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

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.

© 2014 Elsevier Ltd. All rights reserved.

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).

L. Ma et al. / Computers in Biology and Medicine ▮ (▮▮▮▮) ▮▮▮–▮▮▮
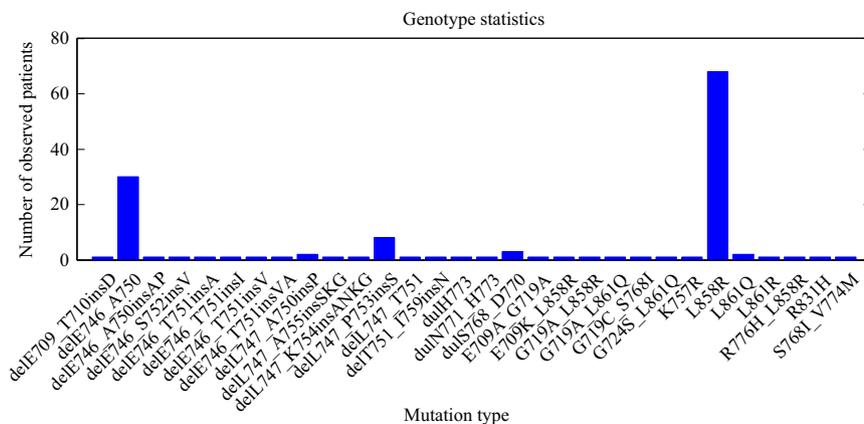
Fig. 1. Statistics of the 30 mutation types of the observed 137 NSCLC patients.

(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 α [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

point p_i can be defined in Eq. (1), where p_i' represents the location and p_i'' is the weight. Two points p_1 and p_2 are defined to be orthogonal if Eq. (2) is satisfied, and sub-orthogonal if Eq. (3) is satisfied. For a given α, the weighted alpha complex is formed by the simplices that meet the condition, under which there exists a weighted point with weight α orthogonal to the weighted points associated with the vertices of the simplex and sub-orthogonal to all other weighted points:

$$p_i = (p_i', p_i'') \tag{1}$$

$$|p_1'p_2'| = p_1'' + p_2'' \tag{2}$$

$$|p_1'p_2'| > p_1'' + p_2''. \tag{3}$$

In this paper, weighted alpha shape modeling was applied to construct the surface of each EGFR mutant structure. We adopted the Computational Geometry Algorithms Library (CGAL) to compute the weighted alpha shape and set the squared common Van der Waals (VDW) radii as the atom weight. The alpha value and probe weight were set to be 0 and squared 1.4 Å respectively, in order to guarantee that the vertices in the weighted alpha shape correspond to the solvent-accessible atoms.

