Journal of Bioinformatics and Sequence Analysis Vol. 2(6), pp. 75-84, December 2010
Available online at http://www.academicjournals.org/jbsa
ISSN 2141-2464 ©2010 Academic Journals

Full Length Research Paper

In silico sequence specific analysis of ERBB2 RTK alterations responsible for neuroectodermal tumors of *Homo sapiens*

Lakshmi P. T. V.* and Subhashini S.

Phytomatics Laboratory, Department of Bioinformatics, Bharathiar University, Coimbatore, Tamil Nadu, India.

Accepted 18 August, 2010

Current theories of oncogenesis suggest that tumors develop as a result of sequential alterations, which activate or inactivate tumor suppressor genes involved in the regulation of the cell cycle, and ultimately over express or suppress the proteins derived from those genes. Mutation is an obvious way of constitutive activation of a kinase. Owing to the importance of ERBB proteins in both development and cellular transformation, a lot of attention has been drawn to characterize the functions of this family of receptor tyrosine kinases. ERBB2 alterations are frequent in human tumors and result from translocation, mutation or amplification. The ERBB2 (Her2) proto-oncogene encodes a receptor tyrosine kinase, which is frequently amplified and over expressed in human tumors. The protein sequence of ERBB2, containing 1255 residues obtained from Swiss-Prot, was induced with a mutation to be characterized by using *in silico* tools and compared with the wild type.

Key words: Oncogenesis, inactive - tumor suppressor genes, mutation, proto-oncogene.

INTRODUCTION

Oncogenes are “gain-of-function” mutations of normal regulatory genes or proto-oncogenes. Proto-oncogenes are involved in signal transduction and execution of mitogenic signals, through their protein products (Todd and Wong, 1999). Usually, oncoproteins destabilize signal transduction pathways at the cell surface, in the cytosol and/or in the nucleus and contribute to human cancer formation by supporting accelerated proliferation, de-regulating cell cycle control or blocking apoptosis. Based on the functional and biochemical properties of protein, products of their normal counterparts can be classified into five groups such as: (i) growth factors (ii) growth factor receptors (RTK) (iii) signal transducers (iv) transcription factors and (v) others that include programmed cell death regulators (Kufe et al., 2003).

The plethora of signal transduction networks mediating the biological processes is regulated in part by polypeptide growth factors that generate signals by activating cell surface receptors either in paracrine or autocrine manner. The primary mediators of such physiological cell responses are receptor tyrosine kinases (RTKs) that couple ligand binding to downstream signalling cascades and gene transcriptions (Zwick et al., 2001). Tyrosine kinases are divided into two main classes: receptor tyrosine kinases, which are trans-membrane proteins with a ligand-binding extracellular domain and a catalytic intracellular kinase domain and non-receptor tyrosine kinases that lacks transmembrane domains and are found in the cytosol, nucleus and the inner surface of the plasma membrane (http://www.utdol.com/patients/content/topic.do?topicKey =~nnx7LcCU_HrU9Ni&source=see_link). The kinase domains of tyrosine kinases (TKs) have a bilobar structure; an N-terminal lobe that binds ATP and magnesium and a C-terminal lobe that contains an activation loop and a cleft between the lobes to which polypeptide substrates are bind (Daniela and Richard, 2005). In the absence of ligand, receptor TKs are unphosphorylated and monomeric, and the conformation of their kinase domains is inactive. In some receptor TKs, the cytoplasmic juxtamembrane region further inhibits the enzyme by interacting with the kinase domain. Receptor

*Corresponding author. E-mail: lakshmiptv@yahoo.co.in.

TKs become activated when ligand binds to the extracellular domain, resulting in receptor oligomerization, disruption of the autoinhibitory juxtamembrane interaction and autophosphorylation of a regulatory tyrosine within the activation loop of the kinase. These changes reorient critical amino acid residues, thereby increasing the catalytic activity of the enzyme and binding sites for signaling proteins and recruiting them to the membrane which activates multiple signaling pathways (Schlessinger, 2000).

The ERBB protein family or epidermal growth factor receptor (EGFR) family consists of four structurally related receptor tyrosine kinases: EGFR (ERBB-1), HER2/c-neu (ERBB-2), HER3 (ERBB-3) and HER4 (ERBB-4), of which HER2 is considered as an orphan receptor whose endogenous ligand has not yet been identified and therefore none of the EGF family of ligands will be able to activate it. However, ERBB receptors dimerise on ligand binding and HER2 is the preferential dimerisation partner of other members of the ERBB family (Olayioye, 2001). The HER2 gene or Erbb2 is a proto-oncogene located at the long arm of human chromosome 17 (17q21-q22) and alterations in these proteins, through mutation or altered regulation by gene amplification contribute to tumor formation. Hence, the excessive ERBB signaling is associates with the development of a wide variety of solid tumor (Cho and Leahy, 2002), while insufficient ERBB signaling in humans are associated with the development of neurodegenerative diseases (Bubil and Yarden, 2007).

