HMRF (RFCID/HHSRF)

Health and Medical Research Fund

Computational Platform for Modelling, Analysis and Prediction of Anti-EGFR Drug Resistance for Lung Cancer

Submitted to the Grant Review Board (October 2012)

Applicants

Hong Yan
Department of Electronic Engineering, City University of Hong Kong

Guangyu Zhu
Department of Biology and Chemistry, City University of Hong Kong

Maria Wong
Department of Pathology, Li Ka Shing Faculty of Medicine, University of Hong Kong

Victor Lee
Department of Clinical Oncology, Li Ka Shing Faculty of Medicine, University of Hong Kong

Contents

	Page
Introduction	3
Methods	2
Results	4
Discussion	6
Limitations	7
Conclusions	8
Implications	9
Dissemination	9
Publications	10
References	10
List of Research Workers	11
Acknowledgements	11
Financial Statement	12

Introduction

Lung cancer has the largest mortality rate among all cancer types and results in 1.6 million deaths in the world each year.[1] In Hong Kong, lung, liver and colorectal cancers are the three leading causes of cancer deaths, and lung cancer related deaths are more than those from liver and colorectal cancers combined (http://www3.ha.org.hk/cancereg/facts.html). About 85% of lung cancer patients have non-small cell lung cancer (NSCLC). Many of the NSCLC cases are caused by a mutation of the epidermal growth factor receptor (EGFR), especially in Asia.[2,3] Several drugs are available in the market and they can be effective initially to make the tumour shrink and help the patients recover. However, after several months to one year, almost every patient develops drug resistance due to one or more additional mutations of EGFR, and the patient's condition becomes worse quickly.[2-8]

In this project, we study EGFR mutations at the molecular and atomic levels (Publications 1 to 12). We have collected EGFR mutations from research publication and from clinical cases at Queen Mary Hospital. Some of the mutations observed locally are rare and have never been reported in the literature. Based on computational models, we analyse how the 3D structure of EGFR will change due to a mutation. Then for each drug, we compute its binding strength with EGFR before and after the secondary mutation. The reduction in the binding strength reflects the degradation of the drug effectiveness. We have built and published a 3D structural database of EGFR mutants. We have analysed the characteristics of all known EGFR mutations at the atomic level. Our work can provide a useful reference to medical doctors for assessment of drug resistance level and planning personalised treatment.

We have investigated one of the most important root reasons of NSCLC drug resistance: EGFR mutation. We have conducted computational modelling of EGFR mutants and analysis of EGFR-drug interaction patterns. In our work,

- Any observed EGFR mutation can be modelled mathematically and its 3D structure can be predicted computationally. The fundamental reason of drug resistance can be found at the atomic level.
- Different drugs can be analysed. Based on our computer model, the binding strength between an EGFR mutant and a drug can be calculated.
- Drug resistance level can be evaluated for each mutation and each drug. Thus, a comprehensive database of EGFR mutation and drug effectiveness is established, which provides a useful reference to medical doctors.
- Our computational framework is much less expensive than wet-lab experiments. It can also be combined with biological experiments and clinical trials to improve our database.

Methods

Prediction of 3D structures of EGFR mutants: Experimentally, the 3D structure (the location of each atom) of a protein is usually obtained using X-ray crystallography or nuclear magnetic resonance (NMR) (Publication 2). There are over 100 EGFR mutations, and it would be too expensive to use these techniques to determine the 3D structures of all mutants and all EGFR mutant and drug complexes. In PubMed, there are over 4000 articles related to EGFR.[1-8] However, in the public protein databank (PDB), only about 30 EGFR and drug related complexes are available. Some of these complexes provide information of different domains of EGFR, so only several EGFR mutant-drug patterns can be used directly. To solve this problem, we predict the 3D structure of each EGFR mutant computationally (Publications 1 and 2). The General AMBER Force Field (GAFF) (http://ambermd.org/ antechamber/gaff.html), which covers most of the organic chemical space, is implemented to generate the topology and coordinate files of inhibitors (such as gefitinib and erlotinib). Based on GAFF, the antechamber program in AMBER24 assigns atomic charges and atom/bond types for the inhibitors, and further constructs their topology files. We translate these mutations from protein sequences into their 3D structures. A mutated protein structure is determined

based on homology modelling. Different types of mutations are then obtained using two programs, scap and loopy (Publication 1). An inhibitor is separately aligned to the binding pocket of each mutant structure, to construct their bound complexes.