2.4. Analysis based on solid angles

We used solid angles to characterize the geometric properties of the EGFR surface atoms, which were extracted using the alpha shape model. The solid angles of surface atoms can represent the surface curvature, which is used as a geometric property in this paper. Concave shapes around the binding sites are more likely to offer opportunities for drug binding than convex shapes. If the drug binds to EGFR mutants very tightly, then it will strengthen the response. So we studied the EGFR mutation-induced drug resistance by analyzing the solid angles around the binding sites. The definition of solid angles [32] is as follows: assume **P**, **A**, **B** and **C** are the vertices of a tetrahedron with an origin at **P** subtended by the triangle ABC, and the dihedral angle between PAC and PBC, PAB and PAC, PAB and PBC, are denoted by φ_{ab}, φ_{bc}, φ_{ac} respectively then the solid angle of **P** can be expressed in Eq. (4). Further, the solid angle Ω of a surface atom **M** can be obtained by summing up all the solid angles with the origin at
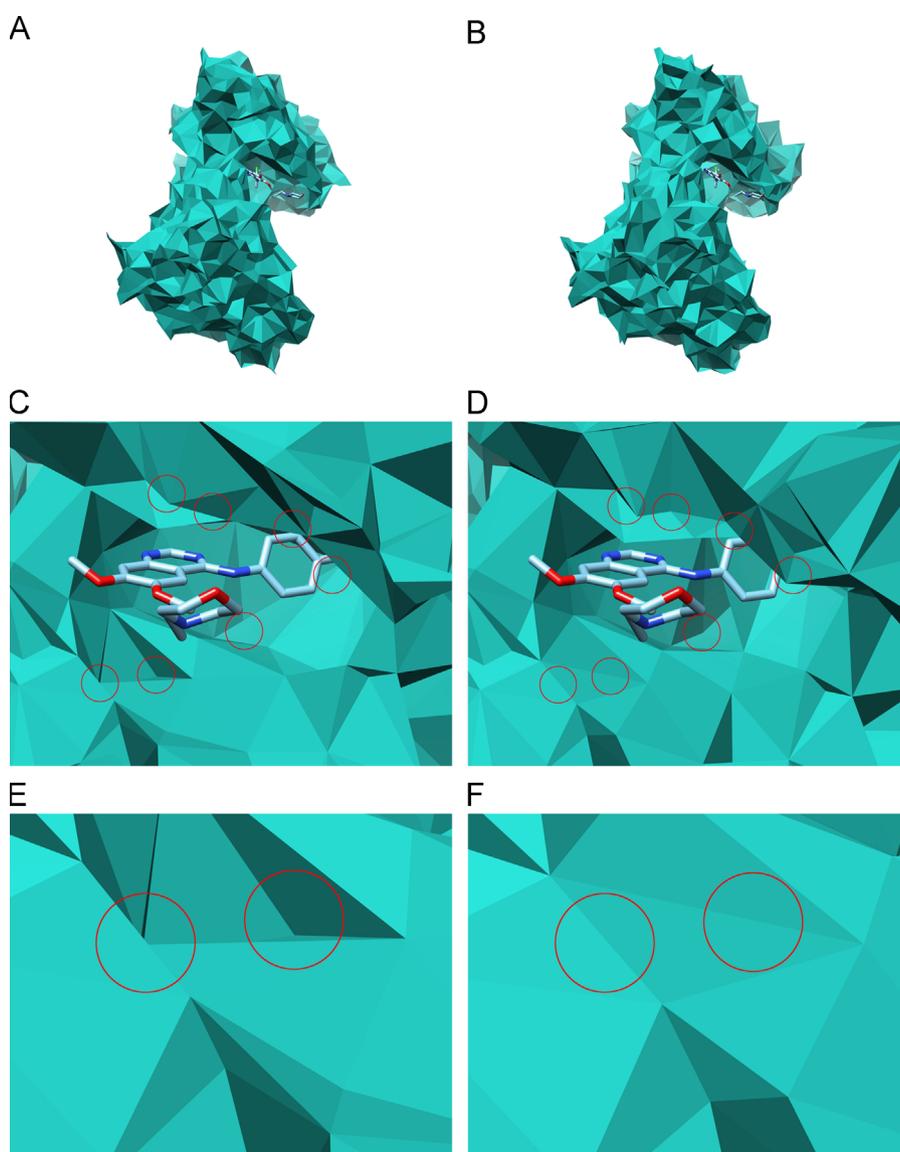
Fig. 2. The 3D alpha shapes. (A) WT EGFR. (B) L858R mutant. (C) Alpha shape surface of the binding sites of WT EGFR. (D) Alpha shape surface of the binding sites of L858R mutant. The major changes around the binding sites compared with WT EGFR are marked with circles. (E) Zooming the two major changes at the left corner of (C). (F) Zooming the two major changes at the left corner of (D).

M in the alpha shape (Eq. (5)). We used Eq. (6) to transform these solid angle values to the range of $[-1, 1]$:

