Appl Biochem Biotechnol (2010) 160:1723–1733
DOI 10.1007/s12010-009-8662-4

In Silico Identification of Significant Detrimental Missense Mutations of EGFR and Their Effect with 4-Anilinoquinazoline-Based Drugs

R. Rajasekaran · Rao Sethumadhavan

Received: 24 February 2009 / Accepted: 28 April 2009 /
Published online: 20 May 2009

Abstract In this work, we identified the detrimental missense mutations (point mutations) in epidermal growth factor receptor (EGFR) and its binding efficiency with the inhibitors namely Erlotinib, Gefitinib, and Lapatinib. Out of 26 point mutations on EGFR, 12 point mutations were commonly less stable, deleterious, and damaged as shown by all the three servers, I-Mutant2.0, SIFT, and PolyPhen. Further, we modeled 12 mutants and superimposed with the native EGFR to get RMSD values. Docking studies showed that Erlotinib had lesser binding affinity against both native and all the 12 mutants. Gefitinib had maximum binding affinity only with two mutants, viz., R748P and L858R. Lapatinib had maximum binding affinity with both native and other 10 mutants. Based on our computational analysis, we recommend that the combined administration of Gefitinib and Lapatinib could give a better effect in combating the disease.

Keywords Point mutation · EGFR · Erlotinib · Gefitinib · Lapatinib

Introduction

The epidermal growth factor receptor (EGFR), also called ERBB1/HER1, is a transmembrane receptor tyrosine kinase that transduces signals critical for cell proliferation, differentiation, and motility. Upon growth-factor binding, EGFR and other ErbB family members undergo homo- or heterodimerization and activate their cytoplasmic tyrosine kinase domains to initiate intracellular signaling [1, 2]. Overexpression or mutational activation of the EGFR is implicated in the development and progression of numerous human malignancies, and a number of small-molecule tyrosine kinase inhibitors (TKIs) have been developed to target the ATP-binding cleft of the EGFR [3]. Several of these inhibitors are currently in clinical trials or have been approved for clinical use, including the 4-anilinoquinazolines Erlotinib (Tarceva), Gefitinib (Iressa), and Lapatinib (GW572016) [4–6].

Gefitinib and Erlotinib are reversible small molecule inhibitors of the intracellular tyrosine kinase domain of EGFR. Gefitinib downregulates EGFR signaling and promotes

R. Rajasekaran · R. Sethumadhavan (✉)
Bioinformatics Division, School of Biotechnology, Chemical and Biomedical Engineering,
Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India
e-mail: rsethumadhavan@vit.ac.in

apoptosis alone and in combination with cytotoxic agents. It has been shown to be active in lung cancer, particularly in a subpopulation of patients with EGFR mutations in the tyrosine kinase domain. Other cancers have also shown some response, such as glioblastoma, cervical, pancreatic, and gastric [7]. Erlotinib is active in a similar tumor subset to Gefitinib. The major side effects of these compounds are acneiform rash and diarrhea [8]. There are current studies investigating the dose to rash effect, as a significant number of patients experiencing a positive effect of the compound on their tumor developed at least a grade 2 rash [9]. Lapatinib is a novel member of the 4-anilinoquinazoline class of kinase inhibitors. It is a potent, reversible, selective dual inhibitor of EGFR and ErbB2 kinases [6, 10, 11].

Conferring the drug-resistance single amino acid polymorphism (missense mutation or point mutation) is a major and challenging task in cancer target (EGFR) studies [12]. So, we computationally identified the detrimental single-point mutations on EGFR receptor and investigated its binding efficiency with the three major inhibitors, viz., Erlotinib, Gefitinib, and Lapatinib.

Materials and Methods

Datasets

The protein sequence and variants (single amino acid polymorphisms) for EGFR were obtained from the Swissprot database available at http://www.expasy.ch/sprot/ to find out the detrimental point mutants. Some polymorphic variants of which may be disease(s) associated with defects in a given protein and most of them are non-synonymous SNPs and single amino acid polymorphisms [13–15]. 3D Cartesian coordinates of EGFR-drug complex with PDB IDs 1m17, 2ity, and 1xkk from Protein Data Bank [16] were obtained for docking studies based on detrimental point mutants.

