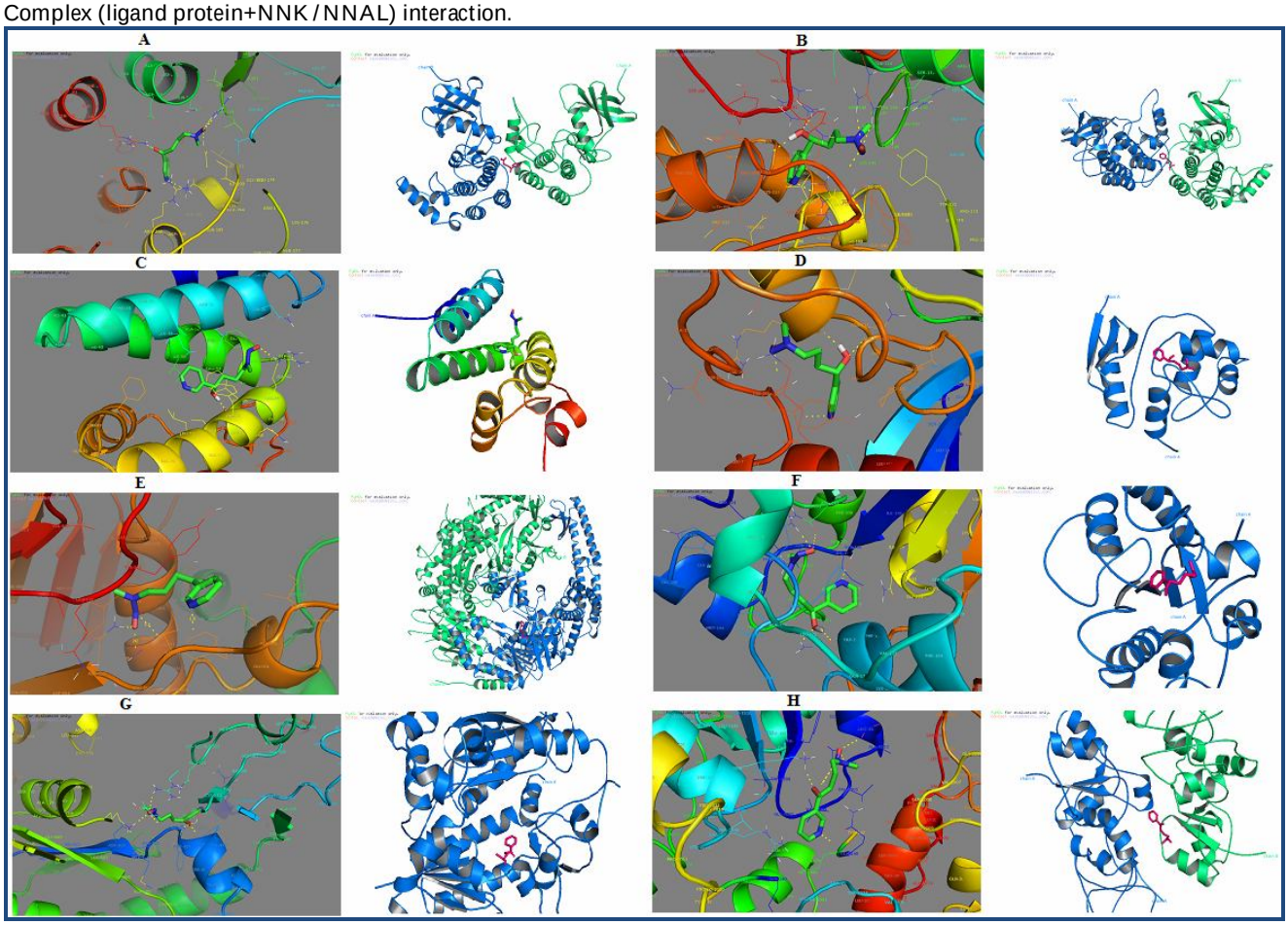

Figure 3: A-H is showing the interaction models of NNK/NNAL and enzymes, Graphics developed by PyMol 3D structure visualizer browser. **(A)** 1CKJ interacted with NNK; **(B)** 1CKJ interacted with NNAL; **(C)** 1Q2Z interacted with NNAL; **(D)** 1T38 interacted with NNAL; **(E)** 2O8B interacted with NNK; **(F)** 2RBA interacted with NNAL; **(G)** 3GQC interacted with NNK; **(H)** 3K05 interacted with NNK.

