IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 21, 2020, accepted December 28, 2020, date of publication December 31, 2020,
date of current version January 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3048427

Geometrical Features of Epidermal Growth Factor Receptor-Related Dimers Reveal the Mechanisms of Drug Resistance in Lung Cancer Patients

MENGXU ZHU, RIZWAN QURESHI, (Graduate Student Member, IEEE), AND HONG YAN, (Fellow, IEEE)
Department of Electrical Engineering, City University of Hong Kong, Hong Kong
Corresponding author: Mengxu Zhu (mengxuzhu2-c@my.cityu.edu.hk)

This work was supported in part by the Hong Kong Research Grants Council under Project 11200818, and in part by the Hong Kong Innovation and Technology Commission (ITC).

ABSTRACT Epidermal growth factor receptor (EGFR) plays an important role in lung cell proliferation. Dimerization of EGFR family members and other receptor tyrosine kinases (RTKs) act as a vital controller for lung cell life cycle signals. Mutations in the kinase domain of EGFR may disorder the signaling networks and lead to cancer. Drug resistance occurs in several generations of EGFR drugs due to genetic mutations, and there is a very less understanding about the mechanism of EGFR-mutated drug resistance. In this work, we investigate the mechanism of wild type EGFR protein and its drug-sensitive and drug-resistant mutations. We performed molecular dynamics (MD) simulation for the 10-ns EGFR-drug mutant complex and investigated the structures' geometrical properties. With features extracted by alpha shape modeling, different geometrical properties, such as matching rates of atom solid angles at interaction sites and centroid distances between interfacial atoms, were calculated to characterize the binding intensity. Wilcoxon rank sum test was applied to reveal the differences between mutations based on extracted properties. We have developed a framework that explains the drug resistance mechanism based on geometrical properties and binding free energy. Results revealed that drug-sensitive mutants have tighter interactions with corresponding RTK in the complex for all protein-drug systems, while drug-resistant mutants are bound looser. The extracted geometrical properties of the drug mutant complex help understand the drug response mechanism at atomic level.

INDEX TERMS Epidermal growth factor receptor (EGFR), drug resistance, molecular dynamics (MD) simulations, alpha shape analysis, geometric properties.

I. INTRODUCTION

A. BACKGROUND

Lung cancer is the most deadly type of cancer, having the lowest survival rate among all types [1], [2]. Non-small cell lung carcinoma (NSCLC) is the primary type of lung cancer that covers about 85% of all the lung cancer cases, and therefore it is an active research problem [3]. Epidermal growth factor receptor (EGFR), a member of the ErbB family, is a vital controller on signal pathways of cell proliferation [4], [5]. As a transmembrane protein, the extracellular domain of EGFR can be activated by binding to a ligand called a growth factor. A homodimer or a heterodimer of EGFR will then be formed, leading to specific residues' phosphorylation on intra-cellular tyrosine kinase domain and switch on the downstream signal pathways. Thus, dimerization proves to be one of the essential parts of signaling [6], [7].

Researches have confirmed that EGFR plays a vital role in lung cancer pathology, and overexpression of EGFR can be found in about 40% 80% of NSCLC patients [8], [9]. Mutations on the EGFR gene are the root cause of the signaling disturbance and will lead to abnormal cell proliferation [10].

Tyrosine kinase inhibitors (TKIs) like Gefitinib or Erlotinib are designed to target EGFR and stop the over-activated signal pathways. The drug molecules can occupy ATP's binding site, blocking the downstream signaling and reducing the abnormal proliferation. These drugs proved to be effective in therapies and are accepted as

The associate editor coordinating the review of this manuscript and approving it for publication was G. R. Sinha.

first-line treatment for NSCLC patients suffering from EGFR mutations [11]. However, in clinical therapies, drug resistance usually appears in about a year [12], [13]. Several studies have been conducted in the past decade to understand the drug resistance mechanism due to a secondary point mutation T790M, which accounts for about 50% of all the resistant cases [14]–[16]. These mutations decreased the binding affinity in the EGFR-TKI system and weaken the effectiveness of TKIs

B. LITERATURE REVIEW

Previous research have been conducted on molecular dynamics simulation (MD) to understand the drug resistance process in lung cancer patients [17]–[19]. Sanders *et al.* simulated the binding of signal ligands to the extracellular domain of EGFR, studying the affinity of binding with free energy [20]. Qureshi *et al.* investigated EGFR mutants' domains using parametric models and complex networks [21], [22]. Principal component analysis can also be used to investigate the structural dynamics [23] and visualization of protein-drug interactions [24]. Dixit *et al.* applied multiscale simulations to study long-range communications in the EGFR kinase domain [25]. Ghosh *et al.* analyzed the hydrogen bonds of the EGFR heterodimer [15].

C. SELECTION OF THE MODELS AND RESEARCH PROCEDURES

This study analyzed the interactions between EGFR mutations and their RTK partners in dimer-drug complexes with computational methods. It is known that apart from ErbB family members, RTKs like Insulin-like growth factor 1 receptor (IGF-1R) [26], [27] can also dimerize with EGFR or its mutants and join the cell signaling network. Thus, we took both EGFR-EGFR homodimers and IGF-1R-EGFR heterodimers for MD simulations. Gefitinib (Iressa, ZD1839) is a first-generation TKI to treat NSCLC patients [28] and was selected to set up the dimer-drug complex. Exon 21 L858R, exon 19 deletion (short as del 19) [29], [30] and G719S [31] mutations are more activate and sensitive to Gefitinib, while secondary mutations T854A, L858R-T790M (short as T790M) [29], [30] and L858R-T790M-C797S (short as C797S) [32], [33] are drug-resistant. It is also shown that the αC-β4 loop is an important part of protein kinases and has a significant influence on protein structures and functions [34]. Insertion mutations on αC-β4 loop of EGFR are found to have a relationship with drug resistance of first generation drugs including Gefitinib. The mutations also influence interfaces of protein-protein interactions and are related with asymmetric dimerization of EGFR, which is the main focus of our work [34], [35]. The αC-β4 loop exists in exon 20 of the protein kinase, and mutations on exon 20 may have a relationship with drug resistance, such as T790M and C797S. Here, we choose the D770_N771insNPG, which is an exon 20 insertion (short as Ins 20 in this work) and is found in αC-β4 loop of EGFR [35]. The crystal structure can be obtained from PDB bank(PDB:4LRM) [36]. Wild type (WT) EGFR is also taken as a reference. The WT EGFR is not sensitive to the drug, and former research showed that it behaves similarly to drug-resistant mutants [37], [38]. These mutations were dimerized with RTK partners, respectively, for comparing.

To simulate the reactions in dimer-drug complexes, MD simulations were implemented for all the mutations in the same environment. Then we used three-dimensional Alpha Shape modeling [39], [40] to extract the geometrical information from all the MD trajectories. For atoms at interaction sites, matching rates of their solid angles and centroid distances between the interaction atoms on different chains are calculated to describe the binding affinities. Wilcoxon rank sum test [41], [42] is then applied to check the difference of extracted properties between two kinds of mutations. After that, as a mature method, binding free energy was also used to compare geometrical analysis. These methods will reveal the influence of EGFR mutations on dimerization in the dimer-drug complex and provide a better understanding of lung cancer drug resistance mechanisms. The paper is organized as follows: In Section 2, protein system preparation and analysis methods are described. Section 3 presents result and discussion, and finally Section 4 concludes the paper with future prospects. The whole research process is shown in Figure 1.

II. METHODS

A. PREPARING MODELS FOR EGFR MUTANTS AND DIMER-DRUG COMPLEXES

EGFR is a multi-domain protein containing a large extracellular domain, transmembrane domain, and kinase domain. The dimerization of the tyrosine kinase (TK) domain is related to our topic. A crystal structure of wild type EGFR kinase domain homodimer (PDB:2GS2, resolution 2.80 Å) [43] is selected as a structural template, and structure of IGF-1R is extracted from PDB:5FXQ (resolution 2.30 Å) [44]. We applied Rosetta to generate all the mutations of EGFR based on one monomer from the EGFR dimer [45]. Mutations include L858R, exon 19 deletion, G719S, T854A, L858R-T790M, L858R-T790M-C797S and exon 20 insertion, D770_N771insNPG in αC-β4 loop. We first filled the template model gaps to obtain a whole chain structure for the alpha shape algorithm, and all necessary fragments including αC-β4 loop is contained in the model. Rosetta provides a fragment library, Robetta [46], to decompose the protein into 9mer and 3mer fragments. Rosetta comparative modeling (CM) protocol generates different possible structures based on fragment files and rank the energy of the output structures. Extractpdb protocol in Rosetta is designed to select the most stable structure, namely the structure with the lowest energy. Exon19 deletion mutant was generated in the same way. Other mutants of EGFR are mostly point mutations. For example, L858R means a replacement from Leucine (L) to Arginine (R) at the 858th position on the chain of residues. The mutated structures can be derived