Tyrosine kinases are dysregulated in cancer cells in several ways. Initial discovery (Slamon et al., 1987) on the amplification of the ERBB2 RTK gene in breast cancers launched the search for other RTK alterations in human. A frequent feature of the partner protein is a domain that causes constitutive oligomerization of the tyrosine kinase in the absence of ligand binding (Nako et al., 1996) or physiologic activating signals, thereby promoting autophosphorylation and activation. A common mechanism could be small deletions and point mutations in the kinase domain of epidermal growth factor receptor (EGFR) that may increase the sensitivity of the receptor to its ligand and alter receptor signaling (Pao et al., 2004).

A second important mechanism of tyrosine kinase activation in hematologic cancers is the fusion of a receptor or non-receptor tyrosine kinase with a partner protein, usually as a consequence of a balanced chromosomal translocation (Smith et al., 2003). A third mechanism of tyrosine kinase dysregulation is the increased or aberrant expression of a receptor tyrosine kinase, its ligand or both. However, ERBB-1 and ERBB-2 are found in many human cancers and their excessive signaling may be a critical factor in the development of malignancies, since the dysregulation of cancer, in particular neuroblastoma, is reported to occur mainly due to mutations (Daniel et al., 2009). Thus, this study aimed at identifying the protein sequence signatures of malignancies and compare both the wild and mutated types of ERBB2 that was attempted by their physico-chemical properties, sequence-structure relationship and active-site geometry, because understanding the mutation dependent activation process would facilitate in identifying the unique drug with potential anti-tumor activity.

METHODOLOGY

Retrieval and activation of mutation to analyze physico-chemical characterization of target ERBB2

The complete protein sequence of *Homo sapiens* ERBB2 (Accession number: P04626) was collected from Swiss-Prot protein sequence database (http://www.uniprot.org/uniprot/P04626&format=html) and saved in FASTA format. The mutations such as: ins 774 (AYVM) and ins 779 (VGS) and the substitutions of L755P, E914K and G776S respectively, was induced in the retrieved sequence by Bio-edit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). The basic physico-chemical properties were calculated using the Prot-Param tool (http://expasy.org/tools/protparam.html).

Multiple sequence alignment of the primary structure

Multiple sequence alignment was performed using the commercial package Geneious Pro software version 4.7 which works with the parameters like Blosum 62 Matrix, Gap open penalty 12, Gap Extension Penalty 3 and Global Alignment Type with free end gaps and 10 Refinement Iterations, thereby deleting the regions of conservation and variation of ancestral relationships between the sequences.

Identification of motifs and domains

The wild type was analyzed for domains by PRODOM (http://prodom.prabi.fr/prodom/current/html/home.php) which is a comprehensive database of protein domain families generated from the global comparison of all available protein sequences in the SWISS-PROT database by DOMAINER algorithm, in order to spot the mutations which play a critical role in the identified domains of wild type ERBB2. PROSITE (http://www.expasy.ch/prosite.html) was used for documentation which described the protein domains, families and functional sites as well as associated patterns and profiles to identify them. Hence, both PRODOM and PROSITE were employed to identify the occurrence of any significant change within the identified domains or functional site, as well as the results obtained, which were compared with the identified mutated sequences.

Secondary structure prediction

SOPMA tool (Geourjon and Deleage, 1995) (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) was used to obtain the secondary structure of protein ERBB2. Default parameters were selected to increase reliability and accuracy of the query sequences.

Protein function prediction

The biological function process of core protein was predicted by protein function prediction (PFP) version 2.0 beta release

>sp|P04626|ERBB2_HUMAN Receptor tyrosine-protein kinase ERBB-2 OS=Homo sapiens GN=ERBB2 PE=1 SV=1

MELAALCRWGLLLALLPPGAASTQVCTGTDMKLRLPASPETHLDMLRHLYQGCQVVQGNLELTYLPTNASLSFLQDIQEVQ
GYVLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCY
QDTILWKDIFHKNNQLALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHEQCAAGC
TGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTDVGSCTLVCPLHNQEVTAED
GTQRCEKCSKPCARVCYGLGMEHLREVRAVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGY
LYISAWPDSLPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGLRSLRELGSGLALIHHNTHLCFVHTVPWDQLFRNPHQA
LLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVEECRVLQGLPREYVNARHCLPCHPECQPQNG
SVTCFGPEADQCVACAHYKDPPFCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSII
SAVVGILLVVVLGVVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKG
IWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSPYVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRLGSQ
DLLNWCMQIAKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPIKWMALESILRRRFTH
QSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYMIMVKCWMIDSECRPRFRELVSEFSRMARDP
QRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGDLVDAEEYLVPQQGFFCPDPAPGAGGMVHHRHRSSSTRSGGGDLTLGLEP
SEEEAPRSPLAPSEGAGSDVFDGDLGMGAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYVNQPDVRPQ
PPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGGAVENPEYLTPQGGAAPQPHPPPAFSPAFDNLYYWDQDPPER
GAPPSTFKGTPTAENPEYLGLDVPV

Figure 1. The FASTA sequence of ERBB2 (P04626).

