Effect of missense mutations on structure and interaction of anaplastic *Lymphoma kinase* (ALK) in neuroblastom

Hafsah Kanwal, Mohammad Haroon Khan and Hamid Rashid
Department of Bioinformatics, Mohammad Ali Jinnah University, Islamabad, Pakistan

Abstract: Neuroblastoma is a cancer of the sympathetic nervous system, accounting for upto 15% of childhood cancer mortality. It can occur in many areas but most of them begin in the abdomen in the adrenal gland and can spread to the bones and other areas. http://en.wikipedia.org/wiki/Neuroblastoma-cite_note-pmid19383347-3. Unfortunately, like other cancers, its causes are still poorly understood. Anaplastic lymphoma kinase (ALK), a membrane associated tyrosine kinase was recently found to be mutated in neuroblastoma. Protein sequence of ALK was retrieved from UniProt and the seven identified mutations were substituted in native sequence to get its mutant proteins. Significant changes were explored in the mutant secondary structures when compared with the native protein. Changes were also observed in the physiochemical properties and it can therefore be inferred that, these changes may be translated in the tertiary structures due to their effects on the folding pattern. Tertiary structure of the protein modeled after refinement and validation was submitted to Protein Model Database (PMDB) and was assigned with the PMDB ID P0077827. RMSD values of the mutant structures were observed deviated from the native structure when compared with probability < 0.05. It was observed that there are a total of 15 Disordered Regions in the protein having a total of 290 Disordered Residues. Protein-ligand interaction analysis was performed to investigate the effects of mutations damaging its interactions and it was observed that the mutations understudy affects its interactions with ATP which ultimately results in causing neuroblastoma. This study was based on the in silico mutation analysis of Seven missense mutations of anaplastic lymphoma kinase which can better explain why missense mutations in ALK protein cause neuroblastoma. Structure and sequence based computations were systematically and comprehensively evaluated applied to the mutants in anaplastic lymphoma kinase and on the basis of our observations a detailed structural explanations have been developed for the measured and predicted impact of these missense substitutions.

Keywords: Neuroblastoma, cancer, ALK, Protein Structure, docking.

INTRODUCTION

Neuroblastoma is an embryonal tumour of the peripheral sympathetic nervous system, accounts for upto 15% of childhood cancer mortility (George *et al*., 2008). Neuroblastoma can occur with a spectrum of disorders related to abnormal development of neural crest derived tissues (Mosse *et al*., 2008) and have the potential to invade other tissues like, bones, liver, lymph nodes etc (Goo, 2010). It is an extra cranial solid cancer in childhood (Schmidt *et al*., 2010) http://en.wikipedia.org/wiki/Neuroblastoma-cite_note-titleeMedicine_-Neuroblastoma:_Article_by_Norman_J_Lacayo.2C_MD-0 known to demonstrate spontaneous regression from an undifferentiated state to a completely benign cellular appearance (Chonpathompikunlert *et al*., 2011). http://en. wikipedia.org/wiki/Neurobla stoma-cite_ note-pmid19383347-3Symptoms depend on the part of the body which is affected. As with most cancers, the causes of neuroblastoma are still poorly understood but most commonly diagnosed in children before the age of five.

Neuroblastoma harbors a variety of genetic changes (Chen *et al*., 2008). The anaplastic lymphoma kinase (ALK) was recently found to be mutated in neuroblastoma (Kwon *et al*., 2011) which is a receptor tyrosine kinase. It was initially identified through the analysis of a specific translocation associated with a rare subtype of non-Hodgkin's lymphoma (Ogawa *et al*., 2011). According to the research, ALK signaling can be activated through different human cancers by creating specific oncogenic fusions of the ALK through translocations at 2p23 (Chiarle *et al*., 2008). These translocation events generate oncogenic ALK which lead to constitutive activation of the kinase domain (Jazii *et al*., 2006; Cheng and Ott, 2010). Proteins are always subjected to different types of mutations affecting them in multiple ways, changing their residues, folding, or interactions. In this study, the impact of missense mutations on the structure, function and folding pattern of anaplastic lymphoma kinase have been investigated at the molecular and 3D level, which will provide a comprehensive insight for the development of future therapeutics against Neuroblastoma.

