The Journal of Animal & Plant Sciences, 28(5): 2018, Page: 1276-1291
ISSN: 1018-7081

PREDICTION OF OGG1 STRUCTURAL AND FUNCTIONAL MOTIFS WITH ACTIVE BINDING SITES FROM *CAMELUS DROMEDARIES*

W. Khan[1], Z. Abduljaleel[2, 3, 4, *], F. A. Al-Allaf[2, 3], N. Shahzad[5], W. El-Huneidi[1, 6], M. Elrobh[7], M. Alanazi[7] and Hani Faidah[8, 9]

[1]Department of Basic Sciences, P.O Box 3660, College of Science and Health Professions, King Saud Bin Abdulaziz University for Health Sciences, Riyadh 11426, Kingdom of Saudi Arabia; [2]Department of Medical Genetics, Faculty of Medicine, Umm Al-Qura University, P.O.Box 715, Makkah 21955, Kingdom of Saudi Arabia; [3]Science and Technology Unit, Umm Al Qura University, P.O. Box 715, Makkah 21955, Kingdom of Saudi Arabia; [3]Molecular Diagnostics Unit, Department of Laboratory and Blood Bank, King Abdullah Medical City, Makkah 21955, Kingdom of Saudi Arabia; [4]Molecular Diagnostics Unit, Al Noor Specialized Hospital P.O.Box: 6251, Makkah, Saudi Arabia; [5]Department of Pharmacology and Toxicology, College of Medicine, Umm Al-Qura University, Makkah, Kingdom of Saudi Arabia; [6]Department of Basic Medical Sciences, College of medicine, University of Sharjah, United Arab Emirates; [7]Genome Research Chair, Department of Biochemistry, College of Science, King Saud University, Riyadh, Kingdom of Saudi Arabia;[8]Molecular Diagnostics Unit, Department of Laboratory and Blood Bank, Al Noor Specialized Hospital P.O.Box: 6251, Makkah, Saudi Arabia.[9]Department of Microbiology, College of Medicine, Umm Al-Qura University, P.O.Box: 715, Makkah 21955, Saudi Arabia.

Corresponding Author's email: zainulbio@gmail.com

ABSTRACT

Many biological mechanisms involve the interaction of proteins or binding of other molecules to proteins. The precise prediction of functionally active binding sites on the protein surface could play an important role in predicting the nature of protein-protein or protein-ligand interactions. The present research was conducted on 8-oxoguanine DNA glycosylase gene*OGG1* sequence from Arabian camel (*camelus dromedaries*) to predict the protein structure. In this study, 1790 long amiono acid (AA) sequences of OGG1 from *C. dromedaries*were used to predict its protein structure based on multiple alignments by Lometes & Illterative ITasser simulations. Because the full structure ofOGG1 protein cannot be predicted based on the homology modeling using conserved regions of other mammalian species, we predicted the 3D structure of two domains of the OGG1 protein. The two regions predicted were OGG1 protein domain 1 (D1) comprised of amino acids from 1-1000 and OGG1 protein domain 2 (D2) comprised of amino acids from 1000-1790. The 3D protein structures were validated using RAMPAGE Ramachandran plot and functional structures were predicted based on the homologous regions from other species including human, rat, mouse and panda. The functional group predications were established using the Eukaryote Linear Motif (ELM) resource. Among the important functions predicted for OGG1 proteins were LIG_BRCT_BRCA1_1 instances of Phosphopeptide motif which directly interacts with the BRCT (carboxy-terminal) domain of the breast cancer gene BRCA1, LIG_FHA_2 of Phosphothreonine motif binding a subset of FHA domains that have a preference for an acidic amino acid at the replication fork, and MOD_TYR_ITSM of the CD150 subfamily of receptors that bind to and are regulated by SH2 adaptor molecule. These sites may constitute target for drug design for many pharmaceutical and biotechnological purposes. The target site of binding region for OGG1 was predicted both in D1 and D2 by using site finder tools of Molecular Operation Environment (MOE). Furthermore, the results also predicted two conserved regions, both 45AA long, having 100.0% similarity with the crystal structure of Calmodium-Dependent Protein Kinase I (CaM kinase I) from rat (Q63450). This is the first report that deals with the Arabian camelOGG1protein structure prediction along with functional motifs and binding sites identification.

Keywords: *Camelus dromedaries, OGG1 gene*, Homology Modeling; Protein Validation; Binding site Prediction, Functional group Prediction, Multiple Sequence Alignments.

INTRODUCTION

Camelusdromedarius (Arabian or one-humped camel) belongs to the family Camelidae and is found in the Arabian deserts and arid and semi-arid areas of the Middle East (Yam and Khomeiri, 2015). The Arabian camel has developed physiological adaptations to deal with extreme environments such as elevated temperature and drought. Camel is persistently under stressful environments that may cause DNA damage and mutation. Several enzymes including DNA glycosylases play a role in the damaged DNA repair mechanism. The 8-oxoguanine DNA glycosylase (*OGG1*) gene, produces the enzyme that is involved in the excision of 8-oxoguanine, an impaired base byproduct formed due to reactive oxygen species (ROS) (Klungland and Bjelland, 2007). If not repaired immediately and correctly, it causes

genomic instability that eventually affects several biological processes (Boiteux and Radicella, 2000).

Deducing biological significance from protein structures is an extremely valuable tool to understand its molecular nature. The strategies to deduce protein structure prediction are divided into three: homology modeling (Schwede*et al.*, 2003), threading (Xu*et al.*, 2003; Soding, 2005; Zhou and Zhou, 2005) and *abinitio* method (Pauling and Corey, 1951; Simons*et al.*, 1997). Homology modeling and threading strategies usually generate accurate protein structure predictions. Protein structure and functional prediction can also be important for disease analysis and may help in the development of new drug targets. When a homologous protein with a recognized structure is identified, it can be used as a template to model the 3D structure for the query protein (Rychlewski*et al.*, 1998) , because homologous proteins usually have quite similar 3D structures (Kinch and Grishin, 2002). The 3D model could then assist to make hypotheses to conduct experiments. The protein sequence identity with other species can be acheived by BLAST search and by using universal protein resource (UniPort) for protein sequence and annotation data (UniProt, 2010). The UniPort meta-genomic and environmental sequences (UniMES) database is a source specially developed for meta-genomic and environmental data. The data were analyzed by multiple alignments using genome workbench (CLCbio) application, (Denmark) (Petrie and Joyce, 2010). The regions with similar structure can be observed using multiple thread alignment of Lometes I-TASSER simulation for functional group prediction of the eukaryotic motif resources (Gould*et al.*, 2010). Linear motifs are short segments of multi-domain proteins that provide regulatory functions independently of protein tertiary structure. A lot of intracellular signaling passes by means of protein modifications at linear motifs. Several linear motif occurrences, most notably phosphorylation sites, have now been reported. Although exact linear motifs are difficult to predict using *de novo* protein sequences due to the difficulty of obtaining robust statistical assessments. The eukaryotic linear motifs (ELM), a useful resource (http://elm.eu.org/), provides an expanding knowledge based on functional prediction (Gould*et al.*, 2010). The binding site prediction is also very important for target drug design. It is important to identify and characterize binding sites using computational methods not only to understand the molecular interactions that exist in nature and in diseased conditions, but also to exploit the protein structural information for drug design in the pharmaceutical and biotechnology industry. The majority of methods that are currently used to identify the protein binding sites are based on Fischer's lock and key model where a substrate binds to an enzyme like the key into a lock. Shape complementarity between the ligand and the protein is an important determinant of binding and small molecules usually bind in concave pockets on the protein surfaces.

This study was based on the *in silico* analysis of the OGG1protein structural prediction from Arabian camel (*C. dromedarius*). The 3D OGG1protein structures of two domains D1 and D2 were predicted and confirmed by using Ramachandran plot. Furthermore, the protein binding sites and functional motifs were also identified because of the important role they play in the signaling pathways through phosphorylation and also by interacting with other proteins.

MATERIALS AND METHODS

Structure prediction: Full length *OGG1* cDNA sequence (Cam-Roo1675, sources ENSP0000-355759, Scaffold 31:2548052) was obtained from the Camel Genome database, King Abdullah City for Science a Technology (KACST), Riyadh, Saudi Arabia. The I-TASSER &Lometes simulation was used to predict the 3D structure of the OGG1 protein. I-TASSER is a program used to predict the protein structure and function annotation. It utilizes the amino acid sequence of target proteins and creates full-length atomic structural models based on multiple threading alignments and iterative structural assembly simulations. The program predicts a structure and provides its function related information (Yang and Zhang, 2015). The I-TASSER initially recognizes homologous structure templates from the protein data base (PDB) library by means of LOMETS (Wu and Zhang, 2007). LOMETS is a meta-threading program which is comprised of multiple individual threading algorithms (Wu and Zhang, 2007). The prediction also shows a correlation between the C-score and the TM-score (a structural similarity measurement) with a correlation coefficient of 0.91. A cutoff > -1.5 of C-score was used for models of correct topology; both false positive and false negative rates were kept below 0.1. Depending on the combination of C-score and protein length, accurate I-TASSER models can be predicted with an error rate of 0.08 for TM-score and 2Å for RMSD (Zhang, 2008; Roy*et al.*, 2010). The I-TASSER procedure, matched the query sequence against a non-redundant sequence database by position-specific iterated BLAST (PSI-BLAST) (Altschul*et al.*, 1997) to identify evolutionary relatives. The sequence profile created was based on the multiple alignments of homologs to predict the secondary structure (PSIPRED) (Jones, 1999). The query sequence was threaded through a representative PDB structure library using LOMETS (Wu and Zhang, 2007). However, the individual threading was carried out by using programs such as FUGUE (Shi*et al.*, 2001), MUSTER (Wu and Zhang, 2008), PROSPECT (Xu and Xu, 2000), PPA(Wu*et al.*, 2007) and SP3 (Zhou and Zhou,

2005)where the templates were ranked based on the sequence and scores.

