Structural Stability and Binding Efficiency of Strong and Weak Interactions in Epidermal Growth Factor Receptor (EGFR) Tyrosine Kinase Domain of Human and Drosophila

Pandian Swarnadeepa, S. Chakkaravarthi[*], Hatti Kaushik, Behera Sanjukta and Kush Anil

Department of Bioinformatics, Vittal Mallya Scientific Research Foundation, Bangalore, India

Abstract: Epidermal growth factor receptor (EGFR) is a cell membrane receptor serving as a molecular target for Non Small Cell Lung cancer (NSCLC). The aim of the study was to analyze the weak ion interactions taking place in the tyrosine kinase domain of Human and Drosophila EGFR. This study throws light on the environment preferability, stabilizing residues taking part in weak and strong interactions of EGFR TK domain. Analysis of short, medium and long range contacts showed that the cation–π interactions are mainly formed by long-range contacts, whereas CH…OC interactions are formed by both long and short range interactions. Except Pro, other non-polar hydrophobic amino acids in Drosophila EGFR TK domain were present in buried environment. However, in humans other than Pro and Ala, which were involved in SS [CH…OC] and SM [NH…OC] interactions, other non-polar hydrophobic amino acid preferred to be in buried environment. The results observed in this study will be useful for understanding the contribution of weak and strong interactions to the stability of EGFR TK domain. Further, the weak interactions have distinct roles in the stability of EGFR TK domain in addition to other conventional strong interactions. The results can lead to the identification of novel targets for NSCLC drugs which in turn will serve as a major break through in the treatment of NSCLC.

INTRODUCTION

Non-small cell lung cancer (NSCLC) is the most common type of lung cancer. There are three major types of NSCLC adenocarcinomas, Ssuamous cell carcinomas and large cell carcinomas [1]. Chemotherapeutic treatment of NSCLC produces objective responses in <30% of patients with advanced-stage disease. Median survival was 8 to 12.5 months [2-4]. The epidermal growth factor receptor (EGFR) and its associated signaling pathways emerged as the leading targets for NSCLC therapy. Detailed steady state modeling of EGFR dynamics has already investigated the binding, internalization and degradation of ligands [5]. EGFR was overexpressed in 80% of NSCLCs and, with other EGFR family members, played a key role in signal transduction pathways, regulating cell proliferation through the mitogen-activated protein kinase pathway and survival through the phosphatidylinositol 3-kinase/AKT pathway [6]. Simplest models of EGFR signaling in tissues must simultaneously account for ligand transport, binding, signal transduction, and gene expression. Given this complexity, such tissue-level models are difficult to test; therefore, they require appropriate experimental paradigms for their validation. The model organisms of developmental genetics, such as the fruit fly (*Drosophila melanogaster*), can be used as experimental systems for the development and validation of computational descriptions of EGFR signaling in tissues [7]. Erlotinib (*N*-(3-ethynylphenyl)-6,7-*bis*(2-methoxyethoxy)quinazolin-4-amine) and Gefitinib (*N*-(3-chloro-4-fluoro-phenyl)-7-

methoxy-6-(3-morpholin-4-ylpropoxy)quinazolin-4-amine), both orally available, low molecular weight, reversible and selective EGFR-TKIs, were used as advanced EGFR-targeted drugs in NSCLC treatment. They blocked signal transduction pathways implicated in proliferation and survival of cancer cells [8, 9]. Computational modeling and quantitative experimentation of the EGFR pathway have therefore addressed reaction kinetics as well as the effect on cell-behavior like proliferation and migration [10]. Improvements in lung cancer treatments have been achieved by novel agents targeting molecular pathways that promote tumor cell growth and survival. The crystal structure of EGFR has been used to perform a virtual screening of a large chemical library. The discovery of *N*-(1,3-benzodioxol-5-ylmethyl)-1-(3-methylphenyl)-1*H*-pyrazolo[3,4-d]pyrimidin-4-amine as a low-micromolar EGFR inhibitor with a unique binding mode opened a new avenue towards the development of novel drug scaffolds as potent and selective inhibitors of EGFR tyrosine kinase activity [11]. Emergence of resistant mutations in drug targets represents a serious problem in the targeted chemotherapy [12]. To resolve the problem of drug resistance, the study of structural stability of the protein is very important as was in this case, the EGFR structural analysis [13].

The stability of a protein was determined by various noncovalent interactions, such as hydrophobic, electrostatic, hydrogen bonds, and van der Waals interactions [14]. Of these, several studies have elucidated that hydrogen bonds have a key role in structure-function relationship of proteins, which include aiding over all folding, maintaining local structure, facilitating protein ligand recognition and enzymatic activity [15]. Apart from these hydrogen bonds, it is now generally accepted that other weak electrostatic interactions termed noncanonical interactions (NCI), such as

*Address correspondence to this author at the Molecular Systems Biology Group, CNC, University of Coimbra, Coimbra, Portugal; E-mail: chakkaravarthi77@gmail.com

those involving C—H...O, C—H...π, and the N—H...π interactions, contribute to structural stability of both small molecules and biological macromolecules [16]. Although recent studies have focused on various aspects of NSCLC, considering the mutation in EGFR, structural analysis but no study has addressed the role and contribution of weak interactions to the overall stability of EGFR. In this article, we have explored the occurrence of the noncanonical and cation–π interactions in EGFR TK domain of humans (2gs2) and Drosophila (1m17). Apart from the above, we systematically investigated the amino acid composition, preference of secondary structural unit, solvent accessibility, and sequential separation of these interaction residues in both human and Drosophila EGFR TK domain. We have also concentrated on the docking of Erlotinib and Gefitinib to TK domain of both Drosophila and human and tried to find out the residues which are not only involved in the weak interactions but also play a role in binding with the ligand molecule.