MATERIALS AND METHODS

The *ALK* gene codes for anaplastic lymphoma kinase protein, activating the kinase through phosphorylation,

**Corresponding author:* e-mail: haroon.khan@jinnah.edu.pk

which can activate other proteins in the cell by transferring a phosphate group to it. This activation continues in a signaling pathway through a series of proteins, which is important in cellular processes like cell growth, differentiation and division. The simple germ line mutation in it will lead towards a serious neuroblastoma. Changes at molecular level in a protein can affect the phenotype of the cells, tissues and finally the organisms. The Protein sequence of anaplastic lymphoma kinase was retrieved from UniProt (http://www.uniprot.org) with accession number AAB71619 as a prerequisite for detailed *in silico* analysis. Functional impacts of a mutation can be better understood through analyzing the relevant information related to sequence, structure and interaction. Prediction of protein secondary structure elements is important for characterizing the 3D structures. Secondary Structure of the seed anaplastic lymphoma kinase protein along with its mutated versions were predicted through the CFSSP (http://www.biogem.org/cgi-bin/cho-fas.pl) and were comparatively analyzed for the impacts of mutations on the helix, sheets, turns and coils. Tertiary structure of the ALK protein was unpredicted and so was predicted through CPH Model 3.0 (Nielsen *et al*., 2010), homology based protein modeling server. It recognizes the template on the basis of profile-profile alignment guided with secondary structure and exposure predictions. The same method was used for the mutated sequences also. For quality assurance, the predicted tertiary structures were refined through ModRefiner (Xu and Zhang, 2011) and validated through ProSA-web (Wiederstein and Sippl, 2007) for Z score, RAMPAGE (Lovell *et al*., 2002) for rotatable angles and WHATIF (Vriend, 1990) for other parameters like Planarity, Anomalous Bond Angle, Hand Check, Name Check etc. Substitution of residues in protein can cause physiochemical differences which in turn results in affecting its interactions, so physiochemical properties were predicted through ProtParam (http://web.expasy.org/protparam), which allows the computation of various physical and chemical parameters for a given protein.

Over the past decade it has become evident that many proteins have disordered regions, even in their native states. There is much interest in characterizing these proteins because disordered protein regions often lead to difficulties in research. It was therefore essential to predict the disordered regions in the anaplastic lymphoma kinase protein and so was performed through FoldIndex (Prilusky *et al*., 2005). The product of ALK is an enzyme so its basic information along with its candidate ligand information was retrieved from the enzyme database (http://www.enzyme-database.org/). Detailed investigation of the impacts of mutations on the interaction of anaplastic lymphoma kinase was the most important step of this study. Protein-ligand interaction analysis was therefore performed through Docking Server (http://www.dockingserver.com/web/docking/) to further understand their complex structure and also to analyze the impact of substitutions on ALK protein interactions at the atomic level.

RESULTS

The sequence of anaplastic lymphoma kinase protein consisting of 1620 amino acids was retrieved from UniProt with accession No. AAB71619 in FASTA format. The seven identified mutations, Y1507F, G1128A, I1171N, F1174L, R1192P, F1245C and R1275Q were substituted through MUTATE_MODEL in the native protein sequence to get the mutants for further analysis. Secondary structure of the native protein consists of 2 Beta sheets, 5 beta hairpins, 2 beta bulges, 8 strands, 14 helices, 16 helix-helix interactions, 28 beta turns and 10 gamma turns which were observed changed in all the mutants. Changes were also observed in the physiochemical properties of the mutants when compared with the native conformation (table 1). It can therefore be inferred that, these changes may be translated in the tertiary structures due to their effects on the folding pattern.

The tertiary structure of ALK protein was previously unpredicted and further analysis cannot be preceded without knowing the details of its tertiary structure so as prerequisite for further detailed analysis, its tertiary structure was modeled through CPH Model 3.0. The accuracy of the modeled structure was significantly acceptable but for increased accuracy, the model was refined and validated. According to the validation results, no atoms were observed having the wrong handedness, all the atoms have proper chirality with bond angles in agreement with the standard using a tolerance of 4 sigma. The structure after all the required validations was submitted to Protein Model Database (PMDB) and was assigned with the PMDB ID P0077827. All the mutant models were compared with the native type anaplastic lymphoma kinase protein using DaliLite server (http://www.ebi.ac.uk/Tools/dalilite/) to analyze the degree of damage caused by amino acids substitutions (fig). When the Native structure was compared with the mutants with probability < 0.05, the RMSD values observed were, R1275Q 0.14 Å, R1192P 0.15 Å, Y1507F 0.14 Å, G1128A 0.17 Å, I1171N 0.14 Å, F1174L 0.17 Å and F1245C 0.15 Å.

