# A Computational Approach for Prioritizing Selection of Therapies Targeting Drug Resistant Variation in Anaplastic Lymphoma Kinase

**Matthew D. McCoy, PhD[1], Subha Madhavan, PhD[1]**
**[1]Innovation Center for Biomedical Informatics, Georgetown University Medical Center, Washington, DC, United States of America**

**Abstract**

*Anaplastic lymphoma kinase (ALK) is a receptor tyrosine kinase implicated as a driver of a number of cancer types, and activates cellular pathways involved in cell proliferation and differentiation. Tyrosine kinase inhibitors (TKIs) are a small molecule therapeutic that blocks ALK function, but tumor evolution leads to the rapid emergence of drug resistant somatic variation and necessitates selection of a new treatment strategy. Computational simulations of protein:drug interactions were used to investigate the impact of seven drug resistant mutations on binding to eleven TKIs approved, or under investigation, for treatment of ALK positive cancers. The results show variant specific disruptions to TKI molecular interactions, and demonstrate the potential to aid prioritization of therapeutic interventions. Validation remains a challenge due to the complex dependence of biomolecular interactions on the local biophysical environment, but improvements to the underlying structural model and continued curation efforts will improve the clinical utility of computational predictions.*

**Introduction**

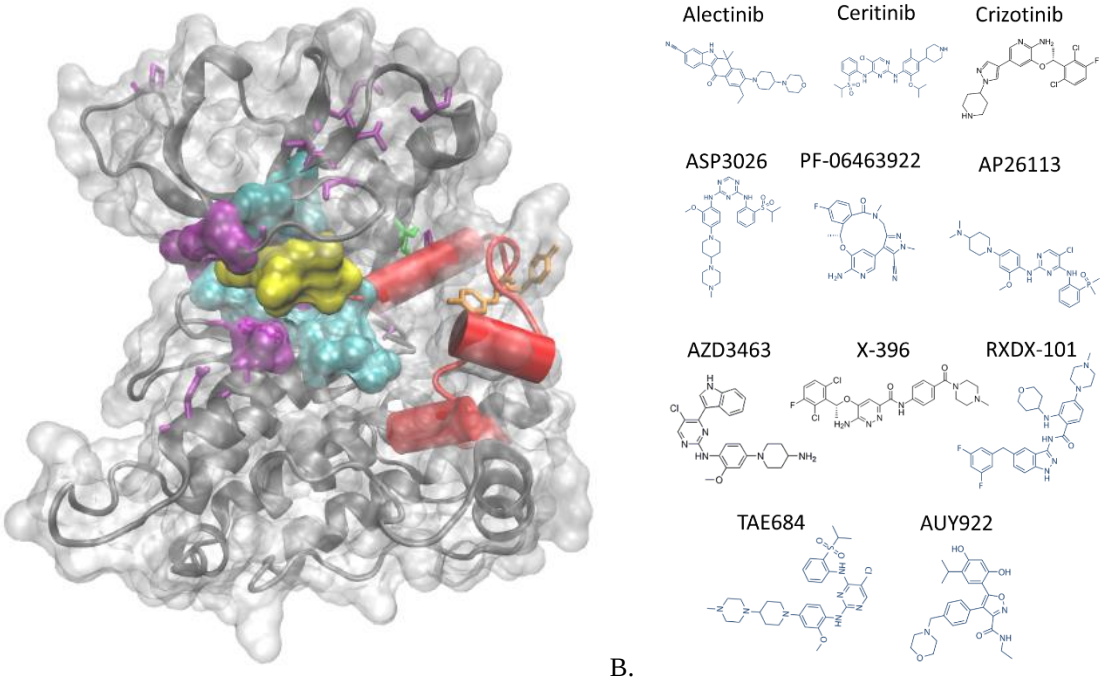
Since its discovery as a fusion partner to a nucleolar protein gene in non-Hodgkin's lymphoma[1], anaplastic lymphoma kinase (ALK) has been found to drive oncogenic pathways in a broad range of cancers[2]. ALK is a receptor tyrosine kinase (RTK), a class of proteins responsible for sensing molecular signals from the environment to regulate internal cellular processes. An RTK monomer links an extracellular receptor domain with an internal tyrosine kinase domain through a transmembrane helix. ALK dimerization occurs on the cell surface in response to a regulatory signal, which leads to autophosphorylation and subsequent activation of the kinase domains[3].