Protein structure validation: For each residue of the OGG1 protein, Ramachandran diagram plots (RAMPAGE) showing phi versus psi dihedral angles was performed. The result output was divided into three categories i.e. favored, allowed and disallowed regions, based on the density dependent smoothing for non-Glycine, non-Proline and non-preProline residues with B < 30 for 500 high-resolution protein structures. Regions were also defined for Glycine, Proline and preProline(Lovell*et al.*, 2003).

OGG1 protein Active site prediction: The Molecular Operating Environment (MOE) program predicts the active site of the protein based on the surface calculations and molecular docking studies depending on the favorable binding configurations of the ligands and the protein target. Typically, the scoring functions were shown as a favorable hydrophobic, ionic and hydrogen bond contacts. Based on the Edelsbrunner's alpha shapes, we detected candidate protein-ligands and protein-protein binding sites using a fast geometric algorithm. The binding sites of the two predicted OGG1protein domains (D1 and D2) were ranked according to their accessible hydrophobic contact surfaces. The active binding sites in the receptor are usually hydrophobic pockets that contain characteristic side chain atoms and only spheres that corresponded to the tight atomic packing within the receptor were retained. Each alpha sphere was assessed as either "hydrophobic" or "hydrophilic" depending on the hydrogen bonding in the receptor or protein. Those hydrophilic spheres, which were not close to a hydrophobic sphere, were eliminated as sometimes they correspond to water sites. The alpha spheres were then clustered using a single-linkage clustering algorithm to generate a group of sites. The individual sites were observed with "dummy atoms" for docking calculations or starting points for de novo ligand designing and the active sites analysis tool to identify polar, hydrophobic, acidic and basic residues. It also visualizes solvent exposed ligand atoms and residues that are in close contact with the ligand atoms including the side chain and the backbone acceptor (Goodford, 1985; Edelsbrunner*et al.*, 1995).

Functional motif prediction: The computational biology resource ELM (The eukaryotic liner motif resources) was utilized for exploring the potential functional sites in the protein structure. The functional sites were predicted based on the “linear motif” using the regular expression rules. To reduce the number of false positives, the predictive power, context-based rules and logical filters were applied. The core functionality was obtained by filtering the cell compartment, phylogeny, globular domain clash (using the SMART/Pfam databases) and structure. In addition, both known ELM instances and any positional conserved matches in sequences similar to ELM instance sequences were also identified. The functional motif in a sequence showing similarity to an ELM sequence and the positional conserved region was predicted depending on the score displayed by the ELM instance mapper and using the structure filter (Lovell*et al.*, 2003).

Protein Structural comparisons: For a partial order graph representation of multiple alignments that recognizes and identifies areas that were conserved in a sub set of OGG1-D2 input structures and indicates acceptable internal rearrangements of the protein structures by using a multiple protein structure alignment program i.e., partial order structure alignment (POSA). The POSA program by visualizing the mosaic nature of multiple structural alignments out performs other programs with regard to structural flexibilities and provides new insights. Such a flexibility alignment parameter is essential for improving the alignment quality as well as for better understanding of the protein structures and superimposed structural configurations (Horikawa*et al.*, 1973).

Multiple Sequence Alignment: The CLCbio Genome Workbench (Denmark) was used to determine the multiple sequence alignments. The interpretation of a multiple sequence alignment was based on the evolutionary relationship. The alignment used was based on the search for homology between sequences or groups of sequences, and mutations were detected. Furthermore, these sequence alignments indicate structural and/or functional characteristics of sequences and when compared with well-described sequences, hence new information may be gained from unknown sequence data. Conserved regions in the sequence alignment (Thompson*et al.*, 1994) can identify conserved domains, which may indicate functionally important sites such as binding sites, active sites or sites that are related to other key functions (Thompson*et al.*, 1994).

RESULTS

Protein structure of *Camelus* OGG1 domains D1 and D2: The structures of the two regions of *OGG1* protein were predicted i.e. D1 & D2, where D1 region was 1-1000 amino acids (1000AA) and D2 region was 1000-1790 amino acids (790AA) using the best models which predicted the protein structures based on the protein structure prediction program I-TASSER. The 3D models were constructed based on multiple-threading alignments by Lometes&IllterativeTasser simulations using state-of-the art algorithms depending on the scores. The OGG1 3D predicted structure D1 model had TM-Score 0.41±0.0.14 along with C-score-2.59, RMSD (Å) 15.6±3.3, number of decoys 158 and cluster density

0.0304 (Fig. 1). Whereas, the other OGG1 3D predicted structure D2 model showed TM-score 0.35±0.12 along with C-score 3.24, RMSD (Å) 16.7±2.9, number of decoys 256 and cluster density 0.0509 (Fig. 2). The structure predication referred to the computational procedure for identifying template proteins from solved structure databases having similar structure or similar motif structure to the query protein sequence. To improve the efficiency of the I-TASSER search, we adopted a reduced model to predict the protein chain along with each residue described by its C-Alpha atom and mass. The modeling of the two regions (D1 and D2) was based on lattice and template fragments during simulations, which assist in predicting the structure.

Protein structure validation: The two predicted structures D1 and D2 of OGG1 protein were also validated by RAMPAGE Ramachandran plot, where the deviation of the observed C-ß atom from the ideal position provides a single measure encapsulating the major structure-validation data contained in the bond angle distortions. The C-ß deviation is usually sensitive to incompatibilities between the side chain and the backbone caused by misfit conformations or inappropriate refinement restraints.The phi, psi plot utilized for density- dependent smoothing of non-Gly, non-Pro, and non-prePro residues from 500 excessive-resolution proteins exhibited sharp boundaries at critical edges and clear delineation between large empty areas and regions allowed but not favored. One such region was the gamma-turn conformation near +75 degrees at -60 level, counted as forbidden by common structure-validation applications; however, it occurred in effectively-ordered parts of the sound structure, it was near the functional sites, and strain was partly compensated by the gamma-turn H-bond. Favored and allowed phi, psi regions were also defined for Pro, pre-Pro, and Gly (necessary as a result of Gly phi, psi angles were somewhat allowed but less accurately determined). A proposed factor explaining this discrepancy was the crowding of the 2-peptide NHs permitted to donate only a single H-bond. The predicted OGG1 protein structure D1 residues favored 98.0% region with allowed 2.0% area and in the 10.4% outlier region (Fig. 3A-C), whereas the predicted OGG1 protein structure D2 residues were also in 98.0% favored region within the 2.0% allowed area but with 6.2% outlier area (Fig. 4).

Prediction of OGG1 active binding sites: The active binding site prediction of OGG1 using MOE site finder showed that the OGG1 predicted D1 residues had two active sites (Fig. 5 A-C), whereas the predicted D2 residues had only one active binding site (Fig. 6 A-B). The first binding site was about column size 679 (Fig. 5B) which indicates the number of contact atoms in the receptor, Hyd-87 (the hydrophobic contact atoms in the receptor), column side 460 showing the number of side chain of the contact atom in the receptor. The second binding site of OGG1 D2 was of Size-563 (Fig. 5C), with Hyd-76, with side column 426. The residues of the first and second binding sites are shown in Table 1. The OGG1 D2 showed one binding site (Fig. 6B) with column size 476, Hyd-132, side column 252 and its amino acid residues are shown in Table 2.

Prediction of Functional group in the OGG1 protein D1 and D2 structures: Based on the eukaryotic motif resources and putative functional sites, protein functional sites can be predicted based on expression patterns. The predictive power guidelines and logical filters are utilized for checking false positives. The predicted functional site of OGG1 protein D1 structure showed highly conserved areas of several functional groups from the different species (Fig. 7). The ELM showed CLV_NDR_NDR_1 motifs matched sequence with amino acid residues RRT (878-880), a functional motif of N-Arg dibasic convertase (nardilysine) cleavage site (Xaa-I-Arg-Lys or Arg-I-Arg-Xaa) in the extracellular Golgi apparatus and cell surface. However, CLV_PCSK_KEX2_1 motifs were matched with amino acid residues KRR (877-879), a functional motif of (Lys-Arg-|- Xaa or Arg-Arg-|-Xaa) in extracellular and Golgi apparatus. The CLV_PCSK_PC1ET2_1 motif matched with amino acid residues KRR (877-879) which is a functional residue of NEC1/NEC2 cleavage site (Lys-Arg-|- Xaa) in the Golgi membrane and extracellular. Similarly, the CLV_PCSK_SKI1_1 motif matched with amino acid residues RDLKV (323-327), a functional motif of Subtilisin/kexin isozyme-1 (SKI1) cleavage site ([RK]-X-[hydrophobic]-[LTKF]-|-X) of endoplasmic reticulum lumen and Golgi apparatus. The LIG_14-3-3_2 motif matched with amino acid residues RPHASLS (572-578), which is a functional motif of longer mode 2 interacting phospho-motif for 14-3-3 proteins in the nucleus, mitochondrion, internal side of plasma membrane and cytosol. The LIG_BRCT_BRCA1_1 motif corresponding with amino acid residues LSFLF (896-900) that is a functional residue of Phosphopeptide motif, which directly interacts with the BRCT (carboxy-terminal) domain of the Breast Cancer Gene BRCA1 with low affinity and BRCA1-BARD1 complex. The motif has the consensus sequence i.e. S.F in the binding pocket of the BRCT domains. The high affinity motif has additional lysine residues (S.F.K.) in the LIG_CYCIN_1 motif that matched with amino acid residues RDLKV (323-327), a function of substrate recognition site that interacts with cyclin and increases phosphorylation of cyclin/cdk complexes. Predicted protein had the MOD_CDK site, which is used by cyclin inhibitors in the cytosol and nucleus. The LIG_EH1_1 motifs matched with amino acid residues SFSILKILL (799-807), a functional domain of homolog domain 1 motif of energetic repressors and different transcription families, and appealed in the