Predicting Stability Change on Single Amino Acid Polymorphism Based on Support Vector Machine (I-Mutant 2.0)

We used the program I-Mutant2.0 available at http://gpcr.biocomp.unibo.it/cgi/predictors/I-Mutant2.0/I-Mutant2.0.cgi. I-Mutant 2.0 is a support vector machine-based tool for the automatic prediction of protein stability changes upon single-point mutations. I-Mutant 2.0 predictions are performed starting either from the protein structure or, more importantly, from the protein sequence [17]. This program was trained and tested on a dataset derived from ProTherm [18], which is presently the most comprehensive available database of thermodynamic experimental data of free energy changes of protein stability upon mutation under different conditions. The output file shows the predicted free energy change value or sign (DDG) which is calculated from the unfolding Gibbs free energy value of the mutated protein minus the unfolding Gibbs free energy value of the native type (kcal/mol). Positive DDG values means that the mutated protein posses high stability and vice versa.

Analysis of Functional Consequences of Point Mutant by Sequence Homology-Based Method (SIFT)

We used the program SIFT [19] which is specifically to detect the deleterious single amino acid polymorphism, available at http://blocks.fhcrc.org/sift/SIFT.html. SIFT is a sequence homology-based tool which presumes that important amino acids will be conserved in the

protein family. Hence, changes at well-conserved positions tend to be predicted as deleterious [19]. We submitted the query in the form of protein sequences. The underlying principle of this program is that SIFT takes a query sequence and uses multiple alignment information to predict tolerated and deleterious substitutions for every position of the query sequence. SIFT is a multistep procedure that, given a protein sequence, (a) searches for similar sequences, (b) chooses closely related sequences that may share similar function, (c) obtains the multiple alignment of these chosen sequences, and (d) calculates normalized probabilities for all possible substitutions at each position from the alignment. Substitutions at each position with normalized probabilities less than a chosen cutoff are predicted to be deleterious and greater than or equal to the cutoff are predicted to be tolerated [20]. The cutoff value in SIFT program is tolerance index of ≥ 0.05. The higher the tolerance index is, the less functional impact a particular amino acid substitution is likely to have.

Simulation for Functional Change in Point Mutant by Structure Homology-Based Method (PolyPhen)

Analyzing the damaged point mutations at the structural level is considered to be very important to understand the functional activity of concerned protein. We used the server PolyPhen [21] which is available at http://coot.embl.de/PolyPhen/ for this purpose. Input options for PolyPhen server is protein sequence or SWALL database ID or accession number together with sequence position with two amino acid variants. We submitted the query in the form of protein sequence with mutational position and two amino acid variants. Sequence-based characterization of the substitution site, profile analysis of homologous sequences, and mapping of substitution site to a known protein 3D structures are the parameters taken into account by PolyPhen server to calculate the score. It calculates position-specific independent counts (PSIC) scores for each of the two variants and then computes the PSIC scores difference between them. The higher the PSIC score difference is, the higher the functional impact a particular amino acid substitution is likely to have.

Modeling Single Amino Acid Polymorphism Location on Protein Structure to Compute RMSD

Structure analysis was performed for evaluating the structural deviation between native type and mutant types by means of root mean square deviation (RMSD). We used the web resource single amino acid polymorphism database (SAAPdb) [22] to identify the 3D structure of EGFR (PDB ID: 2a91). We also confirmed the mutation position and the mutation residue in PDB ID 2a91. The mutation was performed by using SWISSPDB viewer, and energy minimization for 3D structures was performed by NOMAD-Ref server [23]. This server use Gromacs as default force field for energy minimization based on the methods of steepest descent, conjugate gradient, and L-BFGS methods [24]. We used conjugate gradient method to minimize the energy of the 3D structure of EGFR protein. In order to optimize the 3D structure of EGFR protein, we used ifold server [25] for simulated annealing which is based on discrete molecular dynamics and is one of the fastest strategies for simulating protein dynamics. This server is also efficient in sampling the vast conformation space of biomolecules in both length and time scales. Divergence in mutant structure with native structure is due to mutation, deletions, and insertions [26], and the deviation between the two structures could alter the functional activity [27] with respect to binding efficiency of the inhibitors which is evaluated by their RMSD values.