$$\Omega_i = \varphi_{ab} + \varphi_{bc} + \varphi_{ac} - \pi \qquad (4)$$

$$\Omega = \sum_i \Omega_i \qquad (5)$$

$$\Omega' = \cos(\Omega/4). \qquad (6)$$

When the scaled solid angle Ω' falls into $[-1, 0]$ or $(0, 1]$, the corresponding shape centered at **M** is a concave or convex one respectively. Overall, a specific value of Ω' represents the concave or convex degree of an atom.
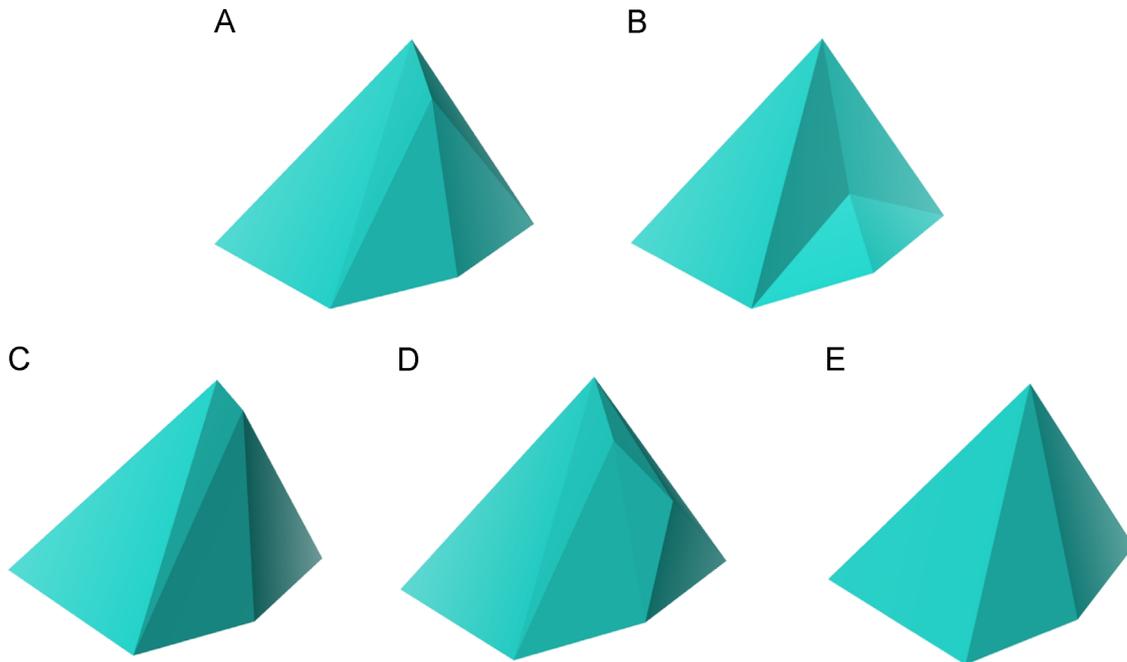
Fig. 3. Different types of local surface changes. (A) The original surface (WT protein), (B) reverse of one atom, (C) convex degree variation, (D) atom emergence and (E) atom disappearance.
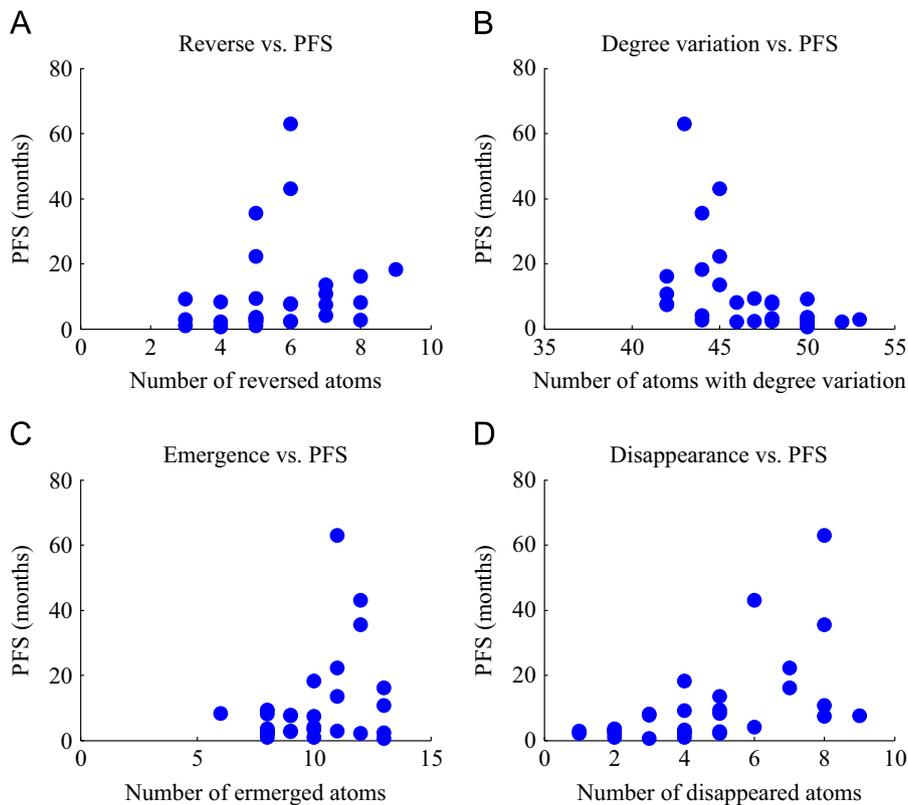
Fig. 4. The relation between the PFS and the four types of local surface changes or the number of (A) reversed, (B) degree-varied, (C) emerged, and (D) disappeared atoms.