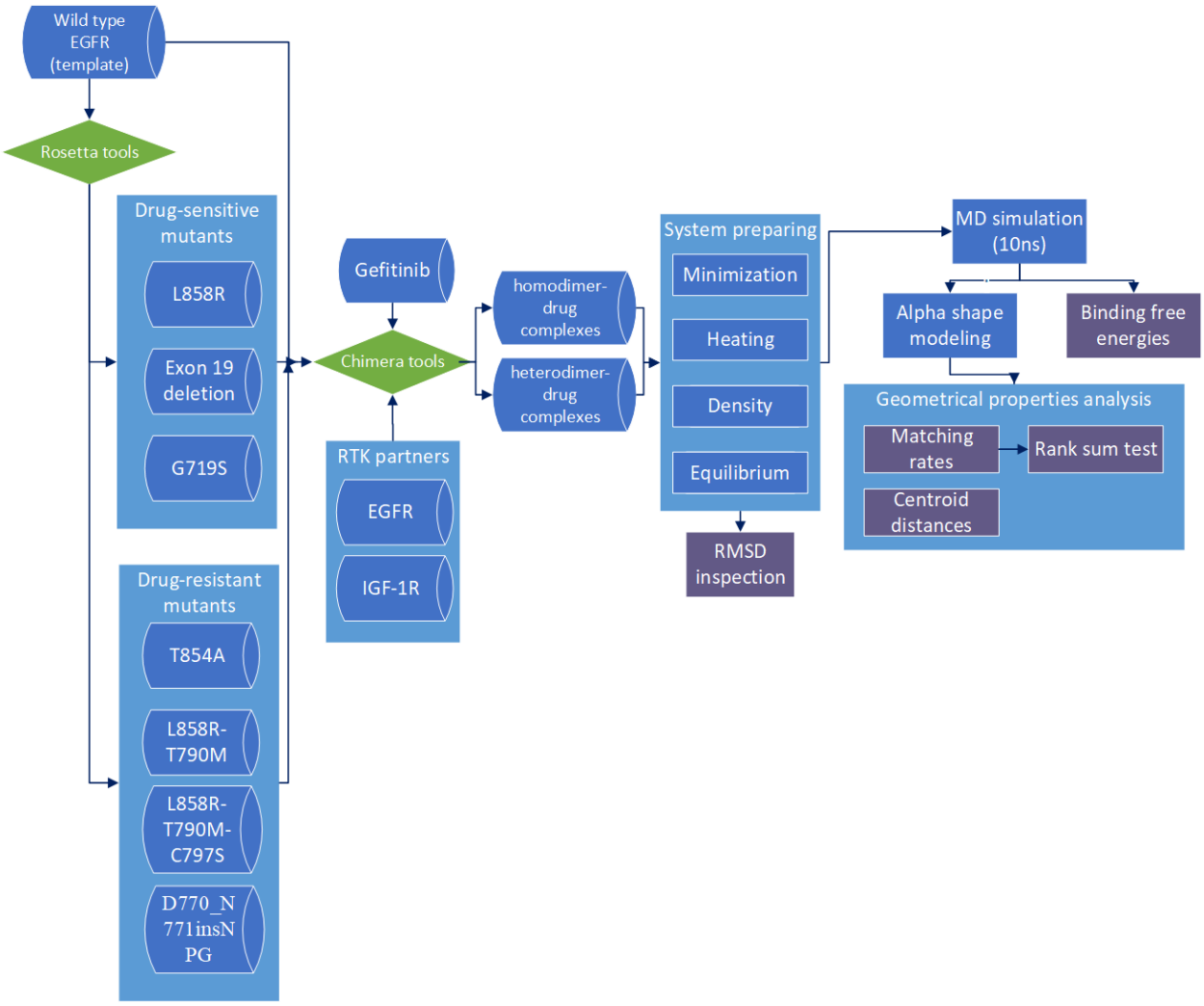

FIGURE 1. Flow diagram that shows the whole research process. We extract the templates from PDB and perform modeling using Chimera. We prepare EGFR homodimers and heterodimers with Gefitinib drug-molecule. Drug-sensitive and drug resistive types characterize the systems. After that, we perform MD simulation for 10-ns and use alpha shape modeling and binding free energies to evaluate the geometrical properties. Simulation results reveal interesting insights that explain the drug resistance mechanism.

by a high-resolution ddgmonomer (HRDM) protocol. In the pre-minimization procedure, we set harmonic distance constraints of 9 Angstroms for all the C-alpha atoms. The minimization procedure is run for three rounds, with a repulsive term set to 100 percent weight and Rosetta parameter file score12 as input parameters. An example of a modeling result is shown in Figures 2A and 2B.

Models of EGFR-RTK dimers were built by the alignment tool of Chimera [47]. Our EGFR-EGFR homodimer template has two peptide chains. The research found that TKI-blocked EGFR with an intact C-terminal lobe (C-lobe) face can be re-activated by the N-lobe and ATP-binding site of its RTK partner [48]. To avoid this phenomenon's interference, we selected one chain with the N-terminal lobe (N-lobe) face at the dimer interaction area, and EGFR (wild type or its mutants) were aligned to this position. The other chain's C-lobe face is near the ATP binding site and located in the interaction area. RTK partners were aligned to this position to establish the whole dimer. The Gefitinib molecule's structure and position are extracted from a crystal structure of the EGFR-drug complex (PDB:2ITY) [43] and aligned to the same drug docking site method. An example of a dimer-drug complex structure is shown in Figure 2C.

B. MOLECULAR DYNAMICS (MD) SIMULATION

MD simulation is a widely used tool for analyzing the properties and behavior of molecules. With initial data, including coordinates, potential energies, and force fields as input, MD algorithm calculates all the target atoms' motions by Newton's equations [49]. In this study, we applied the Amber software suite to execute MD simulations on all the dimer-drug complexes [50] based on the explicit-solvent model. We first solvated the complex into a water-box with Ff99SB and gaff force fields for the simulation procedures. The solvent system was neutralized with Na+ and Cl- ions and the energy of the system was minimized. Procedures

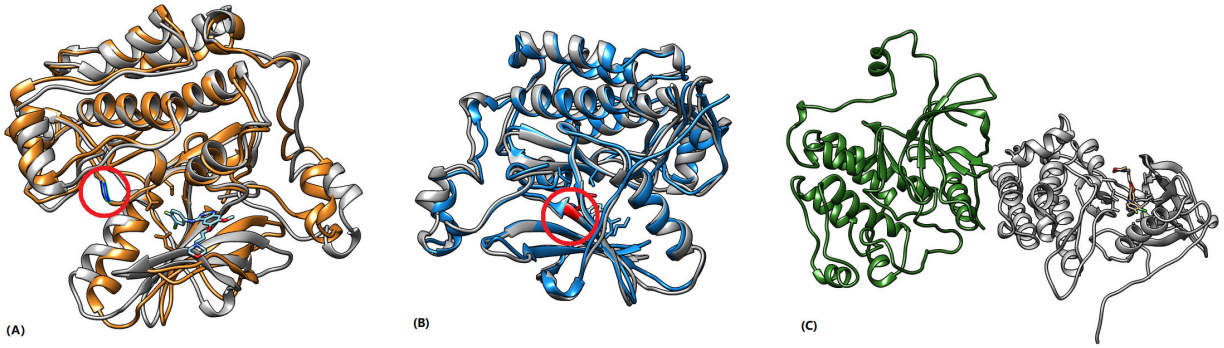

FIGURE 2. Examples of model structures (A) Structure of L858R mutant (colored orange) and wild type EGFR (colored grey), mutation site 858 is colored blue. (B) Structure of L858R-T790M mutant (colored blue) and wild type EGFR (colored grey), secondary mutation site 790 is colored red. (C) Example of a dimer-drug complex. The wild type EGFR is colored grey, with RTK partner colored green. A molecule of Gefitinib is attached to the wild type EGFR.