Disordered regions were predicted for the native protein and it was observed that there are a total of 15 Disordered Regions in it having a total of 290 Disordered Residues. As anaplastic lymphoma kinase is an enzyme, its basic information regarding its substrate and other reaction details necessary for its interaction analysis were retrieved from Enzyme database. Its EC Number is 2.7.10.1, ontology is GO: 00004714, substrate is ATP while cofactors are Ca^{2+} and Hydrogen peroxide. The optimum

Table 1: Secondary structure summary and physiochemical properties for native and mutant anaplastic lymphoma kina protein

	Native protein	Y1507F	R1275Q	G1128A	I1171N	F1174L	R1192P	Y1507F
β sheets	2	3	3	3	3	3	3	3
β hairpins	5	6	6	6	6	6	6	6
β bulges	2	3	3	3	3	3	3	3
Strands	8	10	10	10	10	10	10	10
Helices	14	14	14	14	14	14	14	14
Helix-Helix interactions	16	15	15	15	15	15	25	15
β turns	28	29	29	30	29	30	29	29
γ turns	10	4	4	4	4	4	4	4
-ive residues (Asp + Glu)	161	161	161	157	157	157	157	157
+ive residues (Arg + Lys)	154	153	154	149	149	149	148	149
Total No of Atoms	24612	24606	24611	23709	23701	23705	23697	23697
Ext. coefficient	239005	236630	237515	234535	234535	234535	234535	234535
Instability Index	51.12	51.24	51.25	51.57	51.25	51.47	51.81	51.52
Aliphatic index	78.33	78.33	78.33	78.97	78.65	79.15	78.90	78.90
GRAVY	-0.297	-0.297	-0.295	-0.284	-0.290	-0.285	-0.283	-0.286

Table 2: Docking Analysis of the native and mutants anaplastic lymphoma kina protein

	BE	IME	EE	SI	NHB	Interactions
Native protein	+0.75	-2.77	+0.82	782.716	4	LEU, VAL, LYS, MET, ASP, SER, ARG, ASN, LEU, ASP
Y1507F	-1.30	-4.79	+0.08	852.6	4	LEU, VAL, ALA, LYS, MET, ASP, ARG, ASLEU, ASP, MET
R1275Q	-1.79	-4.71	-0.41	839.363	9	VAL, LYS, ASP, ASN, ARG, LEU, ASP, MET, TYR, PRO
G1128A	-0.59	-4.44	+0.23	868.217	5	VAL, ALA, LYS, MET, ARG, ASN, LEU, ASP, ARG, ASP, TYR, MET
I1171N	+0.57	-3.64	+0.04	879.179	4	LEU, LYS, GLU, MET, ASP, LEU, ASP, ARG
F1174L	-0.27	-3.83	+0.02	737.161	11	VAL, ALA, LYS, MET, ASP, LYS, ARG, LEU, ASP, MET
R1192P	-0.21	-1.87	+0.45	812.224	2	LEY, VAL, ALA, LYS, MET, ASP, LEU, ASP, ARG

BE=Binding Energy, IME=Intermolecular Energy, EE=Electrostatic Energy, SI=Surface Interaction, NHB=No. of Hydrogen bonds

pH for the enzyme to work is 7.4-7.5 while the optimum temperature is around 20-25^{o}C. It reacts with ATP and produce [protein]-L-tyrosine phosphate and ADP as a byproduct.

Residues substitution in a protein affects the interactions of a protein in different ways by affecting its structure. Protein-ligand interaction analysis was therefore performed to investigate the effects of mutations damaging its interactions with the ligand. It was observed that the mutations understudy affects the 3-dimensional structure of the protein and thus its interactions with ATP (Ligand) thus results in causing neuroblastoma (table 2).

DISCUSSION

The gene responsible for neuroblastoma is ALK of length 4,863 bp located on chromosome 2 at the locus 2p23 from 29,415,639 to 30,144,476. The product of ALK, anaplastic lymphoma kinase is a member of the insulin receptor super family and its function is still poorly understood. It has an important role in the development and maintenance of central and peripheral nervous system. It may exert antagonist functions, proapoptotic or antiapoptotic, depending on the absence or presence of a ligand (Mourali *et al*., 2006).

