ORIGINAL ARTICLE

Structural Phylogenomics: Selection Pressure Suggests the Functionally Important Residues Encoded by Cisplatin Resistance Related 9 Gene

Saadia Shakir,[1] Waqas Ahmed Waseem,[1] Syed Khaqan Hasan,[1] Muhammad Munir,[2] Muhammad Rafiq Khanani[2] and Mushtaq Hussain[2]

ABSTRACT

Cisplatin Resistance Related 9 gene, CRR9, contributes towards the efficacy of the chemotherapeutic drug namely cisplatin. Genetic studies have established the association between CRR9 gene mutations and several cancers, but the structure-function aspects of the encoded protein remains largely unaddressed. In the present study, we have constructed a consensus phylogenetic tree of the CRR9 gene using maximum likelihood method after 1000 bootstrap replicates. Multiple sequence alignment of the selected orthologs was undertaken and important variations were analyzed with reference to the clade segregations and spatial locations in the constructed protein structure models. The topology of the phylogenetic tree appears in line with the established phylogenetic relationship of the mammalian lineage. The protein models of selected mammalian representatives suggest strong uniformity as Root Mean Square Deviation which varies from 0.03Å to 0.14Å. Both the DAS server and protein structure suggest the presence of two novel transmembrane regions ranging from Val461 to Ala483 and Thr490 to Tyr506. Multiple sequence alignment of the protein showed primate specific amino acid substitutions. Importantly, these variations are mostly situated in the core part of the protein structure implying their structural and/or functional significance. Conclusively, the present study of structural phylogenomics approach, not only illustrate the architecture of CRR9 protein but also delineate the critically important amino acids of possible structural and/or functional importance. Further studies in the direction of the site direction mutagenesis verify our finding and assist in functional understanding of the protein. Additionally, it also allows to contemplate new drugs for chemotherapy using potentials of CRR9 gene.

Key words: CRR9, cisplatin, structural phylogenomics, molecular evolution.

INTRODUCTION

Since its approval in 1978, cisplatin or cisdiaminodichloroplatinium II (CDDP) is considered as one of the most effective chemotherapeutic drug used in the treatment of variety of cancers for instance malignant tumors of testis, head and neck, oesophagus, lungs, ovaries and bladder.[1] Briefly, the mode of action of the drug suggest that cisplatin cross links with the DNA by interacting with its central platinium atom with the nitrogen base, preferably with the guanine bases, in the DNA.[2] The cross linked cisplatin mediated DNA adducts elicits cellular stress response particularly the DNA repair mechanism, which in turns lead to the apoptosis of the treated cell. Though details of cisplatin induced apoptosis has not been unraveled in greater detail, however, a mitochondrial serine protease namely Omi/Htra2 has been considered to be involved. Additionally, a high mobility proteins (HMG) has also known to play some role in the cytotoxicity of the drug.[3] In addition to that, several other cellular proteins like PARP-1, DNA ligase III, XRCC1 and hMutSâ are also found to interact with the cross linked anti-parallel strands established by cisplatin.[4-5] Though cisplatin is still considered as an effective chemotherapeutic drug but its long term usage has been found hampered by the development of the drug resistance in the treated malignant cells. Some studies suggest that a paralog of ATP-binding cassette (ABC) transporters, multidrug resistance-associated protein (MRP), is involved in the development of resistance in the cancer cells by causing the ATP dependent efflux of the molecules to minimize the accumulation of the drug in the target cells.[6] However, other studies have not substantiated these findings.[7] Alternatively, in CDDP-resistant cells, the loss of molecules associated with DNA repair mechanism like MLH1 and MSH6 has been accounted responsible for the emergence of cisplatin resistance.[7] Yamamoto et al., 2001[8] reported upregulation of a novel gene namely cisplatin resistance regulated gene

Bioinformatics Laboratory, Institute of Pharmaceutical and Environmental Research (IPER),[1] Department of Molecular Pathology,[2] Dow University of Health Sciences, Karachi, Pakistan.