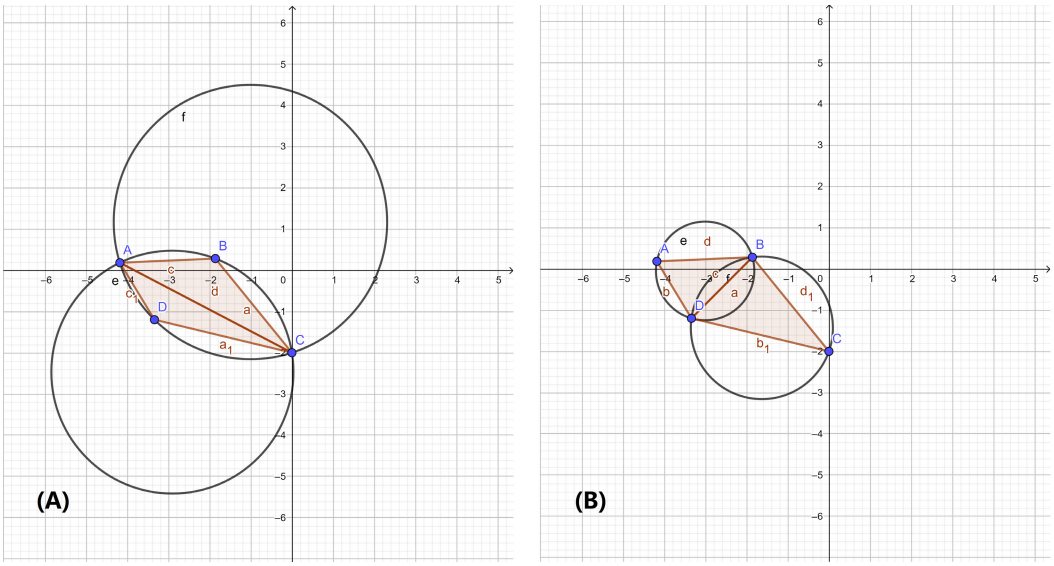

FIGURE 3. Triangulation examples of point set A, B, C, D. The figure is drawn according to fundamental theories [51]. (a) Unsuccessful triangulation that does not satisfy the criterion, as point D in the set, is located in triangle ABC's circumcircle. (b) Successful triangulation after changing the way of triangulation.

of heating for 50 picoseconds(ps), density equilibrium for 50ps, and constant pressure equilibrium for four nanoseconds (ns) were performed in sequence. After all these steps, MD simulation was performed for 10ns, with a sampling interval of 10ps, resulting in 1000 frames' trajectory for each protein systems. The production MD files are analyzed using computational methods.

C. GEOMETRICAL ANALYSIS OF EGFR-RTK INTERACTION IN DIMER-DRUG COMPLEXES

1) ALPHA SHAPE MODELING

Alpha shape modeling is a convenient tool to extract geometrical features. As a linear approximation method, the alpha shape algorithm facilitates the geometrical data and reconstructs the target object's surface to obtain the properties [51]. The alpha shape theory is based on triangulation algorithms. The basic definition of triangulation is that after the triangulation of a particular set of points, no point in the set will be located in the circumcircle of any triangles generated by the algorithm. Figure 3 shows examples on the same point set A, B, C, D. Figure 3A shows a triangulation that does not meet the criterion and needs to be re-generated, while Figure 3B shows a successful triangulation.

Different algorithms are designed to implement the triangulation theory, and Delaunay triangulation is the most popular method. The Delaunay triangulation algorithm aims to maximize the minimum of all angles in the simplexes [52], with criterion as follows:

$$\begin{vmatrix} x_A & y_A & x_A^2+y_A^2 & 1 \\ x_B & y_B & x_B^2+y_B^2 & 1 \\ x_C & y_C & x_C^2+y_C^2 & 1 \\ x_D & y_D & x_D^2+y_D^2 & 1 \end{vmatrix} > 0 \qquad (1)$$

Here, each pair (xi, yi) stands for the coordinates of the corresponding point i. If the determinant's value is positive,

then point D locates in the circumcircle of the triangle formed by points A, B, and C.

The alpha shape algorithm we adopted starts from 3D Delaunay triangulation. Points in dataset l form tetrahedrons instead of 2D triangles, and circumscribed spheres of every tetrahedron are calculated. A value α is defined, and squared radii of the spheres are required to be smaller than or equal to α, while no points in the dataset should appear inside the spheres. The algorithm examines and modifies the tetrahedrons until the requirements are reached, and the corresponding tetrahedrons form the alpha shape at value α.

In the dimer-drug complex, our interest focused on the heavy atoms that composed the molecule's skeleton. As different kinds of atoms have different mass, we implement a weighted alpha shape algorithm, as it considers positions of the points and weights [39]. For this method, we set two points, $p_1 = (\mathrm{p}_1', \mathrm{p}_1'')$ and $p_2 = (\mathrm{p}_2', \mathrm{p}_2'')$ as example. p_1' and p_2' are the coordinates of points, while p_1'' and p_2'' are the values of their weights. Euclidean distance between the two points are set as $\|p_1', p_2'\|$ and the algorithm will compare the Euclidean distance with the sum of weights. The judgement formulas are shown as follows:

$$\|\mathbf{p_1}, \mathbf{p_2}\| = (\mathbf{p_1'} + \mathbf{p_1'}) \tag{2}$$

$$\|\mathbf{p_1}, \mathbf{p_2}\| > (\mathbf{p_1'} + \mathbf{p_1'}) \tag{3}$$

Here, Equation (2) means the two points, p1 and p2, are orthogonal, while Equation (3) means the corresponding points are sub-orthogonal. Then, a point can be defined as:

$$\mathbf{p}_\alpha = (\mathbf{p}', \mathbf{p}' + \alpha) \tag{4}$$

where α is a pre-defined real value.

For every tetrahedron composing the resulting weighted alpha shape model, there is a set of relevant points to the vertices of the tetrahedron. There should be an orthogonal point to all these points and is sub-orthogonal to all the rest points. When the condition is met, the algorithm is successfully ended.

In this study, Computational Geometry Algorithm Library (CGAL) was adopted to calculate the weighted Alpha Shape models of all the MD trajectory frames individually [53]. The value of α was set as 0 for all the complexes. The atoms' coordinates were set as locations of points, and the square values of Van der Waals radii of the atoms were taken as point weights. Resulting weighted alpha shapes were collected as input data for further analysis.

2) MATCHING RATE OF INTERFACIAL ATOMS

As we mainly focused on the EGFR-RTK interactions in the complexes, states of atoms on the interface between two peptide chains are essential. Alpha shape algorithm can help generate the geometrical surface of a specific structure, and we made use of this function to fetch the target atoms. We collected the dimer's surface atoms based on alpha shape model and named these atoms as point set D. Then, we obtain point set E for surface atoms of monomer EGFR and point set R, which correspond to surface atoms of monomer RTK separately. The point set of interfacial atoms, named I, was obtained by set operations. These interfacial atoms can be further divided into atoms on EGFR chain and atoms on RTK chain, named as point set I_E and I_R respectively. S Point sets are obtained as follows:

$$\begin{cases} I = (E \cup R) - D \\ I_E = I \cap E \\ I_R = I \cap R \end{cases} \tag{5}$$

Convex or concave degree of alpha shape surface at interfacial atoms can describe the geometrical properties and characterize monomers' interactions. The solid angle is implemented as a quantization method of curvature [54]. We set the vertices as A, B, C, and D in a tetrahedron, respectively. The dihedral angle between geometrical surfaces ABC and ACD is named as φ_{AC}, dihedral angle between surface ABC and ABD is φ_{AB}, and dihedral angle between surface ACD and ABD is φ_{AD}. Solid angle of point A is calculated by these three values, and the equation is:

$$\Omega = \cos \frac{\sum_i (\varphi_{i,AB} + \varphi_{i,AC} + \varphi_{i,AD} - \pi)}{4} \tag{6}$$

where i stands for the index number of any tetrahedron with A as vertex, and the final result of Ω is scaled into the range of $[-1, 1]$. The value characterizes the surface's shape, where a positive result corresponds to a convex curve, and a negative result corresponds to a concave curve.

With the values that can describe curvatures, we can indicate the interactions between the dimer's two peptide chains by their geometrical properties. For instance, we select an atom A on one chain of the dimer and atom B on the other chain. If these two atoms are on the same edge of any tetrahedron that connects the two chains, we take them as a pair. Suppose curvature at one atom is convex, and the other is concave. In that case, the shapes fit into each other, implying an intensive interaction between the two atoms, and the pair is recorded as matched.

On the contrary, if both curvatures at atoms are both convex or are both concave, then the interaction is weaker, and the pair is marked as unmatched. A two-dimensional example is shown in Figure 4. In the algorithm, we calculate the solid angle of A and B, named as Ω_A and Ω_B. Whether they are matched can be decided and recorded by the equation below:

$$f(A, B) = \begin{cases} 1, & \Omega_A \times \Omega_B < 0 \\ 0, & otherwise \end{cases} \tag{7}$$

For each of the target model, we calculate the matching rate of two chains in the complex based on the following equation:

$$r = \frac{\sum_{(i,j)} f(A_i, B_j)}{N} \tag{8}$$

where (i, j) is the atoms in a pair, and N is the total number of all the pairs. The algorithm was implemented on each frame of the MD trajectory, and the results were collected to analyze the interaction properties.