Phosphorylation of ALK tyrosine residues stabilizes an alternate conformation of the proteins structure to expose a protein interaction surface and orient the ATP binding pocket for efficient phosphate transfer[4]. Once activated, ALK will phosphorylate a number of downstream protein targets to initiate cell signaling cascades associated with cell growth, transformation, proliferation, and apoptotic avoidance[2]. Primarily expressed in the brain and relatively absent elsewhere in the body[5], tissue expression limits the promiscuous phosphorylation function of ALK to specific biological systems. In cancer, the normal regulation of ALK activity is disrupted by gene fusions, overexpression, and activating point mutations[2]. When fusions occur with ALK, the intracellular kinase domain becomes linked to a new protein, essentially conferring a new functional and regulatory domain to the fusion partner. ALK overexpression is another driving mechanism in a number of cancer types, and also disrupts the tissue specific regulatory control of downstream pathways. Additionally somatic mutations which impact the conformational stability of the activation loop and organization of the APT binding pocket can lead to an activated kinase domain in the absence of an initiating signal[6].

RTKs are part of the drugable genome and subject to modulation through small molecule interventions[7], and ALK positive cancers were among the first to benefit from targeted therapy[8]. Tyrosine kinase inhibitors (TKIs) compete with APT in the ALK binding pocket, rendering an activated protein unable to transfer a phosphate group to any downstream target. TKIs vary widely in molecular structure, and selectively bind to a particular kinase based on gene specific variation in the ATP binding domain[9]. Crizotinib was the first TKI to target ALK positive cancers, however evolutionary mechanisms active in tumor growth lead to resistance within 10.5 months, on average[10]. Several variants in ALK have been associated with resistance to selective kinase inhibitors, spurring the development of a number of second line TKIs to consider as therapeutic alternatives[9, 11-14] (Figure 1).

Computational protein-ligand binding is an established method for screening small molecule libraries to identify candidates with an increased binding affinity, and has been successfully used in the design of targeted therapeutic interventions[15, 16]. In general, these methods are designed to provide an efficient search and optimization of the steric, ionic, and hydrophobic molecular interactions that occur during ligand binding. In general, a three dimensional search space is defined around the binding region of the protein, and the interaction energy is calculated for different orientations of the target ligand. Flexibly in both the ligand structure and sidechains bordering the search space is

modeled by defining how individual bonds rotate around their axis, and the results summarize the low energy (most stable) conformations within the protein binding pocket. While typically used to search a library of molecules, the computational approach can be adapted to understand how somatic variation will influence the energetics of TKI interactions with ALK. The research described below illustrates how comparing the output of coarse grained simulation has the potential to produce a rapid assessment of the functional consequences of somatic variation, and demonstrates the technical limitations facing the computational prediction of drug resistant mutations emerging from targeted cancer therapies.




A. B.

**Figure 1. A**. The annotated, graphical representation of ALK (grey) bound to ADP (yellow) shows the distribution of somatic mutations (purple) associated with resistance to tyrosine kinase inhibition. The surface of residues lining the ligand binding pocket, where the those highlighted in blue have not been observed to be mutated in resistant proteins. Additional features include the activation loop in red, along with the activating F1275L mutation in green and phosphorylation sites shown in orange (PDB: 3LCT[17]). **B**. The structural diversity of tyrosine kinase inhibitors used in personalized therapeutic approaches targeted at disrupting ALK function.

**Methods**

Analysis of how drug resistant variants map to the 3D protein structure of ALK began by extracting 51 records from the Protein Data Bank[18] (PDB - www.rcsb.org), representing a variety of sequence variants and bound ligands. Multiple sequence and structure alignment was performed using MUSTANG[19], providing a pairwise calculation of root-mean-square deviation (RMSD) based on the position of alpha carbons in the ALK structures. Hierarchal clustering of the resulting distance matrix revealed the structural diversity of records deposited in PDB (Figure 2A) and was used to select a diverse set of starting templates for the ligand binding simulations.