Identification of Binding Sites and Computation of Atomic Contact Energy Between EGFR and Inhibitors

In order to compute the atomic contact energy (ACE) between EGFR and three inhibitors, we submitted the three PDB IDs 1m17, 2ity, and 1xkk into the Ligand Contact Tool (LCT) program available at http://firedb.bioinfo.cnio.es/Php/Contact.php [28]. This server calculates contacts between the binding residues of EGFR with inhibitors with default parameters. Then, we unbound the inhibitors from EGFR of the three PDB IDs, 1m17, 2ity, and 1xkk to perform point mutations on EGFR by swisspdb viewer, energy minimization by NomadRef, and simulated annealing by ifold. Finally, we used the program PatchDock for docking the native and mutant EGFR with the three inhibitors to compute ACE by using additional option of binding residue parameter. The underlying principle of this server is based on molecular shape representation, surface patch matching plus filtering, and scoring [29]. It is aimed at finding docking transformations that yield good molecular shape complementarity. Such transformations, when applied, induce both wide interface areas and small amounts of steric clashes. A wide interface is ensured to include several matched local features of the docked molecules that have complementary characteristics. The PatchDock algorithm divides the Connolly dot surface representation [30, 31] of the molecules into concave, convex, and flat patches. Then, complementary patches are matched in order to generate candidate transformations. Each candidate transformation is further evaluated by a scoring function that considers both geometric fit and atomic desolvation energy [32, 33]. Finally, an RMSD clustering is applied to the candidate solutions to discard redundant solutions. The main reason behind PatchDock's high efficiency is its fast transformational search, which is driven by local feature matching rather than brute-force searching of the 6D transformation space. It further speeds up the computational processing time by utilizing advanced data structures and spatial pattern detection techniques, such as geometric hashing and pose clustering.

Moreover, ACE should provide a reasonably accurate and rapidly evaluable solvation component of free energy and should, thus, make accessible range of docking, design, and protein-folding calculations. ACE can be considered as pairwise atomic desolvation energies. ACE is based on interresidue contact energies belonging to a family of potentials collectively called statistical (or structure-derived) potentials. Its principle is to estimate the free energy of replacing atom–water contacts by atom–atom contacts for various types of protein atoms. These so-called atomic contact energies were obtained from an analysis of atom-pairing frequencies in known protein structures using a procedure formally analogous to the method of Miyazawa and Jernigan [34], though with some significant departures. The ACE was derived by the following parameters which include computation of contact energies, reference state, energy unit (quasi-chemical approximation), the complete free energy function, electrostatic interactions, side-chain entropy, van der Waals energy, backbone conformation, and rigid body degrees of freedom [33].

Results and Discussion

Single Amino Acid Polymorphism Dataset from Swissprot

The EGFR protein and a total of 26 variants (R98Q, P266R, R521K, V674I, E709A, E709K, G719A, G719C, G719D, G719S, G724S, E734K, L747F, R748P, Q787R, T790M, L833V, V834L, L858M, L858R, L861Q, G873E, R962G, H988P, L1034R, A1210V) investigated in this work were retrieved from Swissprot database [13–15].

Identification of Functional Variants by I-Mutant 2.0

The more negative the DDG value is, the less stable the given point mutation is likely to be, as predicted by I-Mutant 2.0 server [17]. Out of 26 variants, we obtained 25 variants found to be less stable from this server as shown in Table 1. Out of such 25 variants, three variants, viz., G719A, G719C, and L858R showed DDG values of −2.08, −2.08, and −2.18, respectively. The other nine variants, viz., G719D, G719S, G724S, R748P, Q787, L833V, V834L, L861Q, and R962G showed DDG values between −1.79 and −1.22, and the remaining 13 variants showed a DDG values between −0.89 and −0.03 as depicted in Table 1.

Out of the 25 variants which showed negative DDG, three variants namely G719S, G724S, and L861Q changed their amino acid from nonpolar to polar amino acid, three variants, P266R, L858R, and L1034R, changed from nonpolar to positively charged, three variants, R962G, R748P, and H988P, changed from positively charged to nonpolar, two variants, G719D and G873E, changed from nonpolar to negatively charged, two

Table 1 List of variants that were predicted to be functionally significant by I-Mutant 2.0, SIFT, and PolyPhen.