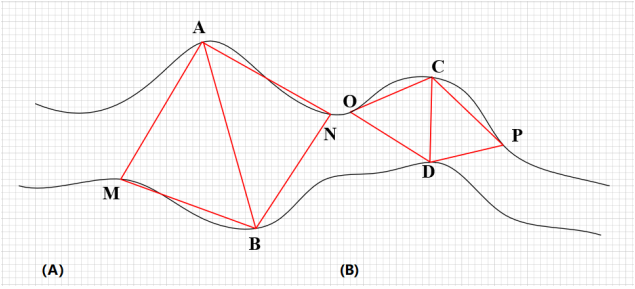

FIGURE 4. 2D example of solid angle matching. (A) At point A, a solid angle is calculated based on all the triangles with A as vertex, namely ABM, ABN, etc. As curvature at points A and B are both concave, surface shapes do not fit into each other. Thus, the interaction here is weak, the atoms do not match. (B) The curvature at point C is concave, but D is convex, leading to a better interaction and the atoms match according to the decision formula.

3) WILCOXON RANK SUM TEST

The Wilcoxon rank sum test, also known as Mann-Whitney U test, is designed to judge whether two data groups are taken out from the same population [55]. First, we make a null hypothesis that the two datasets, named A and B, are from the same population. Then, we combined all the observed values from the two datasets into one group. The number of elements in subset A is recorded as n_A, and number of data in subset B as n_B. The total amount of all the data is N, and data in the new group is re-ranked from 1 to N. Next, the rank scores are divided by the group, and the sum of scores is calculated for each group, recorded as T_A and T_B respectively. The computing method of value U is as following equations:

$$\begin{aligned} &\text{If} \quad n_A > n_B : U = T_A - \frac{n_A(n_A+1)}{2} \\ &\text{If} \quad n_B > n_A : U = T_B - \frac{n_B(n_B+1)}{2} \end{aligned} \tag{9}$$

U is discrete or uniform so that we can set a critical value with assigned probability. U's statistic is compared with this vital value, and if U is larger, we will reject the null hypothesis and declare that the two datasets are not from the same population. Outputs of Wilcoxon rank sum test can quantize the difference between distributions of two datasets and check whether the properties of different mutations are similar or different.

4) CENTROID DISTANCES BETWEEN INTERFACIAL ATOMS

The interfacial atoms are derived as methods above. According to which peptide chain of the dimer they are located on, the interfacial atoms can be divided into two groups, and each group has a mass center. We take a group of point objects as an example. The total number of the objects is N, and their masses are set as $m_1, m_2, m_3, \ldots, m_N$. All these point objects are supposed to be on a specific axis x, with coordinates $x_1, x_2, x_3, \ldots, x_N$. The coordinate of mass center can be calculated as:

$$x_{com} = \frac{\sum_{i=1}^{N} m_i x_i}{\sum_{i=1}^{N} m_i} \tag{10}$$

For a 2D or 3D point object, the algorithm follows similar ideas. In this study, we applied the cpptraj tool in amber to obtain the locations of mass centers [56]. Centroid distances are then calculated on each frame among the MD trajectory. These output data can intuitively evaluate the degree of interactions between two peptide chains in the system.

5) BINDING FREE ENERGY IN THE COMPLEX

Free energy is a widely accepted metric for describing the binding affinity between segments of a system. In a solvent environment, it is quite complicated to calculate the binding free energy value directly. Therefore, the computing method is based on the thermodynamic cycle, with the equation as follows [57], [58]:

$$\begin{aligned} \Delta G_{bind} &= \Delta G_{sol,bind} - \Delta G_{vac,bind} \\ &= \Delta G_{sol,complex} - \Delta G_{sol,receptor} - \Delta G_{sol,ligand} \end{aligned} \tag{11}$$

where $\Delta G_{sol,bind}$ stands for the difference of energy in the solvent environment when the system changes between bounded and unbounded states. $\Delta G_{vac,bind}$ stands for the same energy difference but in the vacuum environment. $\Delta G_{sol,complex}$, $\Delta G_{sol,receptor}$ and $\Delta G_{sol,ligand}$ are the free energies in the solvent environment for the whole dimer-drug complex, receptor component (here is the RTK monomer), and ligand component (here is the EGFR-drug complex), respectively. For a stable system, binding free energy is always negative, and a more negative value of binding free energy represents a more stable binding. Here we implement Molecular mechanics generalized Born and Surface Area (MM/GBSA) protocol provided by Amber to calculate the binding free energy. The output data will state the intensities of interactions in different systems and validate the results obtained from geometrical properties.

III. RESULTS AND DISCUSSIONS

A. ESTABLISHING OF DIMER-DRUG COMPLEXES, MD SIMULATIONS, AND ALPHA SHAPE MODELING

The wild type EGFR, together with all the six mutants, form the target dimers with EGFR or IGF-1R monomer as their partner, respectively. A molecule of Gefitinib is aligned into the 14 binding site residues on the EGFR-Gefitinib crystal structure template. MD simulation of each dimer-drug complexes is executed in an explicit-solvent environment by Amber. The equilibration stage, which lasts for 4ns with a sampling interval of 2ps, is designed to guarantee the prepared system is reliable for further simulation. As an inspection, we applied root-mean-square-deviation (RMSD) on frames of equilibration trajectories, and output data for

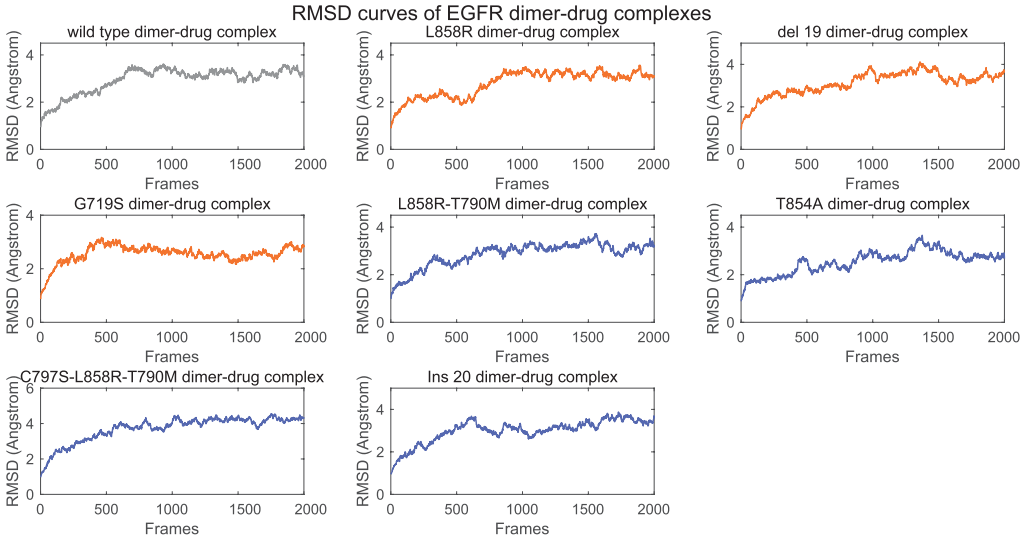

FIGURE 5. RMSD curves of homodimer-drug complexes, with EGFR as RTK partner, plotted on frames of equilibration trajectories. Wild type complex is colored grey, while complexes with drug-sensitive mutants are colored orange, and complexes with drug-resistant mutants are colored blue.

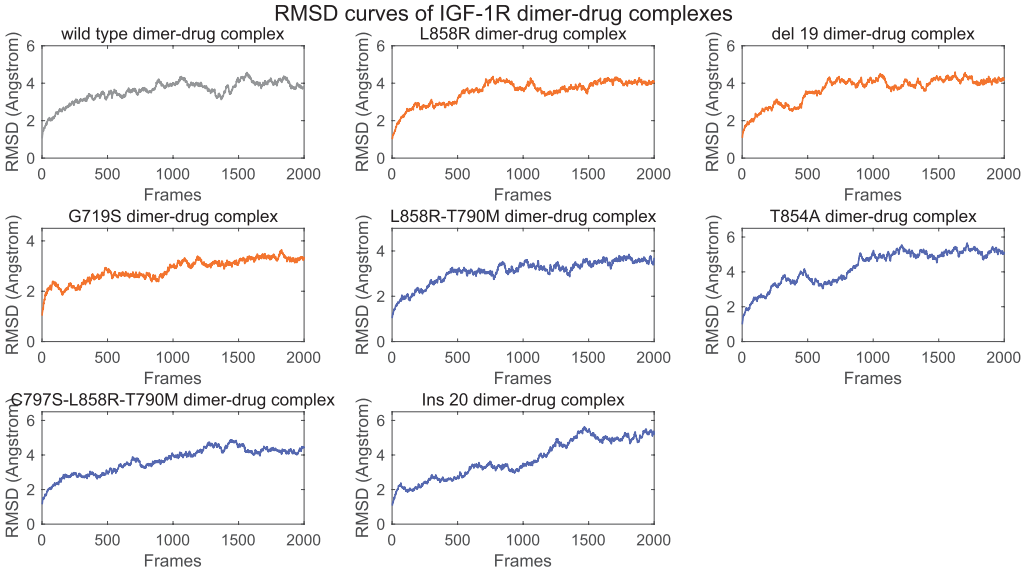

FIGURE 6. RMSD curves of heterodimer-drug complexes, with IGF-1R as RTK partner, plotted on frames of equilibration trajectories. Wild type complex is colored grey, while complexes with drug-sensitive mutants are colored orange, and complexes with drug-resistant mutants are colored blue.

all the complexes are shown in figure 5 and figure 6. It is demonstrated that all the curves are leveling off and are acceptable for subsequent steps. MD simulations are applied to the complexes, and trajectories of 1000 frames are generated for each system.