Molecular dynamics (MD) simulations: An EGFR mutant-drug complex is computationally solvated into a water box. The dynamics of the complex is simulated in this solvent environment (Publications 1 and 2). Prior to the crucial MD simulation, the entire system should be equilibrated to a stable state. We employ sander in AMBER for a series of equilibrating operations, which incorporates a short 1000-step minimization (the first half with the steepest descent steps) to remove bad contacts, a 50-picosecond (ps) heating (0 , 300 K) and a 50 ps density equilibration with weak restraints (weight of 2.0) from a harmonic potential on the mutant-inhibitor complex, and a 500 ps constant pressure equilibration at 300 K. All simulations are performed with SHAKE constraints on hydrogen atoms to remove their bond stretching freedom, and the Langevin dynamics is adopted for an efficient temperature control. The equilibration of each system is verified through observing the temperature, density, energy and backbone root-mean-square deviation (RMSD) of each system. Once each system equilibration is achieved, we generate the production MD simulation for 2 ns, where we collect trajectory frames at a step of 10 ps and 200 frames in each trajectory. A stable backbone RMSD in each system is an apparent indicator of the stabilization of the production MD simulation, which guarantees a posterior reliable calculation of the binding free energy. For each system, the backbone RMSD distribution over the simulation period (2 ns) is investigated.

Computation of binding free energy: The production MD simulations produce the motion trajectories of the solvated mutant-inhibitor systems, and the binding free energies are calculated based on these trajectories (Publications 1, 2 and 7). Binding free energy is a quantitative estimate of the binding affinity of a solvated receptor-ligand system. Based on the computations of different types of free energy differences, MMPBSA in AMBER derives the binding free energies, which encompass energy components of Van der Waals forces (VDW), electrostatic interactions (EEL), and the polar (EPB) and non-polar (ENPOLAR) terms of the solvation free energies. For the wild-type protein and observed mutants, we calculate their binding free energies with each inhibitor.

Computational prediction of drug resistance: The potency of an inhibitor in the treatment of a specific patient can be measured by its survival time or response level. In clinical observations, survival time is generally recorded in the unit of months or days, corresponding to a continuous variable in computation. We use machine learning methods to build a classifier for the prediction of drug resistance levels (Publication 1) Response level, as output of the classifier can be divided into four categories and thus mapped into a discrete variable ranging in [1, 4]. The computed binding free energy is used as an important feature to the classifier. In addition, personal information of each patient is recorded as well, including age, gender, smoking history, performance status, subtypes of the NSCLC, stages describing the development of the NSCLC, brain metastasis, and suspension of drugs. Once trained, the classifier can be used to process new input data.

Analysis of EGFR mutant surface characteristics: We used solid angles to characterize the geometric properties of the EGFR surface atoms, which were extracted using the alpha shape model (Publications 3, 5, 6, 8, 9, and 11). The solid angles of surface atoms can represent the surface curvature, which is used as a geometric property in our analysis. 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 analysing the solid angles around the binding sites.

Study of irreversible inhibitors: Besides reversible tyrosine kinase inhibitors (TKIs), new-generation irreversible inhibitors, such as afatinib, embark on playing an important role in NSCLC treatment. To achieve an optimal application of these inhibitors, the correlation between the EGFR mutation status and the potency of such an inhibitor should be decoded. In this study, the correlation was

profiled for afatinib, based on a cohort of patients with the EGFR-mutated NSCLC. Relying on extracted DNAs from the paraffin-embedded tumour samples, EGFR mutations were detected by direct sequencing (Publication 10). Progression-free survival (PFS) and the response level were recorded as study endpoints. These PFS and response values were analysed and correlated to different mutation types, implying a higher potency of afatinib to classic activation mutations (L858R and deletion 19) and a lower one to T790M-related mutations. To further bridge the mutation status with afatinib-related response or PFS, we conducted a computational study to estimate the binding affinity in a mutant-afatinib system, based on molecular structural modelling and dynamics simulations.

Characterisation of EGFR and ErbB-3 heterodimerization: We have separately investigated the EGFR and ErbB-3 heterodimerization, regarded as the origin of intracellular signaling pathways (Publication 4). On one hand, we combined the molecular interaction in EGFR heterodimerization with that between the EGFR tyrosine kinase and its inhibitor. For 168 clinical subjects, we characterized their corresponding EGFR mutations using molecular interactions, with three potential dimerization partners (ErbB-2, IGF-1R and c-Met) of EGFR and two of its small molecule inhibitors (gefitinib and erlotinib). Based on molecular dynamics simulations and structural analysis, we modelled these mutant-partner or mutant-inhibitor interactions using binding free energy and its components.

