SCIENTIFIC REPORTS

OPEN

Structure and energy based quantitative missense variant effect analysis provides insights into drug resistance mechanisms of anaplastic lymphoma kinase mutations

Received: 2 February 2018
Accepted: 29 June 2018
Published online: 13 July 2018

Jianzong Li[1], Yue Huang[1,2], Miaomiao Wu[1], Chuanfang Wu[1], Xin Li[3] & Jinku Bao[1,3,4]

Anaplastic lymphoma kinase (ALK) is considered as a validated molecular target in multiple malignancies, such as non-small cell lung cancer (NSCLC). However, the effectiveness of molecularly targeted therapies using ALK inhibitors is almost universally limited by drug resistance. Drug resistance to molecularly targeted therapies has now become a major obstacle to effective cancer treatment and personalized medicine. It is of particular importance to provide an improved understanding on the mechanisms of resistance of ALK inhibitors, thus rational new therapeutic strategies can be developed to combat resistance. We used state-of-the-art computational approaches to systematically explore the mutational effects of ALK mutations on drug resistance properties. We found the activation of ALK was increased by substitution with destabilizing mutations, creating the capacity to confer drug resistance to inhibitors. In addition, results implied that evolutionary constraints might affect the drug resistance properties. Moreover, an extensive profile of drugs against ALK mutations was constructed to give better understanding of the mechanism of drug resistance based on structural transitions and energetic variation. Our work hopes to provide an up-to-date mechanistic framework for understanding the mechanisms of drug resistance induced by ALK mutations, thus tailor treatment decisions after the emergence of resistance in ALK-dependent diseases.

Anaplastic lymphoma kinase (ALK), a member of the superfamily of insulin receptor protein-tyrosine kinases, was characterized by the identification of a 2;5 chromosomal translocations in anaplastic large-cell lymphoma (ALCL) cell line[1]. This chromosomal rearrangement generates nucleophosmin (NPM)-ALK fusion protein that has a constitutively activated ALK kinase domain[2]. In addition to NPM-ALK fusion protein, the echinoderm microtubule-associated protein-like 4 (EML4)–ALK fusion detected in NSCLC is the most widely identified[3]. It has been indicated that the ALK fusion proteins play an important role in driving tumorigenesis[2,3]. In contrast to fusion proteins, activation of the full-length ALK is normally regulated by extracellular ligand-binding domain. The full-length ALK consists of an extracellular ligand-binding domain (residues 19–1038), a transmembrane domain (residues1039–1059) and an intracellular tyrosine kinase domain (residues 1116–1392). Experimental genetic evidences indicate that mutated full-length ALK plays an important role in multiple carcinomas, such as neuroblastoma and thyroid cancer, but the mechanisms have not been illuminated very clearly[4–7].

[1]Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, 610065, Sichuan, P. R. China. [2]College of Computer Science, Sichuan University, Chengdu, 610207, Sichuan, P. R. China. [3]State Key Laboratory of Oral Diseases, National Clinical Research Center for Oral Diseases, West China Hospital of Stomatology, Sichuan University, Chengdu, China. [4]State Key Laboratory of Biotherapy/Collaborative Innovation Center for Biotherapy, West China Hospital, Sichuan University, Chengdu, 610041, China. Jianzong Li, Yue Huang and Miaomiao Wu contributed equally to this work. Correspondence and requests for materials should be addressed to X.L. (email: lixin0914071@126.com) or J.B. (email: baojinku@scu.edu.cn)

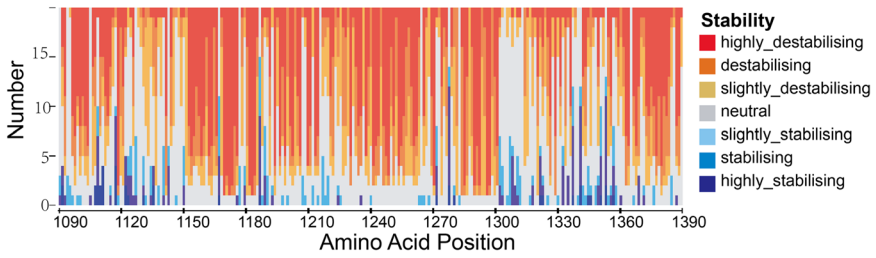

Figure 1. Effect of mutations on ALK kinase domain stability. All 19 possible mutations along with the synonymous mutations that the residue mutates to itself at each position in ALK kinase domain are colored on a vertical bar in terms of their stability relative to wide-type ALK. The values of $\Delta\Delta G_{fold}$ are binned into seven categories: highly stabilizing ($\Delta\Delta G_{fold} < -1.84\,kcal{\cdot}mol^{-1}$) and highly destabilizing ($\Delta\Delta G_{fold} > 1.84\,kcal{\cdot}mol^{-1}$); stabilizing ($-1.84\,kcal{\cdot}mol^{-1} < \Delta\Delta G_{fold} < -0.92\,kcal{\cdot}mol^{-1}$) and destabilizing ($0.92\,kcal{\cdot}mol^{-1} < \Delta\Delta G_{fold} < 1.84\,kcal{\cdot}mol^{-1}$); slightly stabilizing ($-0.92\,kcal{\cdot}mol^{-1} < \Delta\Delta G_{foldd} < -0.46\,kcal{\cdot}mol^{-1}$) and slightly destabilizing ($0.92\,kcal{\cdot}mol^{-1} < \Delta\Delta G_{fold} < 1.84\,kcal{\cdot}mol^{-1}$); and neutral ($-0.46\,kcal\,kcal{\cdot}mol^{-1} < \Delta\Delta G_{fold} < 0.46\,kcal{\cdot}mol^{-1}$).

ALK has been validated as a therapeutic molecular target for the treatment of ALK-rearranged cancer. Substantial efforts among academia and pharmaceutical industry have been made to develop effective ALK inhibitors. Nowadays, crizotinib, ceritinib and alectinib have been approved by the US Food and Drug Administration (US. FDA) for the treatment of patients with advanced "ALK-positive" NSCLC[8–13]. Considerable small-molecular inhibitors targeting ALK are currently in clinical trials, such as AP26113[14] and lorlatinib (PF-06463922)[15]. However, the rapid emergence of inevitable drug resistance is occurring worldwide, endangering the efficacy of chemotherapy involving these drugs. Generally, different ALK inhibitors actually result in the emergence of resistance to ALK inhibitors that is characterized by different mechanisms. Crizotinib is the first-generation ALK inhibitor, resistance to this drug occurs in patients who initially benefited from target therapies. It is reported that about one third of resistance cases are related to the diverse mutations in EML4-ALK fusion protein[16]. Acquired secondary ALK resistance mutations to the crizotinib include I115ITins, L1152P/R, C1156Y/T, I1171T/N/S, F1174C/L/V, V1180L, L1196M, G1202R, S1206C/Y, E1210K, or G1269A/S[17]. Ceritinib and alectinib are the second-generation ALK inhibitors that are developed to overcome the resistance to the first generation ALK inhibitors, but resistant mutation to these drugs are also inevitably reported. Resistant mutations to ceritinib include I115ITins L1152P/R, C1156Y/T, I1171T/N/S, F1174C/L/V, and G1202R. Resistant mutations to alectinib include I1171T/N/S and G1202R. Among which L1196M gatekeeper mutation is the most common resistance mutation to crizotinib[17–19]. The hotspot mutations F1174 (mutated to L, S, I, C or V) in ALK kinase domain are identified in about 85% of the cases with ALK mutations. G1202R is located at the solvent front of the ALK kinase domain and exhibits broad-spectrum resistance to all ALK inhibitors. There may be some other potential resistance harboring in primary ALK mutations. Although the functional research for these mutations are very limited, more and more experimental evidences show that they play an important role in tumorigenesis and may possess potential effects on ALK targeting therapy[20–22].

Numerous studies have been performed to dissect the mechanisms of drug resistance to ALK inhibitors[7,23–25]. It has been widely acknowledged that the drug-resistant mutations cause drug resistance by re-inducing kinase activation and signaling despite the presence of the inhibitors. These mutations can hinder the inhibitor binding to ALK, alter the kinase's conformation, and/or alter the ATP-binding affinity of the kinase[7,23,25]. It has been suggested that evolving paradigms exist in cancer drug resistance and contribute to the evolution process of tumor clones in response to the selection pressure by drug treatments[26,27]. Some interesting works have evaluated the influence of subtle mutations on the shifts of the energetics and function of proteins as well as specific stability-function tradeoffs in the evolution processes of enzymes[20]. Comprehensive quantitative profiling of mutational effects on drug resistance is increasingly important in tailoring treatment decisions after the emergence of resistance in ALK-depend diseases given the promise of novel molecular targeted therapies. Of particular interest is whether the mechanisms that mediate drug resistance in protein are caused by a few mutations occurring in distal regions. Here, we focus on exploring diverse mechanisms of drug resistance induced by secondary or primary mutations in ALK and hope to find some common shared features. In addition, a comprehensive profile of drugs against ALK mutations has been constructed to analyze the energetic variation mediated by mutations, thus providing insights into the molecular basis of drug resistance from a structural perspective. We show how an integration of structural transition and energy variation moves us towards a better understanding of why biological molecules have the properties that they do, such as the emergence of acquired drug resistance.

Results and Discussions

Stability landscape of all possible mutations in ALK kinase domain. In order to estimate the stability effect of all possible mutations of each residue of the ALK kinase domain, the stability changes ($\Delta\Delta G_{fold}$) were computed using FoldX. As shown in Fig. 1, most of the mutations (2730 of 4940 = 55.26%) have moderate thermodynamic stability effects, the values of $\Delta\Delta G_{fold}$ of these mutations range from $-1.84\,kcal{\cdot}mol^{-1}$ to $1.84\,kcal{\cdot}mol^{-1}$. 2170 mutations (43.93%) showed highly destabilizing ($\Delta\Delta G_{fold} > 1.84\,kcal{\cdot}mol^{-1}$). A handful of mutations are found to have strongly stabilizing effect on the structure of ALK kinase domain, only 40 mutations (0.81%) show that the values of $\Delta\Delta G_{fold}$ are less than $-1.84\,kcal{\cdot}mol^{-1}$, most of these mutations are located in or close to the kinase activation segment of ALK, the detailed data was provided in Supplementary Table S1.