recruitment of Groucho/TLE co-repressors in the nucleus. The LIG_FHA_1 motifs matched with amino acid residues RDTHCLG (658-664) that is a functional residue of Phosphothreonine motif binding site, a subset of FHA domains that have a preference for a large aliphatic amino acid at the pT+3 position in the nucleus. The LIG_FHA_2 motif matched with amino acid residues FYTERDA (295-301) residue, which is a functional domain of Phosphothreonine motif binding region a subset of FHA domains that has affinity for acidic amino acid at the pT+3 position in the nucleus and the replication fork. The LIG_MAPK_1 motifs corresponded with amino acid residues RKPFLSF (892-898), which is a functional residue of MAPK interacting molecules (e.g. MAPKKs, substrates, phosphatases) having docking motif and contribute in specific interaction in the MAPK cascade. The classic motif is a hydrophobic residue in the nucleus and cytosol. The LIG_NRBOX motif matched with amino acid residues ILKLLL (802-808) functional residue of nuclear receptor box motif (LXXLL) confers binding to nuclear receptors in the nucleus. The LIG_SH2_STAT5 motifs matched with amino acid residues YTER (296-299), a functional residue of STAT5 Src Homology 2 (SH2) domain binding motif in the cytosol. The LIG_CYCLIN_1 motifs in lined with amino acid residues amino acid residues RDLKV (323-327), which is a substrate recognition site that act together with cyclin and increases phosphorylation of cyclin/cdk complexes. Predicted protein should have the MOD_CDK site. Additionally, used by cyclin inhibitors in cytosol and nucleus. The MOD_GlcNH glycan motifs matched with amino acid residues VSGG (281-284) that are functional residues of Glycosaminoglycan attachment site in extracellular and Golgi apparatus. The MOD_PKA_2 motifs matched with amino acid residues CRVSGGE (279-285) that is part of a functional residue of secondary preference for PKA-type AGC kinase phosphorylation in cytosol, nucleus, and cAMP-dependent protein kinase complex. The TRG_NES_CRM1_1 motifs matched with amino acid residues DLGIVHRDLKVGGE (317-330), which is a functional residue of some proteins, re- exported from the nucleus and contain a Leucine-rich nuclear export signal (NES) that binds to the CRM1 export protein in the nucleus.

The predicted OGG1 protein D2 structure also showed various functional groups such as LIG_FHA_2, MOD_GlcNH glycan, LIG_SH2_STAT5, MOD_PKA_2, TRG_NES_CRM1_1, LIG_CYCLIN_1 and LIG_BRCT_BRCA1_1 (Fig. 8A-B). The CLV_NDR_NDR_1 motifs matched with amino acid residues IRK (351-353) which is functional residue of N-Arg dibasic convertase (nardilysine) cleavage site (Xaa-|-Arg-Lys or Arg-|-Arg-Xaa) present in the extracellular, Golgi apparatus and cell surface. The CLV_PCSK_PC1ET2_1 motifs matched with amino acid residues KRF (175-177) residues with the functional NEC1/NEC2 cleavage site (Lys-Arg-|-Xaa) present in the extracellular and Golgi membrane. The LIG_APCC_Dbox_1 motifs that matched with amino acid residues GRYELAVLE (121-129) functional residues that bind to the Cdh1 and Cdc20 components of APC/C causing protein destruction during the cell cycle in the cytosol. The LIG_BRCT_BRCA1_1 motifs corresponded with amino acid residues ASKTF (249-253), a functional Phosphopeptide motif, which directly interacts with the BRCT (carboxy-terminal) domain of the Breast Cancer Gene BRCA1 with low affinity in nucleus. The BRCT domains are primarily found in Eukaryote, whereas the BRCT domains are usually associated in the DNA damage response. They recognize and bind to the specific phosphorylated serine (pS) sequences and are involved during cell cycle checkpoint and DNA repair mechanisms. Another remarkable characteristic of the BRCT motifs is that the phosphopeptides may also bind across the domain-domain interface (Clapperton*et al.*, 2004). The available data propose that BRCTs may bind exclusively to phosphoserine peptides. However, by contrast FHA domains, which are often found in a similar functional context, recognize phosphothreonine peptides (ELM: LIG_FHA_1). Many of the BRCT ligands are located at the pSQ motifs and are phosphorylated by the checkpoint kinases, ATM, ATR, and DNA-PK(Glover*et al.*, 2004; Zhang*et al.*, 2010). It has been shown that the BRCA1- binding motifs are S.F and K (high affinity) or S.F (lower affinity).

The LIG_MDM2 motifs matched with FSFLWRGL (253-260) functional residue of p53 family members, which confers binding to the N-terminal domain of MDM2 in the nucleus. TheMOD_CK2_1 motifs matched with amino acid residues SVHSAVE (615-621), which is a functional residue of CK2 phosphorylation site present in nucleus, cytosol and protein kinase CK2 complex. The MOD_GSK3_1 motifs matched with amino acid residues SQGSPPRS (11-18) that is a functional residue of GSK3 phosphorylation located in the cytosol. The MOD_TYR_ITSM motifs matched with amino acid residues YCTLYIDV (325-332) which is a functional domain of ITSM (immunoreceptor tyrosine-based switch motif) present within the cytoplasmic region of the CD150 subfamily of CD2 family and bind to the SH2 adaptor molecules of SH2DIA in the cytosol.

Protein structure comparison with other species: Our results indicate that the two regions of the predicted OGG1 protein structures (D1 and D2) showed similarities with other proteins from other species. It was observed that the OGG1 45AA region (281-325) showed 100% similarity with the Calcium/Calmodulin-dependent protein kinase (CDPK) (PDB ID 1A06) region (VAL: 98

– LYS: 142) of *rattusnorvegicus* (Norway rat) (Fig. 9A). The CDPKs are calcium-signaling proteins related to calmodulin-dependent proteins (CaMK), belonging to a large family of serine/threonine kinases, whereas CaMKs are activated by calmodulin-calcium complex. The kinase domain of a CDPK becomes active via interaction with its typically C-Terminally located calcium binding domain. The other region of predicted OGG1 protein structure from 1-45 (45AA) also showed 100% identity with the CDPK region (Fig. 9B). Furthermore we modeled the 3D structure of the OGG1 based on multiple structure alignment and superimposed with the rat CDPK protein to look for database conserved regions using partial order structure alignment server (POSA, http://fatcat.burnham.org/POSA/). The results showed that the superimposed structures based on sequence and RMSD: 0.00Å showed 100.0% identity for the particular region (Fig. 9A). Furthermore, two more regions of the OGG1 protein also showed similarity with CDPK proteins of other mammalian species using UniPort blast search server http://www.uniprot.org/blast/, such as with *Rattusnorvegicus* (Rat, Q63450), *Mus musculus* (Mouse, Q91YSA), *Homo sapiens* (Human, Q14012), *Bostaurus* (Bovine, Q08DQ1), *Ailuropodamelanoleuca* (Giant panda, D2HRA9) (Table 3). Our results showed that all the above-mentioned species have similarity with the two regions of OGG1 protein i.e. 281-324 (45AA) (Fig. 9C) and 98-142 (45AA) (Fig. 9D) based on multiple alignments using CLCbio, whereas the phylogenetic trees show the evolutionary relationships of OGG1 amino acid sequences with protein structures of different mammalian species based on the UPGMA algorithm analyzed using CLCbio (**Fig.** 9E).

Table 1. The two active binding sites amino acid residues within the predicted OGG1 D1 structure using MOE site finder.

Biding Site	Amino acid residues
Binding site #1	ARG232, HIS233, ARG235, TRP237, LEU241, PRO242, HIS243, HIS244, ALA245, VAL247, SER248, GLY249, PRO250, ALA251, PRO252, ALA253, SER254, LEU256, LEU259, PRO260, PRO261, TRP263, PRO264, LEU265, CYS266, LEU267, PRO268, CYS269, SER270, LEU271, GLY272, ASP273, CYS274, SER275, VAL276, LEU309, HIS316, ASP317, LEU318, GLY319, ILE320, VAL321, HIS322, ARG323 ASP324, LEU325, LYS326, VAL327, GLY328, GLY329, GLU330, GLY331, VAL332, TRP333, GLY334, ALA335, GLY336, ALA337, PRO338, ARG339, GLY340, GLY341, ALA342, SER343, HIS354, PRO355, LEU356, ALA357, GLU358, PRO35P, TRP360, LEU361, CYS364, LYS365, ASN366, GLN405, SER431, LEU432, PRO433, ARG434, PRO435.
Binding site #2	ASP381, THR382, GLN383, VAL384, ASP385, THR386, VAL388, ALA391, ILE393, GLU442, VAL444, PRO448, GLY449, ARG480, GLY481, PRO482, TYR483, HIS484, THR485, THR487, ALA488, SER490, ILE491, PHE494, PHE500, ARG502, GLU503, GLU504, VAL505, LYS506, THR507, SER513, PRO519, ILE520, LEU522, SER523, ARG529, VAL530, PRO531, ASP532, PRO533, CYC534, ALA552, ARG557, CYC558, LEU559, GLY560, ASN561, PRO562, CYC563, PRO567, GLU624, ARG632, ALA633, ARG635, HIS637, TRP638, THR639, GLN640, GLY641, TRP642, GLY643, ARG644, SER645, CYC646, ILE647.

