Computers in Biology and Medicine 63 (2015) 293–300

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

# Decoding the EGFR mutation-induced drug resistance in lung cancer treatment by local surface geometric properties

Lichun Ma [a,*], Debby D. Wang [a], Yiqing Huang [b], Maria P. Wong [c], Victor H.F. Lee [c], Hong Yan [a]

[a] Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong, China
[b] School of Computer Science and Technology, Soochow University, Suzhou, China
[c] Li Ka Sing Faculty of Medicne, University of Hong Kong, Pokfulam, Hong Kong, China

ARTICLE INFO

*Article history:*
Received 10 April 2014
Accepted 23 June 2014

*Keywords:*
Drug resistance
EGFR mutation
Lung cancer
Alpha shape
Solid angle
Protein surface geometric properties

ABSTRACT

Epidermal growth factor receptor (EGFR) mutation-induced drug resistance leads to a limited efficacy of tyrosine kinase inhibitors during lung cancer treatments. In this study, we explore the correlations between the local surface geometric properties of EGFR mutants and the progression-free survival (PFS). The geometric properties include local surface changes (four types) of the EGFR mutants compared with the wild-type EGFR, and the convex degrees of these local surfaces. Our analysis results show that the Spearman's rank correlation coefficients between the PFS and three types of local surface properties are all greater than 0.6 with small *P*-values, implying a high significance. Moreover, the number of atoms with solid angles in the ranges of [0.71, 1], [0.61, 1] or [0.5, 1], indicating the convex degree of a local EGFR surface, also shows a strong correlation with the PFS. Overall, these characteristics can be efficiently applied to the prediction of drug resistance in lung cancer treatments, and easily extended to other cancer treatments.



## 1. Introduction

Epidermal growth factor receptor (EGFR) mutations that lead to its overexpression can activate the anti-apoptotic pathways [1,2], eventually result in an aberrant proliferation of cells. The aberrant cell proliferation is a typical cause of human cancers, such as the non-small-cell lung carcinoma (NSCLC) [3–6]. Clinically, gefitinib (*IRESSA*$^{TM}$), a kind of tyrosine kinase inhibitor (TKI), is widely used to interrupt EGFR downstream signals during the treatment of NSCLC patients [7,8]. EGFR mutants such as L858R (substitution of leucine with arginine at residue site 858) show stronger binding affinity with gefitinib than the wild-type (WT) EGFR [9,10]. However, other mutants such as those with an insertion in exon 20 of the tyrosine kinase domain show weak responses to gefitinib [11]. Moreover, for the mutant L858R, the efficacy of gefitinib becomes limited if a second mutation T790M (substitution of threonine with methionine at residue site 790) occurs [12]. It is important to study the characteristics of EGFR mutants in order to understand the mechanism of the mutation-induced drug resistance.

Recently, computational methods have been efficiently applied to the studies of drug resistance [13–16]. Sequence-based and structure-based approaches are the two primary categories of these computational methods. With the rapid development of techniques such as X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, three-dimensional (3D) structural data of proteins become more readily available in structure-related researches [17]. Zhou et al. [11] conducted a structural exploration of EGFR proteins, and predicted anti-EGFR drug resistance based on the binding free energy and hydrogen bond analyses. Wang et al. [18] employed the computational methods in personalized prediction of EGFR mutation-induced drug resistance, using 3D structural data of EGFR proteins. In this work, we investigated the EGFR mutation-induced drug resistance in lung cancer treatment, by analyzing the surface geometric properties of WT EGFR and its mutants. Rosetta [19] was used to generate EGFR mutants and Amber [20] was applied to optimize the structures of obtained mutants. Subsequently, we employed the 3D alpha shape modeling method [21,22] to construct the surfaces of these EGFR structures, after which a solid angle analysis of the atoms at the drug-binding sites of EGFR proteins was conducted to reveal the geometric properties of EGFR surfaces. Finally, we carried out a correlation analysis [23] on these geometric properties and the pre-recorded progression-free survival (PFS)

* Corresponding author. Tel.: +852 34424357.
*E-mail address:* lichunma2-c@my.cityu.edu.hk (L. Ma).