Swiss variant	AA change	DDG	Tolerance index	PSIC SD
VAR_019293	R98Q	−0.89	0.54	0.319
VAR_019294	P266R	−0.88	0.11	0.584
VAR_019295	R521K	−0.69	0.85	0.335
VAR_019296	V674I	−0.05	0.69	0.313
VAR_026084	E709A	0.09	0.02	2.222
VAR_026085	E709K	−0.36	0.11	1.934
VAR_026086	G719A	−2.08	0.00	2.112
VAR_026087	G719C	−2.08	0.00	2.787
VAR_026088	G719D	−1.63	0.00	2.337
VAR_019297	G719S	−1.58	0.00	2.112
VAR_026089	G724S	−1.68	0.00	2.112
VAR_026090	E734K	−0.54	0.31	1.767
VAR_026093	L747F	−0.43	0.02	1.710
VAR_026095	R748P	−1.69	0.05	2.247
VAR_026097	Q787R	−1.22	0.05	2.292
VAR_026098	T790M	−0.17	0.02	2.418
VAR_026099	L833V	−1.64	0.04	0.821
VAR_026100	V834L	−1.79	0.00	1.710
VAR_026101	L858M	−0.32	0.24	1.298
VAR_019298	L858R	−2.18	0.00	2.198
VAR_026102	L861Q	−1.39	0.21	0.788
VAR_026103	G873E	−0.57	0.44	0.538
VAR_019299	R962G	−1.23	0.05	2.261
VAR_019300	H988P	−0.76	0.16	2.382
VAR_042095	L1034R	−0.51	0.58	1.594
VAR_042096	A1210V	−0.03	0.00	1.227

AA change amino acid change

variants, E709K and E734K, changed from negatively charged to positively charged, one variant, R98Q, changed from positively charged to polar (uncharged), one variant, T790M, changed from polar to nonpolar, one variant, Q787R, changed from polar to positively charged, and one variant, L747F, changed from nonpolar amino acids to aromatic amino acids. It is also to be noted that seven other variants, V674I, G719A, G719C, L833V, V834L, L858M, and A1210V, which retained their nonpolar properties, one other variant, R521K, which retained its positively charged property, were found to be less stable by I-Mutant 2.0.

Deleterious Single Point Mutant by SIFT Program

The conservation level of a particular position in a protein was determined by using a sequence homology-based tool, SIFT [19]. Protein sequences of 26 variants were submitted independently to SIFT program to check its tolerance index. The higher the tolerance index is, the lesser the functional impact a particular amino acid substitution is likely to have and vice versa. Among the 26 variants, 15 variants were found to be deleterious having the tolerance index score of ≤0.05. The results are shown in Table 1. We observed that out of 15 such deleterious variants, eight variants, G719A, G719C, G719D, G719S, G724S, V834L, L858R, and A1210V, showed a highly deleterious tolerance index score of 0.00. Three variants, E709A, L747F, and T790M, with a tolerance index score of 0.02 followed by one variant, L833V, with a tolerance index score of 0.04, and the remaining three variants, R748P, Q787R, and R962G, with a tolerance index score of 0.05. It is to be noted that all the deleterious variants except E709A according to SIFT also were seen to be less stable by I-Mutant 2.0 server.

Damaged Single Point Mutant by PolyPhen Server

The structural levels of alteration were determined by applying PolyPhen program [21]. Protein sequence with mutational position and amino acid variants associated to 26 single point mutants investigated in this work was submitted as input to the PolyPhen server, and the results were shown in Table 1. A PSIC score difference of 1.5 and above is considered to be damaging. It can be seen that out of 26 variants, 17 variants were considered to be damaging by PolyPhen. Also, these 17 variants exhibited a PSIC score difference between 1.594 and 2.787. It is to be noted that all the variants except E709A which were considered to be damaging by PolyPhen also were seen to be less stable according to I-Mutant 2.0 server. Similarly, all the variants except E734K, H988P, and L1034R, which were considered to be damaging by PolyPhen, also were seen to be deleterious according to SIFT server.

We rationally considered the most potential detrimental point mutations for further course of investigations which were G719A, G719C, G719D, G719S, G724S, L747F, R748P, Q787R, T790M, V834L, L858R, and R962G as they were commonly found to be less stable, deleterious, and damaging by I-Mutant 2.0, SIFT, and PolyPhen servers also. Among these, three variants, G719S, T790M, and L858R, showed very good agreement with experimental observation performed elsewhere [35].