Table 2. The only active binding site amino acid residues within the predicted OGG1 D2 structure using MOE site finder.

Biding Site	Residues
Binding site #1	ASN36, PHE37, LEU38, SER39, ALA40, SER41, THR42, SER43, GLY44, PRO45, CYC66, SER68, PRO69, PRO70, PRO71, HIS72, ARG73, ASN74, ALA75, PHE76, PRO77, LEU80, SER81, PRO82, THR83, SER84, PHE177, GLY178, GLY179, LYS180, GLN197, ALA198, CYS199, ALA200, TYR246, PRO257, ARG268, HIS269, PHE270, LEU271, LEU272, THR273, ILE276, ILE330, ASP331, VAL332, PHE333, ALA334, PRO337, MET338, LYS341, ALA342, PHE347, GLN348, GLY349, GLN350, ARG352, THR354, ARG355, TYR356, THR357, TRP358, ASP361, HIS383, GLN384, GLY385, SER386, PRO387, GLU388, GLN389, THR390, LEU391, ALA393, VAL394, LEU397, ASN400, THR403, SER404, CYS406, LEU407, GLN408, GLY410, GLU411, ALA412, ILE413, ASP421, GLU422, PRO423, LEU432, HIS433, THR438, ASN440, LYS451, CYC629, PRO630, PRO648, GLU649, PHE684, LEU685, LEU686, HIS687, ARG688, TYR690, SER777, THR778, MET779, PRO780.

Table 3. Uni Port BLAST search result showing similarity of OGG1 protein regions with other mammalian species.

Accession	Name	Protein name	Organism	Length	Identity	Score	E-Value
Q63450	KCC1A_RAT	Calcium/calmodulin-dependent protein kinase type 1	*Rattusnorvegicus* (Rat)	374	100.0%	234	1.0×10^{-1}
Q91YSA	KCC1A_MUS	Calcium/calmodulin-dependent protein kinase type 1	*Mus musculus* (Mouse)	374	100.0%	234	1.0×10^{-1}
Q14012	KCC1A_HUM	Calcium/calmodulin-dependent protein kinase type 1	*Homo sapiens* (Human)	374	100.0%	234	1.0×10^{-1}
Q08DQ1	Q08DQ1_BOVI	Calcium/calmodulin-dependent protein kinase I	*Bostaurus* (Bovine)	370	100.0%	234	1.0×10^{-1}
D2HRA9	D2HRA9_AIME	Putative uncharacterized protein	*Ailuropodamelanoleuca* (Giant panda)	370	100.0%	234	1.0×10^{-1}

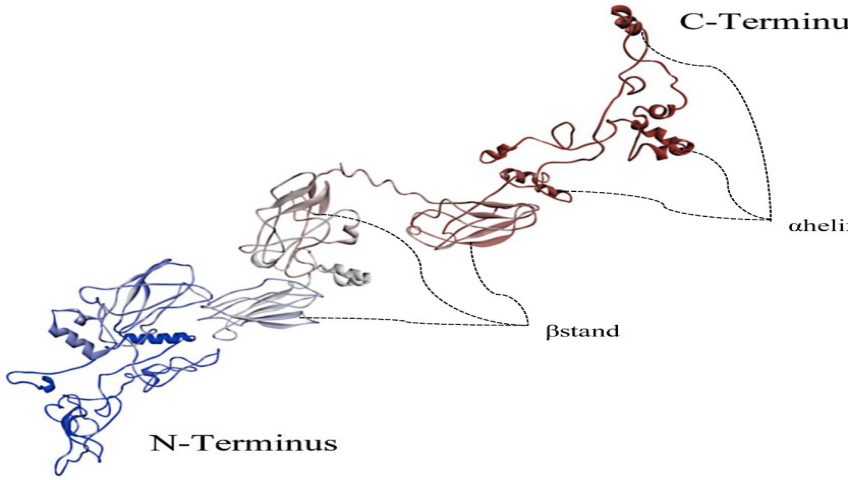

Figure 1. The protein structure of the domain *OGG1*; D1, (region: 1. 1-1000, length: 1000) is shown; the individual domains, colored as in the cartoon below, have been separated for visual clarity (red is C-terminus and blue is N-terminus). Dotted lines indicate the coordination of αhelix (in red), as well as another dotted indicate of βstand (D1) domain. Figure was generated with accelrys.

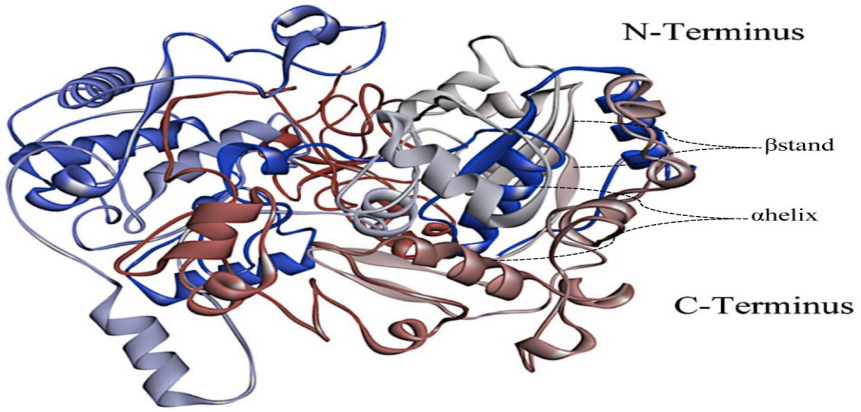

Figure 2. The protein structure of the domain *OGG1*; D2, region: 1000-1790 (790AA) is shown; the individual domains, colored as in the cartoon below, have been separated for visual clarity (red is C-terminus and blue is N-terminus). Dotted lines indicate the coordination of αhelix (in red), as well as another dotted indicate of βstand (D1) domain. Figure was generated with Accelrys.

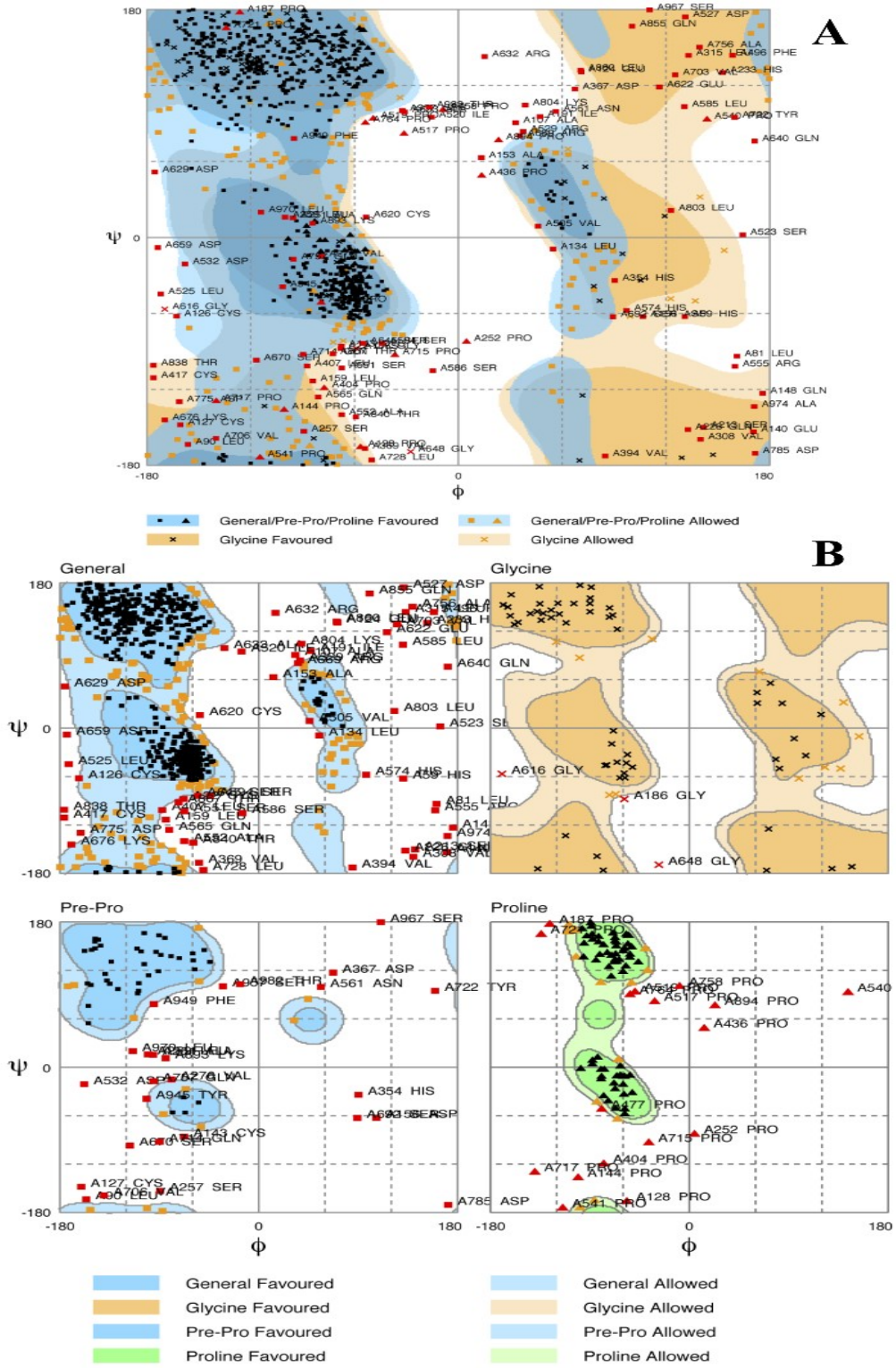

Figure 3 A-B. The OGG1 D1 protein residues.