MATERIALS AND METHODOLOGY

Modeling of Drosophila EGFR TK

With reference to Anuradha and Annadurai [17], Drosophila EGFR TK domain sequence (AAR85240 available at NCBI protein sequence database) was homology modeled using the template, (human EGFR TK domain structures 2gs2 and 1m17) which was obtained from RCSB Protein Data Bank [18]. Swiss Model 8.05 [19] (An Automated Comparative Protein Modeling Server) was used to model this domain. The TK domain consists of 296 amino acids. Pair wise alignment between human and Drosophila TK domain showed 59% similarity.

Identification of Weak Interactions

The type of non-canonical interaction (NCI) was indicated by a two-letter code, in which the first letter indicated the donor atom and the second, the acceptor: M, and S represent the main-chain atom and side-chain atom. The different types of NCI involving main-chain to main-chain (MM)-[C—H…O=C], main-chain to side-chain (MS)-[C—H...O=C; C—H...π; N—H...π], side chain to main-chain (SM)-[C—H...O=C] and side-chain to side-chain (SS)-[C—H...O=C; C—H...π; π...π], were identified using the program Protein Structure Analysis Package (PSAP), that identifies each of the above interactions based on the distances of the atoms under consideration [20]. The number of cation–π interactions in each protein had been calculated using the program CAPTURE developed by Gallivan and Dougherty [21] available at http://capture-.caltech.edu. We have considered only the energetically significant $E_{cat\pi} \leq$ -2 kcal/mole) interactions in the present study.

Contribution of Individual Amino Acid Residues and Their Sequential Separation

The percentage contribution of each amino acid residue that was identified to participate in the different types of non-canonical interaction was calculated as the ratio of the occurrence of a specific amino acid involved in the particular type of non-canonical interaction to the occurrence of the same amino acid in our data. The amino acid residue pairs involved in the interactions were classified as short-range (<±3 residues), medium range (±3 or ±4 residues), or long-range (>± 4 residues) based on their location in the amino acid sequence [22, 23].

Solvent Accessibility and Secondary Structure

Solvent accessibility (ASA) and secondary structure preference of the amino acid residues were among the key factors that were essential to understand the environmental and structure-function relationship of proteins. Secondary structure and accessibility for the amino acid residues were calculated using SOPMA [24] and NETASA [25] respectively. We have systematically analyzed (a) the solvent accessibility pattern of the amino acid residues involved in each type of the non-canonical interaction and (b) the secondary structural preference of cation-π interaction residues. The secondary structural units had been classified as helix, strand, turn, and coil in accordance with Heringa and Argos [26]. Solvent accessibility was divided into three classes, 0–20%, 20–50%, and >50%, indicating the least, moderate, and high accessibility respectively, of the amino acid residues to the solvent [27, 28]. The percentage of an amino acid in a particular ASA class involved in a particular non-canonical interaction was evaluated using the relation

$$\%ASA = \frac{N_{AA}NCI}{N_{AA}} \times 100$$

where $N_{AA}NCI$ and N_{AA} indicate the number of instances a particular amino acid belonging to a specific ASA class involved in NCI and the total number of instances of the same amino acid (found in that ASA class) in the whole protein [16].

Involvement of Stabilizing Residues in Weak Interactions

The stabilizing residues in TK domain of EGFR Human and Drosophila were identified. The occurrence ratio of individual amino acid residue involved in stabilization and also in the non-canonical interaction was calculated in order to assess the probability of the amino acids that are predicted to be involved both as stabilizing residue and in the non-canonical interaction. SRide [29] online server was used to identify the stabilizing residues. A residue was selected as a stabilizing residue if it had high surrounding hydrophobicity, high long-range order, high conservation score and if it belonged to a stabilization center.

Role of Cation-π Interaction on EGFR Ligand Interaction

Docking is a structure-based computational technique that samples complementary fits of a macromolecular receptor and possible ligands. All the docking experiments were done using AutoDock-4 [30, 31]. Erlotinib and Gefitinib were docked to TK domain of both human and Drosophila EGFR. The role played by cation-π interaction in stabilizing the docking was observed along with the positively charged amino acid involved.

Fig. (1). Superimposed structure and pair wise alignment of Human and Drosophila EGFR Tyrosine kinase domain.

RESULTS AND DISCUSSION

Sequence Similarity Analysis of Human and Drosophila EGFR Tyrosine Kinase

ClustalW [32, 33] pairwise alignment algorithm was used to compare human and Drosophila EGFR TK domain. The sequences showed a similarity of 59% and an identity of 46.2%, which infers that the TK domain was conserved in both. Conservative substitution between the positively charged, polar amino acids Arg and Lys, neutral amino acids Val and Leu were observed. The superimposed structure of homology modeled Drosophila EGFR TK domain and human EGFR TK domain was shown in Fig. (**1**). *Drosophila* can be a good genetic model system for screening tyrosine kinase inhibitors, as large numbers of mutants, transgenics were available and EGFR pathway was highly conserved between humans and *Drosophila*. Thus, the ease of dissecting the effect of the inhibitors on the complex signaling pathways *in vivo* in *Drosophila* is far more insightful as compared to the *in vitro* system.