Computing RMSD by Modeling of Mutant Structures

Mapping the 12 potential detrimental variants into protein structure information was obtained from SAAPdb [22]. The available structure for EGFR protein has a PDB ID 2a91.

The mutational position and amino acid variants were mapped in the 2a91 native structure. Mutation at specified position was performed by SWISSPDB viewer independently to get modeled structures. Then, energy minimizations and simulated annealing were performed by NOMAD-Ref server [23] and ifold server [25], respectively, for both the native structure (PDB 2a91) and mutant modeled structures (G719A, G719C, G719D, G719S, G724S, L747F, R748P, Q787R, T790M, V834L, L858R, and R962G).

In order to find out the deviation between the two structures, we superimposed the native structure (PDB 2a91) with all the mutant modeled structures to get RMSD. The higher the RMSD value is, the more the deviation is between the native and mutant structure, which in turn changes their binding efficiency with inhibitors due to deviation in the 3D space of the binding residues of EGFR. Figure 1 shows the RMSD for native structure with all the mutant modeled structures. It can be seen from Fig. 1 that the variant G719C exhibited a high RMSD value of 1.23 Å followed by the variant L858R with RMSD, 1.19 Å, four other variants, G719A, G719D, G724S, and T790M, with RMSD, 1.18 Å, the variant, G719S with RMSD, 1.17 Å, two other variants, V834L and R748P, with RMSD, 1.16 Å, two other variants, Q787 and R962G, with RMSD, 1.15 Å, and finally the variant, L747F, with RMSD, 1.14 Å. It was interesting to note that we found only two variants, G719C and L858R, with higher RMSD (Fig. 2), which were observed to be less stable with DDG value of >-2.0, deleterious with the tolerance index score of ≤ 0.00 and damaging with PSIC score of >2.0 as compared to other variants.

Rationale of Binding Efficiency for Native and Mutant EGFR with its Inhibitors

In order to understand the binding efficiency of EGFR protein with its inhibitors, we selected three PDB IDs 1m17, 2ity, and 1xkk which have EGFR complex with Erlotinib, Gefitinib, and Lapatinib, respectively. The LCT program was used to calculate contacts between the binding residues of EGFR and inhibitor (Table 2). It can be seen from Table 2 that a total of 14 amino acids act as binding residues for EGFR complex with Erlotinib, a total of 16 amino acids act as binding residues for EGFR complex with Gefitinib, and a

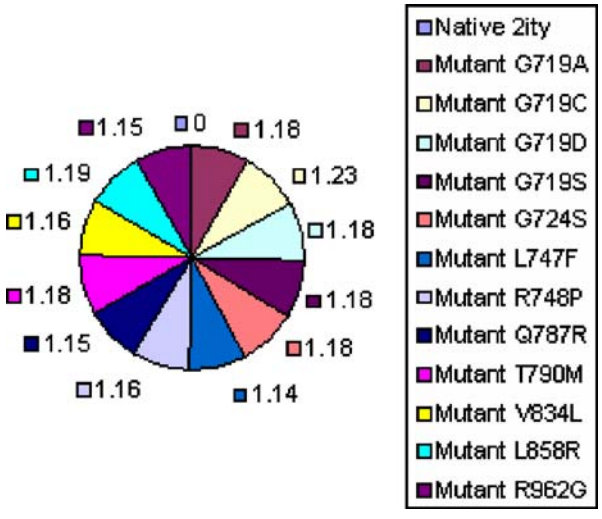

Fig. 1 RMSD for native with mutants of EGFR: the values of RMSD between native- and mutant-modeled structures are represented in Å

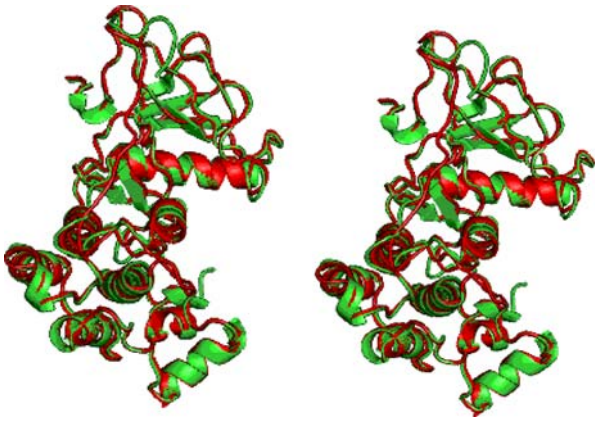

Fig. 2 **a** Superimposed structure of native (*green*) with mutant G719C (*red*) with the RMSD of 1.23 Å. **b** Superimposed structure of native (*green*) with mutant L858R (*red*) with the RMSD of 1.19 Å