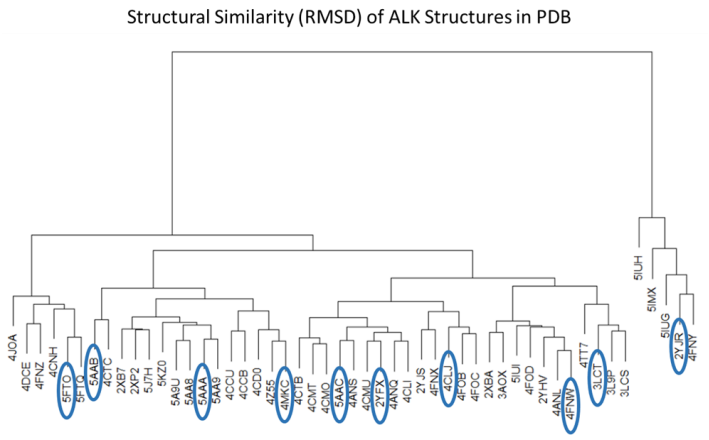
Somatic mutations in ALK with drug interaction annotations were collected from the COSMIC[20] and JAX-Clinical Knowledgebase[21] databases, which provide expert curated variant data from published literature and clinical trial reports. The database included many types of somatic variation, including rearrangements, amplifications, fusions and missense mutations. For the binding simulations, variants occurring at loci participated in ADP binding were identified and the drug interactions were categorized as resistant, decreased, or sensitive depending on the available curation data (

Table **1**). For many of the drug/variant combinations drug interaction data was not available. Additionally, a majority of the variants associated with ALK resistance to TKIs map to locations in the protein structure that do not directly participate in ligand binding and therefore cannot be analyzed by coarse-grained binding simulations.

**Table 1.** Somatic ALK mutations found in the ligand binding domain (COSMIC[20], JAX-Clinical Knowledgebase[21]). Variants can be *sensitive*, show *decreased* affinity, or confer *resistance* to a given TKI. Discrepancies in curation databases are labeled as *conflicting*, and blank spaces correspond to an absence of a curated variant/drug relationship.

| Variant | Crizotinib | Alectinib | Ceritinib | ASP3026 | AP26113 | PF-06463922 | RXDX-101 | X-396 | TAE684 | AZD3463 | AUY922 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| L1122V | | resistant | resistant | resistant | resistant | | | | | | |
| V1180L | resistant | resistant | sensitive | sensitive | sensitive | sensitive | | | sensitive | | |
| L1196Q | resistant | | | sensitive | | | | | | | |
| L1196M | resistant | decreased | conflicting | resistant | sensitive | sensitive | sensitive | sensitive | | | |
| L1198F | sensitive | resistant | resistant | resistant | conflicting | decreased | | | | | |
| D1203N | decreased | sensitive | sensitive | | sensitive | sensitive | | | | | |
| G1269A | resistant | sensitive | conflinctin | conflicting | sensitive | sensitive | decreased | | | sensitive | |

Simulated drug binding is a computational method used to model the interactions of a small molecule bound to a protein target, and require the 3D structure of the protein and small molecule ligand as input. To capture variation in the organization of the ATP binding pocket represented in the PDB database, 10 starting structures from PDB[11, 13, 14, 17, 22-24] were selected to represent the conformational diversity found in the crystallographic structures (Figure 2B). The starting template for 11 TKIs either currently approved for treatment of ALK positive cancers, or under preclinical investigation as second line therapies, were obtained from DrugBank[25]. ALK variant structures were generated for each PDB structure using DeepView[26], expanding the template into 8 variant models (7 somatic and 1 wildtype). Computational ligand binding was performed with AutoDock Vina[27], where torsional flexibility was modeled for molecular bonds in the ligands and sidechains of 14 residues lining the ALK ATP binding region. To quantify predicted drug resistance, the binding affinity of the lowest energy conformations from each variant were compared to the wildtype for all variant/drug combination.




A.

| Simulation Structures | Variant | Ligand |
|---|---|---|
| 3CLT | Wildtype | ADP |
| 5AAA | L1198F | Crizotinib |
| 2YFX | L1196M | Crizotinib |
| 5AAC | C1156Y,L1196M | Crizotinib |
| 4FNW | F1174L | apo |
| 5FTO | Wildtype | entrectinib |
| 5AAB | C1156Y, L1198F | Crizotinib |
| 4CLJ | L1196M | PF-06463922 |
| 2YJR | F1174L | wt |
| 4MKC | S1281G | Ceritinib |

B.

**Figure 2**. **A.** The hierarchal clustering based on RMSD of the protein backbone is shown for all ALK structures available in PDB. Structures chosen as templates in the ligand binding simulations are circled in blue. **B.** The variation and bound ligand for PDB structures chosen for computational binding to TKIs.