Weak Interaction in Tyrosine Kinase Domain of Drosophila and Human EGFR

The data set consisted 12 different weak and strong interactions. The main chain-main chain interactions play a major role in MM-[NH…OC] and MM-[OH…O], contri-

buting about 50% in Drosophila and 48% in human EGFR TK domain. MM-[CH…OC] contributes about 20% incase of Drosophila and 17% in case of Human EGFR TK domain. Further analysis indicated that the main-chain atoms contribute significantly (about 96%, 93%) either as donor or acceptor or both. This analysis on the data indicates that the distribution of the atoms involved in the noncanonical weak interactions were not random, but were more oriented toward the high incident main-chain atoms. This suggests that the NCI could contribute significantly to the stability of the EGFR proteins. In addition to the main-chain to main chain interactions, we find high incidence of side-chain to main chain and side-chain to side-chain interactions in EGFR TK domain; this data was represented in Table **1**.

Sequential Separation of Weak and Strong Interactions

The contribution of NCI in EGFR TK domain could define the local as well as the global stability of the protein. Therefore, to evaluate the contribution of both stabilities, the residual difference between interacting residues, the sequential distance (short-range: $<\pm3$ residues, medium-range: $\pm$ 3 or 4 residues and long-range: $>\pm$ 4 residues) was calculated and the result is depicted (Fig. **2** and Table **6**). It was evident that a greater percentage (85%) of the interactions was defined by long-range contacts in cation-π interactions. Whereas the main chain-main chain NCI

Table 1.   **Percentage Contribution of Weak and Strong Interactions in the Structural Stability of Human and Drosophila EGFR Tyrosine Kinase Domains**

EGFR-TK	CH…OC				NH…OC				OH…O			
	MM	**MS**	**SM**	**SS**	**MM**	**MS**	**SM**	**SS**	**MM**	**MS**	**SM**	**SS**
Drosophila	61.29	3.35	30.83	4.53	87.48	5.55	4.13	2.84	63.55	16.82	16.82	2.80
Human	51.82	3.30	36.52	8.36	85.06	6.42	4.44	4.07	61.37	15.88	15.88	6.87

MM – Main chain- Main chain interaction, **MS** – Main chain – Side chain interaction, **SM**- Side chain –Main chain interaction, **SS**- Side chain –side chain interaction.

Fig. (2). Sequential separation of strong and weak interaction residues of Human and Drosophila EGFR Tyrosine kinase domain.
Hu – Human; Dr – Drosophila; Interaction ranges (Long – Long range; Medium – Medium range; Short – Short range).

Table 2.   Percentage Amino Acid Composition and Contribution of Weak and Strong Interactions in the Structural Stability of Drosophila EGFR Tyrosine Kinase Domain

Amino acids	% composition	% CH…OC				% NH…OC				% OH…O			
		MM	MS	SM	SS	MM	MS	SM	SS	MM	MS	SM	SS
ALA	8.11	100.00	4.17	75.00	25.00	100.00	12.50	-	-	58.33	12.50	12.50	-
ARG	5.41	100.00	6.25	81.25	6.25	100.00	18.75	31.25	37.50	56.25	0.00	0.00	-
ASN	3.38	100.00	40.00	90.00	30.00	100.00	30.00	40.00	70.00	20.00	30.00	30.00	-
ASP	4.39	100.00	30.77	76.92	61.54	100.00	46.15	7.69	15.38	61.54	30.77	30.77	15.38
CYS	2.36	100.00	-	85.71	14.29	100.00	14.29	14.29	-	42.86	14.29	14.29	-
GLN	2.36	100.00	28.57	85.71	57.14	100.00	71.43	42.86	57.14	71.43	28.57	28.57	14.29
GLU	6.42	100.00	63.16	84.21	42.11	100.00	57.89	5.26	36.84	84.21	47.37	47.37	5.26
GLY	6.42	100.00	5.26	63.16	-	100.00	5.26	10.53	-	68.42	0.00	0.00	-
HIS	1.35	100.00	-	100.00	25.00	100.00	50.00	50.00	-	50.00	0.00	0.00	-
ILE	5.74	100.00	23.53	94.12	17.65	100.00	11.76	5.88	-	64.71	11.76	11.76	-
LEU	13.51	100.00	5.00	87.50	10.00	100.00	12.50	15.00	-	75.00	15.00	15.00	-
LYS	7.43	100.00	-	86.36	9.09	100.00	9.09	27.27	18.18	50.00	9.09	9.09	-
MET	3.04	100.00	-	66.67	22.22	100.00	11.11	11.11	-	33.33	0.00	0.00	-
PHE	3.38	100.00	-	100.00	30.00	100.00	-	10.00	-	70.00	20.00	20.00	-
PRO	4.39	100.00	15.38	92.31	38.46	92.31	-	15.38	-	84.62	7.69	7.69	-
SER	6.08	100.00	16.67	77.78	27.78	94.44	61.11	22.22	5.56	72.22	44.44	44.44	-
THR	4.73	100.00	42.86	85.71	50.00	100.00	64.29	7.14	-	64.29	64.29	64.29	14.29
TRP	2.03	100.00	16.67	100.00	16.67	100.00	-	66.67	16.67	50.00	16.67	16.67	-
TYR	2.7	100.00	25.00	100.00	50.00	100.00	25.00	-	12.50	62.50	50.00	50.00	12.50
VAL	6.76	95.00	10.00	85.00	15.00	100.00	5.00	5.00	-	55.00	10.00	10.00	-

The '– 'shows these corresponding amino acid did not contribute to the interaction.

involving atoms indulge in short range interactions, the contribution of the medium-range contacts was comparatively less in all the interactions studied. Long range interactions were evident only in side chain-side chain interaction of NH…OC, CH…OC and side chain-main chain interaction of NH…OC.