Discussion:

The 1IRD **(**Crystal Structure of Human Carbonmonoxy-Haemoglobin at 1.25 Å Resolution) was employed as a positive control and 3CI9 (Human heat shock factor-binding protein 1) as a negative control to validate our docking analysis. Furthermore, docking results of these proteins showed that 1IRD docked with NNK, observed binding energy was -6.68 Kcal/Mol, it docked with NNAL and observed binding energy was -6.31 Kcal/Mol. 3CI9 docked with NNK with observed binding energy of -3.91 Kcal/Mol, it docked with NNAL with binding energy of +2.09 Kcal/Mol.

A total 72 enzymes involved in the DNA repair mechanisms and their interactions with ligands (NNK and NNAL) were analysed during study. NNK showed the binding efficiency with enzymes ranging from +25.41 to -7.47 Kcal/Mol and NNAL showed the binding efficiency ranging from +36.96 to -6.52 Kcal/Mol. Simulations were completed and hydrogen

bonds were built in docked structures. After analysing the binding energy of distinct formed clusters, top 4 enzymes those shown the higher efficiency to bind with ligands (NNK binding efficiency with top 4 enzymes of DNA repair mechanisms ranging from -6.80 to -7.47 Kcal/Mol and NNAL binding efficiency with top 4 enzymes of DNA repair Mechanisms ranging from -6.17 to -6.52 Kcal/Mol were selected on the basis of their binding energy obtained from docked conformation. Hydrogen bond distances calculated by Discovery Studio Visualizer version 2.5.5 and reference RMSD and inhibition constant obtained from docked files are shown in **Table 3 & Table 4 (see supplementary material).**

The study shows that NNK has shown good binding efficiency with proteins 1CKJ (Casein kinases) (**Figure 3A** -6.80 Kcal/Mol.), 3K05 (**Figure 3H** -7.04 Kcal/Mol), 2O8B (centromeric DNA binding; protein binding) **(Figure 3E** -7.47 Kcal/Mol) and 3GQC (Deoxycytidyl transferase (**Figure 3G** -6.94 Kcal/Mol). On the other hand it was observed that NNAL shown good binding efficiency with proteins 1CKJ (**Figure 3B** -6.34 Kcal/Mol), 1Q2Z (Single stranded DNA-dependent ATP-dependent helicase) (**Figure 3C** -6.17 Kcal/Mol.), 1T38 (Methylated-DNA-protein-cysteine methyltransferase) (**Figure 3D** -6.52 Kcal/Mol) and 2RBA (G/T mismatch-specific thymine DNA glycosylase) (**Figure 3F** -6.41 Kcal/Mol).

It was reported that oral cancer developed by smoking and consuming tobacco **[19]**. In most of the cases when people orally consume the tobacco they would be in contact with NNK and derivatives like NNAL present in the cigarette smoke which leads to reaction with Cytochrome P450. After that NNK binds with the DNA and form DNA adduct that is responsible for the tumor progression **[20].** Hence, we have taken an account to elucidate the mechanism of carcinogenesis induced by environmental carcinogens NNK and NNAL with implementation of molecular docking techniques. Z dock scores are varying between Protein-Protein and Protein-Complex scores. Observed highest Zdock score is between 2O8B Vs 3NA3=23.40 which is greater than PC score=22.16. **Table 5 (see supplementary material)** Z dock scores obtained from Protein-Protein docking as well as from Protein-Complex conformations shows that carcinogens are reducing the normal capability of DNA repair enzymes.

Conclusion:

Study indicated the loss of functions of these enzymes, which probably could be a reason for fettering of DNA repair pathways resulting in damage accumulation and finally cancer occurs. Furthermore, our study suggests that carcinogens (NNK and NNAL) definitely alter the mechanisms of DNA repair pathways and enzymes functioning could be affected by carcinogens. NNK showed the binding efficiency with enzymes ranging from +25.41 to -7.47 Kcal/Mol and NNAL shown the binding efficiency ranging from +36.96 to -6.52 Kcal/Mol. We have selected top 4 enzymes those shown the higher efficiency to bind with ligands (NNK: -6.80 to -7.47 Kcal/Mol. and NNAL: -6.17 to -6.52 Kcal/Mol) on the basis of their binding energy obtained from docked conformation. Binding was characterized for the enzymes sharing equivalent or better interaction as compared to +Ve control. After analysing the Zdock scores obtained from Protein-Protein and Protein-complex interactions, we observed that when NNK and NNAL are binding with enzymes the main functioning of enzymes decreases.

Computer based structural analysis of bio macromolecular could play an important role in cancer treatment. At last but not least, a deep analysis is needed to elucidate the failure of DNA repair mechanisms with best suitable techniques and tools. *In vivo* and *In vitro* validation is needed to authenticate *in silico* results obtained from this study.