2.5. Correlation analysis

Spearman's rank correlation coefficient is a nonparametric measure of statistical dependence between two variables [23]. Consider two variables X and Y of size n, they are firstly converted into ranks x_i and y_i ($i = 1, …, n$), after which the spearman's rank correlation coefficient ρ can be computed as follows:

$$\rho = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_i (x_i - \overline{x})^2 \sum_i (y_i - \overline{y})^2}} \tag{7}$$

ρ is in the range of $[-1, 1]$, which indicates the dependence between the two variables. $\rho > 0$ means that X and Y are positively dependent, while $\rho < 0$ shows a negative dependence between them. The larger the absolute value of ρ is, the stronger the two variables are correlated. A perfect Spearman correlation of $+1$ or -1 can occur when each of the variables is a perfect monotone function of the other, under the condition that there are no repeated data values in the dataset. Specifically in our studies, the correlation analysis was applied to the EGFR surface geometric properties and the pre-recorded PFS, aiming to extract valuable dependence between them. Moreover, we employed P-value to test whether ρ was significantly different from zero. The P-value describes the probability that the statistical result is equal or more extreme than the actually observed one, with the assumption of the null hypothesis being true [33]. It was computed using the permutation test, and the null hypothesis was rejected if the P-value was less than 0.05.

2.6. Hierarchical clustering method

The hierarchical clustering method is widely used in pattern classification [34]. There are two approaches, agglomerative clustering and divisive clustering. The agglomerative clustering

Table 1
The Spearman's rank correlation coefficients of four type local surface changes and the PFS.

Types of local surface changes	Reverse			Degree variation	Emergence	Disappearance
	Concave converting to convex	Convex converting to concave	Total			
Correlation coefficients	0.003	**0.61**	0.41	**−0.63**	0.24	**0.63**
P-value	0.99	**3.30E−4**	0.03	**1.85E−4**	0.21	**2.17E−4**

Table 2
The Spearman's rank correlation coefficients of the number of atoms with different convex degree and the PFS.

Solid angle value range	[0.79,1]	[0.71,1]	[0.61,1]	[0.5,1]	[0.38,1]	[0,1]
Correlation coefficients	−0.22	**−0.61**	**−0.61**	**−0.61**	−0.21	−0.40
P-value	0.24	**3.11E−4**	**3.28E−4**	**3.36E−4**	0.25	0.03
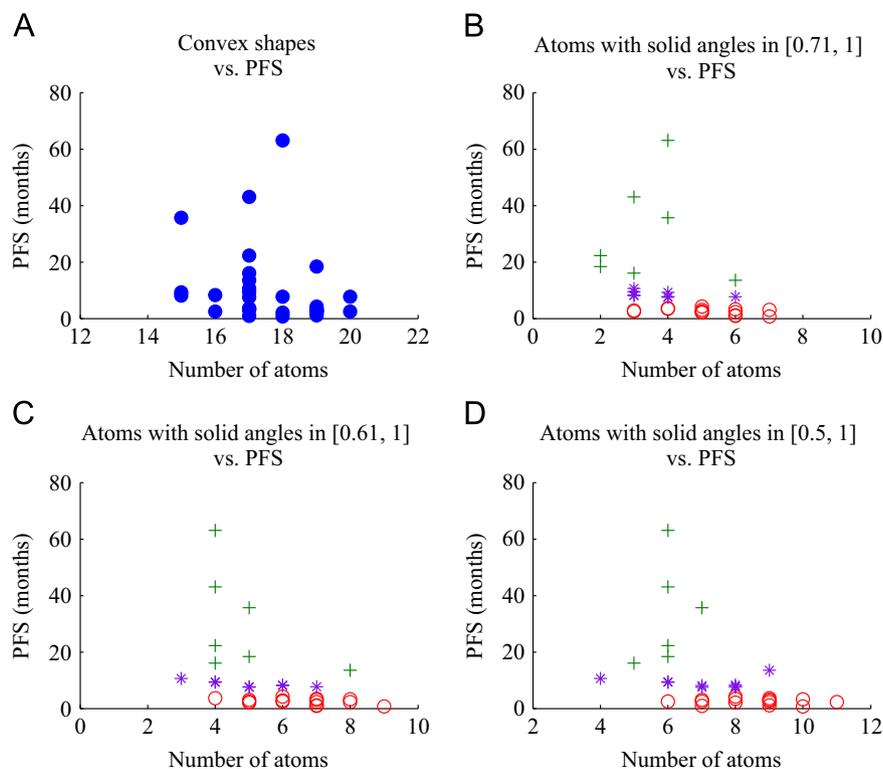
Fig. 5. The relation between the number of convex shapes and the PFS. (A) The total number of convex shapes versus the PFS. (B)–(D) The number of atoms with solid angles in [0.71, 1], [0.61, 1], [0.5, 1] versus the PFS. Two major clusters are obtained by using the hierarchical clustering method. The mutants involved in clusters 1 and 2 are marked with "∘" and "∗" respectively, while other mutants are marked with "+".

method builds a hierarchy of clusters by merging the clusters at the bottom to move up the hierarchy, while the divisive clustering approach splits the top cluster to move down the hierarchy. In this paper, the agglomerative clustering approach was employed to analyze the relationship of surface geometric properties and the PFS. We applied Euclidean distance to measure the pairwise distance between pairs of points, and the points with the shortest distance were merged into the same class. Then the distances between the new merged class and the old classes were calculated to obtain new clusters until all the points were merged into one cluster.