(http://dragon.bio.purdue.edu/pfp). PFP algorithm searched conventional databases with relative probability of gene ontologies (GO) to predict the most probable GO annotations in three biological processes (BP), molecular function (MF) and cellular component (CC) categories. The nature of the kinase to interact with core protein was predicted by NetPhosK as a kinase-specific phosphorylation site predictor (http://www.cbs.dtu.dk/services/NetPhosK/).

Template selection

The standard protein-protein BLAST-blastP (http://blast.ncbi.nlm.nih.gov/Blast.cgi/) was used to identify similar sequences in protein databases for the upload FASTA sequence of both wild and mutated ERBB2 protein sequences.

Modeller for structure prediction

Modeller 9V7 (Andrej, 2009) was used for homology or comparative modeling of protein three-dimensional structures. The target sequence in the PIR format, readable by MODELLER (file "ERBB2.ali"), was loaded and aligned with the structure of templates by align2d mult () command which was stored in PIR (".ali") and PAP (".pap") file formats. The PIR format was further used by MODELLER in the subsequent model building stage, while the PAP alignment format was used to inspect the model visually. After the target-template alignment was constructed, MODELLER calculated a 3D model of the target automatically using its “automodel class” which provided the loop that was further analyzed by “automated loop modeling”. Then the model structure was optimized to ensure the stable structure of the mutated sequence.

Structure validation for the modeled structure

The stereochemical validation of the modeled structures was performed by structural analysis and verification server (SAVES) (http://nihserver.mbi.ucla.edu/SAVES) that validates the models using the programs like Pro-check, What-check, verify 3D, Errat and Prove.

RESULTS

Sequence retrieval

The sequence of the receptor tyrosine-protein kinase ERBB-2 (Accession No: P04626) retrieved and constituted about 1255 amino acids with a molecular weight of 137, 910 Da (Figure 1).

Activation of mutations

Insertional mutations at ins774 (AYVM) and ins779 (VGS), along with L755P, E914K and G776S respectively, were induced on the retrieved sequence using Bioedit software (Table 1).

Physico-chemical characterization of target ERBB2

The parameters computed by Prot-Param revealed (Table 2) that the sequence with the ins774 (AYVM) accounts for the higher molecular weight 138375.0 Da. Some amino acids such as leucine, proline, glycine, alanine, valine and glutamine were found abundant in proteins than other amino acids totaling 37% of all the amino acid residues in the protein. However, tryptophan and methionine together constituted approximately 3% of all amino acids in the protein. The number of positively charged residues was same (142) for all mutations except the substitution E914K that accounted to 141. Correspondingly, the number of negatively charged

Table 1. Activated mutations and their locations.

Mutation	Sequence region		
	Start	Region	End
Ins 774 (AVYM)	771	AYVAYVMMAG	780
Ins 779 (VGS)	771	AYVMAGVGVG	780
	781	SSPYVSRLLG	790
L755P	751	AIKVPRENTS	760
E914K	911	TVWKLMTFGA	920
G776S	771	AYVMASVGSP	780

Table 2. Physico - chemical properties of the mutated ERBB2 RTK.

Property	Wild	Mutant				
	ERBB2	Ins 774 (AYVM)	Ins 776 (VGS)	L755P	E914K	G776S
No. of amino acids	1255	1259	1258	1255	1255	1255
Molecular weight	137910.5	138375.0	138153.	137894.4	137909.5	137894.4
Total number of negatively charged residues (Asp + Glu)	142	142	142	142	141	142
Total number of positively charged residues (Arg + Lys)	110	110	110	110	111	110
Total no. of atoms	19180	19244	19214	19175	19185	19175
Theoretical pI	5.58	5.58	5.58	5.58	5.65	5.58
Estimated half-life (h)	30	30	30	30	30	30
Instability index (unstable)	56.13	55.96	56.11	56.07	56.07	56.07
Aliphatic index	82.35	82.40	82.38	82.04	82.35	82.04
Grand average of hydropathicity (GRAVY)	-0.247	-0.241	-0.245	-0.252	-0.248	-0.252
Ext. coefficient ++ Cys (1)	138275	139765	138275	138275	138275	138275
Ext. Coefficient -- Cys (2)	134650	136140	134650	134650	134650	134650