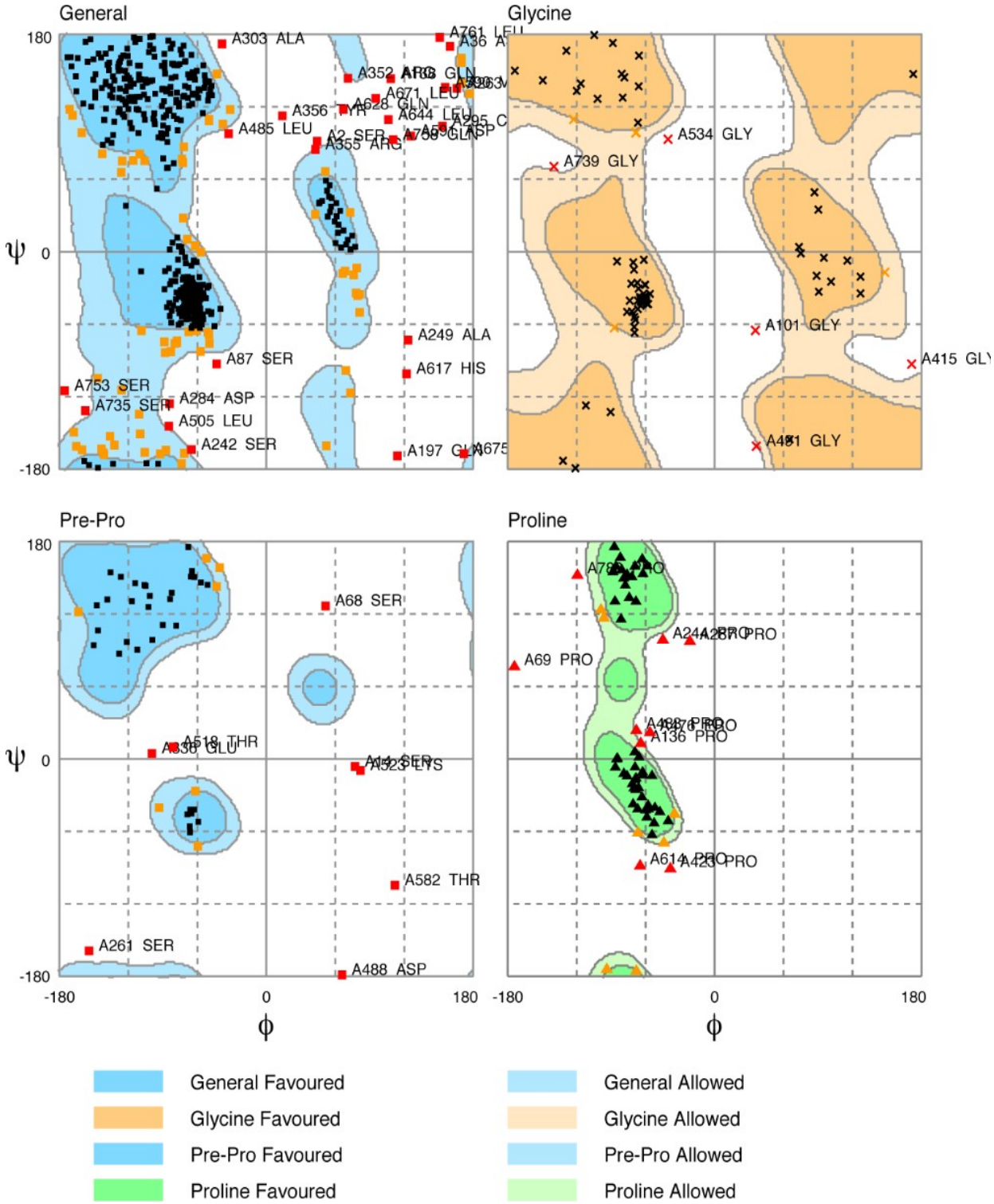

Figure 4. The OGG1 D2 protein residues.

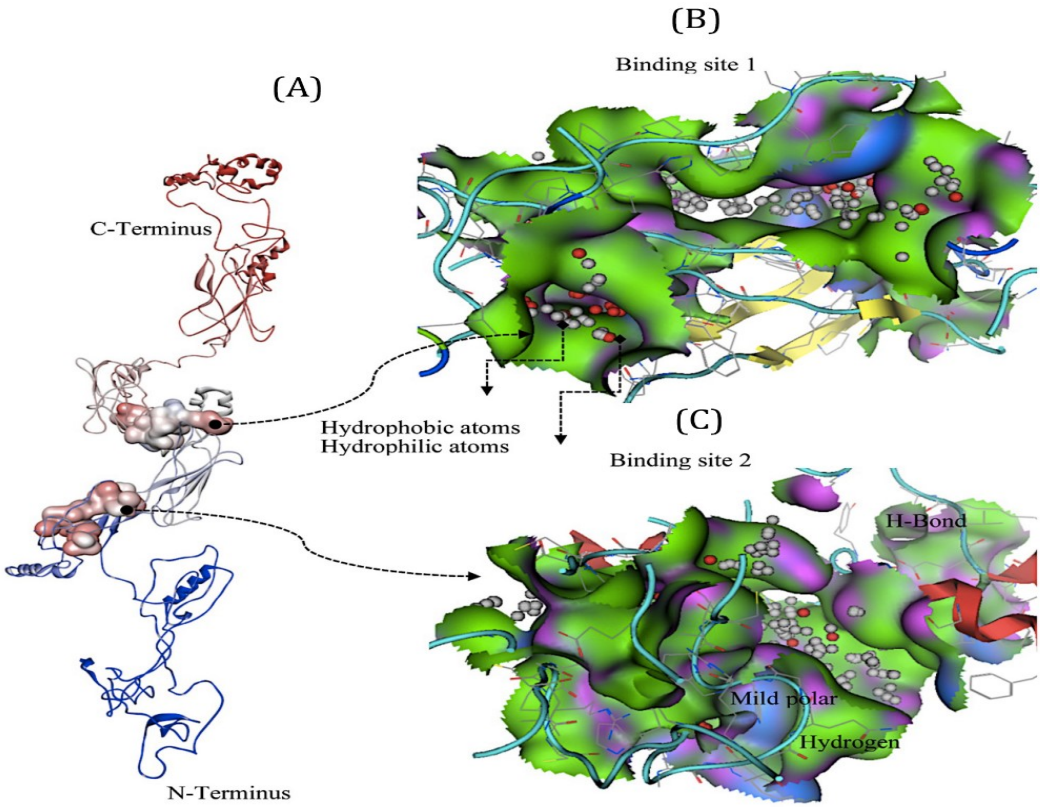

Figure 5. The active OGG1 binding sites. A) The protein structure of the domain *OGG1*; D1, region: 1. 1-1000, length: 1000), red color is C-terminus) and (color is blue N-terminus); B and C) the individual domains have 2 binding sites, separated for visual clarity (red sphere is Hydrophilic and White sphere is Hydrophobic). Dotted lines indicate the coordination of active site (in red), as well as another dotted line indicate active site in surface model for D1 domain and for another binding site (blue-mild polar; green-Hydrogen and pink-H-bond) using Molecular Operting Environment (MOE) and accelrys. Red color spheres indicate LPA (for long pair active), white color spheres HYD (for hydrophobic).