Results

Our research outcomes are summarized on the website http://bcc.ee.cityu.edu.hk/SFBG/. Some key results are presented below.

Computational Platform and EGFR Mutant Structural Database: On the website, under "Computational Platform", we list all known EGFR mutations that cause NSCLC and drug resistance, which are collected from literature and local hospitals. When the user clicks on a mutation name, the computer will show an interactive display of the corresponding EGFR mutant. The user can rotate and zoom in or out the mutant. The computer can also automatically show the mutant viewed at different angles if the user clicks on "spin on". The wild-type EGFR will be superimposed on the mutant if the user clicks on "reference on". Under "Computational Platform", the user can report any newly observed EGFR mutation to us for analysis by clicking on "Report New EGFR Mutations". We also provide a summary of all these mutants in "EGFR Mutant Structural Database: Computationally predicted 3D structures and the corresponding binding free energies with gefitinib and erlotinib" under "Research and Publications", which is illustrated in Figure 1, on our website http://bcc.ee.cityu.edu.hk/SFBG/research_and_publications.html (Publication 2).

Personalized prediction of NSCLC drug resistance: We combine EGFR mutant-drug binding free energy with specific personal features for 168 clinical subjects to construct a personalized drug resistance prediction model (Publication 1). The binding free energy is evaluated based on EGFR 3D structure prediction and MD simulation methods discussed above. Personal features used are shown in Figure 2. The utilization of extreme learning machines (ELMs) and leave-one-out cross-validation provides a successful identification of resistant subjects with high performance. The classification rates for different configurations of ELMs are shown in Figure 2. We have achieved about 90% accuracy for testing samples.

EGFR mutant surface characteristics: We used 3D alpha shape modelling and a solid-angle analysis to reveal the local surface geometric properties of EGFR mutants (Publication 3). Our analysis results show that the number of atoms with the changes of convex shapes converting to concave, degree variation and disappearance has strong correlations with the drug progression time. Moreover, the number of binding-site atoms with solid angles in [0.71, 1], [0.61, 1] or [0.5, 1] also shows a strong correlation with the drug progression time. 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. Figure 3 show the 3D alpha shapes of the wild type EGRF and its mutation with L858R. The mutation changes the convexity or concavity of the local surface regions. Figure 4 illustrates the four types of surface structure changes due to protein mutation. These changes reduce the protein-drug binding strength and reduce the drug effectiveness. Our analysis provides an insight to the physical reasons why drug resistance occurs because of EGFR mutations.

Number	Mutation types	Description	3D structures	Binding free energy with gefitinib (structureG01-gefitinib complex)(kcal/mol)	Binding free energy with erlotinib (structureE01-erlotinib complex)(kcal/mol)
1	A767_TLA_S768	Insertion	PDB files	-42.3641	-31.7393
2	D761_EAFQ_E762	Insertion	PDB files	-31.0127	-28.0779
3	D770_G_N771	Insertion	PDB files	-23.9836	-40.816
4	V769_ASV_D770	Insertion	PDB files	-45.9188	-33.6622
5	V769_CV_D770	Insertion	PDB files	-36.1098	-33.3974
6	V769_Y_D770	Insertion	PDB files	-40.7622	-33.868
7	V774_HV_C775	Insertion	PDB files	-32.6008	-38.1589
8	delE746_A750	Deletion	PDB files	-35.2995	-44.6007
9	delE746_S752	Deletion	PDB files	-31.2015	-33.2221
10	delE749_T751	Deletion	PDB files	-42.381	-38.8129
.....					
106	G719A_L861Q	Substitution	PDB files	-35.8614	-45.7243
107	G719C_S768I	Substitution	PDB files	-44.4563	-41.8073
108	G724S_L861Q	Substitution	PDB files	-35.6351	-45.3965
109	L858R_L861F	Substitution	PDB files	-34.8422	-40.7398
110	R776H_L858R	Substitution	PDB files	-33.2638	-43.246
111	S768I_V774M	Substitution	PDB files	-43.5982	-39.3976
112	T854A_L858R	Substitution	PDB files	-29.8618	-36.6631

Figure 1: The "EGFR Mutant Structural Database" we have developed is available on-line.