total of 23 amino acids act as binding residues for EGFR complex with Lapatinib. Based on this analysis, we believe that Lapatinib might establish a best binding affinity with EGFR, followed by Gefitinib and Erlotinib. It is interesting to note from Table 2 that Thr (766) is found to be a common amino acid in all the three PDB IDs, 1m17, 2ity, and 1xkk. Also, Thr (766) acts as a key binding residue for EGFR with all the three inhibitors supported by an experimental study performed by some other group [36].

To understand the binding efficiency of these three inhibitors with both native and mutant EGFR structures, we unbound the inhibitors from EGFR for all the PDB IDs, 1m17, 2ity, and 1xkk. We once again performed the point mutation by SWISSPDB viewer for the 12 variants as shown in Fig. 1. Energy minimization was performed by Nomad-Ref [23] followed by simulated annealing with ifold server [25] to get optimized structures for both native and mutants.

Docking was performed using PatchDock server between the three inhibitors with both native- and mutant-modeled structures of EGFR to find out the binding efficiency in the form of ACE as could be seen from Table 3. By this analysis, we found that Erlotinib with EGFR of both native and mutants have the ACE values between −151.13 and −165.93. Gefitinib with EGFR of both native and mutants have the ACE values between −168.70 and −203.80. Lapatinib with EGFR of both native and mutants have the ACE values

Table 2 Binding residues of EGFR receptor–inhibitors complex.

PDB IDs	Binding residues in EGFR
1m17	Leu(694), Ala(719), Lys(721), Leu(764), **Thr(766)**, Gln(767), Leu(768), Met(769), Pro(770), Phe(771), Gly(772), Leu(820), Thr(830), Asp(831)
2ity	Leu(718), Gly(719), Val(726), Ala(743), Lys(745), Glu(762), **Thr(766)**, Leu(788), Thr(790), Gln(791), Leu(792), Met(793), Pro(794), Gly(796), Leu(844), Thr(854)
1xkk	Leu(718), Val(726), Ala(743), Lys(745), **Thr(766)**, Cys(775), Arg(776), Leu(777), Leu(788), Thr(790), Gln(791), Leu(792), Met(793), Gly(796), Cys(797), Leu(799), Asp(800), Leu(844), Thr(854), Asp(855), Phe(856), Leu(858), Met(1002)

The common binding residue is shown in bold

Table 3 ACE for native with mutants of EGFR in kcal/mol.

Structures	ACE of Erlotinib	ACE of Gefitinib	ACE of Lapatinib
Native type	−152.24	−202.56	−351.04
Mutant G719A	−159.80	−203.05	−299.56
Mutant G719C	−165.88	−203.22	−325.14
Mutant G719D	−161.49	−199.64	−288.78
Mutant G719S	−151.13	−172.96	−348.00
Mutant G724S	−164.58	−202.89	−282.31
Mutant L747F	−143.44	−178.09	−351.45
Mutant R748P	−163.95	−201.97	−174.11
Mutant Q787R	−155.62	−169.70	−328.70
Mutant T790M	−141.83	−203.80	−300.90
Mutant V834L	−165.93	−173.06	−329.12
Mutant L858R	−153.17	−202.98	−142.50
Mutant R962G	−163.93	−203.22	−351.54

between −142.50 and −351.54. It was interesting to note from Table 3 that Erlotinib had lesser binding affinity with both native and all mutants as compared with Gefitinib and Lapatinib. But Gefitinib had maximum binding affinity only with two mutants, viz., R748P and L858R as compared to Lapatinib, whereas Lapatinib had maximum binding affinity with both native and all the mutants except the two mutants, R748P and L858R, in terms of ACE by docking analysis. We emphasize that the maximum binding of Gefitinib with two mutants namely R748P and L858R might be due to the 3D conformation of Gefitinib which exclusively made comfortable fit with high ACE into the 3D space of the binding residues of these two mutants as compared to other mutants. Similarly, Lapatinib exclusively established the best and complementary fit with maximum binding affinity and high ACE by its nature of 3D conformation with the 3D space of the binding residues of native and all mutants except R748P and L858R. Figure 3 shows the docked complex of Lapatinib with mutant R962G having the highest ACE as an illustrative example of this work. This analysis reveals that EGFR may possibly render resistance with Erlotinib and Gefitinib