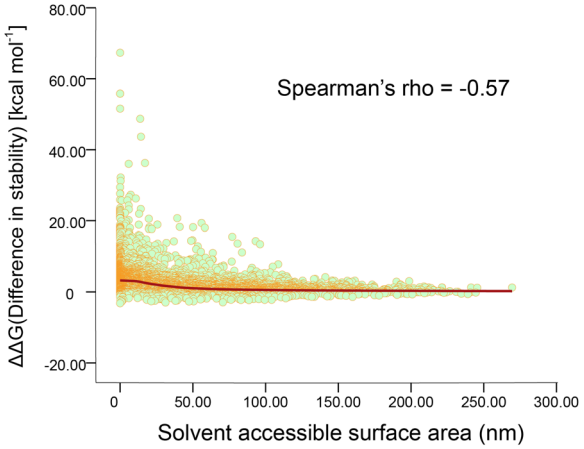

Figure 2. Scatter plot of the correlation between solvent accessible surface area (SASA) and difference in stability ($\Delta\Delta G_{fold}$) of ALK mutations.

Experimental evidences have demonstrated that residues in the active sites of the enzyme are characterized by intrinsic destabilizing due to the high energy state, and the replacement of these residues can actually increase the stability of the enzyme[28–30].

Interest in quantitative analysis of the correlation between the degree of surface exposure and stability effect of residues led us to further investigate the relevance of the stability effect of mutation and the solution accessible surface area (SASA) of each residue. Buried residues are not accessible to the solvent inside a cell, therefore their SASA usually equal zero or closes to zero. Here, SASA of each residue was calculated to estimate the surface exposure level using FreeSASA[31]. Then we simply used Spearman's rank correlation coefficient (Spearman's rho) to investigate how well the relationship between $\Delta\Delta G_{fold}$ and SASA could be described using a monotonic function. As shown in Fig. 2, the Spearman's rho equaled −0.57, indicating that moderate negative correlation existed between $\Delta\Delta G_{fold}$ and SASA. Highly destabilizing mutations ($\Delta\Delta G_{fold} > 1.8$ kcal·mol^{-1}) were more likely to occur in residues that possessed smaller SASA. If mutations occur in buried residues, their $\Delta\Delta G_{fold}$ can be very high. Mutations with larger SASA show smaller volatility of stability effects, and these mutations are usually neutral or slightly destabilizing.

It has been well known that surface residues are not important for protein stability since their interactions with the solvent should be similar in the unfolded and native states[32]. Here, results showed that the changes in stability mediated by the mutation were related to the solvent-accessible area of residues. For example, surface residue Met1290 possessed highest SASA (208.26 nm), and mutations occurring in this position had moderate effects on protein stability, most of their $\Delta\Delta G_{fold}$were lower than 0.92 kcal·mol^{-1}, only M1290V was highly destabilizing (3.79 kcal·mol^{-1}). The SASA of surface residue Lys1114 was 172.55 nm, the changes in stability mediated by the mutation in this position were also insignificant, the average of $\Delta\Delta G_{fold}$ value in this position was 0.17 kcal·mol^{-1}, no mutation occurring in the site was stabilizing or destabilizing. Once residues possess SASA over 70 nm, highly destabilizing or stabilizing mutations are unlikely to occur in these sites (Supplementary Table S2). Results reveal that mutations occur in exterior regions have moderate effects on protein stability, while mutations that occurr in interior regions with low-level surface exposure are more likely to be destabilizing.

Effects of observed ALK mutations on protein evolution and drug resistance. Theoretically, there may be incalculable possible mutational forms in processes of evolution, but ALK has to satisfy diverse constraints, including reasonable folding, thermodynamic stability and solubility, to maintain its specific function[29,33,34]. Only few ALK mutations are compatible with the constraints. Here, 145 observed mutations occurring in ALK kinase domain were collected from Uniprot[35], COSMIC[36] and ClinVar[37] database to investigate the stability-function tradeoffs followed by these mutations and their evolutionary features underlying drug treatment condition. Most of observed mutations are associated with diseases and play an important role in the pathogenesis of neuroblastoma, anaplastic thyroidcarcinoma and NSCLC[20–22,38,39], 131 mutations are the original primary mutation in neuroblastoma and lung adenocarcinoma cell lines, and 14 mutations are acquired as secondary mutations occurring independently or dependently in individual NSCLC cell line (Supplementary Table S3).

The distribution of the $\Delta\Delta G_{fold}$ values of all possible mutations of ALK kinase domain and observed mutations occurring during the evolution, or rather, tumorigenesis was computed. As can be seen in Fig. 3, the distribution of stability effects of all these possible mutations is unimodal and has its largest peak close to a $\Delta\Delta G_{fold}$ of zero, suggesting that the simulated mutations appear to be neutral or have moderate effects on the protein thermodynamic stability. In contrast to all possible mutations, the distribution of the $\Delta\Delta G_{fold}$ values of observed mutations follows a multimodal distribution and has an excess of slight destabilizing than possible mutations. As can be seen, it has two slight distinct peaks including a peak of high destabilizing whose values centered at 3.12 kcal·mol^{-1} and a smaller peak close to zero.

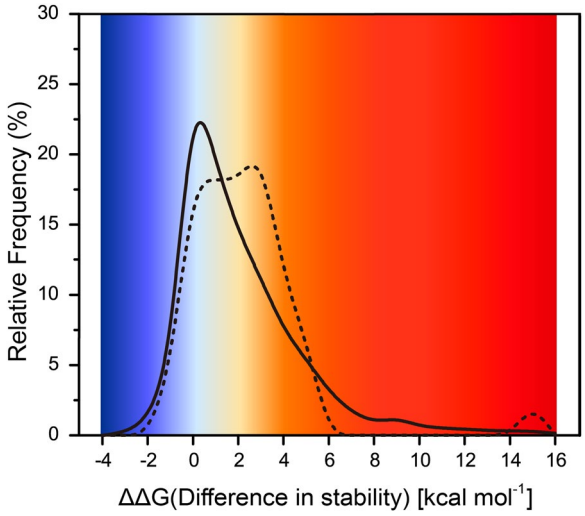

Figure 3. Distribution of stability effects of all possible simulated mutations and those observed in clinic. The distribution of stability changes arising from mutations observed in tumor cell (dashed line) stands in contrast to that of all possible simulated mutations (solid line). The probability distributions shown here are obtained by kernel smoothing of the original data

Although highly destabilizing mutations are unlikely to occur in nature conditions, mutations that observed in clinic appear to be highly destabilizing. The stability-function tradeoffs hypothesis suggests that substitution of key residues in enzymes decreases the stability, resulting in the increasing activity of enzyme[40]. Studies of enzyme steady-state kinetic parameters of ALK showed that F1174L mutant attained 86% of the activity of the phosphorylated ALK, and R1275Q mutant attained 28% of the activity of the phosphorylated ALK[41]. The $\Delta\Delta G_{fold}$ values of F1174L mutant was 1.40 kcal·mol^{-1} and the $\Delta\Delta G_{fold}$ values of R1275Q mutant was 1.90 kcal·mol^{-1}. In addition, mutations occurring in Fhe1174 loci (mutated to Ile, Leu, Ser or Val) also attained higher $\Delta\Delta G_{fold}$ values, they were able to relax the steric clash with F1245 occurring in the catalytic loop and thus facilitated activation[22,23]. These mutations have been described in several diseases as mentioned above, as well as drug resistance in NSCLC. Our results are consistent with the "stability-function" hypothesis[42]. Mutations occurring within or near key regulatory regions can dramatically increase the activity of the enzyme owing to the intervention of destabilizing mutations.

Furthermore, the $\Delta\Delta G_{fold}$ values of total 21 reported drug resistance mutations were investigated here. In particular, we found that ALK drug resistance mutations presented significantly more destabilizing, most of which attained higher $\Delta\Delta G_{fold}$ values than 1.84 kcal·mol^{-1}. They were the main forces to form the peak of high destabilizing whose values centered at 3.12 kcal·mol^{-1} in the distribution of stability effects. Their $\Delta\Delta G_{fold}$ values ranged from −1.02 kcal·mol^{-1} (E1210K/S1206C) to 11.06 kcal·mol^{-1} (L1152P). Mean of the $\Delta\Delta G_{fold}$ values was 3.03 kcal·mol^{-1} and the median was 3.07 kcal·mol^{-1}, see Table 1. Rare drug resistance mutations are stabilizing, only four mutations can slightly stabilize the structure, including the gatekeeper L1196M mutation (−0.16 kcal·mol^{-1}), L1198F mutation (−1.19 kcal·mol^{-1}) and two double mutations E1210K/S1206C and L1198F/C1156Y (−0.26 kcal·mol^{-1}). The gatekeeper mutation L1196M is the most common crizotinib-resistance mutant identified in NSCLC. The replacement of leucine by methionine, which has comparable stabilizing effects on the ALK structure as well as identical electronic properties and steric demands. Our previous study showed that the loss of a non-polar binding energy interaction between the leucine side chain and the methyl substituent of crizotinib might account for the resistance against crizitinib when leucine was mutated to a methionine[43]. The mechanisms of which other drug resistance mutations facilitate activation of ALK in the context of drug treatment conditions are currently unclear. ALK inhibitors approved by the US. FDA are all type I1/2 inhibitors that bind to an inactive protein kinases with the DFG-in conformation[44,45], including cizotinib, ceritinib and alectinib. It has been demonstrated that mutations could facilitate the transition between inactive and active conformations and increased the stability of the active conformation. Here, our analysis suggests that the activation of ALK is increased by the substitution with destabilizing mutations, creating the capacity to confer drug resistance inhibitors.

Interestingly, S1206C mutant has been reported as crizotinib-resistant mutation, but the double ALK mutant (S1206C/E1210K) does not show resistance to crizotinib anymore. The resistance to ALK inhibitors shifts from crizotinib to brigatinib, a new ALK inhibitor that is currently under clinical testing. C1156Y mutant has been reported as crizotinib-resistant mutation and ceritinib-resistant mutation, but the double mutant (L1198F/C1156Y) confers resistance to lorlatinib and causes sensitization to crizotinib[19]. C1156Y, D1203N and S1206C are highly destabilizing mutations underlying our results, L1198F and E1210K are stabilizing mutations that can increase the thermodynamic stability of ALK, their $\Delta\Delta G_{fold}$ values are −1.19 kcal·mol^{-1} and −0.46 kcal·mol^{-1}, respectively. It has been demonstrated that the maintenance of the protein's stability is a strong constraint to evolution[46], drug resistance mutations are actually subject to the constraints[47]. In the scenario of double mutations S1206C/E1210K, C1156Y/L1198F and D1203N/E1210K, the stability effects of high destabilizing mutations on ALK structure are constantly pruned by the stabilizing mutations.