residues was revealed to be the same (110) in all mutations, except E914K which accounted to 111. All the mutated sequences had the theoretical pI value of 5.58 except the mutation E914K which revealed the higher pI value of 5.65. The estimated half-life was 30 h for all the five mutated sequences, while the instability index (II) computed ranged from 55.96 to 56.13, indicating the stability of mutated proteins. However, based on the aliphatic index, ins 774 (AYVM) had the highest thermo stability (82.40), while the substitutions, L755P and G776S, showed the least (82.04) thermo-stability of the mutated sequences, respectively. Grand average of hydropathy (GRAVY) was revealed to be significantly higher -0.241 when compared to other mutated sequences. Similarly, the extinction coefficient under both ++ Cys and - - Cys residues indicated higher value with the induced mutation of ins 774 (AVYM) which represented 139765 and 136140, respectively at 280 nm (Table 2).

(i) Based on the assumption that all pairs of Cys residues form cystines.

(ii) Based on the assumption that all Cys residues are reduced.

Identification of motifs and functional domains

PROSITE results obtained for ERBB2 wild type indicated

Table 3. The identified motif regions by PROSITE.

Found Motif	Position	Description
PROTEIN_KINASE_TYR	841..853	Tyrosine protein kinase specific active-site signature
EF_HAND_1	1011..1023	EF-hand calcium-binding domain
PROTEIN_KINASE_ATP	726..753	Protein kinase ATP-binding region signature

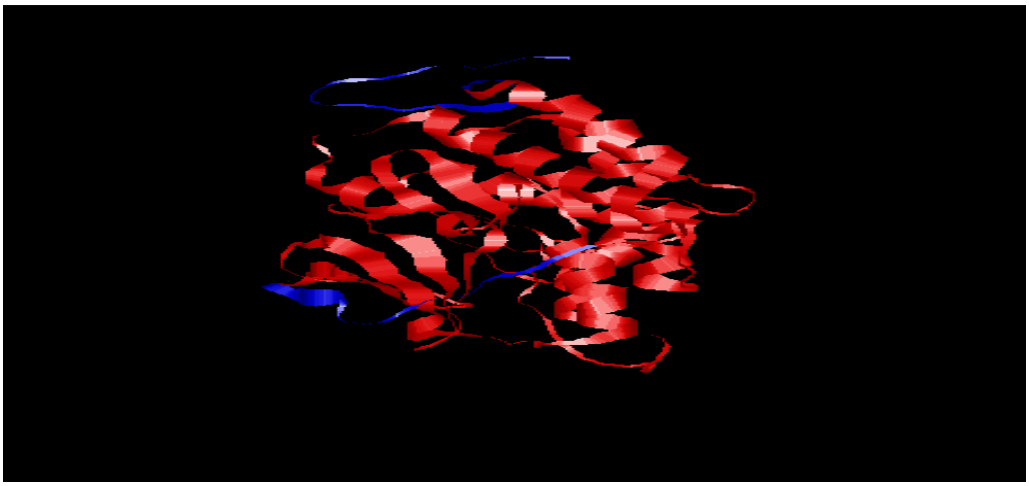

Figure 2. The identified domain region by ProDom. *Red - indicates the region between 724 and 983 of kinase ATP-binding nucleotide domain of model ERBB2; *Blue - indicates the structure of ERBB2 model.

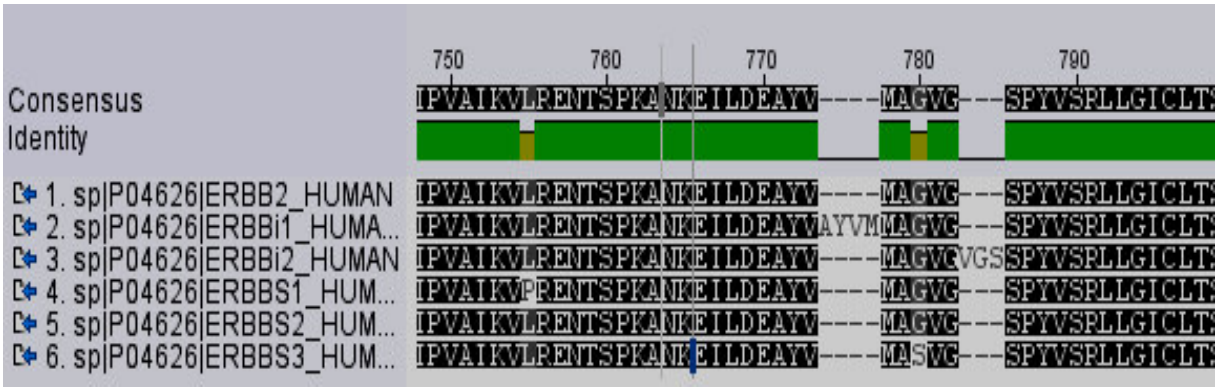

Figure 3. The multiple sequence alignment by Geneious Pro. 1, wild type; 2, ins774 (AVYM); 3, ins779 (VGS); 4, L755P; 5, E 914K; 6, G776S.