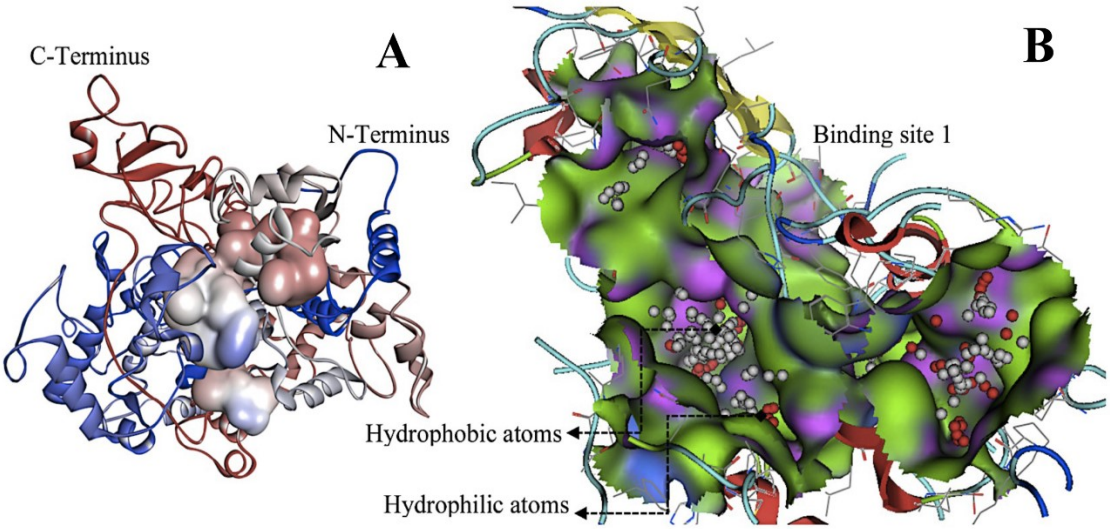

Figure 6. A) The protein structure of the domain D2 of *OGG1* region. The 1000-1790 (790AA) was predicted (red color showing C-terminus and blue color showing N-terminus); B) the individual domains have 1 binding site has been separated for visual clarity (red sphere indicating Hydrophilic region, whereas as the white sphere shows Hydrophobic region) using Molecular Operting Environment (MOE) and accelrys.

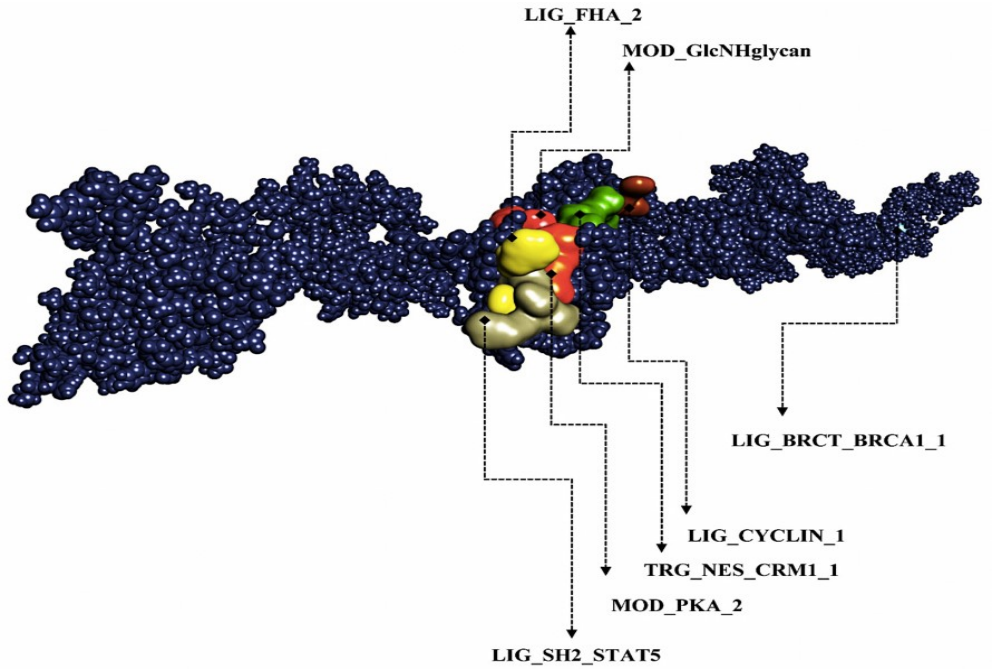

Figure 7. The *OGG1* (D1) predicted protein structure of showing some functional motifs in the molecular surface model. The individual functional domain of LIG_FHA_2 (yellow), MOD_GlcNHglycan (red), LIG_SH2_STAT5 (gray), MOD_PKA_2 (orange), TRG_NES_CRM1_1 (green), LIG_CYCLIN_1 (brown) and LIG_BRCT_BRCA1_1 (blue).

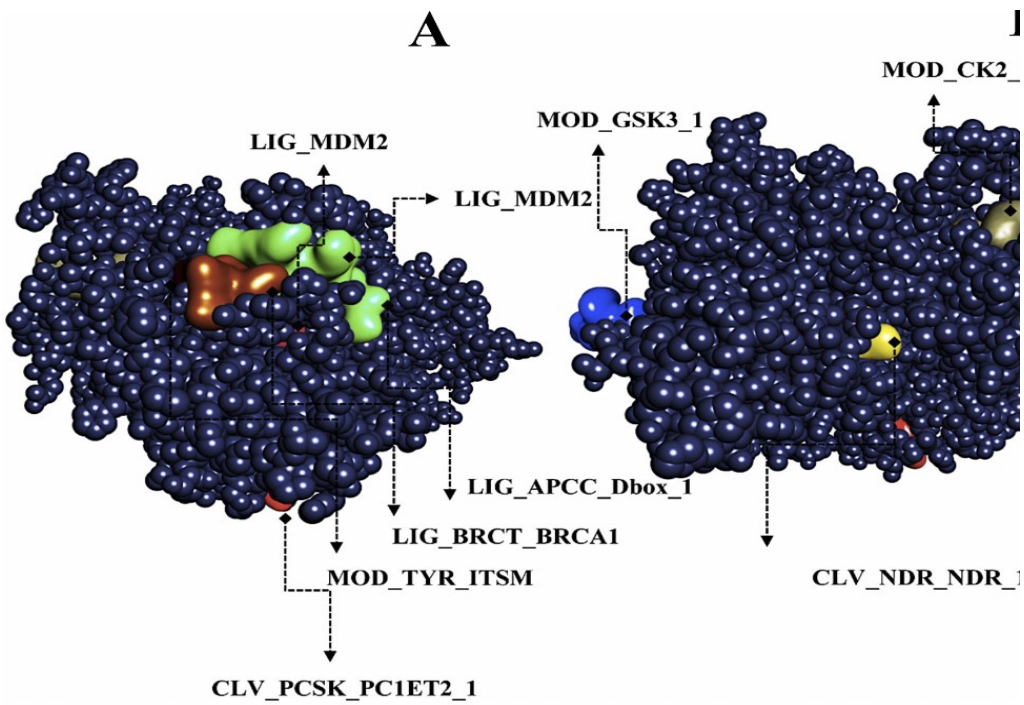

Figure 8. The predicted *OGG1* D2 protein structure showing different functional motifs, A) Front view of *OGG1* D2 surface showing predicted functional groups LIG_APCC_Dbox_1(green); LIG_MDM2 (rose); LIG_BRCT_BRCA1_1 (brown); MOD_TYR_ITSM (pale red); CLV_CK2_1 (red) and MOD_CK2_1 (gray). B) Back view of *OGG1* D2 surface indicating functional groups CLV_NDR_NDR_1(yellow) and MOD_GSK3_1 (blue).