Table 3.　**Percentage Amino Acid Composition and their Contribution to Weak and Strong Interactions in the Structural Stability of Human EGFR Tyrosine Kinase Domain**

Amino Acids	% Composition	% CH...OC				% NH...OC				% OH...O			
		MM	MS	SM	SS	MM	MS	SM	SS	MM	MS	SM	SS
ALA	6.23	100.00	15.79	63.16	15.79	94.74	21.05	15.79	-	73.68	10.53	10.53	-
ARG	4.59	100.00	-	78.57	14.29	100.00	14.29	42.86	35.71	42.86	-	-	-
ASN	2.30	100.00	-	85.71	57.14	100.00	42.86	42.86	42.86	71.43	28.57	28.57	-
ASP	6.23	100.00	15.79	84.21	52.63	89.47	47.37	15.79	31.58	52.63	26.32	26.32	21.05
CYS	1.97	100.00	-	66.67	-	100.00	-	33.33	-	66.67	16.67	16.67	-
GLN	3.28	100.00	10.00	60.00	50.00	100.00	60.00	50.00	50.00	60.00	30.00	30.00	30.00
GLU	7.54	100.00	26.09	65.22	52.17	100.00	26.09	8.70	34.78	60.87	26.09	21.74	17.39
GLY	6.23	100.00	-	63.16	-	100.00	15.79	10.53	-	68.42	5.26	5.26	-
HIS	2.62	100.00	12.50	87.50	25.00	100.00	62.50	62.50	-	62.50	12.50	12.50	-
ILE	8.20	100.00	-	80.00	12.00	96.00	4.00	8.00	-	60.00	8.00	8.00	-
LEU	10.16	100.00	6.45	77.42	9.68	83.87	3.23	9.68	-	54.84	9.68	9.68	-
LYS	7.54	100.00	8.70	86.96	26.09	95.65	8.70	13.04	34.78	69.57	13.04	13.04	-
MET	3.61	100.00	-	100.00	18.18	100.00	-	18.18	-	72.73	-	-	-
PHE	2.30	100.00	-	85.71	28.57	100.00	-	-	-	57.14	-	-	-
PRO	5.25	100.00	12.50	87.50	37.50	81.25	25.00	25.00	-	87.50	6.25	6.25	-
SER	4.59	100.00	21.43	78.57	21.43	92.86	64.29	14.29	7.14	64.29	78.57	78.57	28.57
THR	3.93	100.00	41.67	58.33	33.33	91.67	66.67	8.33	8.33	66.67	75.00	75.00	8.33
TRP	1.97	100.00	-	100.00	33.33	100.00	-	66.67	33.33	66.67	33.33	33.33	-
TYR	3.93	100.00	33.33	91.67	83.33	100.00	8.33	16.67	25.00	66.67	50.00	50.00	16.67
VAL	7.54	100.00	-	100.00	8.70	91.30	17.39	-	-	65.22	-	-	-

The '– 'shows these corresponding amino acid did not contribute to the interaction.

Composition and Contribution of Amino Acids, in Weak and Strong Hydrogen Bonds, in Drosophila and Human EGFR

Amino acid composition and their contribution to weak and strong hydrogen bonds in Drosophila and human EGFR TK domain were analyzed and represented in Tables **2** and **3**, respectively. The percentage of Asp and Glu composition was comparatively higher in human than in Drosophila EGFR TK domain. This might be due to the variation in the number of amino acids between human 322 (amino acids) and Drosophila (296 amino acids) EGFR TK domain. Except Gln, the composition of other hydrophilic amino acids was higher in Drosophila EGFR-TK, which was not the case in human EGFR TK domain, indicating that the EGFR TK of Drosophila was highly composed of hydrophilic amino acids. Both in human and Drosophila EGFR –TK domain, composition of leucine was higher than the other amino acids. As expected in globular proteins, both in human and Drosophila EGFR TK domain, all the weak and strong hydrogen bond interactions mainly occured between main chain atoms, which were studied and well established in detail by Baker and Hubbard [15]. Though the percentage composition of Glu is higher in human EGFR TK domain than Drosophila, but Glu contribution to form MS-

[CH...OC] interaction in Drosophila was higher. More than 40% of acidic amino acids (Ser and Thr) were observed to contribute significantly to strong hydrogen bonds (NH...OC and OH...O) in both human and Drosophila EGFR TK domain. All amino acids were found to be involved in SS-[CH...OC] interactions, except Gly in case of Drosophila and with respect to humans, the exceptions were Gly and Cysteine. However, among the strong side chain - side chain interactions (NH....OC and OH....O), only few of the amino acids were found to be involved, which stands true for both human and Drosophila EGFR-TK domain. This finding indicates that contribution of side chain-side chain weak interaction was much higher than side chain-side chain strong interactions.