No.	Abbreviation	Description	Continuous/Discrete	Range
1	age	Age of the patient	Discrete	[0, 4]
2	sex	Gender of the patient	Discrete	[0, 1]
3	smoke	Smoking history	Discrete	[0, 2]
4	PS	Performance status	Discrete	[0, 3]
5	histolog	Subtypes of the NSCLC	Discrete	[1, 7]
6	t_stage	Stages describing the development of the NSCLC	Discrete	[1, 4]
7	n_stage		Discrete	[0, 3]
8	m_stage		Discrete	[0, 4]
9	o_stage		Continuous	[3.2, 4.0]
10	bra_mets	Brain metastasis	Discrete	[0, 9]
11	susp_ire	Suspension of TKIs	Discrete	[0, 1]

Number of Hidden Nodes $\tilde{N}$	Training Time (Each fold)	Testing Time (Each fold)	Training Accuracy (Average)	Testing Accuracy (Average)
50	0.0145	0.0004	0.8004	0.8333
100	0.0516	0.0002	0.9303	0.8810
150	0.1147	0.0003	0.9762	0.9583
200	0.1461	0.0006	0.97624	0.9286
250	0.1649	0.0005	0.97624	0.9226
300	0.1642	0.0005	0.9763	0.9107
350	0.1729	0.0007	0.97624	0.9048
400	0.1827	0.0007	0.9763	0.8869
450	0.1887	0.0009	0.9762	0.8869
500	0.1978	0.0009	0.9763	0.8810

Figure 2: Personalized prediction of NSCLC drug resistance. Left: Personal features in addition to EGFR mutant-inhibitor binding energy used for classification. Right: Training and testing rates with different numbers of hidden nodes used in neural network based classifiers (ELMs).

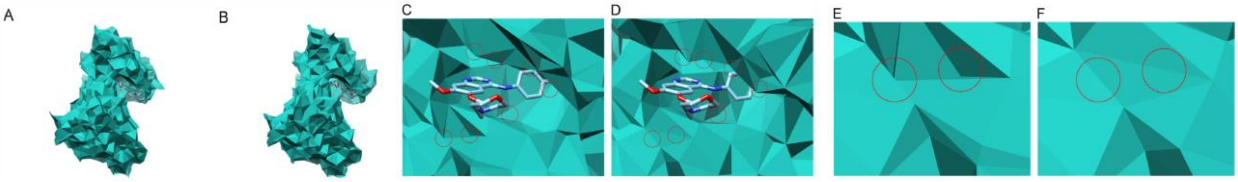

Figure 3: Surface propeerties of EGFR. (A) and (B) show 3D alpha shapes of the wild type EGFR and the mutant with L858R respectively. (C) and (D) display the surface structures at the drug binding site of the wild type EGFR and the mutant respecively. (E) and (F) are the zoomed version of (C) and (D) respectively to illustrate the curvature changes at the binding site.

Selectivity profile of afatinib: A group of patients with EGFR-mutated NSCLC were studied. Their detailed EGFR mutation types were biologically screened prior to their treatment with oral afatinib, and the cancer progression of each patient was carefully followed up (Publication 10). Specifically, PFS and response level were two major indicators of drug effectiveness. Supported by the collected patient data, we detected that the classic activating mutations (L858R and exon 19 deletion) normally resulted in better afatinib-related responses or PFS values, while those involving the T790M mutation mostly perform worse during treatment. Figure 5 shows both experimental and computational results

of potency ranking for afatinib. The GB energy model agrees with the experiments for wild type EGFR and its two mutants (Figure 5, left and middle diagrams). We have also modelled other mutants computational, for which experiment data are not available. Figure 5 (right) shows the ranking profile obtained using the GB energy model.