Correspondence: Mushtaq Hussain, Lab-412, Level 4, Davidson Building University of Glasgow, Glasgow-G128QQ, UK.

Email: mushtaq.hussain@duhs.edu.pk

(CRR9) in the CDDP resistant ovarian tumor cell line. In contrast to the name, the study found that the gene is involved in the CDDP induced apoptosis rather than the conferring resistance against it. The gene, CRR9, is also referred as cleft lip and palate transmembrane 1 like (CLPTM1L) and several genome wide studies have established association of its polymorphism with lung cancer[9-11] and testicular germ cell cancer.[12] Though genetic studies have conspicuously established the association of CRR9 with several cancers, the actual functions of the protein encoded by the gene remain elusive.

In the present study, we have constructed the phylogenetic tree and multiple sequence alignment of CRR9 gene using gene and protein sequences of the different orthoglogs of the selected mammals. We have also proposed a potential tertiary structure of the protein and investigated various structural aspects of the molecules which were not known earlier. Collectively, our finding not only proposed the structural conformation and evolutionary lineages of the CRR9 protein but also suggest the critical residues which may have structural and functional significance in the biological role of the molecule.

MATERIALS AND METHODS

Data mining: DNA and protein sequences of CRR9 gene were retrieved from NCBI [http://www.ncbi.nlm.nhi.gov] and ENSEMBL [http://www.ensembl.org] databases using BLAST [blastn and blastp] with default parameters. The sequences were verified from the genomic information and as well in many cases from the published literature.

Multiple Sequence Alignment: Multiple sequence alignment of CRR9 proteins was carried out using CLUSTALX under default parameters.[13] Wherever necessary (need to be more specific where adjustments were made in the alignment? Please see the mail), manual adjustments in the alignment were conducted. The alignment file was visualized by CLC sequence viewer and identity/similarity matrix plot was developed using GeneDoc.

Phylogenetic Analysis: Total 17 cDNA sequences including the out group of Danio rerio were used to construct the tree by maximum likelihood method based on Kimura-2 evolutionary model[14] using MEGA5.01.[15] Sequences with incomplete information or in some cases not present in both the mentioned databases (NCBI and ENSEMBL (already mentioned in the data mining section) were excluded from the analysis. The consensus tree was developed from 1000 bootstrap replicates[16] to construct the evolutionary history of the taxa under study. The molecular clock hypothesis was rejected by invoking the Gamma distribution to account the difference in the evolutionary rates among different sites. Nearest Neighbor Interchange heuristic method was selected to generate original trees.

Protein Homology Modeling: Selected protein sequences were subjected to protein blast for homologus structures using NCBI database. Residues at conserved domain were verified using CDD and CD search Databases. Atomic coordinates of homologus structures were retrieved from RCSB protein databank.[17] Models of human CRR9 and selected mammalian orthologs were constructed using appropriate template(s) from i-Tasser taken spatial restraints of templates into account.[18] Briefly, i-Tasser constructs the atomic model of the query protein using multiple threading alignments and iterative structural assembly simulations. To further strengthen the acceptability of the model, another structure was constructed using the same templates (short stretches with 10-25 frame width) with Molecular Operating Environment (MOE). Both models were used to construct several plausible models using Modeller 9.10.[19] The models were refined on the basis of thermodynamical parameters like free energy calculation and structural attributes for instance Ramachandran plots (di-hedral angle ratio) and clash score using Swiss Pdb viewer[20] and Molprobity.[21]

RESULTS AND DISCUSSION

Data mining has shown that CRR9 gene is widely distributed among the genomes of animals ranging from eutherian mammals to the even members of invertebrates, suggesting to their house keeping role in the organism physiology. The evolutionary tree based on nucleotide sequence of CRR9 gene was found well supported by the high bootstrap values especially among the major clades of primates, rodentia and artiodactyla. The phylogenetic tree corresponds to the established lineage of mammalian evolution both in terms of common ancestry and time scale divergence[22] (Fig.1). Holistically, the out group (Danio) is well separated from the mammalian lineages. However, a little unorthodox position of the marsupial mammal (Monodelphis) may be due to the presence of the relatively less representative taxa or slow evolutionary rate in comparison to the rest of mammalian lineages.[23]