Environmental Preference of Amino Acids in Human and Drosophila EGFR TK Domain

All the non-polar hydrophobic amino acids, except Pro, in Drosophila EGFR TK domain were observed to be in buried environment as expected. However, in human EGFR TK domain, Pro in all NCI and Ala involved in SS [CH...OC] and SM-[NH...OC] were observed to be in exposed environment. Other than hydrophobic amino acids,

Table 4.   Environment Preferability of Amino Acids Involved in Weak and Strong Interactions of Drosophila EGFR Tyrosine Kinase Domains

Amino acids	% CH…OC				% NH…OC				% OH…O			
	MM	MS	SM	SS	MM	MS	SM	SS	MM	MS	SM	SS
ALA	B	B	B	B	B	B	-	-	B	B	B	-
ARG	E	E	E	E	E	E	E	E	E	-	-	-
ASN	E	E	E	E	E	E	E	E	E	E	E	-
ASP	E	E	E	E	E	E	E	E	E	E	E	E
CYS	B	-	B	B	B	B	B	-	B	B	B	-
GLN	E	E	E	E	E	E	E	E	E	E	E	E
GLU	E	E	E	E	E	E	E	E	E	E	E	E
GLY	E	E	E	-	E	E	Eq	-	E	-	-	-
HIS	E	-	E	E	E	E	E	-	E	-	-	-
ILE	B	B	B	B	B	B	B	-	B	B	B	-
LEU	B	B	B	B	B	B	B	-	B	B	B	-
LYS	E	-	E	E	E	E	E	E	E	E	E	-
MET	B	-	B	B	B	B	B	-	B	-	-	-
PHE	B	-	B	B	B	-	B	-	B	B	B	-
PRO	E	E	E	E	E	-	Eq	-	E	E	E	-
SER	E	E	E	E	E	E	E	E	E	E	E	E
THR	E	E	E	E	E	E	E	-	E	E	E	Eq
TRP	B	B	B	B	B	-	B	B	B	B	B	-
TYR	B	B	B	B	B	B	-	B	B	B	B	B
VAL	B	B	B	B	B	B	B	-	B	B	B	-

B – Buried, E – Exposed, Eq – Equally distributed.

Table 5.   Environment Preferability of Amino Acids Involved in Weak and Strong Interactions of Human EGFR Tyrosine Kinase Domains

Amino acids	% CH…OC				% NH…OC				% OH…O			
	MM	MS	SM	SS	MM	MS	SM	SS	MM	MS	SM	SS
ALA	B	B	B	E	B	Eq	E	-	B	B	B	-
ARG	E	-	E	E	E	E	E	E	E	-	-	-
ASN	E	-	E	E	E	E	E	E	E	E	E	-
ASP	E	E	E	E	E	E	E	E	E	E	E	E
CYS	B	B	B	-	B	-	B	-	B	B	B	-
GLN	E	E	E	E	E	E	E	E	E	E	E	E
GLU	E	E	E	E	E	E	E	E	E	E	E	E
GLY	E	-	E	-	E	E	E	-	E	E	E	-
HIS	E	E	E	E	E	E	E	-	E	E	E	-
ILE	B	-	B	B	B	B	B	-	B	B	B	-
LEU	B	B	B	B	B	B	B	-	B	B	B	-
LYS	E	E	E	E	E	E	E	E	E	E	E	-
MET	B	-	B	B	B	-	B	-	B	-	-	-
PHE	B	-	B	B	B	-	-	-	B	-	-	-
PRO	E	E	E	E	E	E	E	-	E	E	E	-
SER	E	E	E	E	E	E	E	E	E	E	E	E
THR	E	E	E	E	E	E	E	E	E	E	E	E
TRP	B	-	B	B	B	-	B	Eq	B	B	B	-
TYR	B	B	B	B	B	B	Eq	B	Eq	B	B	E
VAL	B	-	B	B	B	B	-	-	B	-	-	-

B – Buried, E – Exposed, Eq – Equally distributed.

Table 6. Solvent Accessibility, Secondary Structure and Cation-π Interaction Energy of Residues Forming Cation-π Interaction

EGFR-TK	Cation	Res	Sec	%ASA	π	Res	Sec	%ASA	-Ees	-Evdw	-Ecatπ	Dseq
DrEGFR-TK	ARG	138	H	32.90	PHE	193	S	6.00	3.71	2.47	6.18	55
	ARG	279	C	28.80	TYR	280	S	13.30	5.83	2.29	8.12	1
	ARG	143	H	28.30	TRP	182	S	6.40	4.33	2.46	6.79	39
	ARG	260	C	23.50	TRP	200	H	7.30	1.26	2.24	3.5	60
	LYS	287	C	55.30	PHE	288	C	6.00	2.93	0.92	3.85	1
	LYS	16	H	73.00	TYR	29	S	10.70	5.6	1.57	7.17	13
DrEGFR-TK-er	ARG	138	H	32.90	PHE	193	S	6.00	3.4	2.98	6.38	55
	ARG	143	H	28.30	TRP	182	S	6.40	5.17	3.11	8.28	39
	ARG	260	C	23.50	TRP	200	H	7.30	2.27	0.71	2.98	60
	LYS	47	H	32.40	PHE	61	H	7.40	2.21	0.72	2.93	14
	LYS	47	H	32.40	PHE	158	C	6.90	2.67	0.49	3.16	111
	LYS	287	C	55.30	PHE	288	C	6.00	3.71	1.12	4.83	1
	LYS	16	H	73.00	TYR	29	S	10.70	6.89	1.49	8.38	13
DrEGFR-TK-gf	ARG	138	H	32.90	PHE	193	S	6.00	3.4	2.98	6.38	55
	ARG	143	H	28.30	TRP	182	S	6.40	5.17	3.11	8.28	39
	ARG	260	C	23.50	TRP	200	H	7.30	2.27	0.71	2.98	60
	LYS	47	H	32.40	PHE	61	H	7.40	2.21	0.72	2.93	14
	LYS	47	H	32.40	PHE	158	C	6.90	2.67	0.49	3.16	111
	LYS	287	C	55.30	PHE	288	C	6.00	3.71	1.12	4.83	1
	LYS	16	H	73.00	TYR	29	S	10.70	6.89	1.49	8.38	13
HuEGFR-TK	ARG	812	C	32.90	TYR	867	S	9.70	3.66	2.95	6.61	55
	ARG	817	H	28.30	TRP	856	H	4.30	5.24	1.74	6.98	39
	LYS	690	S	53.70	TYR	703	S	10.10	2.54	0.77	3.31	13
HuEGFR-er	ARG	812	C	32.90	TYR	867	S	9.70	3.59	3.43	7.02	55
	ARG	817	H	28.30	TRP	856	H	4.30	5.17	1.64	6.81	39
	LYS	690	S	53.70	TYR	703	S	10.10	2.9	0.83	3.73	13
HuEGFR-gf	ARG	812	C	32.90	TYR	867	S	9.70	3.69	0.72	4.41	55
	ARG	953	H	31.90	TYR	954	H	10.50	1.27	3.06	4.33	1
	ARG	817	H	28.30	TRP	856	H	4.30	5.65	1.15	6.8	39
	ARG	934	C	38.70	TRP	874	H	10.40	1.78	1.26	3.04	60
	LYS	690	S	53.70	TYR	703	S	10.10	2.37	0.75	3.12	13