(Table 3) the presence of three 1D functional motif regions such as two protein kinases and one EF-hand, and these were however revealed to be unaffected by the induced mutations.

ProDom significantly identified the region that encodes for the domain which was revealed to occupy the regions between 724 and 983 (Figure 2). However, this specific region corresponded to all the induced mutations and was identified to localize within the Kinase ATP- binding nucleotide, which also plays a significant role in the receptor/domain activity and configuration.

Multiple sequence alignment

The multiple sequence alignment over Geneious Pro for the mutated sequence, revealed an exhibition of 99.3% identity with the wild type sequence through pairwise

Table 4. The results of secondary structure prediction by SOPMA.

Name	Helix (%)	Beta strand (%)	Random coil (%)
ERBB2	24.14	16.25	54.02
ERBB2 Ins(AYVM)	24.15	16.20	54.17
ERBB2 Ins(VGS)	24.09	16.22	54.13
ERBB2 L755P	24.14	16.25	54.02
ERBB2 E914K	24.14	16.10	54.18
ERBB2 G776S	24.14	16.25	54.02

Table 5. The results of molecular function by PFP and kinase specificity (NetPhosK).

Name	Percentage of accuracy	Molecular function	Specific kinase activity (NetPhosK)
ERBB2	100%	Purine nucleotide binding; transmembrane receptor protein kinase activity	0.88 PKC at position 1051

Table 6. The details of template for each mutated sequence ERBB2.

S/No.	Protein	Protein length	Template (PDB ID)	Sequence identity (%)	E - value	Target region	Template region
1	P04626 (Wild type)	1255	3bbtB	77	0	710 - 999	683 - 973
2	Ins 774 (AVYM)	1259	3bbtB	77	0	710 - 1003	683 - 973
3	Ins 779 (VGS)	1258	3bbtB	77	0	710 - 1002	683 - 973
4	L755P	1255	3bbtB	77	0	710 - 999	683 - 973
5	E914K	1255	3bbtB	77	0	710 - 999	683 - 973
6	G776S	1255	3bbtB	77	0	710 - 999	683 - 973

alignment. Perhaps, all the gaps observed were positioned within the residues from 770 to 920 distances (Figure 3). Although it corresponded to a very negligible difference of about 0.7%, it remarkably played a major role in the alteration or change of the domain property or function.

Secondary structure analysis

The secondary structure prediction defined each residue into either alpha helix, beta sheet or random coil secondary structures. SOPMA analysis suggested the mutated sequences to contain more random coils than helices and beta sheets (Table 4).

Function prediction

Prediction of kinase specificity, along an amino acid sequence, is normally done at serine, threonine or tyrosine residues. However, if kinase specificity is additionally predicted at non-STY {serine (Ser), threonine (Thr) and tyrosine (Tyr)} sites, one can notice that the individual kinase predictions tend to cluster amino acid residue at 1051 (Table 5), which could have high score performance value for different kinases as presented by NetPhosKpredictor.

Template selection

The template (PDB ID: 3bbtB) identified, using PSIBLAST, was revealed to be the crystal structure of ERBB4 kinase in complexity with lapatinib (Table 6).

Protein modeling and validation

Three dimensional structures of the wild type ERBB2 was successfully modeled using modeler 9V7. Among the models, the best one was selected based on the model score, z-DOPE (discrete optimized protein energy) score and MPQS (model protein quality score) (Table 7). Optimization of the selected protein models were done to reach the minimum energy level of each model to get the stable structure (Figure 4). The predicted structures were further validated using SAVES and checked for its

Table 7. Scoring pattern indicates model quality and validation of ERBB2 models.

S/N	Protein	Model score	MPQS	Z - Dope	Verify_3D	Pro Check (Ramachandran plot)
1	P04626 (Wild type)	1.00	1.4738	-1.25	95.53% of the residues had an averaged 3D-1D score > 0.2	90.5% core, 8.3% allowed, 1.2% general and 0.0% disallowed
2	Ins 774 (AVYM)	1.00	1.11462	-1.06	89.49% of the residues had an averaged 3D-1D score > 0.2	87.5% core, 10.5% allowed, 1.2% general and 0.8% disallowed
3	Ins 779 (VGS)	1.00	1.14481	-1.25	89.46% of the residues had an averaged 3D-1D score > 0.2	90.6% core, 8.3% allowed, 1.2% general and 0.0% disallowed
4	L755P	1.00	1.14438	-1.19	87.29% of the residues had an averaged 3D-1D score > 0.2	92% core, 7.2% allowed, 0.0% general and 0.8 % disallowed
5	E914K	1.00	1.14928	-1.27	94.16% of the residues had an averaged 3D-1D score > 0.2	90.9% core, 8.3% allowed, 0.8% general and 0.0% disallowed
6	G776S	1.00	1.14788	-1.28	94.50% of the residues had an averaged 3D-1D score > 0.2	91.7% core, 7.1% allowed, 1.2% general and 0.0% disallowed

reliability based on the Ramachandran plot, by fulfilling the required rules.