## Results

The analysis of the available ALK structures found in PDB identified several clusters corresponding to slight differences in the conformational arrangement of flexible loops and relative position of residues in the ATP binding

pocket. The differences were subtle, where the maximum pairwise RMSD between any two structures did not exceed 0.9 angstroms. Two main factors contribute to the segmentation into clusters; missing amino acids from the crystal structure and changes to the binding pocket geometry. Multiple sequence alignment (not shown) found many of the clusters correlated with the absence of similar regions of the protein structure. These unresolved regions correspond to flexible loops in the protein structure and stem from inherent limitations of x-ray crystallography. The other source of structural variation is due to the nature of the bound small molecule, as ligand binding will stabilize conformations of the binding pocket that minimize the overall free energy of the protein.

The ligand binding simulations produce a set of low energy conformations of the small molecule within the ALK binding pocket, which correspond to the most energetically favorable arrangement of the TKI. To quantify the contribution of somatic variation on drug resistance, the lowest energy orientation for each variant/TKI simulation was compared to the wildtype interaction from the corresponding PDB template (Figure 3). Since lower interaction energies correspond to more stable interactions, the sign and magnitude of the difference correlates with how the interaction between ALK and a particular TKI will be altered by somatic variation. Mutations predict to increase the binding energy relative to the wildtype interaction are predicted to disrupt drug binding.

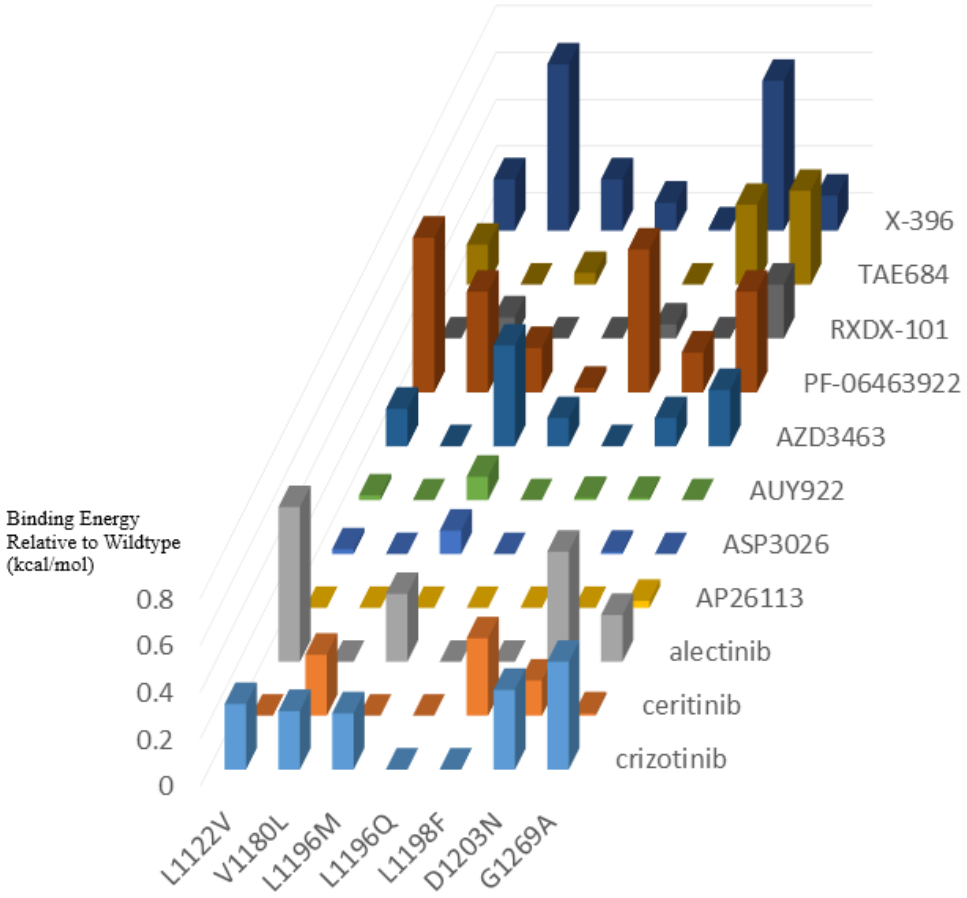



**Figure 3.** The extent a particular somatic variant was predicted to disrupt TKI interactions is quantified by calculating the difference in predicted binding energy relative to the wildtype structure. The positive values correspond to drug resistance in a particular variant, and the plot shows the average difference from 10 starting PDB templates of ALK.