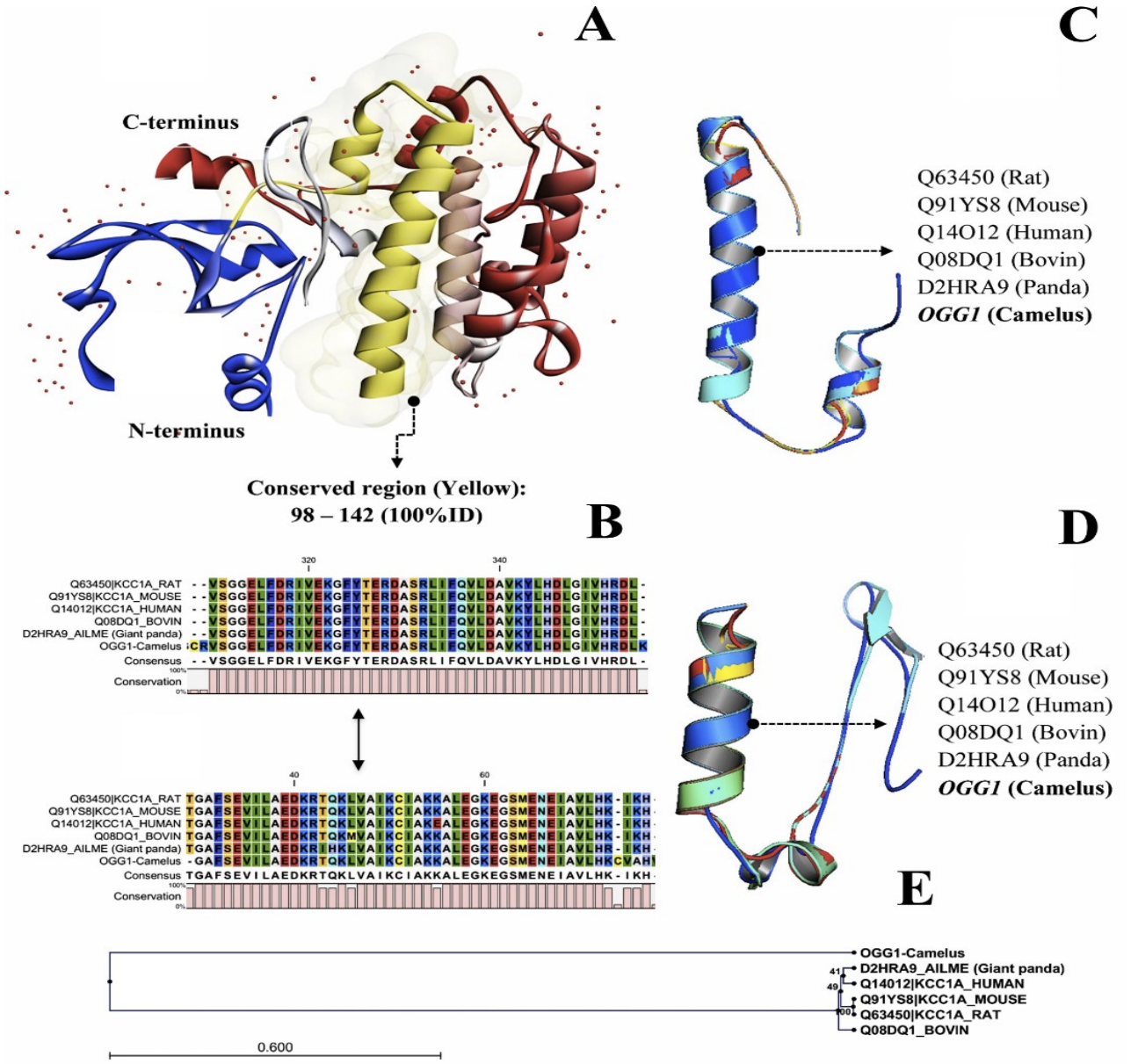

Figure 9. OGG1 Protein structure comparison with other species A)Calmodium- dependent protein kinase (1A06, Q63450) from rat (*Rattusnorvegicus*) showing similarity with OGG1 protein region B) Multiple alignment of OGG1 amino acid sequence with other mammalian species showing identity in the regions 1-45 (45AA) and 281-325 (45AA) of the OGG1 protein. C) 3D structure of OGG1 superimpose with the conserved region (281-324; 45AA) with 100.0% identity of CDPK (PDB: 1A06) rat (*Rattusnorvegicus*). D) Predicted OGG1 protein region superimpose with CDPK (PDB:1A06) region 98-142 (45AA) of rat (*Rattusnorvegicus*) showing 100.0% identity. E) Phylogenetic tree showing taxonomical classification of OGG1 protein sequence using the UPGMA.

DISCUSSION

Studying the DNA repair genes of Arabian camel is important for understanding their roles under extreme desert conditions and possible effect on the animal. The 8-oxoG is a damage lesion resulting due to the exposure of ROS that is repaired by OGG1 enzyme, one of the DNA glycosylase enzymes of the base excision repair mechanism. The structure and functional of OGG1 has previously been observed in several organisms. It is generally assumed that the oxidative DNA lesions are usually tackled by base excision repair (BER) pathway. This multistep repair pathway initiated by a specific DNA glycosylase which identifies and

eliminates the modified base leaving an AP site (apurinic/apyrimidinic site) that is potentially cytotoxic and mutagenic (Seeberg*et al.*, 1995). The 8-oxoG repair is one of the component of a multi-defense pathway, the Gene Ontology (GO) system, and consists of three enzymes; the glycosylases OGG1 and MYH (*MutY*homologue), and the hydrolase MTH (*MutT* homologue). The OGG1, a bifunctional glycosylase protects against mutagenesis through the exclusion of 8-oxoG from the 8-oxoG:C pair and also shows lyase activity, by targeting the basic site following the excision of the 8-oxodG bases (van der Kemp*et al.*, 1996). The OGG1 has previously been described from many eukaryotes and prokaryotes (Radicella*et al.*, 1997). Keeping in view the significance of OGG1 protein in protecting cells from ROS-induced mutagenesis, the role of OGG1 protein active sites was investigated through *in silico* studies, by studying its functional motifs and active sites. This is the first report that deals with the prediction of OGG1protein domains in *C. dromedaries* (Arabian camel). *In silico* analysis of the OGG1 protein has been previously been conducted in *Trypanosoma cruzi* and shown to have putative active sites (El-Sayed*et al.*, 2005). In this study, 1790 AA sequences of *C. dromedaries* OGG1 were used to predict its protein structure based on multiple alignments by Lometes&IllterativeITasser simulations. Because the full structure ofOGG1 protein cannot be predicted based on the homology modeling using conserved regions of other mammalian species, we predicted the 3D structure of two domains (D1 and D2) of the OGG1 protein. The possible active sites of a receptor from the 3D atomic coordinates are helpful for site-directed mutagenesis to look for potential sites for ligand binding docking (Goodford, 1985; Schechner*et al.*, 2004). The interaction energies were used to locate energetically favorable sites between the receptor and different probes. The van der Waals (VdW) energies pointed out sterically accessible regions, however the nature of electrostatic potentials may make the interpretation of energy levels difficult. Alternatively, pure geometric methods can be used to detect "pockets" without using energy models, which is advantageous because proton positions are then not required. LigSite(Hendlich*et al.*, 1997), the active site finder belonged to the category of geometric methods because it is not based on energy models. These techniques depend upon alpha shapes for generalization of convex hulls improvement and identify areas of tight atomic locating pockets, in addition to classifying the sites as hydrophobic and hydrophilic. The chemical sort separates the water sites from all possible hydrophobic sites.

The results showed that the OGG1 protein has several functional motifs predicted using eukaryotic liner motif resources (Lovell*et al.*, 2003). The results showed that the two conserved regions of the OGG1 protein (D1 and D2) showed the best C-scores -2.59 (D1) and 3.24 (D2) by Lometes simulation. The D1 and D2 protein confirmation was carried out by Ramachandran plot (RAMPAGE) to compare the residues allowed, disallowed and favored in the region. The structural validation information was also based on bond angle distortions. The *OGG1* gene D1 amino acid residues showed 98.0% favored region, 2.0% allowed region and 10.4% in outlier region. Whereas the D2 residues showed 98.0 favored regions, 2.0% allowed region and 6.2% outlier region. The functional groups of *C. dromedarius*showed highly specified conserved regions of OGG1 protein from different species for both D1 and D2. The results indicate that similar motifs with functional role may be involved. The D1 region had the most conserved AA residues such as the TRG_NES_CRM1_1 motif matched with amino acid residues DLGIVHRDLKVGGE (317-330) which is a functional site in certain proteins that are re-exported from the nucleus containing Leucine-rich nuclear export signal (NES) binding to the CRM1export among proteins in the nucleus. The MOD_PKA_2 motif matched with CRVSGGE (279-285) functional residue, which has preference for PKA-type AGC kinase phosphorylation in the cytosol, nucleus, and cAMP-dependent protein kinase complex. MOD_CDK_1 motifs matched with LVPTPGR (988-994) functional sites of the substrate motif for phosphorylation used by cyclin-dependent protein kinase holoenzyme complex in nucleus and cycloplasmic. The LIG_CYCLIN_1 motif has resemblance with amino acid residues RDLKV (323-327), a functional site frequently used for substrate recognition that interacts with cyclin for phosphorylation by cyclin/cdk complexes. The D2 also has a large conserved amino acid regions. The LIG_APCC_Dbox_1 residues matched with the amino acid residues GRYELAVLE (121-129) functional motif, which interacts with the Cdh1and Cdc20 components of APC/C directing the protein for destruction in a cell cycle dependent manner. The LIG_MDM2 motifs corresponded with amino acid residues FSFLWRGL (253-260) functional residues of p53 family members, which shows binding to the N-terminal domain of the MDM2 protein in the nucleus. The MOD_CK2_1 residues matched with the amino acid residues SVHSAVE (615-621) functional motif of CK2 phosphorylation region and found in kinase CK2 complexes in nucleus, cytosol and protein. The MOD_GSK3_1 motifs matched with amino acid residues SQGSPPRS (11-18) the functional sites of GSK3 phosphorylation in the cytosol. In general, there was only one major functional group that was conserved in each domain (i.e. D1 and D2). In the D1, LIG_BRCT_BRCA1_1 motif matched with amino acid residues LSFLF (896-900) site, where as in D2 it matched with amino acid residues ASKTF (249-253), which is functional site for Phosphopeptide site and

interacts with the BRCT (carboxy- terminal) domain of the Breast Cancer Gene BRCA1 with low affinity and BRCA1-BARD1 complex. The binding pocket within BRCT domains specially recognizes F and has consensus sequence S. High affinity motif showing binding with lysine residues S.F.K. The phospho-protein facilitated interaction of the BRCT domain has a fundamental role in cell cycle checkpoint and DNA repair mechanism. Our results suggest that similar motifs with functional role may be involved and could play a major part in disease management strategies. Such functional sites may be used in target screening to identify protein residues that may be associated with a disease process (Weston and Weidolf, 2012). Overall, the comparative analysis of OGG1 protein sequence showed 100.0% identity with other mammalian species' protein regions for example amino acids from 98 to 143 of Calcium/Calmodulin-dependent protein kinase (CDPKs, PDB: 1A06) from *rattusnorvegicus*. Likewise, the calcium-signaling proteins related to calmodulin-dependent proteins (CaMK) that belongs to a large family of serine/threonine kinases, where CaMKs are activated by calmodulin-calcium complex also showed similarity. It has been observed that the kinase domain of a CDPK may be active through interaction of C-terminally located calcium binding domain (Zhang*et al.*, 2010). We also carried out the phylogenetic analysis of the OGG1 protein domains (D1 and D2) based on protein sequences from other mammalian species. Based on our results a Predict tree was built using UPGMA algorithm (Horikawa*et al.*, 1973) following alignment of sequences to look for evolutionary relationships among various mammalian species. To the best of our knowledge this is the first report that deals with the predicted protein structure of OGG1 and its functional motifs and binding sites from Arabian camel.