DISCUSSION

Collection and substantiation of novel molecular targets have become of principal importance in light of the surplus new potential therapeutic drug targets. With the increasing number of neuroectodermal tumor deaths, there is an essential need for exploring novel drugs to reduce the impending impact of the emergence of multidrug-resistant. The description of mutated proteins, particularly those that are differing from the wild type provides a new target for drug development. Hence, in this study, the point mutations, namely insertions 774 (AVYM) and 779 (VGS) and substitutions L755P, E914K and G776S, were identified and induced. A protein isoelectric point (pI), that is, the pH at which a protein has no net charge, forms the basis for its isolation. With this criterion, the predicted theoretical pI values of wild and mutated ERBB2 were observed to be at a minimal of 5.58 (acidic), except for mutation E914K which showed the value of 5.65 (acidic). Since at pI, the proteins would be stable and compact, the computed isoelectric point (pI) will be of use for developing buffer system for purification by isoelectric focusing method (Rosenberg and Ian, 2004). Molecular weight provides information not only on full-length of the protein expression, but also on the expression of modified, splice variant, cleavage product and processed proteins. However, any protein modification may lead to a change in the overall protein charge and/or molecular weight (Walsh et al., 2009). The crude bio-computed half life of the ERBB2 was estimated to be 30 h in all wild and induced mutated sequences which indicated stability. All the selected mutations are stable and greater than 30 (II>30) instability index. Perhaps, the higher aliphatic index of the protein with relatively high volume of aliphatic side chains (alanine, valine, isoleucine and leucine), provided the thermo-stability of the proteins.

Prot-Param computed the extension coefficient (EC) for a range of 276, 278, 219, 280 and 282 nm wave lengths and, 280 nm was found to be favorable because proteins were absorbed strongly there (http://www.microspectra.com/component/content/article/35-technical-support/185-protein-absorbance). The extinction coefficient indicates the light absorbed by proteins at a wavelength of 280 nm by considering the Cys residues and this is important in identifying fractions containing protein or estimating the concentration of a purified sample. Amino acids containing aromatic side chains (that is, tyrosine, tryptophan and phenylalanine) exhibit strong UV-light absorption. Consequently, proteins and peptides absorb UV-light in proportion to their aromatic amino acid content of total concentration