3. Results and discussion

In order to represent the surface geometric changes of the EGFR mutants compared to the WT EGFR, we computed the 3D alpha shapes for the WT and those of the mutants (Fig. 2). Since the binding site of the EGFR was our main target, solid angles of the binding-site atoms in the alpha shapes of WT and mutants were calculated. By comparing the obtained solid angles, we captured the local surface changes of the mutants compared to the WT protein.

According to the definition, a concave shape has a solid angle in the range of $[-1, 0]$, and a convex shape in $(0, 1]$. By comparing the solid angle values of the WT and the mutants, we classify the local surface changes into four types (Fig. 3B–E) for further analysis. These four types are Reverse, Degree Variation, Emergence and Disappearance. For an atom with a convex shape in WT, Reverse means that this atom has a concave shape in the mutant; Degree Variation denotes that the shape is still convex, but the solid angle has changed; Emergence indicates that other atoms appear on the surface and their shapes can be concave or convex; and Disappearance means this atom in the mutant is no longer on the alpha shape. For a concave shape, the four types of local surface changes are similarly defined.

There are 14 amino acid residues of 102 atoms at the drug-binding site of the WT EGFR. We calculated the number of atoms having the aforementioned four types of changes for the 30 mutation types, compared to WT. According to the recorded PFS of the 137 observed patients from the Queen Mary Hospital in Hong Kong, we adopted the median of the PFS for patients sharing
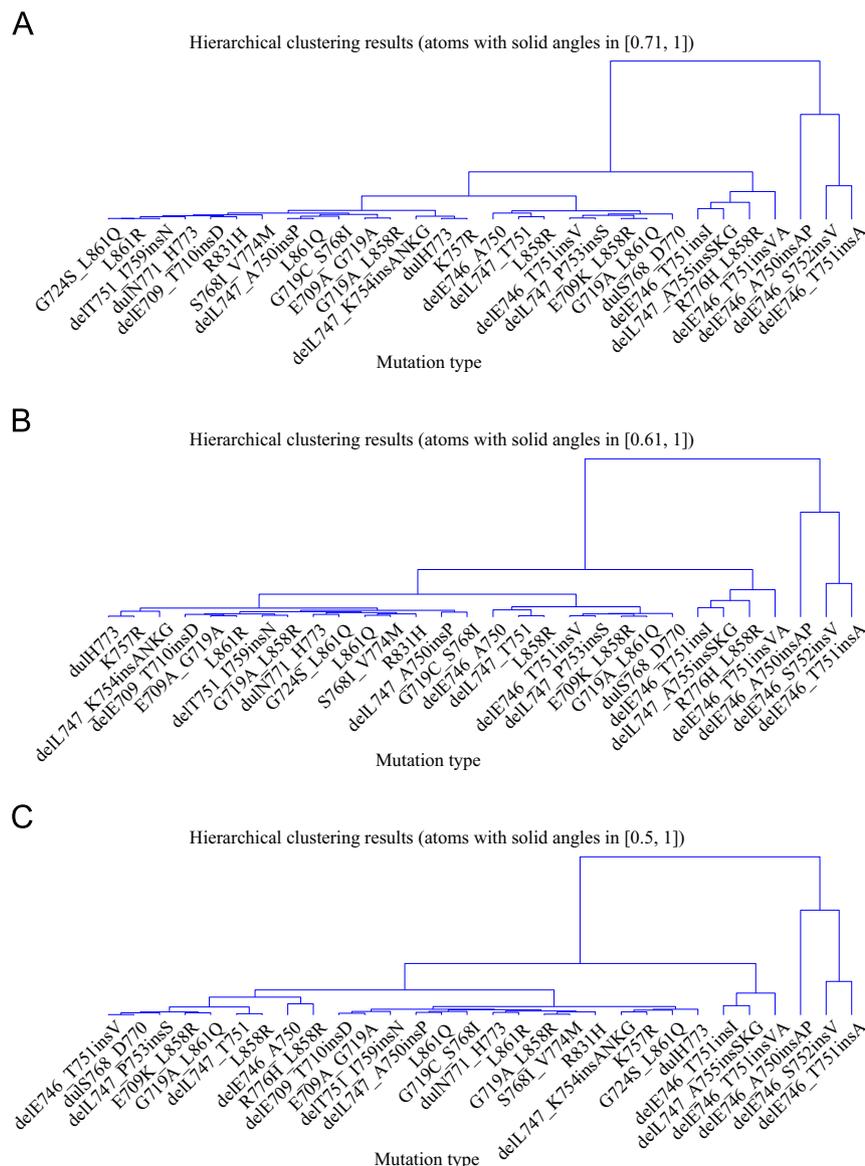
Fig. 6. Hierarchical clustering of the mutants with PFS and the number of atoms with solid angles in (A) [0.71, 1], (B) [0.61, 1] and (C) [0.5, 1].