B. ANALYZING GEOMETRICAL PROPERTIES TO REVEAL THE MECHANISMS OF DRUG RESISTANCE

As dimerization is a vital signaling process, the interactions between EGFR and RTK are essential features that indicate the dimer's stability. Interfacial atoms at the interaction sites between two peptide chains were obtained with a weighted alpha shape algorithm. An example that visually displays the result of alpha shape modeling is showed in Figure 7.

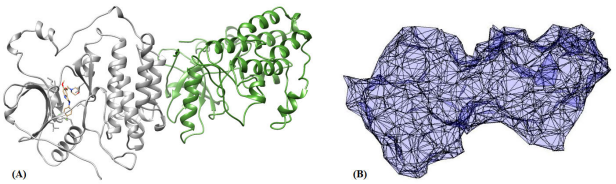

FIGURE 7. An example of alpha shape modeling. (A) A dimer-drug complex with Wild type EGFR. (B) The alpha shape surface we obtained after executing the algorithm on complex in (A).

For each of the dimer-drug complexes, matching rates of all the 1000 frames are calculated. Results are shown in Figures 8 and 9, where the moving average is applied to smooth the curves. In Figure 8, the curves are matching rates between EGFR-EGFR dimer. Figure 8A

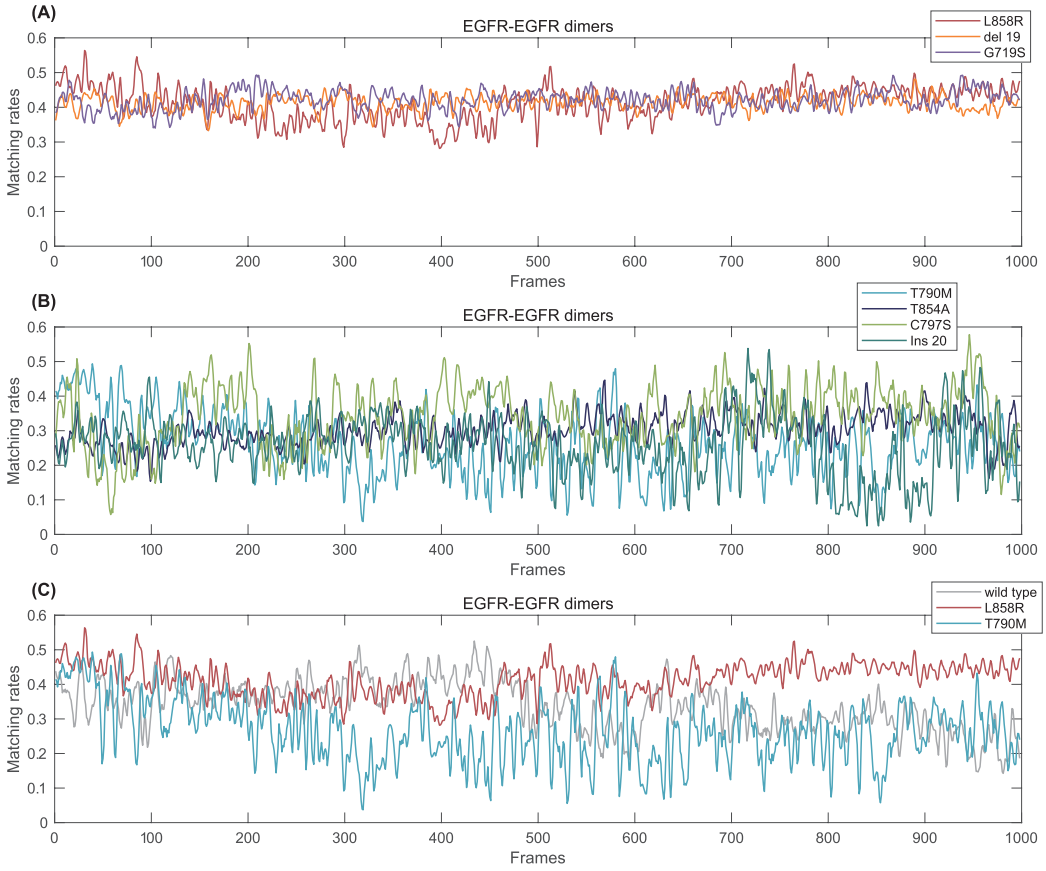

FIGURE 8. Matching rate of atom solid angles obtained from the 1000 trajectory frames of homodimer-drug complexes. Dashed lines are the average value of the corresponding curves. (A) Complexes with drug-sensitive mutants. (B) Complexes with drug-resistant mutants. (C) Comparison between wild type EGFR, L858R complex, and L858R-T790M complex.

shows matching rates of three drug-sensitive mutants (L858R, exon 19 deletion, and G719S) dimerized with EGFR partner, while Figure 8B shows three drug-resistant mutants (T854A, L858R-T790M, L858R-T790M-C797S, and D770_N771insNPG). In Figure 8C, a simple comparison is made. Wild type EGFR is plotted as a reference, and L858R is taken as a representative for drug-resistant dimers. L858R-T790M mutant, with a secondary mutation based on L858R, is selected to represent drug-resistant dimers as they're closely related to each other. Figure 9 shows the curves of IGF-1R-EGFR heterodimer with similar mutants. The high and low order of specific data in the IGF-1R-EGFR results isn't always the same as what appeared in figure 8, but the general regularities are similar.

The average value of frames for each dimer is recorded in table 1. Box plots of the data are shown in Figure 10, providing a more intuitive way to compare and contrast the differences between groups. We can notice from the figures and the table that the drug-sensitive dimers have a higher matching rate than the drug resistance group, which is the same in both homodimer and heterodimer examples. It means that with Gefitinib added to the system, the drug-sensitive dimers, which used to be more active, are now more stable, and more interactions exist in the system. Their EGFR-RTK-drug complex is easier to form and continue to exist. As EGFR and RTK are bound to each other, with drug molecules blocking the ATP binding site, these proteins will not join the signal network. Thus, the over-activation of EGFR is reduced, and downstream signaling will return to normal.

On the contrary, for drug-resistant dimers, the interactions are remarkably weakened by changing geometrical properties due to drug-resistant mutations. The systems become more unstable, giving the drug less chance to come into effect and leading to the loss of efficacy. Results of Wild type EGFR are analogous to drug-resistance examples, which are consistent with the clinical information.

As all the data from complexes are independent from each other, the rank sum test can be executed as a tool for quantizing the differences between datasets. Results reveal the degree of similarity and difference in the dimer groups. Here, we calculated the test result for each pair of the dimers with the ranksum function of MATLAB. The function returns a value for each pair, and the larger the value is, the more consistent the two datasets are. As the values are small, we took a log of them and then filled them in a matrix for