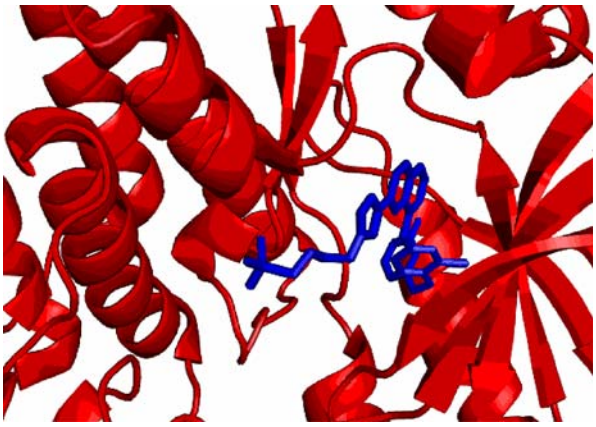

Fig. 3 Docked complex of Lapatinib (*green*) with R962G mutant (*red*) having the highest ACE of −351.54

than Lapatinib based on ACE which is also in good agreement with clinical observations [37–40].

Conclusion

Based on our investigation, we concluded that out of 24 variants, 12 variants showed less stable, deleterious and damaged by I-Mutant2.0, SIFT, and PolyPhen servers, respectively, in which three variants namely G719S, T790M, and L858R are also experimentally proven [35]. Furthermore, Lapatinib could be the potential inhibitor for native and all the mutants except R748P and L858R. Gefitinib could be the potential inhibitor for the latter two mutants in which the variant L858R is sensitive to Gefitinib was proved experimentally [41]. Therefore, the overall conclusion of our work suggests that combined administration of Gefitinib and Lapatinib could be a novel strategy for cancer treatment if the individual drug does not show the desired result. The overall scope of our result is (1) to consider computationally a suitable protocol for single amino acid polymorphism (point mutation) analysis before wet lab experimentation, (2) to provide optimal path for further clinical and experimental studies to characterize EGFR mutants in depth, (3) to consider an appropriate protocol for identifying the drug resistance by point mutations, (4) to validate the potency of drug molecules, and (5) to select the potential drug based on point mutation associated cancer targets.

Acknowledgment The authors thank the management of Vellore Institute of Technology for providing the facilities to carry out this work.

References

1. Yarden, Y., & Sliwkowski, M. X. (2001). *Nature Reviews. Molecular Cell Biology, 2*(2), 127–137. doi:10.1038/35052073.
2. Schlessinger, J. (2004). *Science, 306*(5701), 1506–1507. doi:10.1126/science.1105396.
3. Hynes, N. E., & Lane, H. A. (2005). *Nature Reviews. Cancer, 5*(5), 341–354. doi:10.1038/nrc1609.
4. Pollack, V. A., et al. (1999). *The Journal of Pharmacology and Experimental Therapeutics, 291*(2), 739–748.
5. Wakeling, A. E., et al. (2002). *Cancer Research, 62*(20), 5749–5754.
6. Rusnak, D. W., et al. (2001). *Molecular Cancer Therapeutics, 1*(2), 85–94.
7. Mendelsohn, J., & Baselga, J. (2006). *Seminars in Oncology, 33*(4), 369–385. doi:10.1053/j.seminoncol.2006.04.003.
8. Giaccone, G. (2005). *Future Oncology (London, England), 1*(4), 449–460. doi:10.2217/14796694.1.4.449.
9. Peréz-Soler, R., & Saltz, L. (2005). *Journal of Clinical Oncology, 23*(22), 5235–5246. doi:10.1200/JCO.2005.00.6916.
10. Rusnak, D. W., et al. (2001). *Cancer Research, 61*(19), 7196–7203.
11. Xia, W., et al. (2002). *Oncogene, 21*(41), 6255–6263. doi:10.1038/sj.onc.1205794.
12. Michalczyk, A., et al. (2008). *Bioorganic & Medicinal Chemistry, 16*(7), 3482–3488. doi:10.1016/j.bmc.2008.02.053.
13. Yip, Y. L., et al. (2008). *Human Mutation, 29*(3), 361–366. doi:10.1002/humu.20671.
14. Yip, Y. L., et al. (2004). *Human Mutation, 23*(5), 464–470. doi:10.1002/humu.20021.
15. Boeckmann, B., et al. (2003). *Nucleic Acids Research, 31*(1), 365–370. doi:10.1093/nar/gkg095.
16. Berman, H. M., et al. (2000). *Nucleic Acids Research, 28*(1), 235–242. doi:10.1093/nar/28.1.235.
17. Capriotti, E., et al. (2005). *Nucleic Acids Research, 33*, W306–W310. doi:10.1093/nar/gki375.
18. Bava, K. A., et al. (2004). *Nucleic Acids Research, 32*, D120–D121. doi:10.1093/nar/gkh082.
19. Ng, P. C., & Henikoff, S. (2003). *Nucleic Acids Research, 31*(13), 3812–3814. doi:10.1093/nar/gkg509.
20. Ng, P. C., & Henikoff, S. (2001). *Genome Research, 11*(5), 863–874. doi:10.1101/gr.176601.