The basal time of diversification in terms of rest of mammalian lineage appears to be around 150 MYA (million years ago), which are very close to the earlier studies notifying the basal diversification time of 166.2

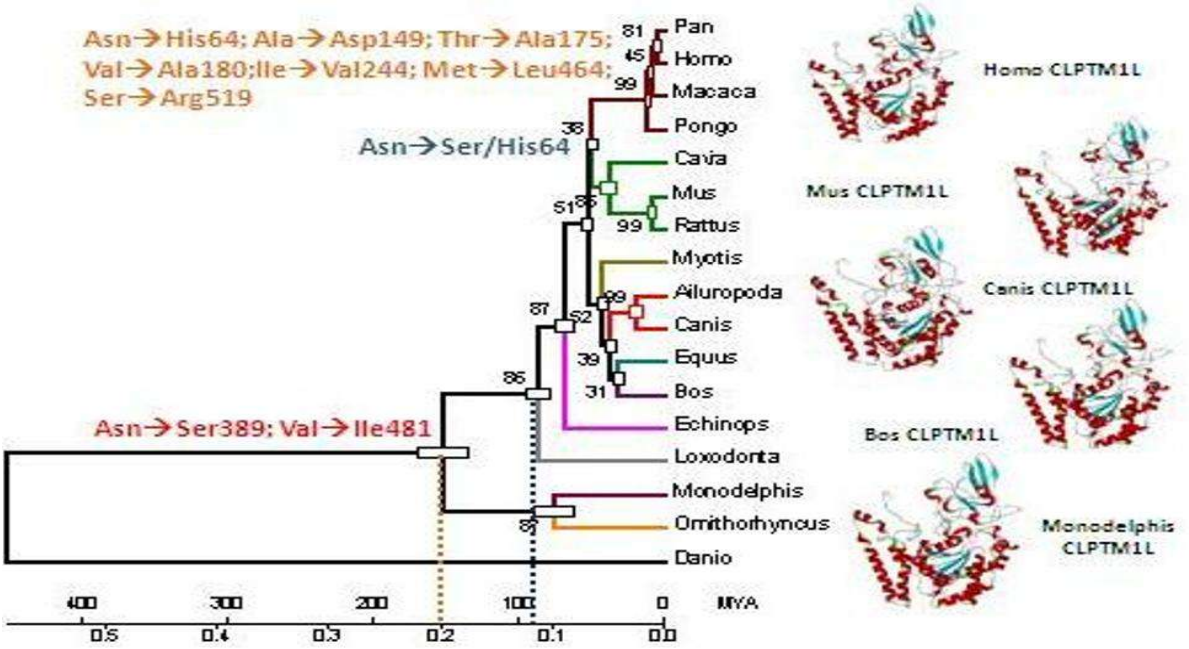
Asn→His64; Ala→Asp149; Thr→Ala175;
Val→Ala180;Ile→Val244; Met→Leu464;
Ser→Arg519
Asn→Ser/His64
Asn→Ser389; Val→Ile481
Pan
Homo
Macaca
Pongo
Cavia
Mus
Rattus
Myotis
Ailuropoda
Canis
Equus
Bos
Echinops
Loxodonta
Monodelphis
Ornithorhynchus
Danio
81
45
99
38
51
99
87
52
39
31
86
88
400
300
200
100
0
MYA
0.5
0.4
0.3
0.2
0.1
0.0
Homo CLPTM1L
Mus CLPTM1L
Canis CLPTM1L
Bos CLPTM1L
Monodelphis CLPTM1L