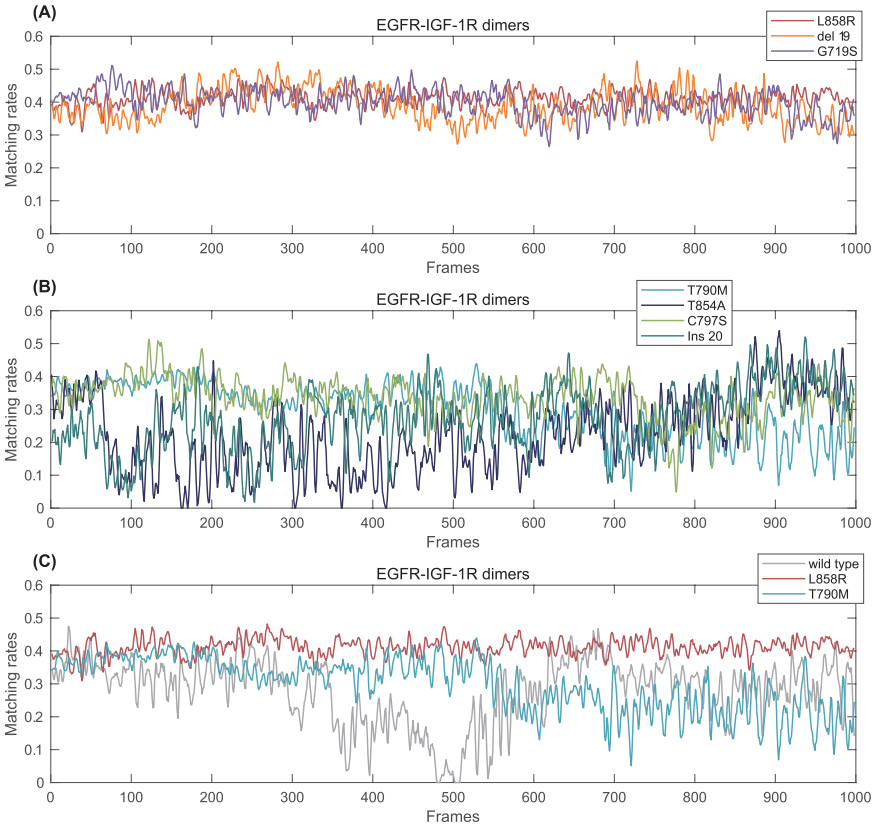

FIGURE 9. Matching rate of atom solid angles obtained from the 1000 trajectory frames of heterodimer-drug complexes. Dashed lines are the average value of the corresponding curves. (A) Complexes with drug-sensitive mutants. (B) Complexes with drug-resistant mutants. (C) Comparison between wild type EGFR, L858R complex, and L858R-T790M complex.

TABLE 1. Average matching rates of the complexes.

Partner \ EGFR	WT	L858R	Del 19	G719S	T790M	T854A	C797S	Ins 20
EGFR	0.3382	0.4136	0.4135	0.4214	0.2634	0.3031	0.3421	0.2586
IGF-IR	0.2793	0.4119	0.3883	0.3990	0.2994	0.2231	0.3370	0.2716

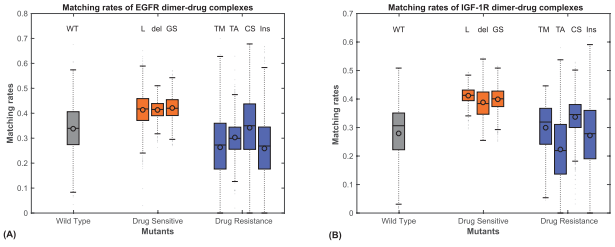

FIGURE 10. Matching rates of different complexes expressed by box plots. As the area is limited, we use L as a short form for complex with L858R mutant. Del stands for exon 19 deletion mutant, GS, TM, TA, CS and Ins corresponding to G719S, T790M, T854A, C797S and D770_N771insNPG mutants. (A) EGFR-EGFR homodimer complexes. (B) IGF-1R-EGFR heterodimer complexes.

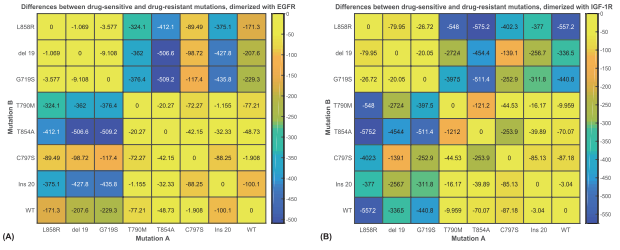

FIGURE 11. Rank sum test results of matching rates, shown as a heatmap. Yellow color corresponds to a larger value, while blue color means a smaller one. (A) Test results of EGFR-EGFR homodimer complexes. (B) Test results of IGF-1R-EGFR heterodimer complexes.

comparison. Heatmap is used to plot the output, as shown in Figure11.

The colors and values directly display results. In general, pairs of dimers are more similar to each other in the left-upper and bottom-right corner, while differences are more extensive in other areas. This indicates that the mutants can be divided into two groups according to similar geometrical properties, with L858R, exon 19 deletions, G719S mutants in one

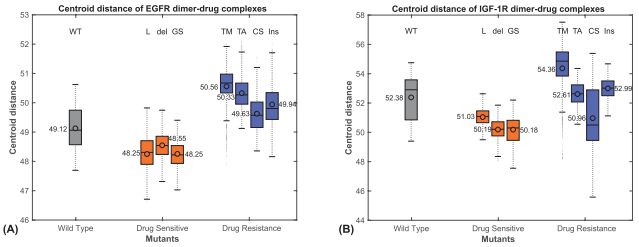

FIGURE 12. Centroid distance of each complex expressed by box plots. The length unit is angstrom. Data of wild-type complexes are colored grey, while drug-sensitive complexes are colored orange, and drug-resistant ones colored blue. (A) EGFR-EGFR homodimer complexes. (B) IGF-1R-EGFR heterodimer complexes.

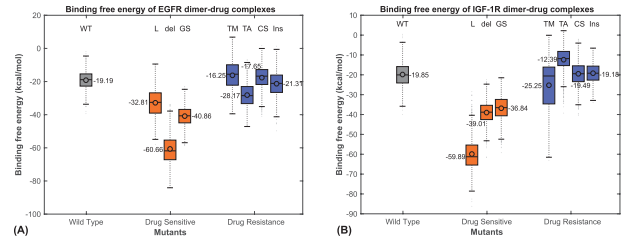

FIGURE 13. Distributions of binding free energy for all the complexes. In general, groups of output data behave the same as what we discussed. (A) Energy for EGFR-EGFR homodimer complexes. (B) Energy for IGF-1R-EGFR heterodimer complexes.

group, and T854A, L858R-T790M, L858R-T790M-C797S, D770_N771insNPG mutants in the other. Wild type EGFR is also not sensitive to the drug and behaves like a drug-resistant mutant. The division meets with the division based on drug resistance, proving that the extracted geometrical features are reliable. The vital information is retained and can distinguish the drug-sensitive mutations from drug-resistance ones. As T790M, C797S and D770_N771insNPG all exist on exon 20 and have similar behavior. Our results demonstrate that the exon 20 mutations have a relationship with the mechanisms of drug resistance. Results of D770_N771insNPG dimers can also show that the αC-β4 loop influences the interactions in both homodimers and heterodimers, indicating that the loop may also play an important role in EGFR dimerization.

C. CENTROID DISTANCE RESULTS

Centroid distances of the interfacial atoms are calculated based on the trajectory frames of all the complexes. The distribution of results is shown in Figure 12. Results show that distances between drug-sensitive mutants and their partners are less than distances in drug-resistant dimers in both homodimer and heterodimer, implying nearer locations and tighter interactions. The results are the same as what we found by matching rates and the previous analysis.

D. BINDING FREE ENERGY OF THE DIMERS

The free energy of binding between the EGFR and RTK partners is calculated by MM/GBSA tool provided by Amber. Distribution of output data corresponding to all the complexes is shown in Figure 13. In both homodimer and heterodimer examples, binding free energy corresponds to drug-sensitive mutants are greater than energy corresponds to drug-resistant ones. The outcomings indicate that binding affinities between drug-sensitive mutants and their partners are tighter, and the complexes are more stable than the drug-resistant ones. These results further confirm our geometrical analysis.

IV. CONCLUSION

In this paper, we focused on interactions between EGFR and its RTK partner, as the dimerization between them plays a crucial role in the cellular signaling network. We established a set of models of drug-sensitive and drug-resistant EGFR mutants, dimerized them with EGFR and IGF-1R, respectively. A TKI, Gefitinib, is also added into each system to form a dimer-drug complex. We applied MD simulations on the complexes to simulate their behaviors for further analysis.

Geometrical properties were extracted with alpha shape modeling. We obtained the solid angles of interface atoms and calculated the matching rate on each phase of MD trajectories to estimate the interactions between partners in the complex. The centroid distance between these interface atoms is also computed as a criterion. It is found that drug-sensitive mutants are bound closely to RTK partners, while binding intensity between drug-resistant mutants and their partners are less. Binding free energy was also taken as a verification method. Results from free energy analysis were similar to results derived from geometrical feathers and supported our findings.

Thus, we can conclude that mutations on EGFR peptide chains result in the changes of geometrical properties on the mutated proteins' surfaces. With the altering of shapes, the interactions between partners are also transformed. Drug-resistant mutations lead to the instability of the TKI blocked dimers, reducing the blocking effect of Gefitinib, and downstream signaling pathways are restored. The three methods are consistent with each other, explaining the mechanism of drug resistance in terms of geometrical properties, distance, and binding free energy. Results showed that mutations on exon 20 are important in drug resistance to Gefitinib, and the αC-β4 loop has a significant association with the asymmetric dimerization of EGFR. Further research can be applied on these topics to provide a deeper understanding of the drug resistance mechanism.