Mutation		$\Delta\Delta G_{fold}$ kcal·mol^{-1}	±SD	ΔE	ALK inhibitors[a]				
					Crizotinib	Ceritinib	Alectinib	Brigatinib	Lorlatinib
Primary mutation	L1152P	11.06	0.024	−5.84	+	+			
	I1171T	3.07	0.026	−2.72	+		+		
	I1171N	3.18	0.030	−4.94	+		+		
	I1171S	2.83	0.009	−4.54	+		+		
	F1174C	4.22	0.027	−6.24	+	+			
	F1174S	5.14	0.069	−7.26	+	+			
	V1180L	0.96	0.027	−3.98	+		+		
	L1198F	−1.19	0.024	−4.14					+
	S1206C	3.23	0.350	−5.28	+				
Secondary mutation	L1152R	3.17	0.409	−4.86	+	+			
	C1156Y	4.47	0.634	−4.41	+				
	F1174L	1.36	0.185	−2.54	+	+			
	L1196M	−0.16	0.039	−4.82	+				
	L1198P	1.46	0.058	−7.18	+				
	G1202R	2.88	0.140	−5.76	+	+	+	+	
	D1203N	3.85	0.149	−4.46	+				
	S1206Y	2.72	0.236	−5.64	+				
	G1269A	0.52	0.376	−1.41	+				
	G1269S	2.92	0.556	−1.72	+				
	E1210K/S1206C	−1.02	0.273	−8.25				+	
	E1210K/D1203N	0.90	0.215	−7.78				+	
	L1198F/C1156Y	−0.26	0.449	−9.17					+

Table 1. Effects of reported resistance mutations to the ALK inhibitors approved by the US. FDA or in clinical testing. [a]Plus sign indicates that the mutation is reported resistance mutation to corresponding ALK inhibitor.

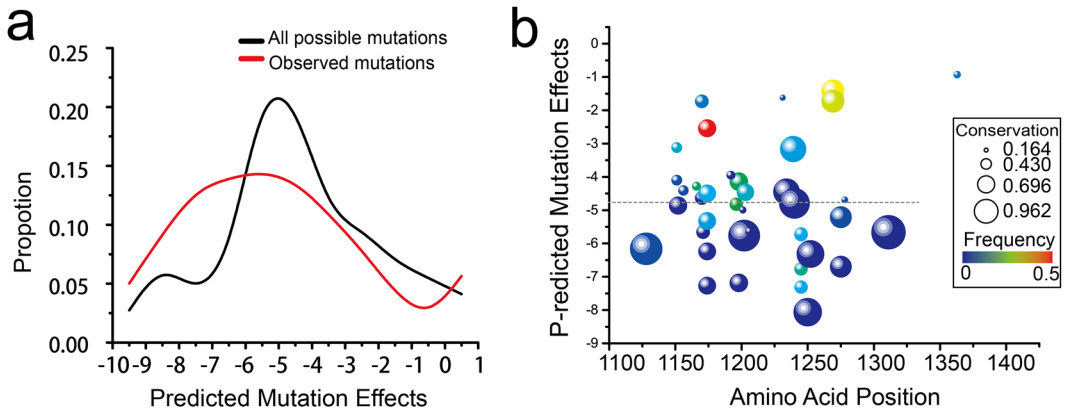

Figure 4. The distribution of mutational effects (ΔE) of ALK mutations. (**a**) The distribution of mutation effect scores of all possible mutations and observed mutation calculated by Evmutation software. (**c**) Scatter plot of mutation effects of pathogenic mutations. The dash line indicates the median value of these scores ($\Delta E = -4.86$). The mutation effects shown here are obtained by bubble and color mapped of the original data.

Accounting for the evolutionary forces that may cause drug resistance can be important for providing reasonable treatment strategies, we employed computational tools to estimate whether resistance mutations in the kinase domain were likely to be benign or damaging using comparative evolutionary considerations. The estimations of the mutation effects (ΔE) of ALK mutations were presented in Fig. 4, the detailed data were provided in Supplementary Table S4. The distributions of mutation effect scores of all possible and observed mutations showed similar variation tendency, the largest peak of ΔE values of the observed mutations were slightly lower than that of all possible mutations. Most observed mutations are identified as pathogenic or likely pathogenic mutations. Only 5 are mutations are not specified, their effects on phenotypes are unclear. The median value of the pathogenic mutations' effect scores is −4.86 taking into account of epistasis, indicating that most of the observed mutations in ALK kinase domain are possibly damaging. The ΔE values of drug resistance mutations ranged from −7.26 (F1174S) to −1.41 (G1269A), the median value was −4.84 and the mean value of the ΔE was −4.64, see Table 1. Results reveals that disease-associated drug resistance mutations are not likely to be tolerated

in the kinase domain of ALK. However, current paradigm of drug resistance mutations indicates that mutations occur prior to treatment are likely to be benign or natural, thus they can escape from the fixing of evolutionary constraints[48]. We further examined whether secondary drug resistance mutations were more likely to be damaging than primary mutations in the ALK kinase domain. We found their ΔE values were very comparable, no remarkable distinction was observed. In the scenario of double mutations, their ΔE values were all higher than that of the constituent single mutations, see Table 1. Although double mutations attain moderate effects on the thermodynamic stability of ALK kinase domain, their effects on the protein fitness are likely to be possibly damaging. Taken together, our results imply that the evolution constraint limits the adaptation of ALK to drug treatment contexts through additional amino acid changes that can shift the drug resistance properties without dramatic effects on the overall thermodynamic stability, resulting in more deleterious effects and eventually failure of novel drug treatments.

Comprehensive profiles for ALK inhibitors against drug resistance mutations. To characterize the effect of drug resistance mutations on ALK inhibitor binding, we built a systematic extensive energy profile of drugs against ALK drug resistance mutations by means of MM/QM-GBSA calculation. In order to test the reliability of different schemes in calculating inhibitor binding affinity change of ALK mutations, 15 pairs of the experimental mutation energies ($\Delta\Delta G_{FR}$ *experimental*) and predicted mutation energies ($\Delta\Delta G_{FR}$ *predicted*) were utilized. $\Delta\Delta G_{FR}$ *experimental* can be estimated by the formula: $\Delta\Delta G_{FR}$ *experimental* = RT·ln·(K_i *mutant*/K_i *wildtype*), where $R = 1.9872 \times 10^{-3}$ kcal·K^{-1}·mol^{-1}, and T = 300 K[49,50]. The experimental data of K_i *mutant* and K_i *wildtype* were collected from studies contributed by Alice T. Shaw and co-workers[19]. Five different schemes were used to calculate the values of $\Delta\Delta G_{FR}$ *predicted* (Supplementary Table S5). $\Delta\Delta G_{FR}$ *predicted* can be derived utilizing the binding free energy of the inhibitor against the ALK mutations (ΔG_{mt}) minus that of wild type (ΔG_{wt}). The correlation between $\Delta\Delta G_{FR}$ *experimental* and $\Delta\Delta G_{FR}$ *Predicted* can be reflected by coefficient of determination (R^2). The closer R^2 approaches 1, the better positive correlation that the two variables possess, thus indicating the better results that the calculating scheme has. As shown in Fig. 5a, the green points were the results calculated by the MM-GBSA method, while the red points stood for the results calculated by MM-PBSA, which both used the molecular mechanics (MM) theory. The R^2 of MM-GBSA was 0.601, while that of MM-PBSA was 0.635, indicating that the latter scheme had the better prediction for the values of $\Delta\Delta G_{FR}$ *predicted* as they were more matched with $\Delta\Delta G_{FR}$ *experimental*. The semi-experiential Hamiltonian methods based on QM/MM-GBSA theory were used here, including AM1, RM1 and PM6[51–53]. As shown in Fig. 5, the R^2 of these schemes were 0.874 (AM1), 0.842 (RM1) and 0.588 (PM6), respectively, indicating that the AM1 scheme possessed the best performance to predict the inhibitor binding affinity change among the five schemes. Thus, by utilizing the AM1 method for calculation, considerable reliability was assured for the further analysis.

Range of binding free energies (ΔG_{mt}) of the 10 ALK inhibitors along with ATP against 21 reported drug resistance mutations were shown in the box-plot (Fig. 6). Obviously, the averages and medians of the ΔG_{mt} values of the 10 inhibitors were way lower than those of ATP (average as −18.07 kcal·mol^{-1}, median as 15.92 kcal·mol^{-1}). Among the inhibitors, the first-generation ALK inhibitor crizotinib had the overall minimal binding affinity (with the average ΔG_{mt} as −35.74 kcal·mol^{-1} and the median ΔG_{mt} as −35.05 kcal·mol^{-1}) against ALK as it was the first small molecule inhibitor approved by US. FDA for the treatment of ALK positive NSCLC and considerable acquired resistances acted upon it. The second-generation inhibitors include alectinib and ceritinib, which both manifested better binding affinity than crizotinib. Lorlatinib, AP26113, ASP3026, belizatinib, CEP-37440, ensartinib and entrectinib all belong to the third-generation inhibitors, of which entrectinib had the overall maximal binding affinity (with the average ΔG_{mt} as −49.07 kcal·mol^{-1} and the median ΔG_{mt} as −49.58 kcal·mol^{-1}) among all 10 inhibitors, suggesting it may be a promising inhibitor against the drug resistance mutations of ALK kinase domain. Generally, the second-generation and the third-generation inhibitors had exhibited better binding affinity than the first-generation inhibitor.

The $\Delta\Delta G_{FR}$ *predicted* of the 10 inhibitors as well as ATP against 21 drug resistance mutations in ALK kinase domain were calculated by the QM/MM-GBSA AM1 scheme (Supplementary Table S6). The results were shown as the heat map in Fig. 7. As can be seen in Fig. 7, the majority of the $\Delta\Delta G_{FR}$ *predicted* values stayed within the range from −6 kcal·mol^{-1} to 6 kcal·mol^{-1} while a few $\Delta\Delta G_{FR}$ *predicted* had considerably great absolute values, indicating the violent affinity changes of these mutations compared with the wild type. The minimum of $\Delta\Delta G_{FR}$ *predicted* occurred at the mutation energy of the inhibitor belizatinib against the double-point mutation L1198F and C1156Y, suggesting that the mutation gained much better affinity towards the inhibitor compared to the wild type. The maximum of $\Delta\Delta G_{FR}$ *predicted* occurred at the mutation energy of the inhibitor ceritinib against the single-point mutation L1152P, indicating that ceritinib bound to the mutation much less closely than to the wild-type ALK and strong drug resistance may occur.