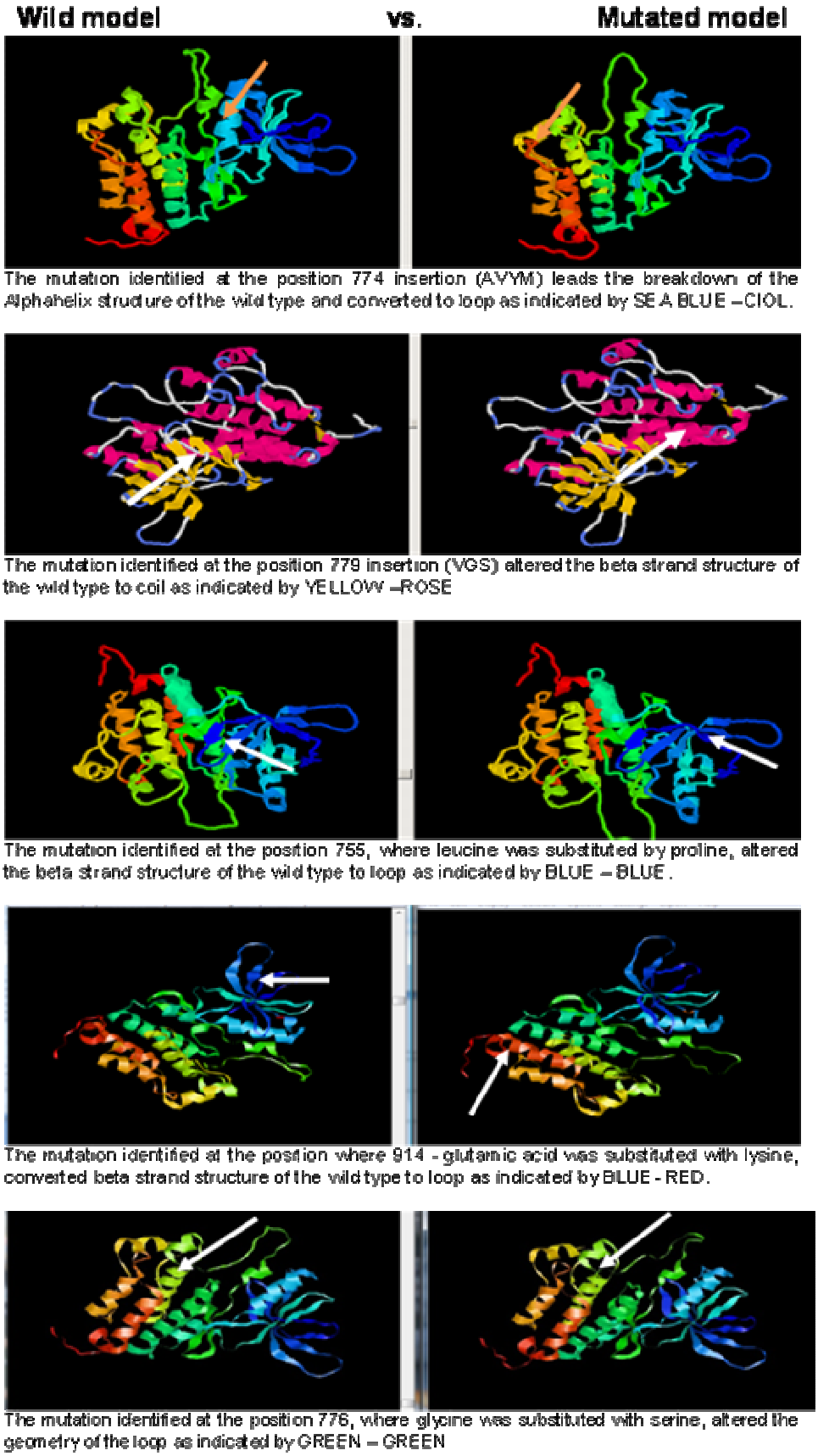

Figure 4. Structural comparison of the wild and mutated types of ERBB2.

(http://www.thermo.com/eThermo/CMA/PDFs/Articles/articlesFile_7178.pdf). Hence, the obtained result also indicated the presence of high concentration of tyrosine, lysine and cysteine in all the mutated sequences. Perhaps, the computed protein concentration and extinction coefficient could help in the quantitative study of protein-protein and protein-ligand interactions in solution. The GRAVY value for a peptide or protein was calculated as the sum of hydropathy values of all the amino acids divided by the number of residues in the sequence. Each amino acid was assigned a hydropathy index, that is, a value reflecting its relative hydrophilicity and hydrophobicity. The sum of the hydropathy indices of a given sequence divided by the number of residues in the sequence generated the GRAVY score, which exhibited a minimum difference, approximately ranging from one to four amino acids for each mutated sequence. Hence, the GRAVY score reflected the different hydropathy indices with negative values representing the presence of more hydrophilic residues (Xiong et al., 2009).

The identified functional domain had the start site from 724 to 923 and corresponded to an important domain namely “kinase ATP- binding nucleotide region” in all the mutated sequences. PROSITE analysis revealed 3 different signatures between the positions (i) 841 to 853 (ii) 726 to 753 and (iii) 1011 to 1023, among which the first two signatures corresponded to tyrosine kinase activity, while the third denoted the EF hand calcium binding domain. On comparing the study’s mutated sequences of ERBB2 with the wild type, approximately 99.3% similarity was obtained within the same subtypes and that the difference of 0.7% could be attributed to the change in the polarity or hydrophobicity. Furthermore, not only the polarity or hydrophobicity was significantly altered, but also the propensity of each amino acid residue could be involved in stabilizing the secondary structure (Lingxiu and Curtis, 1992).

Since the structure of the mutated protein was not available as a crystal structure, the structure of the mutated sequences was modeled. During template selection, it should be noted that the lower the E-value, or the closer it is to “0”, the higher the “significance” of the match would be. However, it is important to note that searches with short sequences can be virtually identical and have relatively high E-value. In the course of generation of protein models, using modeller, the loop and side-chain conformations did not correspond to energetically reasonable structures. Hence, idealization of bond geometry and removal of unfavourable non-bonded contacts was performed by optimization. Since, the quality of the 3D structure of a modeled protein strongly depends on the accuracy of the template structure (Branden and Jones, 1990), a more suitable template (3bbtB) was identified. However, to verify the stereochemical quality of a model-built structure, the parameters like bond lengths, bond angles, torsion angles and correctness of the amino acid chirality have to be provided. Since, a manual inspection of all stereo-chemical parameters is tedious and time-consuming, PROCHECK was used to automatically check all stereochemical properties in the present investigation. This revealed that it contained more than 90% of the amino acids at the core region of Ramachandran plot exhibiting concomitance to that of Laskowski et al. (1993), proving the quality of the models predicted to be highly significant. One important indicator of stereo-chemical quality is the distribution of the main chain torsion angles φ and ψ that can be examined by the Ramachandran plot.