Fig.1

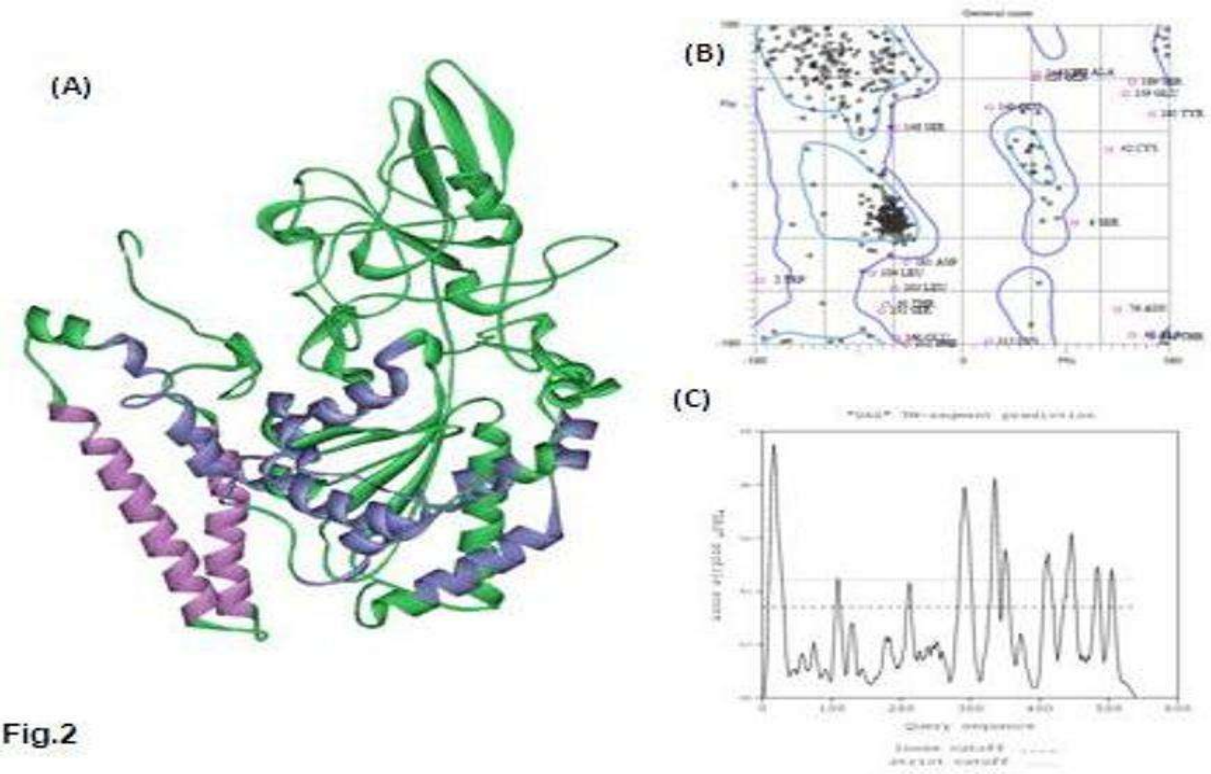
(A)
(B)
(C)