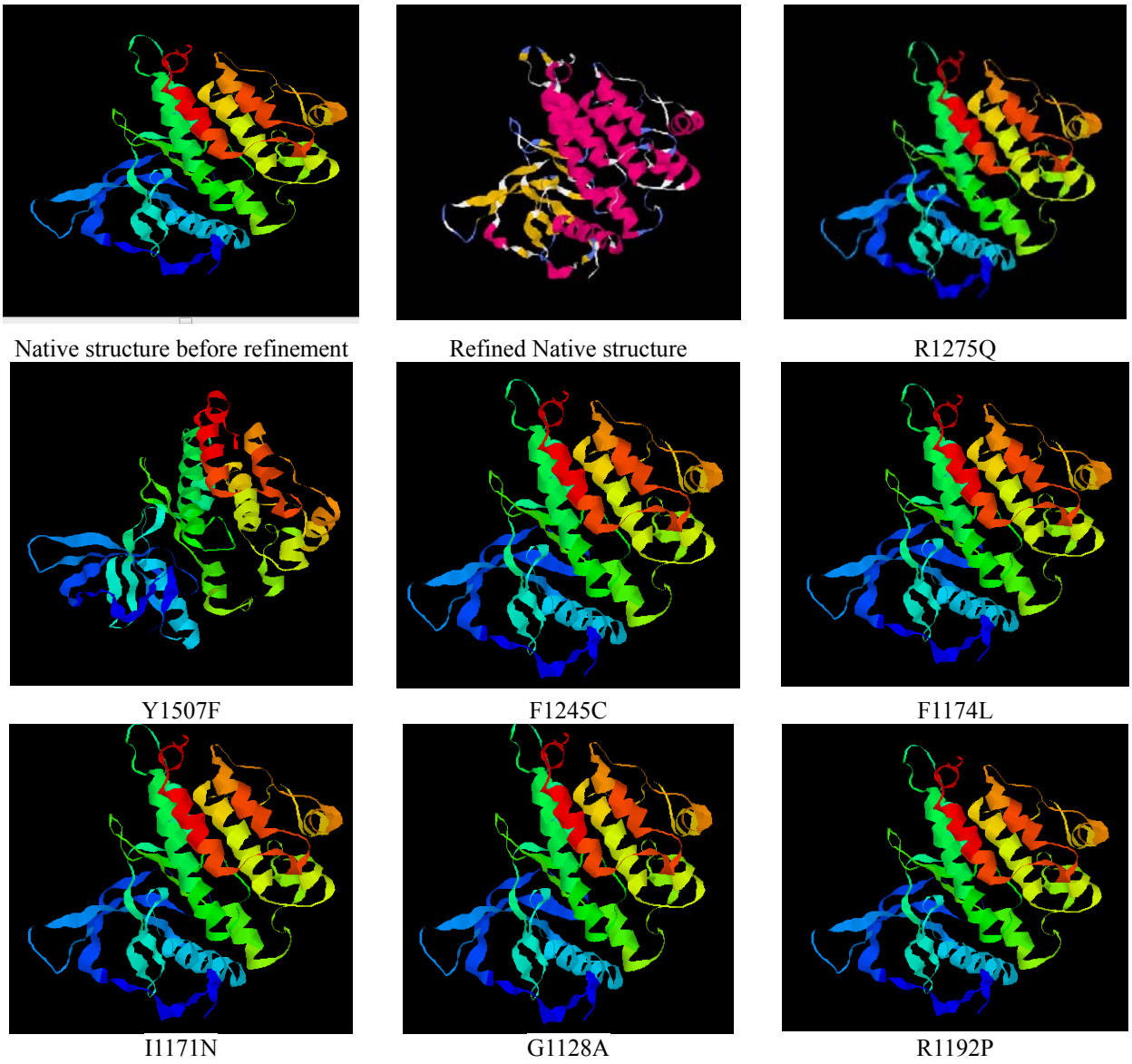

Fig. Tertiary structures of native and mutants anaplastic lymphoma kina protein predicted through CPH Model 3.0