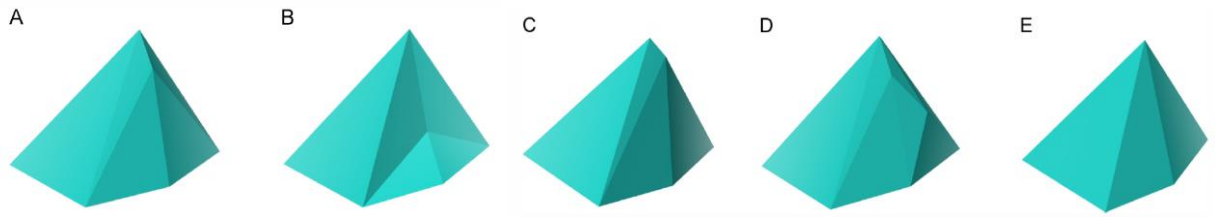

Figure 4: Illustration of different types of local surface curvature changes based on alpha shape analysis. (A) The original surface structure. (B) Convexity/concavity change. (C) Convexity/concavity degree change. (D) Addition of another atom. (E) Disappearance of an atom.

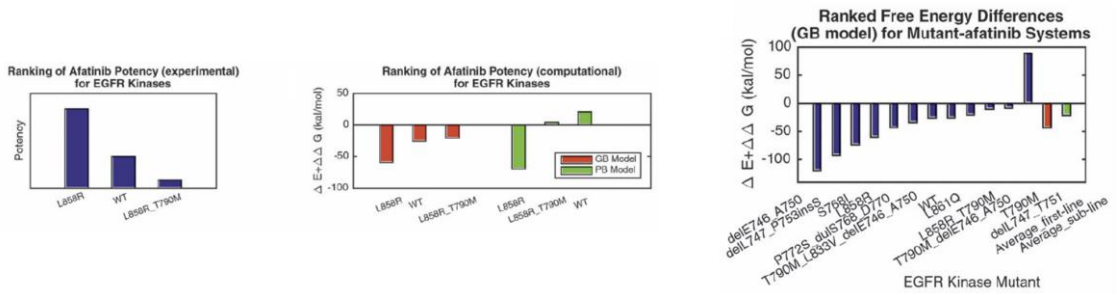

Figure 5: Ranking of afatinib potency for EGFR kinases from experimental (left) and computational (middle and right) studies.

Characterisation of EGFR and ErbB-3 heterodimerization: Based on the data from 168 clinical subjects, we characterized their corresponding EGFR mutations using molecular interactions, with three potential dimerization partners (ErbB-2, IGF-1R and c-Met) of EGFR and two of its small molecule inhibitors (gefitinib and erlotinib) (Figure 6) (Publication 4). We comparably examined the interactions between ErbB-3 and its partners (EGFR mutants, IGF-1R, ErbB-2 and c-Met). Compared to others, c-Met shows a remarkably-strong binding with ErbB-3, implying its significant role in regulating ErbB-3 signaling. Moreover, EGFR mutants corresponding to poor clinical outcomes, such as L858R_T790M, possess lower binding affinities with ErbB-3 than c-Met does. This may promote the communication between ErbB-3 and c-Met in these cancer cells. The analysis verified the important contribution of IGF-1R or c-Met in the drug resistance mechanism developed in lung cancer treatments.

Discussion

This project is carried out with close collaborations among medical doctors, chemists and computer scientists. Medical doctors from Queen Mary Hospital have provided much needed clinical data. Although a large number of EGFR mutations are well known and have been published, there are also many so-called "rare" mutations unique to Hong Kong patients. It would be too expensive to study all the mutations through biological experiments. In this project, we model EGFR mutants and their interactions with drug using computational methods (Publications 1 to 12). Our approach is more flexible and can model almost any EGFR mutations and mutant/drug complexes.

Another advantage of using locally collected clinical data is that we have patient information. A patient's physical status, response level to a drug and survival time are all useful for the assessment of drug effectiveness. These patient data are also valuable for computer algorithm design and verification. For example, we can evaluate the correlation between the EGFR/inhibitor binding free

energy and the patient survival time to determine whether the mathematical model using the energy measurement is effective.

One might argue that drug resistance levels can be evaluated directly from patient data. Indeed, we may be able to know which mutation can cause short or long patient survival time. However, this approach is meaningful only if we have a large of number patients for each mutation so that the result is statistically significant. In practice, we may just have one or two cases for a rare mutation. In addition, the direct observations do not provide any insight into the physical reasons of drug resistance. Using our computational method, we can evaluate the change in the protein/drug binding strength and identify surface curvature variations, which can lead to the degradation of the drug effectiveness.