As for the first-generation inhibitor crizotinib, the first drug resistance reported was the L1196M gatekeeper mutation[54]. It alters the residue at the bottom of the ATP-binding pocket and impairs inhibitor binding[25]. $\Delta\Delta G_{FR}$ predicted of crizotinib against L1196M was 8.59 kcal·mol^{-1}, suggesting that crizotinib had lost considerable affinity towards the ALK kinase domain after ALK mutated. The loss of a non-polar binding energy interaction (ΔG_{surf}) might account for the resistance against crizitinib when leucine was mutated to a methionine[43]. Here, binding free energy of crizotinib against L1196M calculated by QM/MM-GBSA demonstrated that electrostatic interaction (ΔE_{ele} was more reduced than ΔG_{surf} (Supplementary Table S7). G1202R is highly resistant to first- and second-generation inhibitors[25] and it also occurs in the ATP-binding pocket, see Fig. 8. In the mutation, glycine was replaced with arginine, causing the cleft becoming narrower and hindering the binding of crizotinib with increased steric clash. The $\Delta\Delta G_{FR}$ predicted values of crizotinib against G1202R was 10.08 kcal·mol^{-1}, suggesting that considerable binding affinity was lost. The loss of van der Waals interaction energy (ΔE_{vdw}), about 8.78 kcal·mol^{-1}, was the dominant factor in causing such variation.

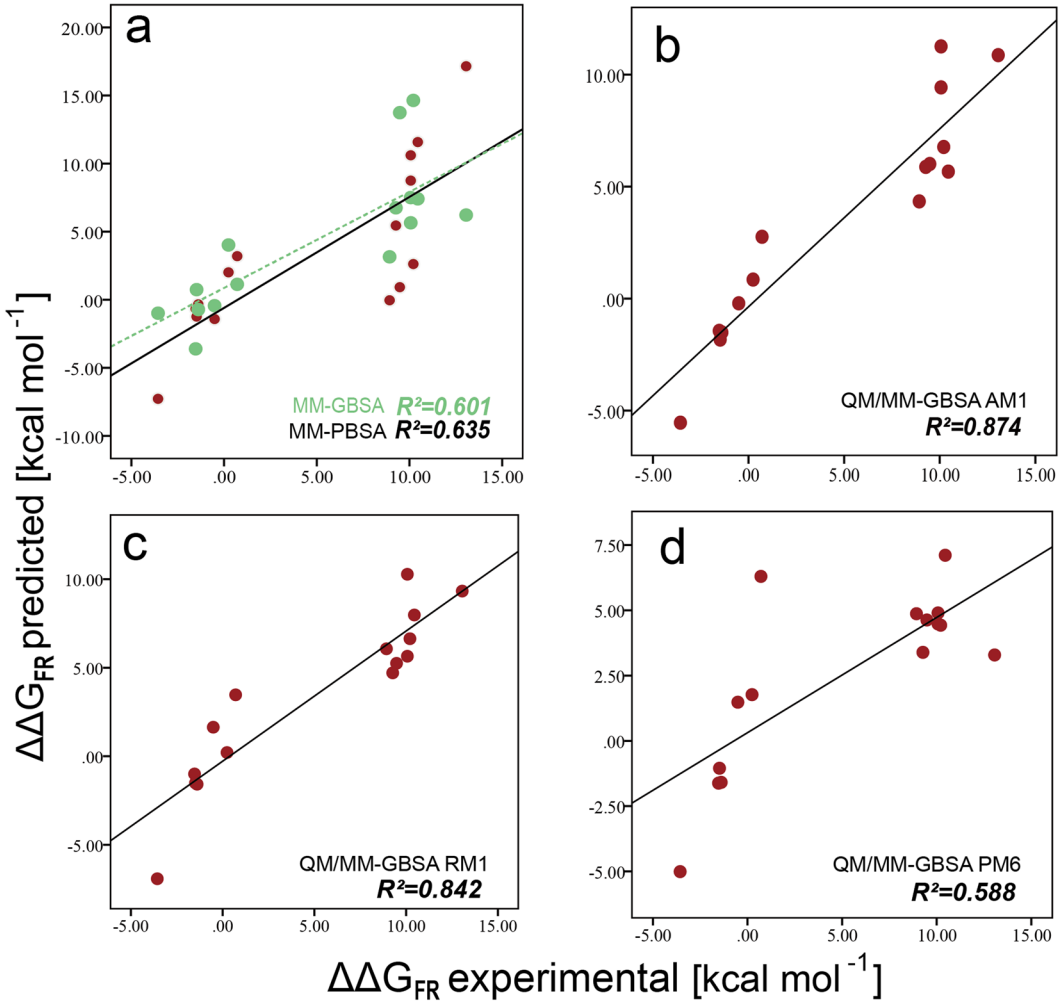

Figure 5. Inhibition constants of ALK Inhibitors with Wild-type and mutant ALK kinase. Predicted FR energies ($\Delta\Delta G_{FR}$ predicted) vs experimental FR energies ($\Delta\Delta G_{FR}$ experimental) for inhibitors with ALK and mutations. (**a**) MM-GBSA (0.601); (**b**) MM-PBSA (0.635); (**c**) AM1 (0.874); (**d**) RM1 (0.842); (**d**) PM6 (0.588).

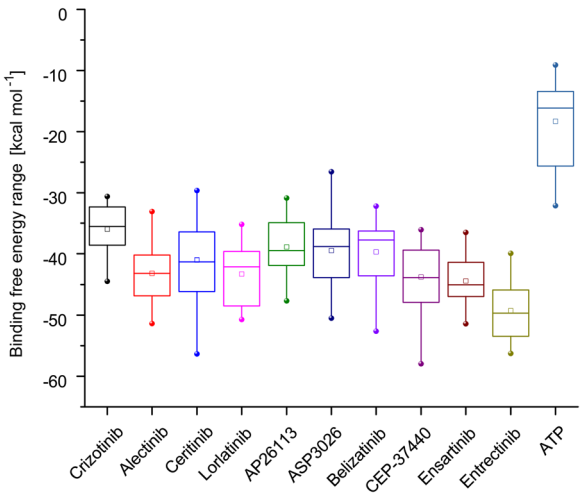

Figure 6. Range of binding free energies of 10 ALK inhibitors and a substrate ATP against 21 reported drug resistance mutations occurring in ALK kinase domain.

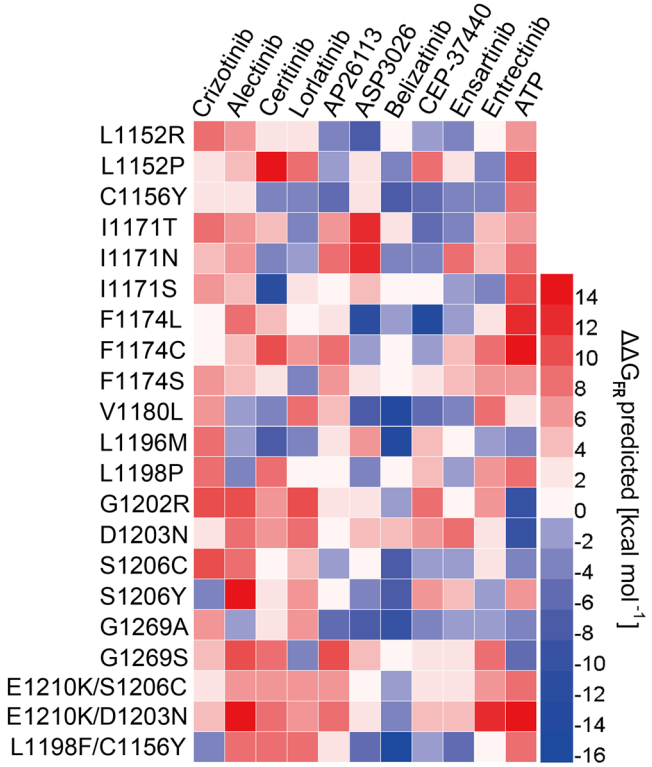

Figure 7. Heat map of the mutation energy profile of 10 ALK inhibitors as well as a substrate ATP against 21 reported drug resistance mutations occurring in ALK kinase domain.

Other reported single-point drug resistance mutations against crizotinib include L1152P/R, C1156Y, I1171T/N/S, F1174C/L/V, V1180L, S1206C/Y and G1269A/S[25], where V1180L and G1269S located in the ATP-binding pocket. L1152P, L1152R and C1156Y affect residues adjacent to the N-terminus while F1174C/L/V affect residues adjacent to the C-terminus of the αC helix[16,18,54,55]. The $\Delta\Delta G_{FR}$ predicted values of 13 mutations were positive, indicating worse affinity after mutations. While only that of S1206Y was negative, which was −2.97 kcal·mol^{-1}. S1206Y belongs to the solvent-front mutations which impair drug binding likely through steric hindrance[16,17]. The ΔE_{ele}, ΔE_{vdw}, ΔG_{surf} and quantum mechanics energy (ΔG_{qm}) provided favorable contribution to the inhibitor bindings. The unfavorable polar solvation energy (ΔG_{gb}) partially impaired these favorable factors. Especially, we found these drug mutations also attained weaker ΔE_{ele} than that of wild type, except S1206C, G1269S and C1156Y/L1198F. The ΔE_{ele} is significant in ligand-receptor binding and plays an key role in the stability of biomolecules. The destabilizing mutations can obviously reduce electrostatic interactions which turn in changing the ligand binding affinity. It should be noticed that the double-point mutation L1198F and C1156Y had a negative $\Delta\Delta G_{FR}$ predicted value (−3.51 kcal·mol^{-1}) while the $\Delta\Delta G_{FR}$ predicted values of single-point mutation C1156Y was 2.36 kcal·mol^{-1}. Previous study suggest that L1198F enhanced binding to crizotinib, negating the effect of C1156Y and re-sensitizing resistant mutations to crizotinib[19]. The ΔE_{ele} of C1156Y/L1198F (−146.84 kcal·mol^{-1}) was slightly lower than that of wild type (−144.18 kcal·mol^{-1}), its ΔE_{vdw} (−46.57 kcal·mol^{-1}) was also slightly lower than wild type (−45.45 kcal·mol^{-1}). Their ΔG_{qm}were very comparable. The ΔE_{qm} of wild type was −21.21 kcal·mol^{-1} and C1156Y/L1198F was −21.68 kcal·mol^{-1}.