The amino acids in bold show the additional residues forming cation-π interaction during docking of Erlotinib and Gefitinib to EGFR TK domain of Drosophila and Human along with their energy values. The last column shows the sequential distance between the cation-π bond forming residues based on which they are categorized into Short, Medium and long range interaction.

irrespective of interactions, the amino acid prefers to be in exposed environment only, as depicted in Tables **4** and **5**.

Effect of Erlotinib and Geftinib in the Cation-π Interaction of Human and Drosophila EGFR-TK

Analysis of energetically significant cation-π interaction in both human and Drosophila EGFR-TK (Table **6**) revealed that only half of the observed cation-π interactions in Drosophila EGFR-TK were observed in the human EGFR-TK. This might be due to the conformational difference, specifically in the cationic and aromatic positions of both of

these domains. As explained previously, all cation-π complexes were not equally favorable, their level of stability crucially depends on the type of partners and on the environment [34]. Binding of Erlotinib and Gefitinib to the Drosophila EGFR-TK leads to the deletion of a cation-π interaction (Arg279–Tyr280) and formation of two new cation-π interactions with an over all decrease in the cation-π interaction energy level of -1.33 kcal/mol. This change in the cation-π interaction energy helps to identify that the binding of Erlotinib and Geftinib will enhance the stability of EGFR-TK domain through these interactions. Earlier, quantum chemistry calculations were used to probe the binding of
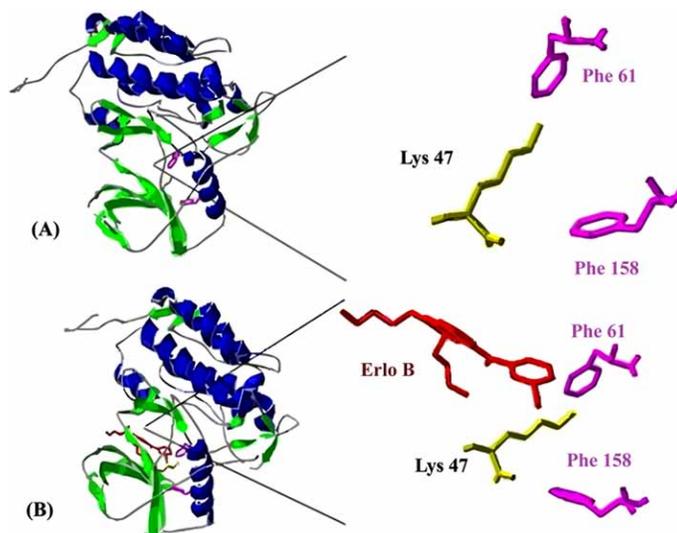
Fig. (3). Effect of Erlotinib in the formation of cation-π interactions on the Drosophila EGFR tyrosine kinase. (**A**) The formation of cation-π interaction in Drosophila EGFR TK domain before Erlotinib binds to the structure. (**B**) Change in the position of the residues - to facilitate the formation of cation-π interaction after the binding of Erlotinib.

adenine cofactor with Arg and Lys side chains of proteins through cation-π interactions. This analysis led to the identification of Arg-Ade base as the most favorable cation-π complex, with an approximate minimum interaction free energy of -7 kcal/mol [34]. Binding of Erlotinib to the Drosophila EGFR-TK forms a new cation-π interaction, where Lys 47 forms interaction with both Phe 61 and Phe 158. Interestingly, this Lys 47 also forms cation-π interaction with Erlotinib, with a cation-π interaction distance of 5.1Å (Fig. **3**). Whereas binding of Erlotinib to Human EGFR-TK does not affect the existing cation-π interaction in this