http://dx.doi.org/10.1016/j.compbiomed.2014.06.016
0010-4825/

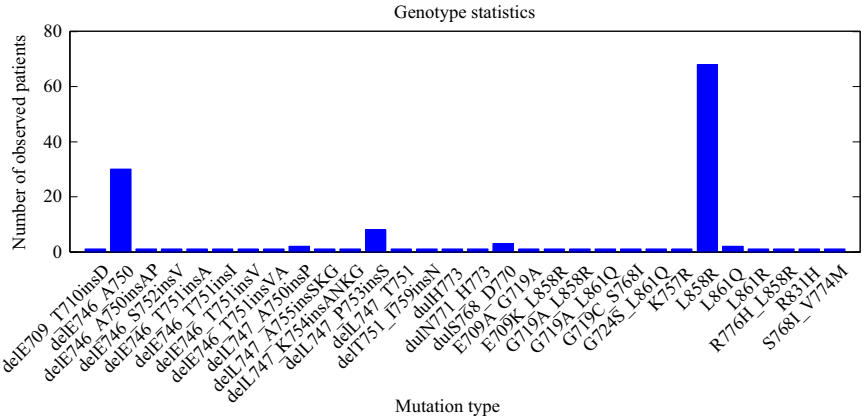


**Fig. 1.** Statistics of the 30 mutation types of the observed 137 NSCLC patients.

(the length of time between using TKI and the disease getting worse [10]) in the treatments. Experiment results show that our proposed surface geometric properties provide valuable information for studying and predicting drug resistance levels in lung cancer treatment.

## 2. Materials and methods

### 2.1. Data collection

The experimental data used in this paper were obtained from the Queen Mary Hospital in Hong Kong [18]. This data set consists of clinical observations on 137 NSCLC patients, with gefinitib applied in their treatments. These patients share a total of 30 EGFR mutation types, and their statistics are shown in Fig. 1. The mutation types are notated by their corresponding changes in protein sequences relative to the WT EGFR [18]. For example, L858R, delE746_A750, dulH773 and delE709_T710insD respectively represent amino acid substitution, deletion, duplication and modification (deletion plus insertion). All the EGFR mutants were generated based on the template structure "2ITY" downloaded from the Protein Data Bank (PDB) [17]. The progression time (in the unit of month) of gefitinib in the treatment of each patient was recorded, revealing the drug resistance levels in each treatment.

### 2.2. Generating EGFR mutants

We employed Rosetta to generate the EGFR mutants. Rosetta is a molecular modeling software package for protein structure prediction and analysis of protein structures [24]. It is a popular software for computational modeling in computational biology and very successful for *de novo* protein structure prediction (protein tertiary structure predicting from the primary sequence) [25]. EGFR mutants were generated based on the crystal structure "2ITY", with the drug molecule removed. For point mutations, the ddg_monomer protocol in Rosetta 3.4 [26] was adopted using inputs of the template structure and the protein sequence of each mutant. For deletions, insertions and duplications, the comparative modeling (CM) protocol [27] in Rosetta was selected. This procedure includes target–template alignment, model construction and model assessment. We applied ClustalW [28], a method for multiple-sequence alignment, to align the target sequence to the template. In the model construction, a fragment library should be prepared first. We employed Psipred [29] to predict secondary structures for each target sequence, and the fragment picker protocol in Rosetta was selected to pick fragments (3- or 9-residue long) for the sequence. Such 3-mers and 9-mers fragment files form a fragment library that can be used in fragment insertion during the structure prediction. Subsequently, we constructed the 3D structures of the mutants using the CM protocol, and a conservative modeling step was adopted to keep the backbone atoms consistent with those of the template in well-aligned regions. The 3D structures of the mutants were assessed by their physics-based energies, and the one with the minimum energy was selected. In this process, we used the full atom energy scoring function to identify accurate structures. The scoring function is a model generated with various scoring terms and the corresponding weights for each term [25]. The scoring terms include van der Waals, Lennard–Jones interactions, residue pair interactions, solvation, rotamer self-energy, Ramachandran torsion preferences, hydrogen bonding and unfolded state reference energy. The total score of a predicted structure is obtained by computing the weighted sum of the scoring terms. Although the predicted mutations generated with software simulation methods cannot always correspond exceptionally well with true structures, based on the observation that similar protein sequences usually lead to similar 3D protein structures, the comparative modeling is very valuable for predicting mutants and it is regarded as the most accurate prediction method currently available [27]. After the 3D structures of all the mutants are obtained, Amber 12 [20] was applied to carry out a minimization for each structure [11]. Each optimized structure was then aligned to the template to construct a mutant–drug complex, using the structure alignment tool of the UCSF Chimera [30].

Using Amber 12, the mutant–drug complex was computationally solvated into an octahedron water box (TIP3P model) with a 10.0 Å buffer around the complex in each direction. The ff99SB force field was adopted in these simulations. After solvating the complex, we conducted 1000 steps of minimization for the entire system to remove bad contacts and find the nearest local minima. This optimized structure was our final structure for analysis.

### 2.3. 3D alpha shape modeling

The alpha shape, first defined by Edelsbrunner and Mücke, is a linear approximation of the original shape [31]. The basic alpha shape and the weighted alpha shape are the two primary types. The basic one is based on Delaunay triangulation while the weighted one is derived from the regular triangulation. Considering a set of points, the basic alpha shape consists of all the simplices in the Delaunay triangulation that have an empty circumscribing sphere with a squared radius equal or smaller than $\alpha$ [21], where 'empty' means that the open sphere includes no points. On the other hand, the definition of a weighted alpha shape is based on a set of weighted points [22]. For example, a weighted