The mutation L1196M is the most common mutation that confers crizotinib resistance[16], which promoted the development of the second-generation inhibitors (such as alectinib and ceritinib)[56]. The mutation energies of alectinib and ceritinib against L1196M were −0.04 kcal·mol^{-1} and −6.54 kcal·mol^{-1} respectively while that of crizotinib was 9.3 kcal·mol^{-1}, suggesting the better affinity of alectinib and ceritinib. Besides L1196M, ceritinib was also reported to overcome inhibitor-bearing ALK mutations (such as I1171T and S1206Y), while some mutations were reported resistant to ceritinib (such as F1174C and G1202R)[57–59]. The mutation energies of ceritinib against these mutations were 5.07 kcal·mol^{-1} (I1171T), 3.77 kcal·mol^{-1}(S1206Y), 10.65 kcal·mol^{-1} (F1174C) and 6.51 kcal·mol^{-1} (G1202R) respectively. It is not hard to explain why ceritinib is resistant to F1174C, as the $\Delta\Delta G_{FR}$ *predicted* value was rather high, suggesting that considerable affinity had been lost. However, the $\Delta\Delta G_{FR}$ *predicted* values of I1171T and S1206Y were all positive while they were sensitive to ceritinib. It seemed strange on the first sight, but if the mutation energies of ceritinib's competitor (which was ATP) were considered at the same time, then it may not be as paradoxical as it seemed to be. The $\Delta\Delta G_{FR}$ *predicted* of ATP against I1171T and S1206Y were 7.31 and 6.88 kcal·mol^{-1} respectively, which were all higher than those of ceritinib (which were 5.07 and 3.77 kcal·mol^{-1} respectively). The affinity changes of ceritinib were not very significant, and that of ATP was more violent. As result, although ceritinib lost some of its affinity towards the mutations, it still showed more affinity than its competitor ATP, thus performing its function as an effective inhibitor towards the mutations. The third-generation inhibitor lorlatinib was developed as a more potent ALK antagonist, which was

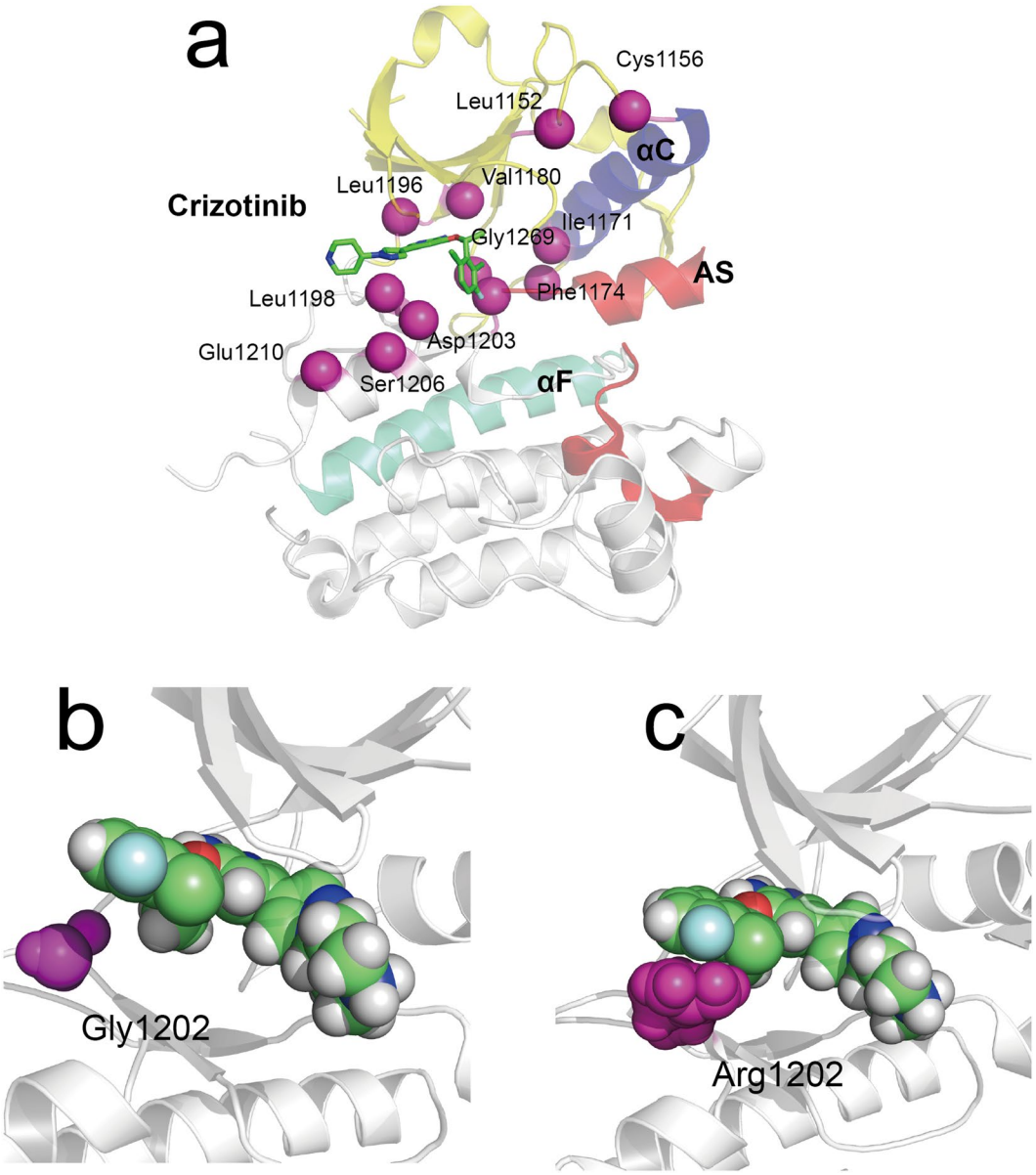

Figure 8. Structural Basis for drug resistance to crizotinib. (**a**) The location of ALK drug resistance mutations. AS refers to activation segment. The co-crystal structures of crizotinib bound to the wild type (panel b) and Arg1202 mutant (panel c).

reported sensitive to L1196M, L1152R, C1156Y, F1174L and S1206Y[60]. Their $\Delta\Delta G_{FR}$ *predicted* values were not very high and all of them were lower than that of ATP, suggesting its effectiveness as an inhibitor. It should be noticed that although C1156Y is sensitive to lorlatinib, adding L1198F disrupts the binding of drug and leads to resistance[25]. The $\Delta\Delta G_{FR}$ *predicted* value of lorlatinib against the double mutation L1198F/C1156Y was 8.64 kcal·mol^{-1}, suggesting that considerable binding affinity was lost and strong drug resistance towards lorlatinib was gained.

From the discussion above, an assumption could be concluded: if the $\Delta\Delta G_{FR}$ *predicted* value of the inhibitor against the mutation is significantly high (suggesting that considerable affinity was lost), chances are that the mutation may be strongly resistant to the inhibitor. If the $\Delta\Delta G_{FR}$ *predicted* value is not very high, then it should be compared with that of ATP to determine which (the inhibitor or ATP) has the better affinity towards the mutation. If the inhibitor gains better affinity (has higher $\Delta\Delta G_{FR}$ *predicted* value) than ATP after mutation, then chances are that the mutation could be sensitive to the inhibitor. Therefore, bold speculations could be made that belizatinib, CEP-37440 and ensartinib could be highly promising inhibitors as they gained overall better affinity towards many drug-resistant mutations compared to ATP. Besides, the mutations D1203N and G1202R could be quite intractable antagonists as many inhibitors fail to gain better affinity than ATP towards them.

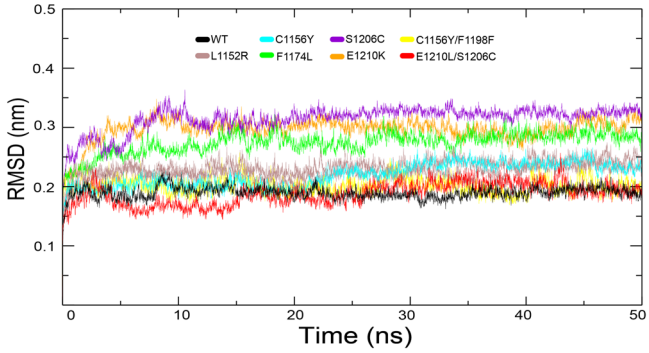

Figure 9. RMSD with respect to the initial structure as a function of time for the various simulations over 50 ns.

Structure basis for drug resistance mediated by ALK mutations. Our previous analysis demonstrated that mutations at the active site could inference the steric hindrance and change ligand binding affinity, thus resulting in the emergence of resistance to ALK inhibitors. In addition, increased ATP binding affinity has been discussed for several ALK mutants. However, the structural mechanism by which non-active site mutations confers resistance against ALK inhibitors still remains unknown. Several studies have been conducted to explore the structural mechanism of drug resistance mediated by the mutations occurring in non-active site. They found that these mutations could regulate the dislocation of ALK inhibitors and the indirect conformational changes in the binding cavity[61,62]. To provide a structural framework for understanding drug resistance mutations, 50-ns MD simulations were performed for wild type and mutant systems. The inherent flexibility profile of a protein is a key factor for its full functional activity and ligand binding. The root-mean-square deviation (RMSD) trace was used here to evaluate the dynamics of mutation structure during inhibitor binding, see Fig. 9. It can be seen that the RMSD showed obvious differences between the various simulations. Wild type ALK attained an average of RMSD at 0.197 nm (1.97 Å). The initial 10 ns were excluded as burn-in when average RMSD were calculated. The single mutations showed larger RMSD than wild type system. These were the L1152R, C1156Y, F1174L, S1206Y and E1210K. Their average RMSD values were 2.38 Å, 2.32 Å, 2.86 Å, 3.27 Å and 3.12 Å, respectively. While double mutations C1156Y/L1198F and E1210K/SC1206C attained similar RMSD with wild type. Their average RMSD values were 2.01 Å and 1.98 Å, respectively. The RMSD traces of the double mutations were similar with wild type ALK, indicating that these mutations possessed identical flexibility with wild-type ALK. In previous analysis, we raised that the changed of protein thermostability induced by high destabilizing mutations were fixed by additional stabilizing mutations. Here, results suggested that the flexibility of protein was also constrained by co-evolution of residues.