Acknowledgement:

We gratefully acknowledge the University Grant commission, New Delhi, India for providing Maulana Azad National Fellowship and we are also thankful to *Prof.* S.W. Akhtar, the Hon'ble Vice-Chancellor of Integral University, Lucknow, U.P., India for his support and for providing necessary facility.

References:

[1] http://www.ncbi.nlm.nih.gov/books/NBK21475/
[2] R-Y Huang *et al*. *Oncogene*. 2011 **30**: 116 [PMID: 20818420]
[3] Hecht SS, *Chem Res Toxicol*. 1998 **11**: 559 [PMID:20818420]
[4] Hecht SS & Hoffmann D, *Cancer Surveys*. 1989 **8**: 273 [PMID:2696581]
[5] Gina Manda *et al. Current Chemical Biology*. 2009 **3:** 22
[6] Samuel Saito *et al. Int J Mol Sci*. 2012 **13**: 2846 [PMCID: PMC3317691]
[7] Yang Y *et al*. *J Chromatogr A*. 2011 **1218**: 6505 [PMID: 21791338]
[8] Xinyuan Duan *et al. Int J Mol Sci*. 2012 **13**: 2846 [PMC: 2993495]
[9] Morris GM *et al*. *J Comput Aided Mol Des*. 1996 **10**: 293 [PMID: 8877701]
[10] Pettersen EF *et al*. *J Comput Chem*. 2004 **25**: 1605 [PMID: 15264254]
[11] Ursula Pieper *et al. Nucleic Acids Research*. 2011 **39**: 465 [PMID: 21097780]
[12] Wang J *et al*. *J Mol Graph Model*. 2006 **25**: 247 [PMID: 16458552]
[13] Wang J *et al*. *J Comput Chem*. 2004 **25**: 1157 [PMID: 15116359]
[14] Morris GM *et al. J Comp Chem*. 2009 **19:** 1639
[15] Rarey M *et al. Journal of Molecular Biology*. 1996 **261**: 470 [PMID: 8780787]
[16] Goodsell DS *et al*. *J Mol Recognit*. 1998 **9**: 1 [PMID: 8723313]
[17] Krissinel E and Henrick K, *J Mol Biol*. 2007 **372**: 797 [PMID: 17681537]
[18] Chen R *et al*. *Proteins*. 2003 **52**: 80 [PMID:12784371]
[19] Daniel Keyler *et al*. *Cancer Letters*. 2003 **202**: 1 [PMID: 14643020]
[20] Gharavi N & El-Kadi, *Current Drug Metabolism*. 2004 **5**: 203 [PMID: 15078197]

Edited by P Kangueane
Citation: Jamal *et al*. Bioinformation 8(17): 795-800 (2012)

BIOINFORMATION open access

Supplementary material:

Table 1: Details of carcinogens

S.No	Compound	IUPAC Name	Molecular formula	Molecular Weight	SMILES
1	NNK	4-(Methylnitrosamino)-1-(3-pyridyl)-1-butanone	$C_{10}H_{13}N_3O_2$	207.22912 g / mol	CN(CCCC(=O)C1=CN=CC=C1)N=O
2	NNAL	4-(methylnitrosamino)-1-(3-pyridyl)-1-butan-1-ol	$C_{10}H_{15}N_3O_2$	209.245000 g / mol	CN(CCCC(C1=CN=CC=C1)O)N=O

Table 2: Details of selected enzymes for active site characterization

S.No	Name of the Pathways	Gene Name (synonyms) linked to the GeneCards UK Mirror at Cancer Research UK	PDB ID	Chromosome location linked to NCBI MapView	Accession number linked to NCBI Entrez
1	Base excision repair (BER)	TDG	2RBA	12q23.3	NM_003211
2	Non-homologous end-joining	XRCC5 (Ku80)	1Q2Z □	2q35	NM_021141
		REV1L (REV1)	3GQC □	2q11.2	NM_016316
		PRKDC	1CKJ	8q11.21	NM_006904
3	Direct reversal of damage	MGMT	1T38	10q26.3	NM_002412
4	Other conserved DNA damage response genes	MDC1	3K05	6p21.3	NM_014641
5	Mismatch excision repair (MMR)	MLH3	2O8B	14q24.3	NM_014381