Fig.2

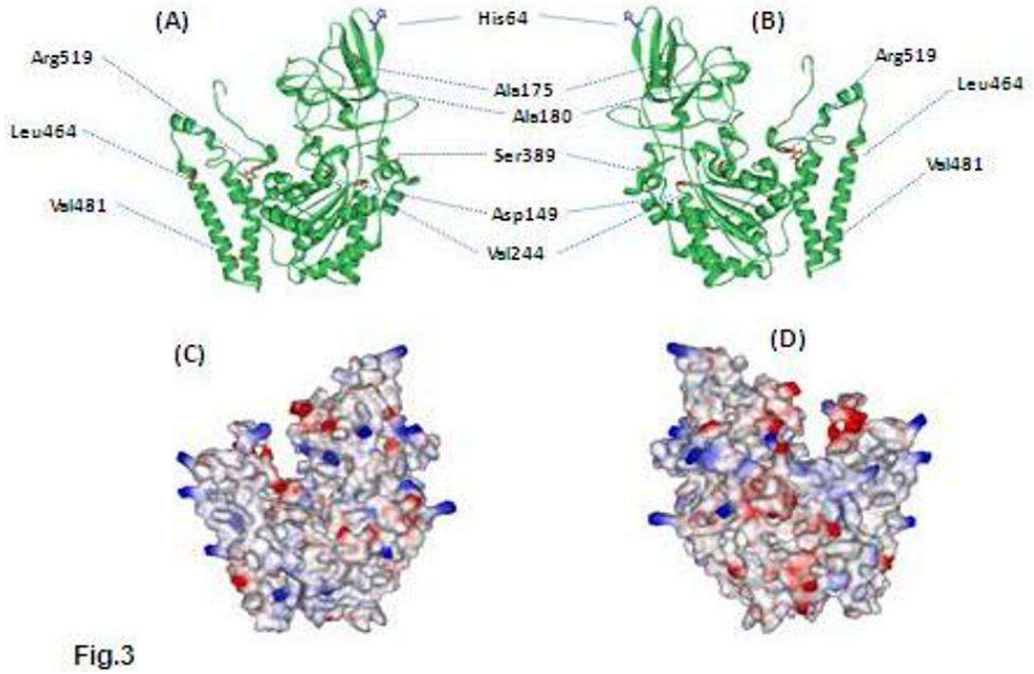

Fig.3

MYA.[24] Albeit, the placental mammals appears to be separated from the non placental mammals around 150 MYA but their diversification, so called crown group origin, appeared to be around 90 MYA, which suggests the presence of increased density of the extant mammalian lineage before the K/T boundary extinction (65 MYA), marked by the extinction of Dinosaur and most of the marine fauna. Interestingly, our results not only relegates the more conventional stand point related to the diversification of mammals based of the dates deduced by paleontological evidences[25-26] but also demands the exploration and/or proposal for different rationale related to the origin of mammalian diversity. It is generally proposed that extinction of the dinosaur around 65 MYA leads to the removal of purifying selection which consequently favors the speciation of mammals. However, in light of our findings and many other molecular data, radiation of mammals at least to crown group extent many have occurred before 65MYA.[27-28] Indeed, the pre K/T boundry extinction period (around 90 MYA) is notable in terms of many profound geo-climatic changes. Most notable among such events are rise of angiosperms (fruit bearing relatively short heighted plants), oceanic anoxia (93 MYA) and significant reduction in the earth temperature.[28] If excluded the chance of the co-incidence, it is conceivable to hypothesize that the rise of angiosperms may compete the gymnosperms (seed bearing relatively tall plants), which may have adversely affected the population of pre dominant group (dinosaurs) at that time allowing mammals to radiate under relatively lenient selection pressure. Moreover, Springer et al., (2003)[28] proposed the 43 lineages of placental mammals that have successfully passed through the K/T mass extinction.

When compared with the multiple sequence alignment (data not shown), several clade specific amino acid substitutions were noted among the compared orthologs. For instance all eutherian mammals have a substitution at position 389 and 481 where Asn was found replaced by Ser and Val was replaced by Ile, respectively. Similarly, in the rodentia specific lineage Asn64 is endemically replaced by Ser. A significantly high number of the substitutions have been noted in the primate clade where 7 substitutions were observed as Asn64His, Ala149Asp, Thr175Ala, Val180Ala, Ile244Val, Met464Leu and Ser519Arg. The clade specific substitutions indicate the presence of specific selection pressure that sculpted and subsequently locked the protein in the specific residual permutation. The presence of CRR9 homologs among invertebrates and specific changes as being seized in the primate lineage could collectively be inferred in terms of acquiring additional functions in addition to the basic one rendered by the mentioned substitutions. Combining the phylogenetic information with the amino acid sequence alignment has recently being used to explore the structurally and functionally critical residues in several proteins.[29-31]