Validation of the predictions is complicated by factors discussed below, but the results illustrate the utility of variant specific predictions on selecting appropriate therapeutic options. Certain TKIs are predicted to have a therapeutic impact on most or all variants, while others may be limited to a specific variant profile. However, these results represent only a snapshot of a dynamic molecular system, and should not be used to inform clinical decision making

until it is verified the simulations capture all the relevant biological context and are corroborated through preclinical research.

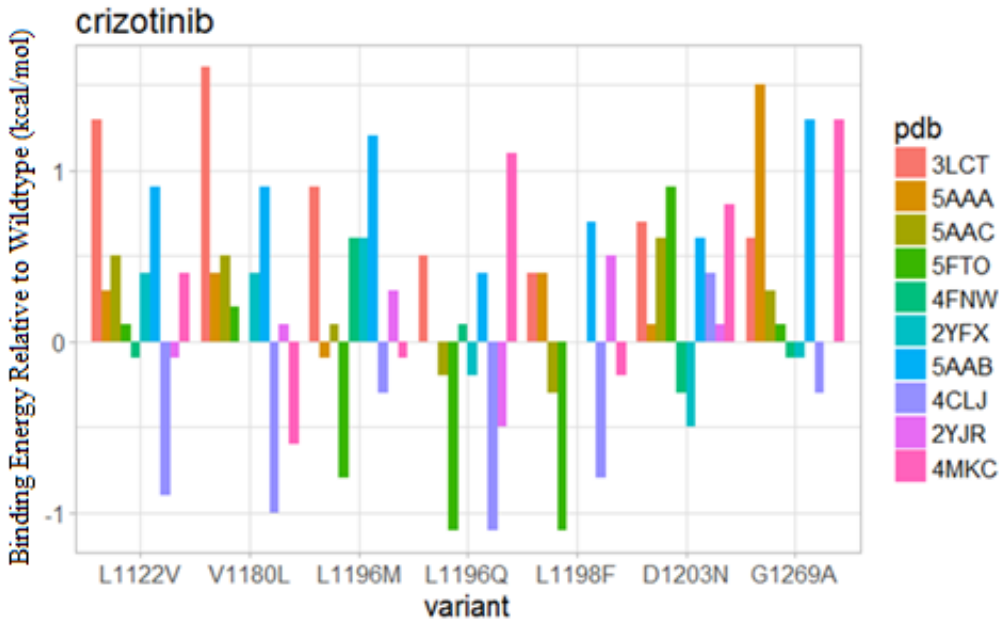



**Figure 4.** The relative predicted binding affinity for Crizotinib of TKI resistant ALK variants in the ligand binding pocket for 10 PDB structures. The predictions are highly dependent on the starting structure used, predicting both stronger and weaker interactions for the same variant/drug combination.

The predicted impact on drug resistance is highly dependent on the protein scaffold used for the simulation and varies widely between different PDB templates. As seen in the results for Crizotinib (Figure 4), the simulations predict interactions to be influenced in both directions, and illustrates the importance small differences in the binding pocket geometry have on calculating binding energetics. Certain PDB templates seem to be outliers from the overall distribution, and the range of predictions seen for Crizotinib is observed for the other TKI molecules (not shown).