the same EGFR mutation type. This median value was recorded as the PFS for the corresponding mutant. Then we analyzed the correlations of the four types of local surface changes and the PFS for all mutation types. The relevance of the four types of local surface changes and PFS is shown in Fig. 4. Fig. 4A and C shows a weak correlation between the number of surface atoms with curvature changes and the PFS. However, Fig. 4B and D shows a clear correlation. In Fig. 4B, as the number of atoms with degree variations increases, the PFS decreases in general. In an overall view of Fig. 4D, the PFS increases when the number of disappeared atoms increases.

Subsequently, we calculated the Spearman's rank correlation coefficients for the four types of local surface changes and the PFS (Table 1). The absolute values of the coefficients for degree variation and disappearance are greater than 0.6, and the P-values are much smaller than 0.05. Thus, the coefficients are significantly different from zero and the null hypothesis is rejected. The disappearance type has a positive correlation with the PFS, which indicates that the drug has a longer progression time with the number of disappeared atoms increasing. However, the degree variation type shows a negative correlation with the PFS, meaning that the PFS will decrease when more atoms stay convex or concave. The reason is that the WT protein has a weaker binding affinity with the drug molecule compared with many mutants, and more changes from the WT may result in a stronger binding affinity with the drug. For the types of emergence and reverse, they show a weak correlation with the PFS. However, the number of reverse atoms, in the case where convex shapes are changed to concave ones, has a strong correlation with the PFS. This means that if there are more convex shapes becoming concave ones, the drug molecule can have a stronger binding affinity with the EGFR mutant, and it will have a longer progression time.

From the correlations between the four types of local surface changes and the PFS, we can predict the EGFR mutation-induced drug resistance. If the EGFR mutants have more disappeared atoms, more atoms with convex shapes converting to concave ones, and fewer atoms with degree variations, they most probably have strong binding affinities with the inhibitor gefitinib.

On the other hand, the concave or convex degree of the drug-binding site of an EGFR structure can be an important factor for the binding with a drug molecule, thus we conducted an additional exploration on the correlation between the convex degree of the binding site and the PFS. Specifically, the convex degree of the binding site was represented by the total number of convex shapes or the number of atoms with different solid angle values in $[t, 1]$, where t is a threshold. A larger t value means that the surface atoms with solid angle values in $[t, 1]$ are at a higher convex degree, while a smaller t indicates a lower convex degree. Table 2 shows the scenarios for different t values. The correlation coefficients are all negative, indicating that the drug tends to have a shorter progression time when the number of convex shapes increases. For the cases concerning solid angle values in $[0.71, 1]$ ($[0°, 180°]$), the solid angle values of the corresponding angles are calculated with Eq. (6), $[0.61, 1]$ ($[0°, 210°]$) and $[0.5, 1]$ ($[0°, 240°]$), the absolute values of the correlation coefficients are more than 0.6 and the P-values are much smaller than 0.05. This implies strong evidence to believe the highly correlation between convex shapes and PFS. Therefore, the null hypothesis can be rejected. The scatter plot of the total number of convex shapes versus the PFS is shown in Fig. 5A, while the plot of the number of atoms with solid angle values in $[0.71, 1]$, $[0.61, 1]$, $[0.5, 1]$ versus the PFS is shown in Fig. 5B–D respectively.

Moreover, the hierarchical clustering method was employed to analyze the relationship between the PFS and the number of atoms with solid angle values in $[0.71, 1]$, $[0.61, 1]$ and $[0.5, 1]$ (Fig. 6). Two major clusters are obtained from Fig. 6A and B, cluster 1

(G724S_L861Q, …, K757R) and cluster 2 (delE746_A750, …, dulS768_D770). The clusters in Fig. 6C are different, with the mutant R776H_L858R in cluster 2. From Fig. 5B, cluster 1 mainly contains the mutants which have more atoms with solid angle values in $[0.71, 1]$ and smaller PFS. On the contrary, cluster 2 includes the mutants with fewer atoms in $[0.71, 1]$ and larger PFS. In addition, similar clusters are found in Fig. 5C and D. Thus, the hierarchical clustering results are consistent with those from the Spearman's correlation coefficients.