Each structural level of a protein depends strongly on the level below it. The primary structure determines how the chain twists and turns. Tertiary structures are description of the whole polypeptide including the secondary structures which folds itself into its final 3D conformation, ultimately defining the protein functions. From the structural validations, it was observed that, all the residue chains that have an intact planar group and all the atoms connected to planar aromatic rings in side chains residues are in the plane were within the expected RMSD. No errors were detected in torsion angles, valine, threonine, isoleucine, leucine, arginine, tyrosine, phenylalanine, aspartic acid, glutamic acid, phosphate group naming conventions etc. The calculated Z-score for the native structure was -7.96. According to the Ramachandran plot statistics, the native structure has 88.6% residues in the Favored Region, 9.8% residues in the Allowed Region and only 1.6% of the residues in the Outlier Region. The longest disordered region was observed having 68 residues. The disordered segments are from the amino acid residues 198-204, 270-310, 473-484, 486-492, 501-507, 647-687, 689-693, 925-932, 1080-1085, 1347-1353, 1363-1372, 1381-1398, 1403-1444, 1446-1456, 1470-1537 with their respective scores of -0.03 ± 0.02, -0.09 ± 0.05, -0.06 ± 0.02, 0.06 ± 0.03, -0.04 ± 0.03, -0.08 ± 0.03, -0.01 ± 0.02, -0.06 ± 0.03, -0.06 ± 0.04, -0.05 ± 0.02, -0.03 ± 0.02, -0.09 ± 0.04, -0.08 ± 0.04, -0.04 ± 0.03 and -0.14 ± 0.11.

The molecular function depends directly on the overall 3-dimensional shape of the protein which determines its interactions with other molecules. Macromolecular docking is the computational modeling of complexes formed by two or more interacting biological molecules.

Docking was carried out for the prediction of the three dimensional structure of the macromolecular complex as it would occur in a living organism. It can produce only plausible candidate structures.

CONCLUSION

ALK has emerged as a strong biomarker and therapeutic target for a significant, percentage of cancer patients who may benefit from ALK-targeting agents (Grande, Bolos and Arriola, 2011).

This study is based on the *in silico* mutation analysis of anaplastic lymphoma kinase which can better explain why missense mutations in ALK protein causing neuro-blastoma. Seven missense mutations were focused in the studies which were compared for their structures at different levels, properties and interactions. Lee *et al*. (2010) also showed from their wet lab results that ALK was only able to confer transforming growth potential to RIE cells with the missense mutations due to the enhanced ALK catalytic activity, resulting in an uncontrolled cell proliferation. Schonherr *et al* (2011) explained that the I1250T mutation results in a much smaller side chain and introduces a polar functionality which probably results in weakening of the hydrophobic contact presumably leading to the destabilization of the entire active site of ALK or the mutation could modulate the interaction of the side chain with the carbonyl oxygen with a possible destabilizing effect impairing the activation of the kinase. This study bridges computational biology to molecular, structural and experimental biology, which may deepen our understanding towards advanced and novel developments against neuroblastoma in the near future. Structure and sequence based computations were systematically and comprehensively evaluated applied to the mutants in anaplastic lymphoma kinase during neuroblastoma. On the basis of our observation, detailed structural explanations have been developed for the measured and predicted impact of the above seven missense substitutions. It is obvious from the results that these bioinformatics approaches can be significantly used against other genetic disorders threatening humanity.

REFERENCES

Chen Y, Takita J, Choi YL, Kato M, Ohira M, Sanada M, Wang L, Soda M, Kikuchi A, Igarashi T, Nakagawara A, Hayashi Y, Mano H and Ogawa S (2008). Oncogenic mutations of ALK kinase in neuroblastoma. *Nature*, **455**: 971-974.

Cheng M and Ott GR (2010). Anaplastic Lymphoma Kinase as a Therapeutic Target in Anaplastic Large Cell Lymphoma, Non-Small Cell Lung Cancer and Neuroblastoma. *Anticancer Agents Med Chem*., **10**(3): 236-249.

Chiarle R, Voena C, Ambrogio C, Piva R and Inghirami G (2008). The anaplastic lymphoma kinase in the pathogenesis of cancer. *Nat. Rev. Cancer*, **8**: 11-23.