It has been well established that the conformational motions activated the dynamical transition of conformational sub-states, resulting in protein relaxation and movements of the ligand within the protein[63,64]. We determined whether the transition could be responsible for the resistance against ALK inhibitors when mutations occur in non-active sites, such as L1152R, C1156Y, F1174L, S1206Y and E1210K. Principal component analysis (PCA) was performed to represent the motion of the native and mutant forms in phase space during crizotinib binding. We extracted eigenvector indexes from the covariance matrix and estimated the correlations between protein motion and eigenvector indexes. The top two dominant principal components PC1 and PC2 accounted for a significant amount of the overall motion, and they can explain more than 85% of the variance together in each case (Fig. 10). The overall motions represented by PC1 and PC2 for wild type and mutant systems were presented in Fig. 11. As shown here, the overall motions between wild-type and most of drug-resistant mutations were extremely diverse. C1156Y showed slightly similar motion with F1174L and S1206C showed slightly similar motion with E1210K. C1156Y/L1198F showed comparable overall motions compared with wild-type as well as E1210K/S1206C, indicating double mutations possessed identical flexibility and similar conformation transition compared with wild-type ALK. This may be one of the reason why double mutations confer sensitization to crizotinib[19].

In addition, we compared the average structure of wild-type and mutations during the last 10 ns of the simulation. Figure 12a showed that crizotinib attained an RMSD value of 1.38 Å when the structure of C1156Y mutation aligned to wild-type, confirming that the C1156Y mutation caused conformational changes and ligand motion. Activation segment of ALK is a key regulation component in determining the more active less active kinase state[65,66]. We compared the average structures of activation segment of different mutations with wild-type extracted from the last 10-ns trajectory, see Fig. 12b. The structures of activation segments of the mutations were obviously heterogeneous compared with wild-type. Their activation segment showed more open and incompact conformations than that of wild-type, such as L1152R, C1156Y and F1174L, which were considered as the features of more active state form that allow protein/peptide binding. These mutations are too far to directly block inhibitor binding, their $\Delta\Delta G_{fold}$ values were all higher than $-1.84\,\text{kcal}\cdot mol^{-1}$, see Table 1. It is interesting to note that an energy dissipation-cum-signaling mechanism by which distal residues were thermodynamically coupled to activate sites had been established[67–70]. The extent of conformational changes induced by these mutations can be up to over 20 Å[69], thus resulting in identifiable functional consequences for ligand binding and/or activity of kinase. The propagation of distal mutational perturbation may play a vital role in drug resistance induced by these mutations. Our previous analysis found that highly destabilizing mutations increase kinase activity. Taken together, we suggest that highly destabilizing mutations alter the kinase's conformation via propagative effect, possibly re-inducing kinase activity despite the presence of inhibitor, in turn causing drug resistance to ALK inhibitors.

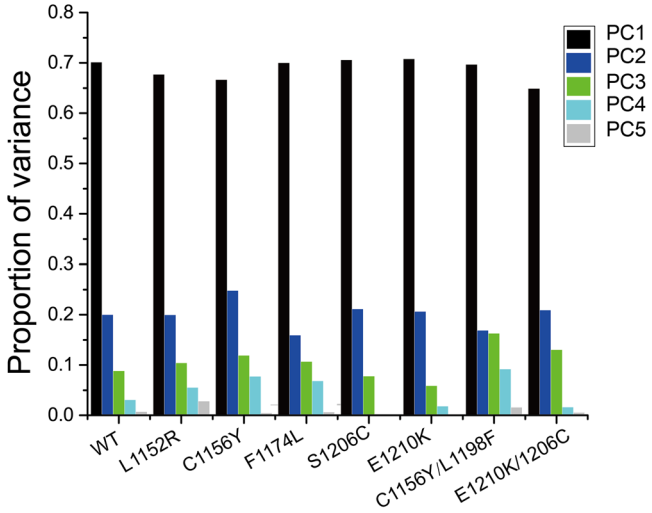

Figure 10. The histogram indicates the variance proportion of each principal component from PC1 to PC5.

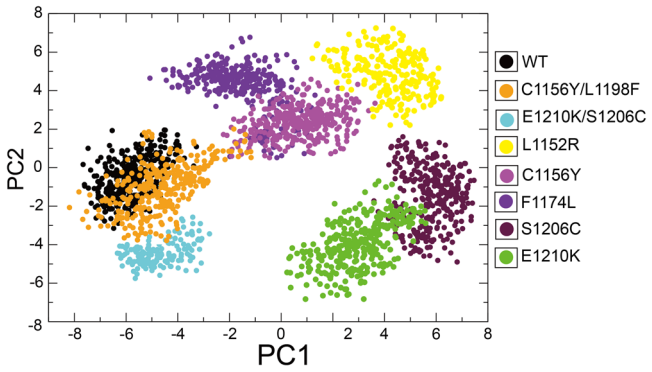

Figure 11. PCA scatter plot along first two principal components, eigenvector 1 (PC1) and eigenvector 2 (PC2) showing difference between the wild type and mutant systems.

Conclusion

Our primary goal was to quantitatively analyze mutational effects of the ALK drug resistance mutations and provide an up-to-date mechanistic framework for understanding the mechanisms of resistance mediated by these mutations. Firstly, the stability changes ($\Delta\Delta G_{fold}$) were computed to estimate the mutational effects of the ligand-free structure of the ALK mutations. The protein fitness of the drug resistance mutations was investigated using comparative evolutionary considerations. We found that drug resistance mutations were significantly more destabilizing, creating the capacity to confer drug resistance to ALK inhibitors. The stability effects of high destabilizing mutations on ALK structure were constantly pruned by the stabilizing mutations. We suggest that co-evolution may limit the ALK's fitness to drug treatment conditions through additional amino acid changes that can shift the drug resistance properties without dramatic effects on the overall thermodynamic stability, causing failure of drug treatments. Secondly, QM/MM-GBSA calculation was performed to investigate the mutational effects of the drug resistance mutations upon ligand binding, we showed evidence that mutations at the active site inferenced the steric hindrance and changed ligand binding affinity, resulting in the emergence of resistance to the ALK inhibitors. In addition, increased ATP binding affinity had been discussed for several drug-resistant ALK mutations. Thirdly, we explored the conformation transition mediated by drug-resistant mutations. We found that highly destabilizing mutations did alter the kinase's conformation via propagative effect despite the presence of ALK inhibitors, resulting in the dislocation of the inhibitors and possibly re-inducing kinase activity. Taken together, our work provides comprehensively quantitative analysis for unraveling drug resistance mechanisms of ALK mutations from multiple perspectives.

Methods

Structure Preparation. All solved ALK mutation structures were obtained from PDB database. The non-solved *in silico* mutant structures of ALK were generated from a minimized structure of ALK kinase domain (PDB code: 2XP2) using the PyMOL mutagenesis tool. To refine the solved models and the *in silico* mutant, we performed energy minimizations with the GROMACS package using the GROMOS96 force field[71].

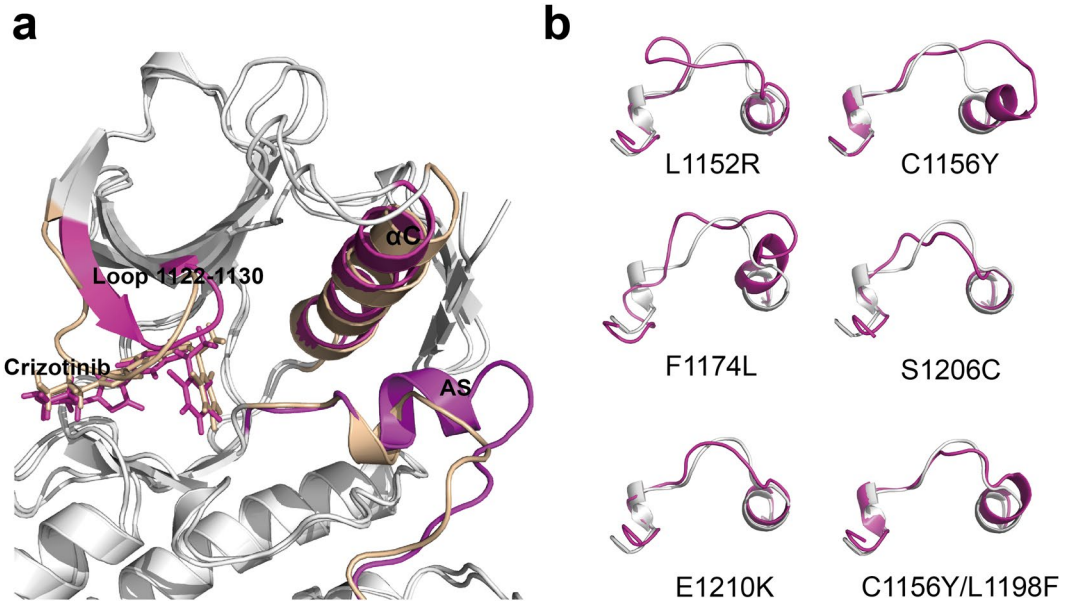

Figure 12. Alignment of structures of wild-type ALK and the non-active sites mutations. (**a**) Comparison of structures of wild-type ALK and C1156Y mutation. C1156Y results in marked conformational changes in activation segment (AS), αC helix and loop 1122–1130. (**b**) Comparison of activation segment structures of wild-type ALK (white) and different mutations (light magenta).

Stability effect calculation. To compute the free-energy change of the mutations of ALK, FoldX4[72] was used as a practical tool. Calculations are based on a minimized structure of ALK kinase domain (PDB code: 2XP2). By applying FoldX's *BuildModel* function, 19 possible point mutations can be introduced to each residue one at a time and 3 independent runs can be carried out calculating respective value of $\Delta\Delta G$ ($\Delta\Delta G = \Delta G_{mut} - \Delta G_{wt}$), which were then averaged getting the final value $\Delta\Delta G_{fold}$. The reported accuracy of FoldX is 0.46 kcal·mol^{-1}, and $\Delta\Delta G_{fold}$ values were then binned into seven categories accordingly, see Fig. 1.

Mutation effect score calculation. The pairwise model of sequences was computed from a family sequence alignment of ALK using *plmc*. The alignment was in aligned A2M format. The strength of l_2-regularization λ_j on the couplings based on the length of the model was set to 52.2. Other parameters were set as default. Then mutation effect scores (ΔE) were predicted by Evmutation[73] using epistatic model.