The focus of our work in this project is to study the resistance of NSCLC patients to the first generation (reversible) drugs. When we submitted the application for this project, the second generation (irreversible) drugs were still under clinical trials. Although we also did some work on second generation drugs, we had very limited clinical data and much more research would be needed in the future. Now the third generation drugs are available to overcome the weak selectivity of the second generation drugs. However, there are also reports that there can be resistance to these new drugs as well. Thus, there is a long battle to fight for NSCLC drug resistance.[4-8]

In this project, we have studied NSCLC drug resistance based on molecular modelling. The computer algorithms can also be extended to the investigation of other diseases if drug resistance can arise from protein mutations.

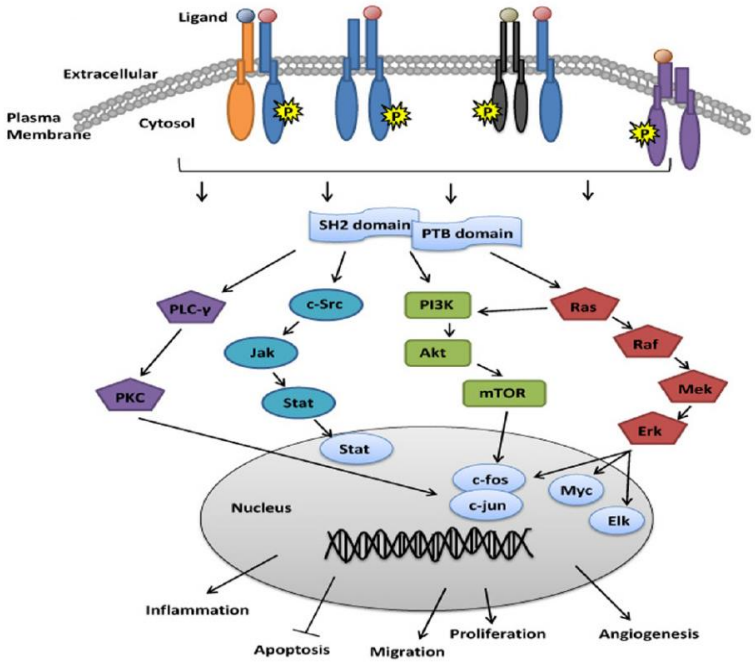

Figure 6: EGRF related dimers and activation and transduction of downstream signals.

Limitations

As discussed above, our computational analysis can be used to deal with almost any EGFR mutation because we can predict the 3D structures of a mutant. The approach is flexible and does not require expensive wet-lab biological experiments. However, our method also has several limitations:

The prediction of the 3D structure of an EGFR-drug complex may have a limited accuracy. This prediction is based on non-linear numerical optimisation and the procedure can be easily trapped to a local optimal solution. We may need to compute the 3D structure multiple times with different

starting points and different parameter settings and select the best result. Even so, the final solution may not be the globally optimal one. With better mathematical models and faster computers and more robust algorithms, we shall be able to improve the accuracy of computational work significantly in our future projects.

We have only observed "rare" EGFR mutations locally in Hong Kong. There can be many other mutations that we have not discovered so far. To overcome this limitation, we can collaborate with hospitals in mainland China and other countries/regions, where there are much more patients and a larger mutation database can be built.

In this project, we have only analysed EGFR mutations as a cause of NSCLC drug resistance. Along the EGFR related signalling pathways, mutations of other proteins can also lead to drug resistance. An investigation of all these protein mutations require much more work, including obtaining gene sequence data from patients, modelling the 3D structures of these mutants, and analysing the EGFR related protein-protein interactions and signalling network (Publication 12).

Conclusions

In this project, we have investigated mathematical models and computational methods for the study of anti-EGFR drug resistance at the molecular and atomic levels. Our project team includes medical doctors, chemists, and computer scientists and engineers. NSCLC can be caused by an EGFR mutation and another mutation can lead to drug resistance. We have collected EGFR mutation data based on reports in the literature and from local hospitals. Many of the mutations observed locally are "rare" ones and are unique to the Hong Kong population. Thus, our research work in this project is especially useful for Hong Kong to help the treatment of NSCLC patients.

It would be too expensive to determine the 3D structures of all EGFR mutants (over 100 in our database) through wet-lab experiments. To solve this problem, we have developed computer methods to predict the 3D structures of EGFR mutants and their interaction complexes with NSCLC drugs based on molecular modelling and numerical optimisation (Publication 1). Starting from the 3D structure of wild type EGFR, which can found in PDB, we build and optimise the structure of each EGFGR mutant and drug complex. From this structure, we compute the binding affinity between EGFR and the drug. A reduction in the binding strength due to EGFR mutation reflects the occurrence of drug resistance. Thus, our method can provide a quantitative measure of drug resistance level.