Conclusions: There is no published report that deals with the OGG1 structure prediction and functional motifs with its active sites. This study successfully predicted the OGG1 protein structure's two domains D1 and D2. Furthermore, we have also predicted the proteins' active sites and functional groups and compared them with the conserved regions found in other mammalian species. The results of this study may be useful and will help to cure OGG1 related diseases in human and other species. Such information is essential and common among biotechnology and pharma based studies for drug target designing. Taken together, the results showed that the OGG1 from Arabian camel has many important functional groups motifs for strong binding and interaction with OGG1 domains.

Acknowledgments: The Authors extend their appreciation to the Deanship of Scientific Research at King Saud University for funding the work through the research group project No: RGP-VPP-200.

REFERENCES

Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D.J. Lipman (1997). Gapped blast and psi-blast: A new generation of protein database search programs. Nucleic Acids Res. 25(17): 3389-3402.

Boiteux, S. and J.P. Radicella (2000). The human ogg1 gene: Structure, functions, and its implication in the process of carcinogenesis. Arch. Biochem. Biophys. 377(1): 1-8.

Clapperton, J.A., I.A. Manke, D.M. Lowery, T. Ho, L.F. Haire, M.B. Yaffe and S.J. Smerdon (2004). Structure and mechanism of brca1 brct domain recognition of phosphorylated bach1 with implications for cancer. Nat. Struct. Mol. Biol. 11(6): 512-518.

Edelsbrunner, H., M. Facello, P. Fu and J. Liang(1995). Measuring proteins and voids in proteins. In: System Sciences, 1995. Proceedings of the Twenty-Eighth Hawaii International Conference on. IEEE: pp: 256-264.

El-Sayed, N.M., P.J. Myler, D.C. Bartholomeu, D. Nilsson, G. Aggarwal, A.N. Tran, E. Ghedin, E.A. Worthey, A.L. Delcher, G. Blandin, S.J. Westenberger, E. Caler, G.C. Cerqueira, C. Branche, B. Haas, A. Anupama, E. Arner, L. Aslund, P. Attipoe, E. Bontempi, F. Bringaud, P. Burton, E. Cadag, D.A. Campbell, M. Carrington, J. Crabtree, H. Darban, J.F. da Silveira, P. de Jong, K. Edwards, P.T. Englund, G. Fazelina, T. Feldblyum, M. Ferella, A.C. Frasch, K. Gull, D. Horn, L. Hou, Y. Huang, E. Kindlund, M. Klingbeil, S. Kluge, H. Koo, D. Lacerda, M.J. Levin, H. Lorenzi, T. Louie, C.R. Machado, R. McCulloch, A. McKenna, Y. Mizuno, J.C. Mottram, S. Nelson, S. Ochaya, K. Osoegawa, G. Pai, M. Parsons, M. Pentony, U. Pettersson, M. Pop, J.L. Ramirez, J. Rinta, L. Robertson, S.L. Salzberg, D.O. Sanchez, A. Seyler, R. Sharma, J. Shetty, A.J. Simpson, E. Sisk, M.T. Tammi, R. Tarleton, S. Teixeira, S. Van Aken, C. Vogt, P.N. Ward, B. Wickstead, J. Wortman, O. White, C.M. Fraser, K.D. Stuart and B. Andersson (2005). The genome sequence of trypanosoma cruzi, etiologic agent of chagas disease. Science. 309(5733): 409-415.

Glover, J.N., R.S. Williams and M.S. Lee (2004). Interactions between brct repeats and phosphoproteins: Tangled up in two. Trends Biochem. Sci. 29(11): 579-585.

Goodford, P.J. (1985). A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J. Med. Chem. 28(7): 849-857.

Gould, C.M., F. Diella, A. Via, P. Puntervoll, C. Gemund, S. Chabanis-Davidson, S. Michael, A. Sayadi, J.C. Bryne, C. Chica, M. Seiler, N.E. Davey, N. Haslam, R.J. Weatheritt, A. Budd, T. Hughes, J. Pas, L. Rychlewski, G. Trave, R. Aasland, M. Helmer-Citterich, R. Linding and T.J. Gibson (2010). Elm: The status of the 2010 eukaryotic linear motif resource. Nucleic Acids Res. 38(Database issue): D167-180.

Hendlich, M., F. Rippmann and G. Barnickel (1997). Ligsite: Automatic and efficient detection of potential small molecule-binding sites in proteins. J. Mol. Graph. Model. 15(6): 359-363, 389.

Horikawa, Y., T. Tsubaki and M. Nakajima (1973). Rubella antibody in multiple sclerosis. Lancet. 1(7810): 996-997.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292(2): 195-202.

Kinch, L.N. and N.V. Grishin (2002). Evolution of protein structures and functions. Curr. Opin. Struct. Biol. 12(3): 400-408.

Klungland, A. and S. Bjelland (2007). Oxidative damage to purines in DNA: role of mammalian Ogg1. *DNA Repair* (Amst.)6, 481–488.

Lovell, S.C., I.W. Davis, W.B. Arendall, 3rd, P.I. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson and D.C. Richardson (2003). Structure validation by calpha geometry: Phi,psi and cbeta deviation. Proteins. 50(3): 437-450.

Pauling, L. and R.B. Corey (1951). Configurations of polypeptide chains with favored orientations around single bonds: Two new pleated sheets. Proc. Natl. Acad. Sc.i U. S. A. 37(11): 729-740.

Petrie, K.L. and G.F. Joyce (2010). Deep sequencing analysis of mutations resulting from the incorporation of dntp analogs. Nucleic Acids Res. 38(22): 8095-8104.

Radicella, J.P., C. Dherin, C. Desmaze, M.S. Fox and S. Boiteux (1997). Cloning and characterization of hogg1, a human homolog of the ogg1 gene of saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U. S. A. 94(15): 8010-8015.

Roy, A., A. Kucukural and Y. Zhang (2010). I-tasser: A unified platform for automated protein structure and function prediction. Nat. Protoc. 5(4): 725-738.

Rychlewski, L., B. Zhang and A. Godzik (1998). Fold and function predictions for mycoplasma genitalium proteins. Fold. Des. 3(4): 229-238.

Schechner, M., F. Sirockin, R.H. Stote and A.P. Dejaegere (2004). Functionality maps of the atp binding site of DNA gyrase b: Generation of a consensus model of ligand binding. J. Med. Chem. 47(18): 4373-4390.

Schwede, T., J. Kopp, N. Guex and M.C. Peitsch (2003). Swiss-model: An automated protein homology-modeling server. Nucleic Acids Res. 31(13): 3381-3385.

Seeberg, E., L. Eide and M. Bjoras (1995). The base excision repair pathway. Trends Biochem. Sci. 20(10): 391-397.

Shi, J., T.L. Blundell and K. Mizuguchi (2001). Fugue: Sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. J. Mol. Biol. 310(1): 243-257.

Simons, K.T., C. Kooperberg, E. Huang and D. Baker (1997). Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions. J. Mol. Biol. 268(1): 209-225.

Soding, J. (2005). Protein homology detection by hmm-hmm comparison. Bioinformatics. 21(7): 951-960.

Thompson, J.D., D. G. Higgins and T.J. Gibson (1994). Clustal w: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22(22): 4673-4680.

UniProt Consortium (2010). The universal protein resource (uniprot) in 2010. Nucleic Acids Res. 38(Database issue): D142-148.

van der Kemp, P.A., D. Thomas, R. Barbey, R. de Oliveira and S. Boiteux (1996). Cloning and expression in escherichia coli of the ogg1 gene of saccharomyces cerevisiae, which codes for a DNA glycosylase that excises 7,8-dihydro-8-oxoguanine and 2,6-diamino-4-hydroxy-5-n-methylformamidopyrimidine. Proc. Natl. Acad. Sci. U S A. 93(11): 5197-5202.

Weston, D.J. and L. Weidolf (2012). Conference report: High-resolution ms in drug discovery and development: Current applications and future perspectives. Bioanalysis. 4(5): 481-486.

Wu, S., J. Skolnick and Y. Zhang (2007). Ab initio modeling of small proteins by iterative tasser simulations. BMC Biol. 5: 17.

Wu, S. and Y. Zhang (2007). Lomets: A local meta-threading-server for protein structure prediction. Nucleic Acids Res. 35(10): 3375-3382.

Wu, S. and Y. Zhang (2008). Muster: Improving protein sequence profile-profile alignments by using multiple sources of structure information. Proteins. 72(2): 547-556.

Xu, J., M. Li, G. Lin, D. Kim and Y. Xu (2003). Protein threading by linear programming. Pac. Symp. Biocomput.: 264-275.

Xu, Y. and D. Xu (2000). Protein threading using prospect: Design and evaluation. Proteins. 40(3): 343-354.

Yam, B. Z. A., & Khomeiri, M. (2015). Introduction to Camel origin, history, raising, characteristics, and wool, hair and skin: A Review. Research J. Agriculture and Environmental Management, 4(11), 496-508.

Zhang, Y. (2008). I-tasser server for protein 3d structure prediction. BMC Bioinformatics. 9: 40.

Zhang, Y., H. Tan, G. Chen and Z. Jia (2010). Investigating the disorder-order transition of calmodulin binding domain upon binding calmodulin using molecular dynamics simulation. J. Mol. Recognit. 23(4): 360-368.

Zhou, H. and Y. Zhou (2005). Fold recognition by combining sequence profiles derived from evolution and from depth-dependent structural alignment of fragments. Proteins. 58(2): 321-328.