To further ascertain our predictions of critical residues in the CRR9 protein, we have constructed the protein model of CRR9 by multiple threading alignment and iterative structural assembly simulations. After removing the inter atomic clashes by manual rotation of the residual side chains, without altering the Cá back bone, most of the residues in the secondary structure component of the proteins congregated in the allowed region in Ramachandran plot which signify the plausibility of the model. Collectively, for constructing the model using i-Tasser and MOE, 10 templates were used namely transferrin receptor (1de4),[32] Drosophila apoptosome (1vt4),[33] Human a glucuronidase (1bhg),[34] Escherichia coli DNA polymerase II (1q8i), Karyopherin nuclear transport complex (1qbk),[35] regulatory domain of human PP2A (1b3u),[36] C protein of Listera monocytogenes (1y9i),[37] angiogenic complex (1a4y)[38] and alginate lyase of Agrobacterium (3a0o).[39] Collectively the protein has two main domain, a C-terminal core domain which is mainly composed of á-helices and the N-terminal extended loop domain with intervening coils and â sheets. In CRR9 protein, six trans-membrane helices have already proposed ranging from Leu11-Val31, Tyr285-Phe305, Ala325-Asp342, Leu347-Val364, Tyr403-Ile423 and Tyr429-Leu449.[40]

However, DAS transmembrane prediction server has predicted some additional transmembrane regions ranging from Val481-Ile486 and Phe503-Tyr506.[41] In our protein model, we noticed that indeed such residues along with the amino acids present in their proximity could attain the helical conformations ranging from Val461-Ala483 and Thr490-Tyr506 and could consequently localized within the cell membrane. To the best of our information, the present study is the first in terms of proposing the two additional (novel) transmembrane helices in CRR9. Considering the number of helices and possible conformation of the protein it is likely that the protein may act as a channel for the drug molecules and may also have some additional role of structural nature in the membrane of the cells (Fig. 2).

To explore the effect of clade specific amino acid substitution on the holistic architecture of the CRR9 protein, homology models of representative of each clade (Human (Homo sapiens), Mouse (Mus musculus), Dog (Canis lupus), Cow (Bos taurus) and Opposum (Monodelphis domestica)) were built using the refined human CRR9 structure as a template. All CRR9 structures were superimposed over each other in order to delineate structural polymorphism among the models. Root Mean Square Deviation (RMSD) among the compared structures was found between 0.03Å in case of Monodelphis and Bos and 0.14Å in case of Mus and Homo. Collectively, the difference as shown by the RMSD is not significantly high suggesting that the changes in the primary structure of the CRR9 are not translated into the tertiary structure. However, the RMSD values between human and opossum (0.11Å) suggest relatively fast rate of structural evolution of CRR9 from marsupials to primates as compared to monotrem and bovidae (Cattle family; Cow; Bos) and carnivores (Dog family; Dog; Canis) where RMSD is about 0.03Å (Fig.1,2). Additionally, the strong conservation of the secondary structure also implicates towards the housekeeping role of the CRR9 gene. Spatially, in human CRR9, out of the seven critical amino acids substitutions as found in the primate specific lineage four (Asp149, Val244, Leu464 and Arg519) are stationed in the core C-terminal part of the protein suggesting its main functional role however, three substitutions like His64, Ala175 and Ala180 which are present in the extended N-terminal region may be involved in some signal transduction or ligand binding due to their potential extracellular or intracellular localization respectively.

Briefly, in the present script, we have not only proposed the potential tertiary structure of the CRR9 protein but have also suggested the structurally and/or functionally important residues in the protein. Additionally, we have also mentioned about the novel structural elements (transmembrane regions) in the said protein. We expect that our findings will facilitate the site directed mutagenesis studies in order to establish the functions of CRR9 empirically and also aid in designing novel drug that can potentially use CRR9 molecular cascade for their efficacy in the treatment of respective gene associated anomalies.

Figure Legends:

Fig.1. Evolutionary tree based on CRR9 nucleotide sequence: Consensus phylogenetic tree was constructed using maximum likelihood method with bootstrap values mentioned at the nodal branches of the tree. Taxonomic distinct groups are differentially colored. Also presented are the proposed models of CRR9 protein of the representative of each clade. The evolutionary time scale is present at the bottom of the tree. Along with each clades lineage specific amino acid substitutions are also mentioned.