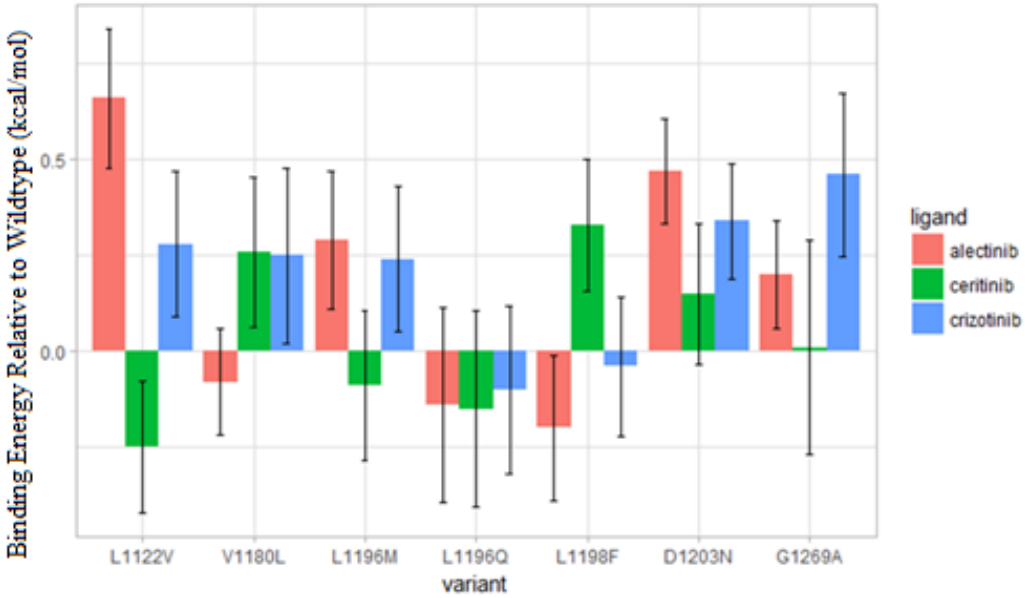


**Figure 5.** The predicted binding affinity is averaged over the 10 starting templates of ALK structure, and provide an indication of the consensus between the simulation results. The error bars show the 95% confidence interval.

To quantify consensus of binding predictions from the different PDB templates, the results for each variant were averaged. The 95% confidence interval provides an indication of the agreement between the simulation results from different PDB templates (Figure 5 and Figure 6), and can be used to inform the overall prediction of drug resistance. If the confidence interval for a variant crosses the x-axis, the impact of a TKI may be more difficult to assign compared to a variant that falls squarely in the positive or negative region.