According to the hierarchical clustering results and the correlation between the number of convex shapes and the PFS, we can predict whether drug resistance occurs for the mutants. If the EGFR mutants have more convex shapes with solid angle values in $[0.71, 1]$, $[0.61, 1]$ or $[0.5, 1]$, they are more likely to be resistant to the inhibitor gefitinib. However, if they have less convex ones in these ranges, the drug molecule may have a long progression time.

4. Conclusions

In this paper, we developed a method to predict the EGFR mutation-induced drug resistance in lung cancer treatment, by analyzing the local surface geometric properties of EGFR mutants. Rosetta was employed to construct the 3D structures of these EGFR mutants. Then 3D alpha shape modeling and a solid-angle analysis were implemented to reveal the local surface geometric properties of these EGFR proteins. We classified the local surface changes of these mutants compared to the WT EGFR into four types, and the Spearman's rank correlation coefficients were applied to calculate the correlations between these local surface changes and the PFS. Our analysis results show that the number of atoms with the changes of convex shapes converting to concave, degree variation and disappearance have strong correlations with the PFS. Moreover, the number of binding-site atoms with solid angles in $[0.71, 1]$, $[0.61, 1]$ or $[0.5, 1]$ show a strong correlation with the PFS as well. By using the hierarchical clustering method, we also analyzed the relationship of PFS and the number of atoms with solid angle values in $[0.71, 1]$, $[0.61, 1]$ and $[0.5, 1]$. The clustering results are consistent with those from the Spearman's correlation coefficients. These characteristics can be applied to the prediction of EGFR mutation-induced drug resistance in lung cancer treatment. Our study indicates that surface geometric properties of the binding site on a key protein such as EGFR play an important role in drug resistance prediction, and this method can also be extended to the investigation of other drug resistance mechanisms for other diseases.

Conflict of interest statement

The authors declare no conflict of interests.

References

[1] R. Sordella, D.W. Bell, D.A. Haber, J. Settleman, Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways, Science 305 (5687) (2004) 1163–1167.

[2] W. Pao, V. Miller, M. Zakowski, J. Doherty, K. Politi, I. Sarkaria, et al., 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 (36) (2004) 13306–13311.

[3] T.J. Lynch, D.W. Bell, R. Sordella, S. Gurubhagavatula, R.A. Okimoto, B.W. Brannigan, et al., Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib, N. Engl. J. Med. 350 (21) (2004) 2129–2139.

[4] D.A. Eberhard, B.E. Johnson, L.C. Amler, A.D. Goddard, S.L. Heldens, R.S. Herbst, et al., Mutations in the epidermal growth factor receptor and in KRAS are predictive and prognostic indicators in patients with non-small-cell lung cancer treated with chemotherapy alone and in combination with erlotinib, J. Clin. Oncol. 23 (25) (2005) 5900–5909.

[5] T. Mitsudomi, T. Kosaka, H. Endoh, Y. Horio, T. Hida, S. Mori, et al., Mutations of the epidermal growth factor receptor gene predict prolonged survival after gefitinib treatment in patients with non-small-cell lung cancer with post-operative recurrence, J. Clin. Oncol. 23 (11) (2005) 2513–2520.

[6] R.S. Herbst, A.M. Maddox, M.L. Rothenberg, E.J. Small, E.H. Rubin, J. Baselga, et al., Selective oral epidermal growth factor receptor tyrosine kinase inhibitor ZD1839 is generally well-tolerated and has activity in non-small-cell lung cancer and other solid tumors: results of a phase i trial, J. Clin. Oncol. 20 (18) (2002) 3815–3825.

[7] M.G. Kris, R.B. Natale, R.S. Herbst, T.J. Lynch Jr, D. Prager, C.P. Belani, et al., Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer, J. Am. Med. Assoc. 290 (16) (2003) 2149–2158.

[8] S. Kobayashi, T.J. Boggon, T. Dayaram, P.A. Jänne, O. Kocher, M. Meyerson, et al., EGFR mutation and resistance of non-small-cell lung cancer to gefitinib, N. Engl. J. Med. 352 (8) (2005) 786–792.

[9] C.H. Yun, T.J. Boggon, Y. Li, M.S. Woo, H. Greulich, M. Meyerson, et al., Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity, Cancer cell 11 (3) (2007) 217–227.