This study provides valuable insights into the EGFR mutation-induced drug resistance, helping to improve the targeted drug therapies for the treatments of NSCLC. The idea and algorithm, we designed to extract features and analyze the geometrical properties are proved to be reliable. These characterization methods have the potential to accomplish relevant tasks in further research. In the future, we will work with newer generation drugs and finding other useful features for drug resistance analysis.

REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, Nov. 2018.

[2] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA, Cancer J. Clin.*, vol. 69, no. 1, pp. 7–34, 2019.

[3] T. Kawaguchi, M. Ando, K. Asami, Y. Okano, M. Fukuda, H. Nakagawa, H. Ibata, T. Kozuki, T. Endo, A. Tamura, M. Kamimura, K. Sakamoto, M. Yoshimi, Y. Soejima, Y. Tomizawa, S.-I. Isa, M. Takada, H. Saka, and A. Kubo, "Randomized phase III trial of erlotinib versus docetaxel as second-or third-line therapy in patients with advanced non–small-cell lung cancer: Docetaxel and erlotinib lung cancer trial (delta)," *J. Clin. Oncol.*, vol. 32, no. 18, pp. 1902–1908, 2014.

[4] N. Normanno, A. De Luca, C. Bianco, L. Strizzi, M. Mancino, M. R. Maiello, A. Carotenuto, G. De Feo, F. Caponigro, and D. S. Salomon, "Epidermal growth factor receptor (EGFR) signaling in cancer," *Gene*, vol. 366, no. 1, pp. 2–16, 2006.

[5] W. Pao, V. Miller, M. Zakowski, J. Doherty, K. Politi, I. Sarkaria, B. Singh, R. Heelan, V. Rusch, L. Fulton, E. Mardis, D. Kupfer, R. Wilson, M. Kris, and H. Varmus, "EGF receptor gene mutations are common in lung cancers from 'never smokers' and are associated with sensitivity of tumors to gefitinib and erlotinib," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 36, pp. 13306–13311, 2004.

[6] P. Blume-Jensen and T. Hunter, "Oncogenic kinase signalling," *Nature*, vol. 411, no. 6835, pp. 355–365, May 2001.

[7] K. Rikova *et al.*, "Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer," *Cell*, vol. 131, no. 6, pp. 1190–1203, Dec. 2007.

[8] S. Fujino, T. Enokibori, N. Tezuka, Y. Asada, S. Inoue, H. Kato, and A. Mori, "A comparison of epidermal growth factor receptor levels and other prognostic parameters in non-small cell lung cancer," *Eur. J. Cancer*, vol. 32, no. 12, pp. 2070–2074, Nov. 1996.

[9] S. V. Sharma, D. W. Bell, J. Settleman, and D. A. Haber, "Epidermal growth factor receptor mutations in lung cancer," *Nature Rev. Cancer*, vol. 7, no. 3, pp. 169–181, 2007.

[10] B. Vogelstein and K. W. Kinzler, "Cancer genes and the pathways they control," *Nature Med.*, vol. 10, no. 8, pp. 789–799, Aug. 2004.

[11] S. Khozin, G. M. Blumenthal, X. Jiang, K. He, K. Boyd, A. Murgo, R. Justice, P. Keegan, and R. Pazdur, "US food and drug administration approval summary: Erlotinib for the first-line treatment of metastatic non-small cell lung cancer with epidermal growth factor receptor exon 19 deletions or exon 21 (L858R) substitution mutations," *Oncologist*, vol. 19, no. 7, p. 774, 2014.

[12] R. Sordella, "Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways," *Science*, vol. 305, no. 5687, pp. 1163–1167, Aug. 2004.

[13] B. Zou, V. H. F. Lee, L. Chen, L. Ma, D. D. Wang, and H. Yan, "Deciphering mechanisms of acquired T790M mutation after EGFR inhibitors for NSCLC by computational simulations," *Sci. Rep.*, vol. 7, no. 1, pp. 1–13, Dec. 2017.

[14] S. Kobayashi, T. J. Boggon, T. Dayaram, P. A. Jänne, O. Kocher, M. Meyerson, B. E. Johnson, M. J. Eck, D. G. Tenen, and B. Halmos, "EGFR mutation and resistance of non–small-cell lung cancer to gefitinib," *New England J. Med.*, vol. 352, no. 8, pp. 786–792, 2005.

[15] A. Ghosh and H. Yan, "Hydrogen bond analysis of the EGFR-ErbB3 heterodimer related to non-small cell lung cancer and drug resistance," *J. Theor. Biol.*, vol. 464, pp. 63–71, Mar. 2019.

[16] C.-H. Yun, K. E. Mengwasser, A. V. Toms, M. S. Woo, H. Greulich, K.-K. Wong, M. Meyerson, and M. J. Eck, "The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP," *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 6, pp. 2070–2075, Feb. 2008.

[17] D. D. Wang, L. Ma, M. P. Wong, V. H. F. Lee, and H. Yan, "Contribution of EGFR and ErbB-3 heterodimerization to the EGFR mutation-induced gefitinib- and erlotinib-resistance in non-small-cell lung carcinoma treatments," *PLoS ONE*, vol. 10, no. 5, May 2015, Art. no. e0128360.

[18] Y. Shan, M. P. Eastwood, X. Zhang, E. T. Kim, A. Arkhipov, R. O. Dror, J. Jumper, J. Kuriyan, and D. E. Shaw, "Oncogenic mutations counteract intrinsic disorder in the EGFR kinase and promote receptor dimerization," *Cell*, vol. 149, no. 4, pp. 860–870, May 2012.

[19] L. Ma, D. D. Wang, Y. Huang, M. P. Wong, V. H. F. Lee, and H. Yan, "Decoding the EGFR mutation-induced drug resistance in lung cancer treatment by local surface geometric properties," *Comput. Biol. Med.*, vol. 63, pp. 293–300, Aug. 2015.

[20] J. M. Sanders, M. E. Wampole, M. L. Thakur, and E. Wickstrom, "Molecular determinants of epidermal growth factor binding: A molecular dynamics study," *PLoS ONE*, vol. 8, no. 1, Jan. 2013, Art. no. e54136.

[21] R. Qureshi, M. Nawaz, A. Ghosh, and H. Yan, "Parametric models for understanding atomic trajectories in different domains of lung cancer causing protein," *IEEE Access*, vol. 7, pp. 67551–67563, 2019.

[22] R. Qureshi, A. Ghosh, and H. Yan, "Correlated motions and dynamics in different domains of EGFR with L858R and T790M mutations," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, early access, May 18, 2020, doi: 10.1109/TCBB.2020.2995569.

[23] R. Qureshi, M. Zhu, A. Ghosh, and H. Yan, "Computational analysis of structural dynamics of EGFR and its mutants," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2019, pp. 2784–2791.

[24] R. Qureshi, M. Zhu, and H. Yan, "Visualization of protein-drug interactions for the analysis of lung cancer drug resistance," *IEEE J. Biomed. Health Informat.*, early access, Sep. 29, 2020, doi: 10.1109/JBHI.2020.3027511.

[25] A. Dixit and G. M. Verkhivker, "Computational modeling of allosteric communication reveals organizing principles of mutation-induced signaling in ABL and EGFR kinases," *PLoS Comput. Biol.*, vol. 7, no. 10, Oct. 2011, Art. no. e1002179.

[26] H. E. Jones, L. Goddard, J. M. W. Gee, S. Hiscox, M. Rubini, D. Barrow, J. M. Knowlden, S. Williams, A. E. Wakeling, and R. I. Nicholson, "Insulin-like growth factor-I receptor signalling and acquired resistance to gefitinib (ZD1839; Iressa) in human breast and prostate cancer cells," *Endocrine-Rel. Cancer*, vol. 11, no. 4, pp. 793–814, Dec. 2004.

[27] T. Ahmad, G. Farnie, N. J. Bundred, and N. G. Anderson, "The mitogenic action of insulin-like growth factor I in normal human mammary epithelial cells requires the epidermal growth factor receptor tyrosine kinase," *J. Biol. Chem.*, vol. 279, no. 3, pp. 1713–1719, Jan. 2004.

[28] S. Kakiuchi, Y. Daigo, N. Ishikawa, C. Furukawa, T. Tsunoda, S. Yano, K. Nakagawa, T. Tsuruo, N. Kohno, M. Fukuoka, S. Sone, and Y. Nakamura, "Prediction of sensitivity of advanced non-small cell lung cancers to gefitinib (Iressa, ZD1839)," *Hum. Mol. Genet.*, vol. 13, no. 24, pp. 3029–3043, Dec. 2004.

[29] M. Chiba, Y. Togashi, E. Bannno, Y. Kobayashi, Y. Nakamura, H. Hayashi, M. Terashima, M. A. De Velasco, K. Sakai, Y. Fujita, T. Mitsudomi, and K. Nishio, "Efficacy of irreversible EGFR-TKIs for the uncommon secondary resistant EGFR mutations L747S, D761Y, and T854A," *BMC Cancer*, vol. 17, no. 1, pp. 1–10, Dec. 2017.