domain. However, binding of Geftinib enhances the cation-π interaction energy for about -4.8 kcal/mol of this domain through the formation of two new short and long range interactions (Arg 953 – Tyr 954 and Arg 934 – Trp 874). Further to this, binding of Erlotinib to Human EGFR-TK is also stabilized through the formation of cation-π interaction between this ligand molecule and Arg 807 in the EGFR-TK domain (Fig. **4**). From Table **6**, it can be seen that most of the cation-π interaction forming residues environment was not affected due to the binding of both Erlotinib and Geftinib. The secondary structures of most of the cation-π
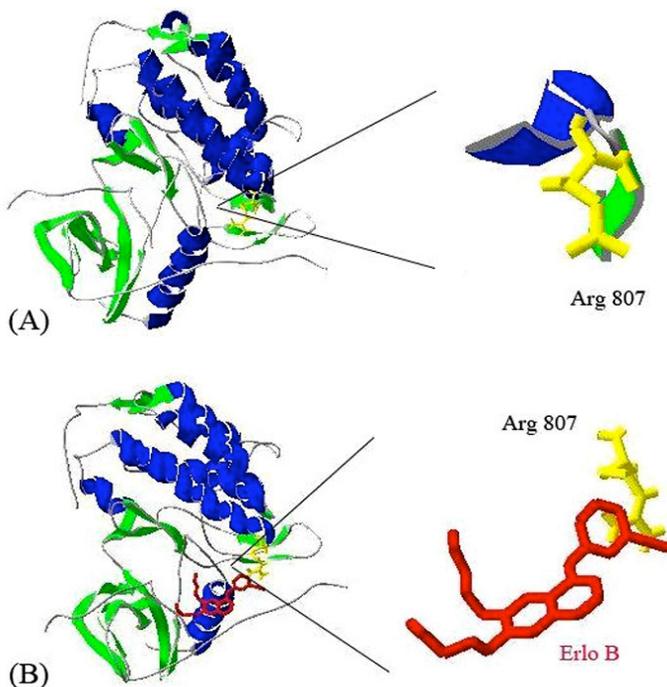
Fig. (4). Effect of Erlotinib in the formation of cation-π interactions on the Human EGFR tyrosine kinase. (**A**) The formation of cation-π interaction in the absence of Erlotinib docked to Human EGFR tyrosine kinase. (**B**) Positively charged aminoacid Arginine at position 807 is involved in cation-π interaction with Erlotinib.

Table 7.   **Stabilizing Residues of Drosophila and Human EGFR Tyrosine Kinase**

Protein	Stabilizing Residues
Drosophila EGFRTK	Gly31, Val44, Ala45, Leu90, Thr92, Gln93, Val147, Lys154
Drosophila EgfrTkErlotinib	Gly31, Val44, Ala45, Glu48, Asn75, Leu90, Thr92, Gln93, Val147, Lys154, Ala184, **Trp200**, Cys252
Drosophila EGFRTKGefitinib	Gly31, Val44, Ala45, Glu48, Asn75, Leu90, Thr92, Gln93, Val147, Lys154, Ala184, **Trp200**, Cys252
Human EGFRTK	Lys704, GLY705, ALA719, LYS828
Human EgfrTkErlotinib	GLY705, VAL718, ALA719, GLN767, LYS828
Human EGFRTKGefitinib	GLY705, VAL718, ALA719, THR766, GLN767, VAL810, VAL821, LYS828, CYS926

Bold residue indicate the occurrence in cation-π interactions.

interacting cationic residues prefer either helix or coil, whereas aromatic residues prefer strand or coil. EGFR-TK of both Human and Drosophila prefers the long range interaction, like membrane protein [35].

Stabilization Effect of Erlotinib and Gefitinib on Drosophila and Human EGFR TK Domain

The stabilizing residues in human and Drosophila EGFR TK domains were identified and presented in Table **7**. The modification in stabilizing residues due to the effective binding of Erlotinib and Gefitinib on these domains was also represented in the same Table **7**. From this table, it has been clearly understood that binding of Erlotininb and Geftinib enhances the stability of Drosophila EGFR-TK through the formation of new stabilizing residues. However, some of the residues which were stabilizing the EGFR TK domain of human no longer stabilized the structure after binding to Erlotinib and Geftinib, instead, new stabilizing residues were formed to stabilize the docked molecule. Trp 200 (a stabilization residue) in Drosophila EGFR TK domain was also involved in cation π interaction suggesting its importance in structural stability of the protein.

CONCLUSION

This study provides insights into the relative importance of strong (NH…O, OH…O) and weak (CH…O) hydrogen bonds in Drosophila and Human EGFR Tyrosine kinase domain. Greater contribution of the NCI is observed for main chain to main chain in CH…OC as well as NH…OC and OH…O. A majority of the cation-π interactions tend to be long range interactions. In cation-π interactions, the cationic residue has a secondary structure preferability of Helix about 55%, whereas the π residue prefers to be strand 54%. Seventy percent of the interacting residue in Drosophila EGFR TK domain involved in cation-π interactions prefers exposed environment. Except some of the hydrophobic amino acids, all the amino acids involved in weak and strong hydrogen bonds prefer to be in exposed environment in both Human and Drosophila EGFR TK domain. Binding of Erlotinib and Geftinib enhances the stability of EGFR-TK domain through these weak interactions. Overall, the study indicates that weak interactions like the conventional hydrogen bonds are environment specific and could greatly influence the binding of drug.

ACKNOWLEDGEMENTS

REFERENCE

[1]   Radiant Lung Study [Homepage on internet] OSI Pharmaceuticals, Inc. 2008. Available from: http://www.radiantlungstudy.com/nsclc

[2]   Bunn PA, Kelly K, Jr. New chemotherapeutic agents prolong survival and improve quality of life in nonsmall cell lung cancer: a review of the literature and future directions. Clin Cancer Res 1998; 4: 1087-100.

[3]   Schiller JH, Harrington D, Belani CP, *et al*. Comparison of four chemotherapy regimens for advanced non-small cell lung cancer. N Engl J Med 2002; 346: 92-8.