Chonpathompikunlert P, Yoshitomi T, Han J, Toh K, Isoda H and Nagasaki Y (2011). Chemical nanotherapy: Nitroxyl radical-containing nanoparticle protects neuroblastoma SH-SY5Y cells from Aβ-induced oxidative stress. *Therapeutic Delivery,* **2**(5): 585-597.

Kwon MJ, Choi YL, Sung KW, Kang SY, Park SM, Choi SY, Kim JS and Suh YL (2011). Oncogenic anaplastic lymphoma kinase (ALK) mutation in neuroblastomas and other pediatric tumors. *Pathol. Res. Pract*., **207**(10): 634-639

George RE, Sanda T, Hanna M, Fröhling S, Luther W, Zhang J, Ahn Y, Zhou W, London WB, McGrady P, Xue L, Zozulya S, Gregor VE, Webb TR, Gray NS, Gilliland DG, Diller L, Greulich H, Morris SW, Meyerson M and Look AT (2008). Activating mutations in ALK provide a therapeutic target in neuroblastoma. *Nature,* **455**: 975-978.

Goo HW (2010). Whole-body MRI of neuroblastoma. *Eur. J. Radio.,* **75**(3): 306-314.

Grande E, Bolos MV and Arriola E (2011). Targeting Oncogenic ALK: A Promising Strategy for Cancer Treatment. *Mol. Cancer Ther.,* **10**: 569-579.

Jazii FR, Najafi Z, Malekzadeh R, Conrads TP, Ziaee AA, Abnet C, Yazdznbod M, Karkhane AA and Salekdeh GH (2006). Identification of squamous cell carcinoma associated proteins by proteomics and loss of beta tropomyosin expression in esophageal cancer. *World J. Gastroenterol*., **12**: 7104-7112.

Lee CC, Jia Y, Li N, Sun X, Ng K, Ambing E, Gao MY, Hua S, Chen C, Kim S, Michellys PY, Lesley SA, Harris JL and Spraggon G (2010). Crystal structure of the ALK (anaplastic lymphoma kinase) catalytic domain. *Biochem. J.,* **430**: 425-437.

Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS and Richardson DC (2002). Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins,* **50**: 437-450.

Mosse YP, Laudenslager M, Longo L, Cole KA, Wood A, Attiyeh EF, Laquaglia MJ, Sennett R, Lynch JE, Perri P, Laureys G, Speleman F, Kim C, Hou C, Hakonarson H, Torkamani A, Schork NJ, Brodeur GM, Tonini GP, Rappaport E, Devoto M and Maris JM (2008). Identification of ALK as a major familial neuroblastoma predisposition gene. *Nature,* **455**(7215)*:* 930-935.

Mourali J, Benard A, Lourenco FC, Monnet C, Greenland C, Moog-Lutz C, Racaud-Sultan C, Gonzalez-Dunia D, Vigny M, Mehlen P, Delsol G and Allouche M (2006). Anaplastic Lymphoma Kinase is a dependence receptor whose proapoptotic functions are activated by caspase cleavage. *Mol. Cell Biol*., **26**(16): 6209-6222.

Nielsen M, Lundegaard C, Lund O and Petersen TN (2010). CPHmodels-3.0-Remote homology modeling using structure guided sequence profiles. *Nucleic Acids*

Res., **38**: 576-581.

Ogawa, S, Takita J, Sanada M and Hayashi Y (2011). Oncogenic mutations of *ALK* in neuroblastoma. *Cancer Sci.,* **102**(2): 302-308.

Prilusky J, Felder CR, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I and Sussman JL (2005). FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics,* **21**(16): 3435-3438.

Schmidt M, Baum RP Simon T and Howman-Giles R (2010). Therapeutic nuclear medicine in pediatric malignancy. *QJ Nucl. Med. Mol. Imaging,* **54**(4): 411-428.

Schonherr C, Ruuth K, Eriksson T, Yamazaki Y, Ottmann C, Combaret V, Vigny M, Kamaraj S, Palmer RH and Hallberg B (2011). The Neuroblastoma ALK (I1250T) Mutation is a kinase-dead RTK *in vitro* and *in vivo*. *Transl. Oncol.,* **4**(4): 248-255.

Vriend G (1990). A molecular modeling and drug design program. *J. Mol. Graph*., **8**: 52-56.

Wiederstein M and Sippl MJ (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.,* **35**: 407-410.

Xu D and Zhang Y (2011). Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Bio-phys. J.,* **101**: 2525-2534.