[30] B. A. Helfrich, D. Raben, M. Varella-Garcia, D. Gustafson, D. C. Chan, L. Bemis, C. Coldren, A. Barón, C. Zeng, W. A. Franklin, F. R. Hirsch, A. Gazdar, J. Minna, and P. A. Bunn, "Antitumor activity of the epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor gefitinib (ZD1839, Iressa) in non–small cell lung cancer cell lines correlates with gene copy number and EGFR mutations but not EGFR protein levels," *Clin. Cancer Res.*, vol. 12, no. 23, pp. 7117–7125, Dec. 2006.

[31] Y.-R. Chen, Y.-N. Fu, C.-H. Lin, S.-T. Yang, S.-F. Hu, Y.-T. Chen, S.-F. Tsai, and S.-F. Huang, "Distinctive activation patterns in constitutively active and gefitinib-sensitive EGFR mutants," *Oncogene*, vol. 25, no. 8, pp. 1205–1215, Feb. 2006.

[32] L. Chen, W. Fu, L. Zheng, Z. Liu, and G. Liang, "Recent progress of small-molecule epidermal growth factor receptor (EGFR) inhibitors against C797S resistance in non-small-cell lung cancer: Miniperspective," *J. Med. Chem.*, vol. 61, no. 10, pp. 4290–4300, 2017.

[33] K. S. Thress, C. P. Paweletz, E. Felip, B. C. Cho, D. Stetson, B. Dougherty, Z. Lai, A. Markovets, A. Vivancos, Y. Kuang, D. Ercan, S. E. Matthews, M. Cantarini, J. C. Barrett, P. A. Jänne, and G. R. Oxnard, "Acquired EGFR C797S mutation mediates resistance to AZD9291 in non–small cell lung cancer harboring EGFR T790M," *Nature Med.*, vol. 21, no. 6, pp. 560–562, Jun. 2015.

[34] W. Yeung, Z. Ruan, and N. Kannan, "Emerging roles of the αc-β4 loop in protein kinase structure, function, evolution, and disease," *IUBMB Life*, vol. 72, no. 6, pp. 1189–1202, 2020.

[35] Z. Ruan and N. Kannan, "Altered conformational landscape and dimerization dependency underpins the activation of EGFR by αc–β4 loop insertion mutations," *Proc. Nat. Acad. Sci. USA*, vol. 115, no. 35, pp. E8162–E8171, 2018.

[36] H. Yasuda *et al.*, "Structural, biochemical, and clinical characterization of epidermal growth factor receptor (EGFR) exon 20 insertion mutations in lung cancer," *Sci. Transl. Med.*, vol. 5, no. 216, 2013, Art. no. 216ra177.

[37] J. G. Paez, "EGFR mutations in lung cancer: Correlation with clinical response to gefitinib therapy," *Science*, vol. 304, no. 5676, pp. 1497–1500, Jun. 2004.

[38] Y. Li, X. Hu, Q. Li, F. Wang, B. Zhang, K. Ding, B. Tan, N. Lin, and C. Zhang, "Shikonin sensitizes wild-type EGFR NSCLC cells to erlotinib and gefitinib therapy," *Mol. Med. Rep.*, vol. 18, pp. 3882–3890, Aug. 2018.

[39] H. Edelsbrunner, *Weighted Alpha Shapes*. Champaign, IL, USA: Univ. Illinois Urbana-Champaign, 1992.
[40] H. Edelsbrunner and E. P. Mücke, "Three-dimensional alpha shapes," *ACM Trans. Graph.*, vol. 13, no. 1, pp. 43–72, Jan. 1994.
[41] F. Wilcoxon, "Individual comparisons by ranking methods," in *Breakthroughs in Statistics*. New York, NY, USA: Springer, 1992, pp. 196–202.
[42] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *Ann. Math. Statist.*, vol. 18, no. 1, pp. 50–60, Mar. 1947.
[43] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The protein data bank," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 235–242, 2000.
[44] S. L. Degorce, S. Boyd, J. O. Curwen, R. Ducray, C. T. Halsall, C. D. Jones, F. Lach, E. M. Lenz, M. Pass, S. Pass, and C. Trigwell, "Discovery of a potent, selective, orally bioavailable, and efficacious novel 2-(pyrazol-4-ylamino)-pyrimidine inhibitor of the insulin-like growth factor-1 receptor (IGF-1R)," *J. Med. Chem.*, vol. 59, no. 10, pp. 4859–4866, May 2016.
[45] K. T. Simons, C. Kooperberg, E. Huang, and D. Baker, "Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions," *J. Mol. Biol.*, vol. 268, no. 1, pp. 209–225, Apr. 1997.
[46] Y. Song, F. DiMaio, R. Y.-R. Wang, D. Kim, C. Miles, T. Brunette, J. Thompson, and D. Baker, "High-resolution comparative modeling with RosettaCM," *Structure*, vol. 21, no. 10, pp. 1735–1742, Oct. 2013.
[47] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, "UCSF chimera—A visualization system for exploratory research and analysis," *J. Comput. Chem.*, vol. 25, no. 13, pp. 1605–1612, 2004.
[48] X. Zhang, J. Gureasko, K. Shen, P. A. Cole, and J. Kuriyan, "An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor," *Cell*, vol. 125, no. 6, pp. 1137–1149, Jun. 2006.
[49] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, vol. 1. Amsterdam, The Netherlands: Elsevier, 2001.
[50] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. R. Crowley, R. C. Walker, W. Zhang, and K. M. Merz, *Amber 10*. Oakland, CA, USA: Univ. California, 2008.
[51] F. Bernardini and C. L. Bajaj, "Sampling and reconstructing manifolds using alpha-shapes," Dept. Comput. Sci., Purdue Univ., West Lafayette, IN, USA, Tech. Rep. CSD-TR 97-013, 1997.
[52] L. Guibas and J. Stolfi, "Primitives for the manipulation of general subdivisions and the computation of Voronoi," *ACM Trans. Graph.*, vol. 4, no. 2, pp. 74–123, Apr. 1985.
[53] CGAL Consortium, "CGAL: Computational geometry algorithms library," Tech. Rep., 1996. [Online]. Available: https://www.cgal.org/
[54] W. Zhou, H. Yan, and Q. Hao, "Analysis of surface structures of hydrogen bonding in protein–ligand interactions using the alpha shape model," *Chem. Phys. Lett.*, vol. 545, pp. 125–131, Aug. 2012.
[55] P. E. McKnight and J. Najab, "Mann-Whitney U test," *Corsini Encyclopedia Psychol.*, p. 1, Jan. 2010, doi: 10.1002/9780470479216.corpsy0524.
[56] D. R. Roe and T. E. Cheatham, "PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data," *J. Chem. Theory Comput.*, vol. 9, no. 7, pp. 3084–3095, Jul. 2013.
[57] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and T. E. Cheatham, "Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models," *Accounts Chem. Res.*, vol. 33, no. 12, pp. 889–897, Dec. 2000.
[58] H. K. Srivastava and G. N. Sastry, "Molecular dynamics investigation on a series of HIV protease inhibitors: Assessing the performance of MM-PBSA and MM-GBSA approaches," *J. Chem. Inf. Model.*, vol. 52, no. 11, pp. 3088–3098, Nov. 2012.

MENGXU ZHU received the B.Eng. degree in biomedical from Zheijiang University, and the M.Sc. degree in electrical engineering from The Hong Kong University of Science and Technology, Hong Kong. He is currently pursuing the Ph.D. degree in electrical engineering with the City University of Hong Kong, Hong Kong. His current research interests include bioinformatics, machine learning, and signal processing.

RIZWAN QURESHI (Graduate Student Member, IEEE) received the B.Eng. degree in electronic engineering from the Mehran University of Engineering & Technology, Jamshoro, in 2010, and the M.Sc. degree in electrical engineering from the Institute of Space Technology, Islamabad, Pakistan, in 2015. He is currently pursuing the Ph.D. degree in electrical engineering with the City University of Hong Kong. Before joining City University of Hong Kong, he was a Lecturer with the Electrical Engineering Department, COMSATS University Islamabad–Wah, Pakistan. His research interests include bioinformatics, signal and image processing, machine learning, and spectral imaging.

HONG YAN (Fellow, IEEE) received the Ph.D. degree from Yale University. He was a Professor of imaging science with The University of Sydney. He is currently the Chair Professor of computer engineering with the City University of Hong Kong. His research interests include image processing, pattern recognition, and bioinformatics, and he has over 600 journal and conference publications in these areas. He is also a Fellow of IAPR. He received the 2016 Norbert Wiener Award from the IEEE SMC Society for contributions to image and biomolecular pattern recognition techniques.