[4]   Herbst RS, Johnson DH, Mininberg E, *et al*. Phase I/II trial evaluating the anti-vascular endothelial growth factor monoclonal antibody bevacizumab in combination with the HER-1/epidermal growth factor receptor tyrosine kinase inhibitor erlotinib for patients with recurrent non-small cell lung cancer. J Clin Oncol 2005; 23: 1-12.

[5]   Wiley H, Cunningham D. A steady state model for analyzing the cellular binding, internalization and degradation of polypeptide ligands. Cell 1981; 25: 433-40.

[6]   Hirsch FR, Varella-Garcia M, Bunn PA, *et al.* Epidermal growth factor receptor in non-small cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. J Clin Oncol 2003; 21: 3798-807.

[7]   Goentoro LA, Shvartsman SY. Complex Science System in Biomedicine- Patterning by EGF Receptor: Models from *Drosophila* Development. 1st ed. Spinger: US 2006.

[8]   Ciardiello F, Tortora G. A novel approach in the treatment of cancer: targeting the epidermal growth factor receptor. Clin Cancer Res 2001; 7: 2958-70.

[9]   Grunwald V, Hidalgo M. Developing inhibitors of the epidermal growth factor receptor for cancer treatment. J Nat Cancer Inst 2003; 95: 851-67.

[10]   Lauffenburger DA, Linderman J. Receptors: Models for Binding, Trafficking, and Signaling. Oxford University Press: Oxford 1996.

[11]   Cavasotto CN, Ortiz MA, Abagyan RA, Piedrafita FJ. In silico identification of novel EGFR inhibitors with antiproliferative activity against cancer Cells. Bioorg Med Chem Lett 2006; 16: 1969-74.

[12]   Liu B, Bernard B, Wu JH. Impact of EGFR Point Mutations on the Sensitivity to Gefitinib: Insights From Comparative Structural Analyses and Molecular Dynamics Simulations. Proteins Struct Funct Bioinform 2006; 65: 331-46.

[13]   Verkhivker GM. In silico profiling of tyrosine kinases binding specificity and drug resistance using Monte Carlo simulations with the ensembles of protein kinase crystal structures. Biopolymers 2007; 85: 333-38.

[14]   Gromiha MM, Selvaraj S. Recent Research Developments in Protein Folding, Stability, and Design. Research Signpost: Trivandrum, India 2002.

[15]   Baker EN, Hubbard RE. Hydrogen bonding in globular proteins. Prog Biophys Mol Biol 1984; 44: 97-179.

[16]   Chakkaravarthi S, Babu MM, Gromiha MM, Jayaraman G, Sethumadhavan R. Exploring the environmental preference of weak interactions in (α/ß)8 barrel proteins. Proteins Struct Funct Bioinform 2006; 65: 75-86.

[17]   Aritakula A, Ramasamy A. Drosophila-based *in vivo* assay for the validation of inhibitors of the epidermal growth factor receptor/ RasPathway. J Biosci 2008; 33: 731-742.

[18]   Berman HM, Westbrook J, Feng Z, *et al.* The protein data bank. Nucleic Acids Res 2000; 28: 235-42.

[19]   Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 2006; 22: 195-201.

[20]   Balamurugan B, Roshan MNAMd, Hameed BS, *et al.* PSAP: protein structure analysis package. J Appl Cryst 2007; 40: 773-7.

[21]   Gallivan JP, Dougherty DA. Cation-pi interactions in structural biology. Proc Natl Acad Sci USA 1999; 96: 9459-64.

[22]   Gromiha MM, Selvaraj S. Influence of medium and long range interactions in different structural classes of globular proteins. J Biol Phys 1997; 23: 151-62.

[23]   Gromiha MM, Selvaraj S. Inter residue interaction in protein folding and stability. Prog Biophys Mol 2004; 86: 235-77.

[24]   Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 1995; 11: 681-4.

[25]   Ahmad S, Gromiha MM, Sarai A. RVP-net: online prediction of real valued accessible surface area of proteins from single sequences. Bioinformatics 2003; 19: 1849-51.

[26]   Heringa J, Argos P. Strain in protein structures as viewed through nonrotameric side chains: I. Their position and interaction. Proteins 1999; 37: 30-43.

[27]   Gilis D, Rooman M. Stability changes upon mutation of solvent accessible residues in proteins evaluated by database-derived potentials. J Mol Biol 1996; 257: 1112-26.

[28]   Gilis D, Rooman M. Predicting protein stability changes upon mutation using database-derived potentials: solvent accessibility determines the importance of local *versus* non-local interactions along the sequence. J Mol Biol 1997; 272: 276-90.

[29]   Magyar C, Gromiha MM. SRide: a server for identifying stabilizing residues in proteins. Nucleic Acids Res 2005; 33: 303-5.

[30]   Morris GM, Goodsell DS, Halliday RS, *et al.* Automated docking using a lamarckian genetic algorithm and and empirical binding free energy function. J Comp Chem 1998; 19: 1639-62.

[31]   Huey R, Morris GM, Olson AJ, Goodsell DS. A semiempirical free energy force field with charge-based desolvation. J Comp Chem 2007; 28: 1145-52.

[32]   Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 1970; 48: 443-53.

[33]   Smith TF, Waterman MS. Identification of common molecular subsequences. J Mol Biol 1981; 147: 195-7.

[34]   Boit C, BuiSine E, Rooman M. Free-energy calculations of protein−ligand cation−π and amino−π interactions: from vacuum to proteinlike environments. J Am Chem Soc 2003; 125: 13988-94.

[35]   Gromiha MM, Selvaraj S. Role of medium- and long-range interactions in discriminating globular and membrane proteins. Int J Biol Macromol 2001; 29: 25-34.