[10] V.H. Lee, V.P. Tin, T.S. Choy, K.O. Lam, C.W. Choi, L.P. Chung, et al., Association of Exon 19 and 21 EGFR mutation patterns with treatment outcome after first-line tyrosine kinase inhibitor in metastatic non-small-cell lung cancer, J. Thorac. Oncol. 8 (9) (2013) 1148–1155.

[11] W. Zhou, D.D. Wang, H. Yan, M. Wong, V. Lee, Prediction of anti-EGFR drug resistance base on binding free energy and hydrogen bond analysis, Comput. Intell. Bioinform. Comput. Biol. (2013) 193–197.

[12] W. Pao, V.A. Miller, K.A. Politi, G.J. Riely, R. Somwar, M.F. Zakowski, et al., Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain, Plos Med. 2 (3) (2005) e73.

[13] Z.W. Cao, L.Y. Han, C.J. Zheng, Z.L. Ji, X. Chen, H.H. Lin, Y.Z. Chen, Computer prediction of drug resistance mutations in proteins, Drug Discov. Today 10 (7) (2005) 521–529.

[14] G.F. Hao, G.F. Yang, C.G. Zhan, Structure-based methods for predicting target mutation-induced drug resistance and rational drug design to overcome the problem, Drug Discov. Today 17 (19) (2012) 1121–1126.

[15] T. Hou, W. Zhang, J. Wang, W. Wang, Predicting drug resistance of the HIV-1 protease using molecular interaction energy components, Proteins: Struct. Funct. Bioinform. 74 (4) (2009) 837–846.

[16] N. Beerenwinkel, T. Sing, T. Lengauer, J. Rahnenführer, K. Roomp, I. Savenkov, et al., Computational methods for the design of effective therapies against drug resistant HIV strains, Bioinformatics 21 (21) (2005) 3943–3950.

[17] The Protein Data Bank, ⟨http://www.rcsb.org⟩.

[18] D.D. Wang, W. Zhou, H. Yan, M. Wong, V. Lee, Personalized prediction of EGFR mutation-induced drug resistance in lung cancer, Sci. Rep. 3 (2013) 2855.

[19] A. Leaver-Fay, M. Tyka, S.M. Lewis, O.F. Lange, J. Thompso, R. Jacak, et al., ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules, Methods Enzymol. 487 (2011) 545–574.

[20] D.A. Case, et al., AMBER 12, University of California, San Francisco, 2012.

[21] H. Edelsbrunner, E.P. Mücke, Three-dimensional alpha shapes, ACM Trans. Graphics 13 (1) (1994) 43–72.

[22] H. Edelsbrunner, Weighted Alpha Shapes, Department of Computer Science, University of Illinois at Urbana-Champaign, 1992.

[23] C. Spearman, The proof and measurement of association between two things, Am. J. Psychol. 15 (1) (1904) 72–101.

[24] K.W. Kaufmann, G.H. Lemmon, S.L. DeLuca, J.H. Sheehan, J. Meiler, Practically useful: what the Rosetta protein modeling suite can do for you, Biochemistry 49 (14) (2010) 2987–2998.

[25] C.A. Rohl, C.E. Strauss, K.M. Misura, D. Baker, Protein structure prediction using Rosetta, Methods Enzymol. 383 (2004) 66–93.

[26] E.H. Kellogg, A. Leaver-Fay, D. Baker, Role of conformational sampling in computing mutation-induced changes in protein structure and stability, Proteins: Struct. Funct. Bioinform. 79 (3) (2011) 830–838.

[27] M.A. Martí-Renom, A.C. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Šali, Comparative protein structure modeling of genes and genomes, Annu. Rev. Biophys. Biomol. Struct. 29 (1) (2000) 291–325.

[28] J.D. Thompson, T. Gibson, D.G. Higgins, Multiple sequence alignment using ClustalW and ClustalX, Curr. Protoc. Bioinform. (2002) 2–3.

[29] L.J. McGuffin, K. Bryson, D.T. Jones, The PSIPRED protein structure prediction server, Bioinformatics 16 (4) (2000) 404–405.

[30] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF chimera – a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (13) (2004) 1605–1612.

[31] F. Bernardini, C.L. Bajaj, Sampling and Reconstructing Manifolds Using Alpha-Shapes, 1997.

[32] W. Zhou, H. Yan, Q. Hao, Analysis of surface structures of hydrogen bonding in protein–ligand interactions using the alpha shape model, Chem. Phys. Lett. 545 (2012) 125–131.

[33] S.N. Goodman, Toward evidence-based medical statistics. I. The *P* value fallacy, Ann. Intern. Med. 130 (12) (1999) 995–1004.

[34] P. Langfelder, B. Zhang, S. Horvath, Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R, Bioinformatics 24 (5) (2008) 719–720.