21. Ramensky, V., et al. (2002). *Nucleic Acids Research, 30*(17), 3894–3900. doi:10.1093/nar/gkf493.
22. Cavallo, A., & Martin, A. C. (2005). *Bioinformatics (Oxford, England), 21*(8), 1443–1450. doi:10.1093/bioinformatics/bti220.
23. Lindahl, E., et al. (2006). *Nucleic Acids Research, 34*, W52–W56. doi:10.1093/nar/gkl082.
24. Delarue, M., & Dumas, P. (2004). *Proceedings of the National Academy of Sciences of the United States of America, 101*(18), 6957–6962. doi:10.1073/pnas.0400301101.
25. Sharma, S., et al. (2006). *Bioinformatics (Oxford, England), 22*(21), 2693–2694. doi:10.1093/bioinformatics/btl460.
26. Han, J. H., et al. (2006). *Structure (London, England), 14*(5), 935–945. doi:10.1016/j.str.2006.01.016.
27. Varfolomeev, S. D., et al. (2002). *Biochemistry. Biokhimiia, 67*(10), 1099–1108. doi:10.1023/A:1020907122341.
28. López, G., et al. (2007). *Nucleic Acids Research, 35*, W573–577.
29. Duhovny, D., et al. (2002). *Lecture Notes in Computer Science 2452* (pp. 185–200). New York: Springer.
30. Connolly, M. L. (1983). *Science, 221*(4612), 709–713. doi:10.1126/science.6879170.
31. Connolly, M. L. (1983). *Journal of Applied Crystallography, 16*, 548–558. doi:10.1107/S0021889883010985.
32. Schneidman-Duhovny, D., et al. (2005). *Nucleic Acids Research, 33*, W363–367. doi:10.1093/nar/gki481.
33. Zhang, C., et al. (1997). *Journal of Molecular Biology, 267*(3), 707–726. doi:10.1006/jmbi.1996.0859.
34. Miyazawa, S., & Jernigan, R. L. (1985). *Macromolecules, 18*(3), 534–552. doi:10.1021/ma00145a039.
35. Yun, C. H., et al. (2007). *Cancer Cell, 11*(3), 217–227. doi:10.1016/j.ccr.2006.12.017.
36. Blencke, S., et al. (2003). *The Journal of Biological Chemistry, 278*(17), 15435–15440. doi:10.1074/jbc.M211158200.
37. Kobayashi, S. (2005). *The New England Journal of Medicine, 352*(8), 786–792. doi:10.1056/NEJMoa044238.
38. Yamasaki, F., et al. (2007). *Cancer Research, 67*(12), 5779–5788. doi:10.1158/0008-5472.CAN-06-3020.
39. Jänne, P. A. (2008). *Lung Cancer (Amsterdam, Netherlands), 60*, S3–S9. doi:10.1016/S0169-5002(08)70099-0.
40. Pao, W., et al. (2005). *PLoS Medicine, 2*(3), e73. doi:10.1371/journal.pmed.0020073.
41. Gow, C. H., et al. (2005). *PLoS Medicine, 2*(9), e269. doi:10.1371/journal.pmed.0020269.