Molecular Dynamics simulation. In order to investigate the stability of the ALK-ligand complexes, MD simulations were carried out by GROMACS 5.1[71]. By using the pdb2gmx module, the structure files were converted into the topology file with AMBERff99SB[74] force field and TIP3P water model[75] applied. As for the ligands' topology files, they were generated by the antechamber module in AmberTools[76] with general amber force field[77] applied. Then, the dodecahedron periodic box was created with the minimal distance between the molecules and the box set to 1.0 nm. The pre-equilibrium steps include system energy minimization, location-constrained pre-simulation, NVT (constant number of particles, volume, and temperature) pre-equilibrium and NPT (constant number of particles, pressure, and temperature) pre-equilibrium, whose purpose is to let the system approach equilibrium as near as possible for the better MD simulations. To be more specific, in the energy minimization period, water molecules along with 0.15 M Na^+ and Cl^- were added into the already energy-minimized system in vacuo to get the new energy-minimized state in solution. In the location-constrained period, the position of the atoms of the receptor-ligand complex remained unchanged while solvent molecules and icons move randomly for further reaching equilibrium. The NVT period lasts for 1000 ps with the system temperature set constantly to 300 K and the sampling time step set to 2 fs. In this period, cutoff method was used to calculate short-range electrostatic interactions and Particle Mesh Ewald method (PME)[78] was used to calculate long-range electrostatic interactions (rlist = 1 nm, VdW distance = 0.9 nm, cutoff distance = 0.9 nm, and PME order = 4)[79]. Linear Constraint algorithm[80] was adopted to fix the length of all bonds in the system. In the NPT pre-equilibrium, some parameters (duration, system temperature and time step) stayed the same as those in NVT period, while the system pressure maintains constantly to 1 atm in NPT. Finally, the MD simulation continued for 50 ns to analyze the bonding affinity of the complex system while trajectory information was collected every 10 fs for further analysis.

Binding affinity calculation. The QM/MM-GBSA approach in Amber was utilized to calculate the binding free energies between the receptors (wild type or mutant) and the inhibitor. Relevant formulas are as follows:

$$\begin{aligned}\Delta G_{mt} &= G_{mt}^{complex} - (G_{mt}^{receptor} + G^{inhibitor})\\ \Delta G_{wt} &= G_{wt}^{complex} - (G_{wt}^{receptor} + G^{inhibitor})\\ G &= H - TS\\ H &= E_{vdw} + E_{ele} + E_{int} + G_{gb} + G_{surf}\end{aligned}$$

ΔG_{mt} is the interaction free energy of the mutant-inhibitor system, where $G_{mt}^{complex}$, $G_{mt}^{receptor}$, $G^{inhibitor}$ are the free energies of the mutant-inhibitor complex, mutant ALK and the inhibitor. Similarly, ΔG_{wt} is the interaction free energy of the wildtype ALK-inhibitor system. G stands for the Gibbs free energy, and H, T, S stand for enthalpy (SI unit: joule), temperature (SI unit: kelvin), entropy (SI unit: joule per kelvin) respectively. The terms E_{vdw}, E_{ele}, E_{int}, G_{gb}, G_{surf} stand for van der Waals interaction, electrostatic interaction, internal energy, polar and non-polar solvation free energy respectively. G_{gb} is calculated by the modified GB model developed by Onufriev *et al.*[81]. G_{surf} is estimated by the empirical formula $G_{surf} = \gamma \times SASA + \beta$ where γ and β were set to $0.0072\,\text{kcal}\cdot mol^{-1}\cdot\text{Å}^{-2}$ and $0.0\,\text{kcal}\cdot mol^{-1}$, respectively[82]. From ΔG_{mt} and ΔG_{wt}, inhibitor affinity change of the ALK ($\Delta\Delta G$) can be derived as $\Delta\Delta G = \Delta G_{mt} - \Delta G_{wt}$. It is reported that QM/MM-GBSA approach often overestimated the interaction free energies[83]. However, since the change of the interaction free energies ($\Delta\Delta G$)is what we care and the deviation in $\Delta\Delta G$ is already considerably eliminated by ΔG_{mt} minus ΔG_{wt}, Thus, the system variation in QM/MM-GBSA approach would not bring about serious impacts to our results. In the QM/MM method, the system was divided into two layers: the inner layer and the outer layer. The inner layer contains the residues involving the hydrogen binding interactions and the residues under mutations, which was treated with the quantum mechanics (QM) theory by applying the semi-experiential Hamiltonian methods. The outer layer consists of the rest of the system, which was treated with the molecular mechanics (MM) theory. As for the entropy, it was calculated by applying a normal-mode analysis at the MM level[84].

Principal component analysis. PCA is a useful tool to determine the correlated motions of the residues to a set of linearly uncorrelated variables called principal components. This method is based on the construction of the covariance matrix of the coordinate fluctuations of the simulated proteins. The eigenvectors and eigenvalues are obtained by diagonalizing the covariance matrix, which provides information about correlated motions throughout the protein. *gmx-covar* from GROMACS was used to perform PCA on 50-ns MD trajectories. Overall rotational and translational motions were removed by fitting the protein structure of each time frame to the initial frame.

References

1. Morris, S. W. *et al*. Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science-New York Then Washington-*, 1281–1281 (1994).
2. Shiota, M. *et al*. Hyperphosphorylation of a novel 80 kDa protein-tyrosine kinase similar to Ltk in a human Ki-1 lymphoma cell line, AMS3. *Oncogene* **9**, 1567–1574 (1994).
3. Soda, M. *et al*. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* **448**, 561–566 (2007).
4. Kruczynski, A., Delsol, G., Laurent, C., Brousset, P. & Lamant, L. Anaplastic lymphoma kinase as a therapeutic target. *Expert opinion on therapeutic targets* **16**, 1127–1138 (2012).
5. Mossé, Y. P., Wood, A. & Maris, J. M. Inhibition of ALK signaling for cancer therapy. *Clinical Cancer Research* **15**, 5609–5614 (2009).
6. Palmer, R. H., Vernersson, E., Grabbe, C. & Hallberg, B. Anaplastic lymphoma kinase: signalling in development and disease. *Biochemical Journal* **420**, 345–361 (2009).
7. Hallberg, B. & Palmer, R. H. Mechanistic insight into ALK receptor tyrosine kinase in human cancer biology. *Nature Reviews Cancer* **13**, 685–700 (2013).
8. Shaw, A. T. *et al*. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *New England Journal of Medicine* **368**, 2385–2394 (2013).
9. Solomon, B. J. *et al*. First-line crizotinib versus chemotherapy in ALK-positive lung cancer. *New England Journal of Medicine* **371**, 2167–2177 (2014).
10. Shaw, A. T. *et al*. Ceritinib in ALK-rearranged non–small-cell lung cancer. *New England Journal of Medicine* **370**, 1189–1197 (2014).
11. Kim, D.-W. *et al*. Activity and safety of ceritinib in patients with ALK-rearranged non-small-cell lung cancer (ASCEND-1): updated results from the multicentre, open-label, phase 1 trial. *The Lancet Oncology* **17**, 452–463 (2016).
12. Ou, S.-H. I. *et al*. Alectinib in crizotinib-refractory ALK-rearranged non–small-cell lung cancer: a phase II global study. *Journal of clinical oncology* **34**, 661–668 (2015).
13. Shaw, A. T. *et al*. Alectinib in ALK-positive, crizotinib-resistant, non-small-cell lung cancer: a single-group, multicentre, phase 2 trial. *The lancet oncology* **17**, 234–242 (2016).
14. Katayama, R. *et al*. Therapeutic strategies to overcome crizotinib resistance in non-small cell lung cancers harboring the fusion oncogene EML4-ALK. *Proceedings of the National Academy of Sciences* **108**, 7535–7540 (2011).
15. Johnson, T. W. *et al*. Discovery of (10 R)-7-Amino-12-fluoro-2, 10,16-trimethyl-15-oxo-10, 15,16,17-tetrahydro-2H-8, 4-(metheno) pyrazolo [4,3-h][2,5,11]-benzoxadiazacyclotetradecine-3-carbonitrile (PF-06463922), a macrocyclic inhibitor of anaplastic lymphoma kinase (ALK) and c-ros oncogene 1 (ROS1) with preclinical brain exposure and broad-spectrum potency against ALK-resistant mutations. *Journal of medicinal chemistry* **57**, 4720–4744 (2014).
16. Katayama, R. *et al*. Mechanisms of acquired crizotinib resistance in ALK-rearranged lung cancers. *Science translational medicine* **4**, 120ra117–120ra117 (2012).
17. Gainor, J. F. *et al*. Molecular mechanisms of resistance to first-and second-generation ALK inhibitors in ALK-rearranged lung cancer. *Cancer discovery* **6**, 1118–1133 (2016).
18. Friboulet, L. *et al*. The ALK inhibitor ceritinib overcomes crizotinib resistance in non–small cell lung cancer. *Cancer discovery* **4**, 662–673 (2014).
19. Shaw, A. T. *et al*. Resensitization to crizotinib by the lorlatinib ALK resistance mutation L1198F. *New England Journal of Medicine* **374**, 54–61 (2016).
20. George, R. E. *et al*. Activating mutations in ALK provide a therapeutic target in neuroblastoma. *Nature* **455**, 975–978 (2008).
21. Janoueix-Lerosey, I. *et al*. Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma. *Nature* **455**, 967–970 (2008).
22. Mossé, Y. P. *et al*. Identification of ALK as a major familial neuroblastoma predisposition gene. *Nature* **455**, 930–935 (2008).