The structures were validated using the SAVES server which showed more than 90% amino acids to occupy the most favored region; hence, it was found to be reliable. By visual comparisons of the models (wild vs mutated types), different levels of changes have been observed. For instance, Ins774 (AYVM) indicated that the structure around the residue 774 of the wild type was predicted to be an alpha helix that was observed in the loop. The structure around the residue 779 of the wild type was predicted to be a beta strand, which was converted to the loop with the insertion of the residues VGS at that position. However, replacement of leucine by proline at residue 755 removed the β-strand, while replacing glutamic acid by lysine at residue 914 removed part of the alpha helical structure. Perhaps, substitution of glycine by serine at residue 776 also altered the geometry of the loop.

In the dearth of experimental data, model-building on the basis of the known three dimensional structure of a homologous protein is the only dependable method to obtain structural information which is based on sequential information. Computational methods play a vital role in accelerating the drug development process, since the present study focused on the characterization of mutated ERBB2 RTK sequences of *Homo sapiens* and provided a new insight into the molecular biologics and functional properties for rational designing of inhibitors in its battle against neuroblastoma!

ACKNOWLEDGEMENTS

The authors are grateful to the University Grants Commission for financial support and encouragement to purchase the commercial package “Geneious Pro Version 4.7.” Also, Subashini thank the UGC for the award of Project Fellowship.

REFERENCES

Branden CJ, Jones TA (1990). Between objectivity and subjectivity. Nature, 343: 687-689.

Bubil EM, Yarden Y (2007). The EGF receptor family: spearheading a merger of signaling and therapeutics. Curr. Opin. Cell Biol., 19(2): 124-34.

Cho HS, Leahy DJ (2002). Structure of the extracellular region of HER3 reveals an interdomain tether. Science, 297(5585): 1330–1333.

Daniel B, Fabrice S, Jean I (2009). A reason why the ERBB2 gene is amplified and not mutated in breast cancer. Cancer Cell Internl. 9: 5-1475-2867.
Daniela SK, Richard AV (2005). Tyrosine Kinases as Targets for Cancer Therapy. New Eng J Med. 2(353): 172-187.
Rosenberg I, Ian M (2005). Protein Analysis And Purification: Benchtop Techniques, Birkhäuser Boston book, 2nd ed., 25, 526 pp.
Kufe Pollock DW, Weichselbaum RE, Bast RR, Robert C, Gansler, Holland TS, James F, Frei III Emil (2003). Cancer Medicine 6th edition. BC Decker, Inc. Part II, Section 1.
Laskowski RA, Macarthur MW, Moss DS, Thornton (1993). PROCHECK: a program to check the stereochemical quality of proteinstructures. J. Appl. Crystal, 26: 283-291.
Lingxiu Z, Johnson WC (1992). Environment affects amino acid preference for secondary structure. Proc. Nat. Acad. Sci., 89: 4462-4465.
Olayioye MA (2001). Update on HER-2 as a target for cancer therapy: intracellular signaling pathways of ERBB2/HER-2 and family members. Breast Cancer Res., 3(6): 385–389.
Pao W, Miller V, Zakowski M, Doherty J, Politi K, Sarkaria I, Singh B, Heelan R, Rusch V, Fulton L, Mardis E, Kupfer D, Wilson R, Kris M, Varmus HE (2004). EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. Proc. Natl. Acad. Sci. USA, 101: 13306-13311.
Schlessinger J (2000). Cell signaling by receptor tyrosine kinases. Cell, 13: 211-225.
Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL (1987). Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science, 235: 177-182.
Smith KM, Yacobi R, Van Etten RA (2003). Autoinhibition of Bcr-Abl through its SH3 domain. Mol. Cell., 12:27-37.
Todd R, Wong DT (2009). Oncogenes. Anticancer Res.. 19: 4729–4746.
Xiong X, Huang S, Zhang H, Jianjun L, Jianying S, Jixian X, Yong L, Liping J, Xianchun W, Sonping L (2009). Enrichment and proteomic analysis of plasma membrane from rat dorsal root ganglions. Proteome Sci., 7: 41-1477-5956.
Zwick E, Bange J, Ullrich A (2001). Receptor tyrosine kinase signalling as a target for cancer intervention strategies. Endocr. Relat. Cancer, 8(3): 161-73.
http://www.utdol.com/patients/content/topic.do?topicKey=~nnx7LcCU_HrUNi&source=see_link
http://www.uniprot.org/uniprot/P04626&format=html
http://www.mbio.ncsu.edu/BioEdit/bioedit.html
http://expasy.org/tools/protparam.html
http://prodom.prabi.fr/prodom/current/html/home.php
http://www.expasy.ch/prosite.html
http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html
http://dragon.bio.purdue.edu/pfp
http://www.cbs.dtu.dk/services/NetPhosK
http://blast.ncbi. nlm.nih.gov/Blast.cgi/