We have built and published the web-based "EGFR Mutant Structural Database", which is available on online to the public (http://bcc.ee.cityu.edu.hk/SFBG/research_and_publications.html) (Publication 2). This database provides information on EGFR mutation, computer predicted 3D structure of each EGFR mutant-gefitinib/erlotinib complex, and the drug binding affinity. Once the mutation of EGFR is known, one can find relevant entries in the database and determine the drug resistance level. Currently the database contains 112 EGFR mutants reported in the literature or observed locally. New entries can be added if additional mutations are found clinically.

Protein related biomolecular interactions often take places through surface contact. A high binding affinity often involves the match of a convex shape with a concave one. In this project, we have studied the surface characteristics of EGFR mutants (Publication 3). A so-called 3D alpha shape model is used to reconstruct the outer surface of an EGFR mutant. At each atom on the surface, we can evaluate the curvature in terms of convexity or concavity. An important reason of drug resistance is the change in curvature. For example, a convex shape can become convex after EGFR mutation and reduce the drug binding strength. Our studies provide the root reasons of drug resistance at the atomic level.

In addition to the 3D structures and binding patterns between EGFR mutants and the first generation drugs gefitinib and erlotinib, which are the main objectives of this project, we have also investigated the effectiveness of the second generation drug afatinib (Publication 10). Furthermore, we have analysed the hetero-dimer formation involving other proteins, such as IGF-1R, ErbB-2 and c-Met, along the EGFR related signalling pathways (Publication 4). Much more work is needed in these areas once more clinical data are available for new drugs, including the second and third generation ones, and for the mutations in other proteins. These problems will be studied in our future projects.

The research results of this project provide deep insight to the fundamental reasons of NSCLC drug resistance in terms of molecular structures and EGFR mutant surface characteristics. The findings, especially the EGFR Mutant Structural Database, provide a useful reference for medical doctors to plan personal treatment of a patient. The computational framework we have developed in the project is also useful for the study of other diseases.

Implications

In addition to clinical observations and wet-lab experiments, computational techniques can provide powerful tools to study drug resistance. With computer methods, we can often reduce the need and expenses of lab experiments. In our approach, we can model almost any protein mutation and compute the binding strength between the protein mutant and a drug. We study NSCLC drug resistance at the molecular and atomic levels, thus the root cause can be investigated. Although we have focused on the first generation drugs in this project, similar principles and the mathematical and computational frameworks can be employed to study new drugs as well.

Our computational results can provide a good reference to medical doctors for optimal drug selection and personalised treatment planning. Because we can study all combinations of EGFR mutations with all drugs, a doctor can easily assess the drug resistance level and design the best therapy for every patient by looking up the numerical results we have produced. In addition, from the EGFR mutant and drug binding strength, the doctor can estimate the survival time of a patient. Such information is useful for the patient as well as the hospital and help them use the correct medicine and save the medical cost.

Much research is still needed to study the NSCLC drug resistance problem. Several suggested topics are listed below:

- Collection of mutation data for EGFR and proteins in related signalling pathways locally and in other cities/countries
- Study of the mechanisms of resistance to second, third and newer generations of drugs at molecular and atomic levels
- Analysis of EGFR related pathways and network on how EGFR and other protein mutations affect drug resistance
- Identify possible multiple pathways to block in order to combat drug resistance

Dissemination

Our research findings have already been published in a number of scientific journals. A list of our publications is provide below.

We have built a website (http://bcc.ee.cityu.edu.hk/SFBG) to present our research results. The website shows our research topics and outcomes. In particular, we publish our "EGFR Mutant Structural Database" online. The reader can download the optimised 3D structures of EGFR mutant and first generation drug complexes.

We are currently also discussing the possibility of collaborating with a local biotech company on the analysis of resistance to the second and third generation drugs based on computational methods.