23. Roskoski, R. Anaplastic lymphoma kinase (ALK): structure, oncogenic activation, and pharmacological inhibition. *Pharmacological research* **68**, 68–94 (2013).
24. Katayama, R., Lovly, C. M. & Shaw, A. T. Therapeutic targeting of anaplastic lymphoma kinase in lung cancer: a paradigm for precision cancer medicine[J]. *Clinical Cancer Research* **21**(10), 2227–2235 (2015).
25. Lin, J. J., Riely, G. J. & Shaw, A. T. Targeting ALK: precision medicine takes on drug resistance. *Cancer discovery* **7**, 137–155 (2017).
26. Holohan, C., Van Schaeybroeck, S., Longley, D. B. & Johnston, P. G. Cancer drug resistance: an evolving paradigm. *Nature Reviews Cancer* **13**, 714–726 (2013).
27. Gerlinger, M. *et al*. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *New England journal of medicine* **366**, 883–892 (2012).
28. Meiering, E. M., Serrano, L. & Fersht, A. R. Effect of active site residues in barnase on activity and stability. *Journal of molecular biology* **225**, 585–589 (1992).
29. Schreiber, G., Buckle, A. M. & Fersht, A. R. Stability and function: two constraints in the evolution of barstar and other proteins. *Structure* **2**, 945–951 (1994).
30. Shoichet, B. K., Baase, W. A., Kuroki, R. & Matthews, B. W. A relationship between protein stability and protein function. *Proceedings of the National Academy of Sciences* **92**, 452–456 (1995).
31. Mitternacht, S. FreeSASA: An open source C library for solvent accessible surface area calculations. *F1000Research* **5** (2016).
32. Villegas, V., Viguera, A. R., Avilés, F. X. & Serrano, L. Stabilization of proteins by rational design of α-helix stability using helix/coil transition theory. *Folding and Design* **1**, 29–34 (1996).
33. Godoy-Ruiz, R. *et al*. Natural selection for kinetic stability is a likely origin of correlations between mutational effects on protein energetics and frequencies of amino acid occurrences in sequence alignments. *Journal of molecular biology* **362**, 966–978 (2006).
34. Worth, C. L., Gong, S. & Blundell, T. L. Structural and functional constraints in the evolution of protein families. *Nature Reviews Molecular Cell Biology* **10**, 709–720 (2009).
35. Consortium, U. UniProt: a hub for protein information. *Nucleic acids research* **43**, D204–D212 (2014).
36. Forbes, S. A. *et al*. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic acids research* **43**, D805–D811 (2014).
37. Landrum, M. J. *et al*. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic acids research* **44**, D862–D868 (2015).
38. Chen, Y. *et al*. Oncogenic mutations of ALK kinase in neuroblastoma. *Nature* **455**, 971–974 (2008).
39. Murugan, A. K. & Xing, M. Anaplastic thyroid cancers harbor novel oncogenic mutations of the ALK gene. *Cancer research* **71**, 4403–4411 (2011).
40. Tokuriki, N., Stricher, F., Serrano, L. & Tawfik, D. S. How protein stability and new functions trade off. *PLoS Computational Biology* **4**, e1000002 (2008).
41. Bresler, S. C. *et al*. Differential inhibitor sensitivity of anaplastic lymphoma kinase variants found in neuroblastoma. *Science translational medicine* **3**, 108ra114–108ra114 (2011).
42. Beadle, B. M. & Shoichet, B. K. Structural bases of stability–function tradeoffs in enzymes. *Journal of molecular biology* **321**, 285–296 (2002).
43. Li, J., Liu, W., Luo, H. & Bao, J. Insight into drug resistance mechanisms and discovery of potential inhibitors against wild-type and L1196M mutant ALK from FDA-approved drugs. *Journal of molecular modeling* **22**, 231 (2016).
44. Roskoski, R. Anaplastic lymphoma kinase (ALK) inhibitors in the treatment of ALK-driven lung cancers. *Pharmacological research* (2017).
45. Zuccotto, F., Ardini, E., Casale, E. & Angiolini, M. Through the "gatekeeper door": exploiting the active kinase conformation. *Journal of medicinal chemistry* **53**, 2681–2694 (2009).
46. Studer, R. A., Christin, P.-A., Williams, M. A. & Orengo, C. A. Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. *Proceedings of the National Academy of Sciences* **111**, 2223–2228 (2014).
47. Friedman, R. Drug resistance missense mutations in cancer are subject to evolutionary constraints. *Plos One* **8**, e82059 (2013).
48. Ohta, T. Slightly deleterious mutant substitutions in evolution. *Nature* **246**, 96 (1973).
49. Balius, T. E. & Rizzo, R. C. Quantitative prediction of fold resistance for inhibitors of EGFR. *Biochemistry* **48**, 8435–8448 (2009).
50. Su, P. C., Tsai, C. C., Mehboob, S., Hevener, K. E. & Johnson, M. E. Comparison of radii sets, entropy, QM methods, and sampling on MM-PBSA, MM-GBSA, and QM/MM-GBSA ligand binding energies of F. tularensis enoyl-ACP reductase (FabI). *Journal of computational chemistry* **36**, 1859–1873 (2015).
51. Park, K., Götz, A. W., Walker, R. C. & Paesani, F. Application of adaptive QM/MM methods to molecular dynamics simulations of aqueous systems. *Journal of chemical theory and computation* **8**, 2868–2877 (2012).
52. Götz, A. W., Clark, M. A. & Walker, R. C. An extensible interface for QM/MM molecular dynamics simulations with AMBER. *Journal of computational chemistry* **35**, 95–108 (2014).
53. Seabra, Gd. M., Walker, R. C., Elstner, M., Case, D. A. & Roitberg, A. E. Implementation of the SCC-DFTB method for hybrid QM/MM simulations within the Amber molecular dynamics package. *The Journal of Physical Chemistry A* **111**, 5655–5664 (2007).
54. Choi, Y. L. *et al*. EML4-ALK mutations in lung cancer that confer resistance to ALK inhibitors. *New England Journal of Medicine* **363**, 1734–1739 (2010).
55. Sasaki, T. *et al*. A novel ALK secondary mutation and EGFR signaling cause resistance to ALK kinase inhibitors. *Cancer research* **71**, 6051–6060 (2011).
56. Roskoski, R. Jr. Anaplastic lymphoma kinase (ALK) inhibitors in the treatment of ALK-driven lung cancers. *Pharmacological research* **117**, 343–356 (2017).
57. Sullivan, I. & Planchard, D. ALK inhibitors in non-small cell lung cancer: the latest evidence and developments. *Therapeutic advances in medical oncology* **8**, 32–47 (2016).
58. Fontana, D., Ceccon, M., Gambacorti-Passerini, C. & Mologni, L. Activity of second-generation ALK inhibitors against crizotinib-resistant mutants in an NPM-ALK model compared to EML4-ALK. *Cancer medicine* **4**, 953–965 (2015).
59. Dagogo-Jack, I. & Shaw, A. Crizotinib resistance: implications for therapeutic strategies. *Annals of Oncology* **27**, iii42–iii50 (2016).
60. Zou, H. Y. *et al*. PF-06463922, an ALK/ROS1 inhibitor, overcomes resistance to first and second generation ALK inhibitors in preclinical models. *Cancer cell* **28**, 70–81 (2015).
61. Sun, H.-Y. & Ji, F.-Q. A molecular dynamics investigation on the crizotinib resistance mechanism of C1156Y mutation in ALK. *Biochemical and biophysical research communications* **423**, 319–324 (2012).
62. Ni, Z., Wang, X., Zhang, T. & Jin, R. Z. Molecular dynamics simulations reveal the allosteric effect of F1174C resistance mutation to ceritinib in ALK-associated lung cancer. *Computational biology and chemistry* **65**, 54–60 (2016).
63. Ostermann, A., Waschipky, R., Parak, F. G. & Nienhaus, G. U. Ligand binding and conformational motions in myoglobin. *Nature* **404**, 205–208 (2000).
64. Tournier, A. L. & Smith, J. C. Principal components of the protein dynamical transition. *Physical review letters* **91**, 208106 (2003).
65. Donella-Deana, A. *et al*. Unique substrate specificity of anaplastic lymphoma kinase (ALK): development of phosphoacceptor peptides for the assay of ALK activity. *Biochemistry* **44**, 8533–8542 (2005).
66. Lee, C. C. *et al*. Crystal structure of the ALK (anaplastic lymphoma kinase) catalytic domain. *Biochemical Journal* **430**, 425–437 (2010).

67. Rajasekaran, N., Suresh, S., Gopi, S., Raman, K. & Naganathan, A. N. A general mechanism for the propagation of mutational effects in proteins. *Biochemistry* **56**, 294–305 (2016).
68. Rajasekaran, N. & Naganathan, A. N. A self-consistent structural perturbation approach for determining the magnitude and extent of allosteric coupling in proteins. *Biochemical Journal* **474**, 2379–2388 (2017).
69. Rajasekaran, N., Sekhar, A. & Naganathan, A. N. A Universal Pattern in the Percolation and Dissipation of Protein Structural Perturbations. *The journal of physical chemistry letters* **8**, 4779–4784 (2017).
70. Gopi, S., Paul, S., Ranu, S. & Naganathan, A. N. Extracting the Hidden Distributions Underlying the Mean Transition State Structures in Protein Folding. *The journal of physical chemistry letters* **9**, 1771–1777 (2018).
71. Pronk, S. *et al*. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
72. Schymkowitz, J. *et al*. The FoldX web server: an online force field. *Nucleic acids research* **33**, W382–W388 (2005).
73. Hopf, T. A. *et al*. Mutation effects predicted from sequence co-variation. *Nature biotechnology* **35**, 128 (2017).
74. Hornak, V. *et al*. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* **65**, 712–725 (2006).
75. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **79**, 926–935 (1983).
76. Case, D. *et al*. *AMBER*, University of California at San Francisco, San Francisco (2012).
77. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *Journal of computational chemistry* **25**, 1157–1174 (2004).
78. Fujii, S., Kono, H., Takenaka, S., Go, N. & Sarai, A. Sequence-dependent DNA deformability studied using molecular dynamics simulations. *Nucleic acids research* **35**, 6063–6074 (2007).
79. Darden, T., Perera, L., Li, L. & Pedersen, L. New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure* **7**, R55–R60 (1999).
80. Nagpal, S., Tiwari, S., Mapa, K. & Thukral, L. Decoding structural properties of a partially unfolded protein substrate: en route to chaperone binding. *PLoS computational biology* **11**, e1004496 (2015).
81. Onufriev, A., Bashford, D. & Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics* **55**, 383–394 (2004).
82. Gohlke, H., Kiel, C. & Case, D. A. Insights into protein–protein binding by binding free energy calculation and free energy decomposition for the Ras–Raf and Ras–RalGDS complexes. *Journal of molecular biology* **330**, 891–913 (2003).
83. Tian, F., Yang, L., Lv, F., Luo, X. & Pan, Y. Why OppA protein can bind sequence-independent peptides? A combination of QM/MM, PB/SA, and structure-based QSAR analyses. *Amino Acids* **40**, 493–503 (2011).
84. Xu, B., Shen, H., Zhu, X. & Li, G. Fast and accurate computation schemes for evaluating vibrational entropy of proteins. *Journal of computational chemistry* **32**, 3188–3193 (2011).

Acknowledgements

We are grateful to our colleagues for providing constructive suggestions. This work was supported in part by National Natural Science Foundation of China (Nos 81373311, 31300674, 81173093, 30970643 and J1103518), the Youth Science Foundation of West China Hospital of Stomatology (No. 2016-3), the Full-time Postdoctoral Research Foundation of Sichuan University (2018SCU12024), the Special Program for Youth Science and the Technology Innovative Research Group of Sichuan Province, China (No. 2011JTD0026), and Graduate Student's Research and Innovation Found of Sichuan University (No. 2018YJSY055).

Author Contributions

Jianzong Li conceived the original idea and Jinku Bao supervised the project; Xin Li and Chuanfang Wu helped supervise the project; Yue Huang and Miaomiao Wu performed the experiments and analyzed the data; Jianzong Li and Yue Huang wrote this manuscript. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-28752-9.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.