Fig.2. Structure of human CRR9 protein (A), Rama chandran plot (B) and DAS transmembrane server prediction (C). Potentially novel transmembrane helices are indicated with purple color while known trans-membrane helices are colored blue. Extracellular/ Intracelluar extended N-terminal domain is represented by green color.

Fig.3. Distribution of potentially critical amino acids in human CRR9 protein (A & B). Electrostatic surface topology of the CRR9 protein. Positively and negatively charged regions with blue and red color respectively (C & D).

REFERENCES

1 Jamieson ER, Lippard SJ. Structure, Recognition, and Processing of Cisplatin-DNA Adducts. Chem Rev 1999; 99:2467-98.

2 Alderden RA, Mellor HR, Modok S, Hambley TW, Callaghan R. Cytotoxic efficacy of an anthraquinone linked platinum anticancer drug. Biochem Pharmacol 2006; 71:1136-45.

3 Ohndorf UM, Whitehead JP, Raju NL, Lippard SJ. Binding of tsHMG, a mouse testis-specific HMG-domain protein, to cisplatin-DNA adducts. Biochemistry 1997; 36:14807-15.

4 Zhu G, Lippard SJ. Photoaffinity labeling reveals nuclear proteins that uniquely recognize cisplatin-DNA interstrand cross-links. Biochemistry 2009; 48:4916-25.

5 Zhu W, Li Y, Liu L, Chen Y, Wang C, Xi F. Supramolecular Hydrogels from Cisplatin-Loaded Block Copolymer Nanoparticles and alpha-Cyclodextrins with a Stepwise Delivery Property. Biomacromolecules 2010.

6 Taniguchi K, Wada M, Kohno K, Nakamura T, Kawabe T, Kawakami M, et al. A human canalicular multispecific organic anion transporter (cMOAT) gene is overexpressed in cisplatin-resistant human cancer cell lines with decreased drug accumulation. Cancer Res 1996; 56:4124-9.

7 Vaisman A, Varchenko M, Umar A, Kunkel TA, Risinger JI, Barrett JC, et al. The role of hMLH1, hMSH3, and hMSH6 defects in cisplatin and oxaliplatin resistance: correlation with replicative bypass of platinum-DNA adducts. Cancer Res 1998; 58:3579-85.

8 Yamamoto K, Okamoto A, Isonishi S, Ochiai K, Ohtake Y. A novel gene, CRR9, which was up-regulated in CDDP-resistant ovarian tumor cell line, was associated with apoptosis. Biochem Biophys Res Commun 2001; 280:1148-54.

9 Zienolddiny S, Skaug V, Landvik NE, Ryberg D, Phillips DH, Houlston R, et al. The TERT-CLPTM1L lung cancer susceptibility variant associates with higher DNA adduct formation in the lung. Carcinogenesis 2009; 30:1368-71.

10 Hu Z, Wu C, Shi Y, Guo H, Zhao X, Yin Z, et al. A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. Nat Genet 2011; 43:792-6.

11 Pande M, Spitz MR, Wu X, Gorlov IP, Chen WV, Amos CI. Novel genetic variants in the chromosome 5p15.33 region associate with lung cancer risk. Carcinogenesis 2011; 32:1493-9.

12 Turnbull C, Rahman N. Genome-wide association studies provide new insights into the genetic basis of testicular germ-cell tumour. Int J Androl 2011; 34:e86-96.

13 Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997; 25:4876-82.

14 Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 1980; 16:111-20.

15 Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011; 28:2731-9.

16 Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 1985; 39:783-91.

17 Rose PW, Beran B, Bi C, Bluhm WF, Dimitropoulos D, Goodsell DS, et al. The RCSB Protein Data Bank: redesigned web site and web services. Nucleic Acids Res 2011; 39:D392-401.