Publications

1. Wang DD, Zhou W, Yan H, Wong M, Lee V. Personalized prediction of EGFR mutation-induced drug resistance in lung cancer. Scientific Reports. 2013;3:2855.
2. Ma L, Wang DD, Huang Y, Yan H, Wong MP, Lee VHF. EGFR Mutant Structural Database: computationally predicted 3D structures and the corresponding binding free energies with gefitinib and erlotinib. BMC Bioinformatics. 2015;16:85.
3. Ma L, Wang DD, Huang Y, Wong MP, Lee VHF, Yan H, Decoding the EGFR mutation-induced drug resistance in lung cancer treatment by local surface geometric properties. Computers in Biology and Medicine. 2015;63:293-300.
4. Wang DD, Ma L, Wong MP, Lee VHF, Yan H. Contribution of EGFR and ErbB-3 heterodimerization to the EGFR mutation-induced gefitinib- and erlotinib-resistance in non-small-cell lung carcinoma treatment. PLoS ONE. 2015;10(5):e0128360.
5. Duan B, Zou B, Wang DD, Yan H, Han L. Computational evaluation of EGFR dynamic characteristics in mutation-induced drug resistance prediction. Proceedings of IEEE International Conference on Systems, Man, and Cybernetics. 2015;2299-2304.
6. Zou B, Wang DD, Ma L, Chen L, Yan H. Analysis of the relationship between lung cancer drug response level and atom connectivity dynamics based on trimmed Delaunay triangulation. Chemical Physics Letters, 2016;652:117-122.
7. Shen Z, Wang DD, Ma L, Yan H, Wong MP, Lee VHF. A computational study of three frequent mutations of EGFR and their effects on protein dimer formation and non-small cell lung cancer drug resistance. Current Bioinformatics. 2016;11(3):382-391.
8. Ma L, Zou B, Yan H. Identifying EGFR mutation-induced drug resistance based on alpha shape model analysis of the dynamics. Proteome Science. 2016;14:12.
9. Ma L, Wang DD, Zou B, Yan H. An eigen-binding site based method for the analysis of anti-EGFR drug resistance in lung cancer treatment. IEEE/ACM Transactions on Computational Biology and Bioinformatics. In press.
10. Wang DD, Lee VHF, Zhu G, Zou B, Ma L, Yan H. Selectivity profile of afatinib for EGFR-mutated non-small-cell lung cancer. Molecular BioSystems. 2016;12:1552-1563.
11. Ma L, Zou B, Yan H. Identifying EGFR mutation-induced drug resistance based on alpha shape model analysis of the dynamics. Proteome Science. 2016;14:12.
12. Zhang X, Ou-Yang L, Yan H. Incorporating prior information into differential network analysis using nonparanormal graphical models. Bioinformatics. In press.

References

1. Stewart BW, Wild CP (eds.) World Cancer Report 2014. WHO Press, 2014.
2. Oda K, Matsuoka Y, Funahashi A, Kitano H. A comprehensive pathway map of epidermal growth factor receptor signalling. Molecular Systems Biology. 2005;0010.
3. Haley JD, Gullick WJ. Cancer Drug Discovery and Development: EGFR Signaling Networks in Cancer Therapy. Springer, 2008.
4. Yarden Y, Pines G. The ERBB network: at last, cancer therapy meets systems biology. Nature Reviews Cancer. 2012;12:553-563.
5. Singh I, Amin H, Rah B, Goswami A. Targeting EGFR and IGF 1R: a promising combination therapy for metastatic cancer. Frontiers in Bioscience. 2013;5:231-246.
6. Kwapiszewski R, Pawlak SD, Adamkiewicz K. Anti-EGFR agents: current status, forecasts and future directions. Targeted Oncology. 2016;11(6):739-752.
7. Temraz S, Mukherji D, Shamseddine A. Dual targeting of HER3 and EGFR in colorectal tumors might overcome anti-EGFR resistance. Critical Reviews in Oncology/Hematology. 2016;101:151-157.

8. Liao BC, Lin CC, Yang JCH, Novel EGFR inhibitors in non-small cell lung cancer: current status of Afatinib. Current Oncology Reports. 2017;19:4.

List of Research Workers

Project investigators: Hong Yan, Guangyu Zhu, Maria Wong, Victor Lee

Postgraduate students: Debby Wang, Lichun Ma, Long Chen, Bin Zou

Research assistants: Yiqing Huang, Baobin Duan, Zhiyun Shen, Le Ou-Yang, Xiaofei Zhang, Hu Zhu, Lizhu Sun, Yi Liang, Guxia Xu

Acknowledgements

This project is supported by Hong Kong Health and Medical Research Fund (Project 01121986). We also would like to thank the staff at the Signal Processing and Biocomputing Lab at City University of Hong Kong for their help with computer hardware and software and maintenance of our website.