Table 3: Enzymes docked with NNK results obtained from Autodock 4.0

S.No.	Enzyme	PDB ID	Residues involved in H- Bond	Distance of Hydrogen bond (Å)	Residues creating hydrophobic region	Binding energy Kcal/Mol	Estimated Inhibition Constant (uM)
1	Casein kinase I isoform delta (EC 2.7.1.-) (CKI-delta) (CKId)	1CKJ	1CKJ:A:ASN188:HD22: 43038:O:N5	2.027	ASP113,GLN114,SER117,ARG118,LEU149,TYR172,ASN188,LEU191,ARG198,GLU202,ARG261,LEU263,ARG279,TYR291	-6.80	22.67
			1CKJ:B:ARG118:HN21: 43038::O:N4,O2	2.122			
			1CKJ:B:ARG118:HH11:43038::O:O2	1.947			
			1CKJ:B:ARG279:HH21:43038:O:O1	2.132			
			1CKJ:A:TYR172:HH:43038:O:O2	1.799			
2	DNA mismatch repair protein	2O8B	2O8B:A:ILE651:HN:43038:O:N5	2.2	GLN68,VAL642,ILE648,ALA649,PHE650,ASN653,GLY674,THR677,TYR678,GLN681	-7.47	3.32
			2O8B:A:ASN653:HD21:43038:O:O2	1.7			
3	Mediator of DNA damage checkpoint protein 1	3K05	3K05:B:ASP1902:HN: 43038::ON4,O2	1.955	GLN140,GLN141,PHE1897,THR1898,GLY1899,VAL1900,VAL1901,ASP1902,LYS1936	-7.04	6.92
			3K05:B:LYS1936:H23:43038::O:O1	2.213			
			3K05:C:GLU141:HN:43038::O:N5	2.073			
4	DNA repair protein REV1	3GQC	3GQC:A:LYS625:HZ3:43038::O:O2	2.095	ARG357,ILE361,ASP423,MET424,CYS424,ASP425,PHE427,PHE428,VAL429,ALA509,SER510,TYR513,ARG516,ASN522,LYS625	-6.94	8.25
			3GQC:A:ARG516:HH21:43038::O,N4	2.212			
			3GQC:A:PHE428:HN:43038::O:O1	1.747			

Table 4: Enzymes docked with NNAL results obtained from Autodock 4.0

S.No.	Enzyme	PDB ID	Residues involved in H- Bond	Distance of Hydrogen bond (Å)	Residues creating hydrophobic region	Binding energy Kcal/Mol	Estimated Inhibition Constant (uM)
1	Casein kinase I isoform delta (EC 2.7.1.-) (CKI-delta) (CKId)	1CKJ	1CKJ:A:AS188:HD:22:94646:O:N4	2.043	ASP113,SER117,TYR172,ASN188,ARG198,GLU202,ARG261,SER262,LEU263,ARG264,PHE265,ASP266,ARG275,ARG279	-6.34	22.64
			1CKJ:A:ARG261:O:94646:O:H24	2.168			
			1CKJ:B:ARG275:HH21:94646:O:O2	2.192			
			1CKJ:A:PHE:265:HN:94646O:N5:O2	2.088			
2	ATP-dependent DNA helicase 2 subunit 2	1Q2Z	1Q2Z:A:PHE66:O:94646::O:H:24	2.188	SER24,ILE28,ILE31,ILE50,ARG65,PHE66,PHE69,LEU70,LEU73	-6.17	29.93
			1Q2Z:A:ARG65:HE:94646::O:O2	2.069			
3	Methylated-DNA protein-cysteine methyltransferase	1T38	1T38:A:SER159:HN:94646::O:N5,O2	2.200	LEU33,TYR114,MET134,ARG135,ASN137,PRO140,SER145,ASN157,TYR158,SER159,GLY160	-6.52	16.52
			1T38:A:SER145:HG:94646::O:O1	1.979			
			1T38:A:MET134:O:94646::O:H24	1.775			
	G/T mismatch-specific thymine DNA glycosylase (EC 3.2.2.-)	2RBA	2RBA:A:ILE139:HN:94646:O:N4	2.086	LEU124,GLY138,ILE139,ASN140,PRO141,GLY142,ALA145,HIS151,TYR152,ASN157,ASN191,SER271,SER272,SER273	-6.41	19.86

Table 5: Zdock scores obtained from Discovery Studio 2.5

S.No	Enzymes PDB ID	Protein-Protein Docking	Protein-Complex* docking
		Z dock score	Z dock score
1	1CKJ_NNAL vs. 1UL6	16.26	15.94
	1T38 vs. 1HCP	16.92	15.54
	2RBA vs. 1A5R	15.64	15.80
	2O8B vs. 3NA3	23.40	22.16
	3GQC vs. 3ABE	20.04	19.7
	3K05 vs. 2DYP	25.44	22.46
	1Q2Z vs. 1JEQ	17.78	18.06
	1CKJ_NNK vs. 1UL6	16.26	15.94

Note: *complex contains NNK/NNAL and enzymes docked conformation