18 Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 2010; 5:725-38.

19 Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, et al. Comparative protein structure modeling using Modeller. Curr Protoc Bioinformatics 2006; Chapter 5:Unit 5 6.

20 Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 1997; 18:2714-23.

21 Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 2010; 66:12-21.

22 Venditti C, Meade A, Pagel M. Multiple routes to mammalian diversity. Nature 2011.

23 Heath TA, Zwickl DJ, Kim J, Hillis DM. Taxon sampling affects inferences of macroevolutionary processes from phylogenetic trees. Syst Biol 2008; 57:160-6.

24 Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, Grenyer R, et al. The delayed rise of present-day mammals. Nature 2007; 446:507-12.

25 Archibald JD, Deutschmann DH. Quantitative analysis of the timing of the origin and diversification of extant placental orders. J Mamm Evol 2001; 8:107-24.

26 Asher RJ, Meng J, Wible JR, McKenna MC, Rougier GW, Dashzeveg D, et al. Stem Lagomorpha and the antiquity of Glires. Science 2005; 307:1091-4.

27 Bromham L, Phillips MJ, Penny D. Growing up with dinosaurs: molecular dates and the mammalian radiation. Trends Ecol Evol 1999; 14:113-8.

28 Springer MS, Murphy WJ, Eizirik E, O'Brien SJ. Placental mammal diversification and the Cretaceous-Tertiary boundary. Proc Natl Acad Sci 2003; 100:1056-61.

29 Sankararaman S, Kolaczkowski B, Sjolander K. Intrepid: a web server for prediction of functionally important residues by evolutionary analysis. Nucleic Acids Res 2009; 37:W390-5.

30 Sankararaman S, Sha F, Kirsch JF, Jordan MI, Sjolander K. Active site prediction using evolutionary and structural information. Bioinformatics 2010; 26:617-24.

31 Gaston D, Susko E, Roger AJ. A phylogenetic mixture model for the identification of functionally divergent protein residues. Bioinformatics 2011; 27:2655-63.

32 Bennett MJ, Lebron JA, Bjorkman PJ. Crystal structure of the hereditary haemochromatosis protein HFE complexed with transferrin receptor. Nature 2000; 403:46-53.

33 Yuan S, Yu X, Topf M, Dorstyn L, Kumar S, Ludtke SJ, et al. Structure of the Drosophila apoptosome at 6.9 a resolution. Structure 2011; 19:128-40.

34 Jain S, Drendel WB, Chen ZW, Mathews FS, Sly WS, Grubb JH. Structure of human beta-glucuronidase reveals candidate lysosomal targeting and active-site motifs. Nat Struct Biol 1996; 3:375-81.

35 Chook YM, Blobel G. Structure of the nuclear transport complex karyopherin-beta2-Ran x GppNHp. Nature 1999; 399:230-7.

36 Groves MR, Hanlon N, Turowski P, Hemmings BA, Barford D. The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs. Cell 1999; 96:99-110.

37 Kumaran D, Bonanno JB, Burley SK, Swaminathan S. Crystal structure of phosphatidylglycerophosphatase (PGPase), a putative membrane-bound lipid phosphatase, reveals a novel binuclear metal binding site and two “proton wires”. Proteins 2006; 64:851-62.

38 Papageorgiou AC, Shapiro R, Acharya KR. Molecular recognition of human angiogenin by placental ribonuclease inhibitor-an X-ray crystallographic study at 2.0 A resolution. EMBO J 1997; 16:5162-77.

39 Ochiai A, Yamasaki M, Mikami B, Hashimoto W, Murata K. Crystal structure of exotype alginate lyase Atu3025 from Agrobacterium tumefaciens. J Biol Chem 2010; 285:24519-28.

40 Bairoch A, Apweiler R, Wu CH, et al. The Universal Protein Resource (UniProt). Nucleic Acids Res 2005; 33:154-9.

41 Cserzo M, Wallin E, Simon I, von Heijne G, Elofsson A. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. Protein Eng 1997; 10:673-6.