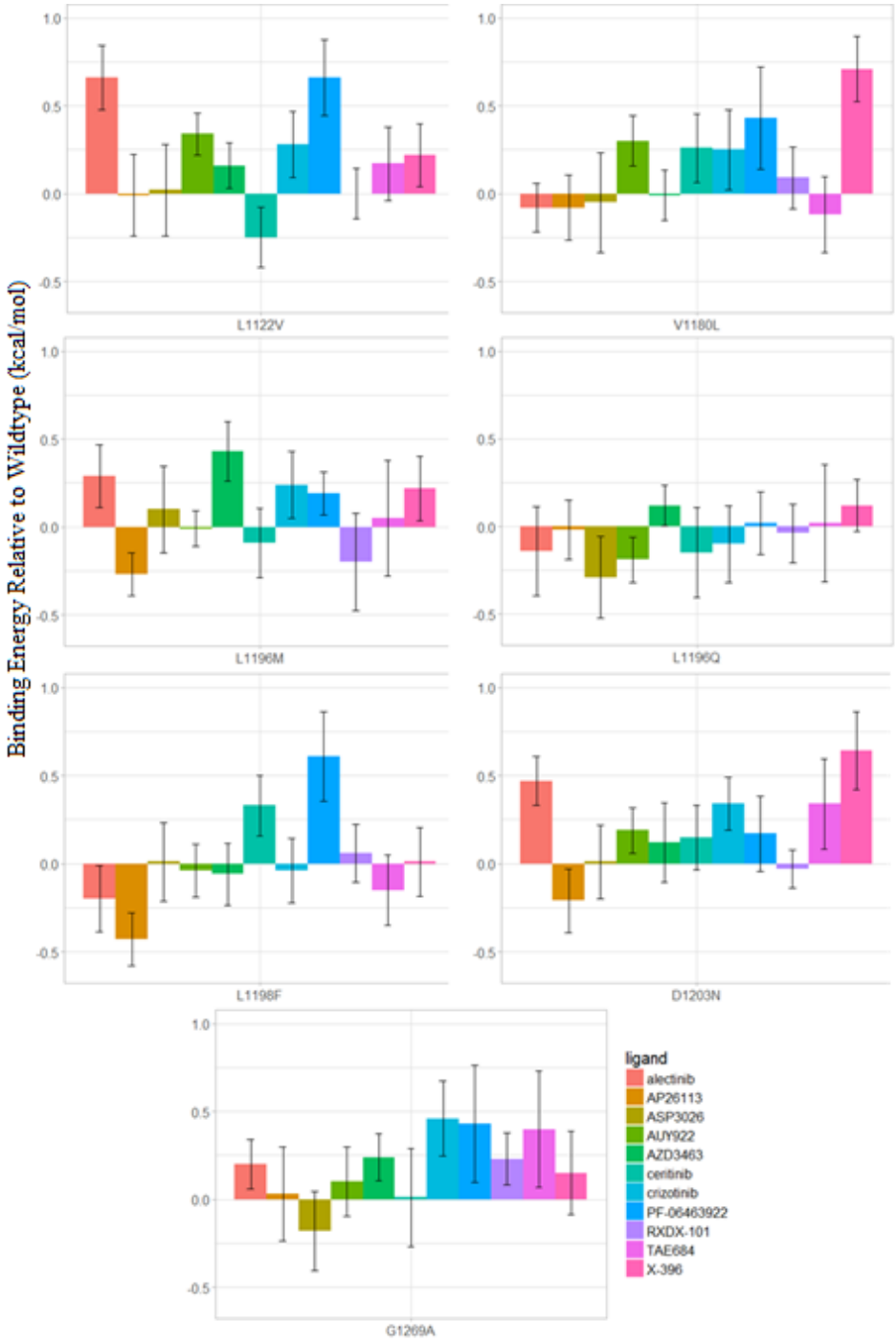


**Figure 6.** The average difference in binding energy across 10 ALK structures is shown for each drug resistant variant bound to each TKI.

**Discussion**

Overall, the utility of the simulation results is limited by the absence of rigorous experimental data to properly validate the underlying model of molecular interactions. The currently curated variants for ALK occur in a variety of biological contexts that may influence the observed correlations with drug resistance. Structural interactions with domains from gene fusion partners, conditions of *in vitro* and *in vivo* characterization, and multiple missense mutations in a single sequence will all influence the conformational dynamics and functional impact. Further validation efforts are complicated by comparing the categorical classifications in the curation databases with the continuous value predicted by the simulations.

Proteins are dynamic structures, and future predicative efforts should incorporate better representations of variant specific contributions to functional conformations of the ALK binding pocket. The specific geometry will be influenced by both changes to the protein sequence and energetic interactions with the ligand, and would benefit from molecular dynamics simulations to define variant/drug specific protein conformations. An additional benefit of all atom simulations is the ability to consider contributions from variants that do not directly participate in ligand binding. Another consideration these simulations fail to address is the competitive binding with the natural ligands, ATP and ADP. Including a relative comparison to the drug binding, those simulations may also provide insight into how a variant influences drug resistance without disrupting its native function. Drug resistance emerges through evolutionary mechanisms, and the true power of a predictive model of variant/drug interactions has incredible potential as a therapeutic tool. This becomes especially important to direct therapies toward cases where rare mutations have not been observed or the when resistance has emerged from multiple rounds of TKI therapies.

Still, the simulations results broadly reflect the trends observed in the curation databases. This is especially true for Crizotinib, where five of the seven variants were predicted to be resistant to the drug. The confidence interval for the two remaining variants indicates a more neutral impact, which is interesting considering the association of that loci with resensitization to Crizotinib after resistance to a second line treatment developed[24]. Another feature of the predictions that aligns with expectation is that certain class of TKIs are predicted to be effective against all of the variants (AP26113, ASP3062, and AUY922). These represent molecules developed specifically to overcome resistance in first line therapies, so it is expected they would strongly interact with drug resistant ALK variants and indicates those TKIs would perform than other therapeutic options. Further investigation into the similarities in drug:ALK binding, more refined computational models underlying the predictions, and experimental validation through molecular oncology methods would increase the confidence in the clinical application of the results.

**Conclusions**

The emergence of drug resistance is a byproduct of the selective pressure applied by a targeted therapeutic on a proliferating tumor, and often occurs when somatic missense mutations disrupt molecular interactions between the drug and the target protein. As observed in ALK, drug resistant variation is not limited to residues directly participating in drug binding and occurs throughout the protein at loci influencing functional conformational dynamics, and future predictive models should incorporate their influence on the geometry of the binding pocket. Due to the costs associated with wet lab experiments and preclinical trials, characterization of all drug resistant variants is not feasible. This is especially true when considering the combinatorial expansion of multiple mutations and the optimal selection of a targeted therapeutic will benefit from accurate computational predictions.

**References**


1. Morris SW, Kirstein MN, Valentine MB, Dittmer KG, Shapiro DN, Saltman DL, et al. Fusion of a kinase gene, alk, to a nucleolar protein gene, npm, in non-hodgkin's lymphoma. Science. 1994;263(5151):1281-4.
2. Hallberg B, Palmer RH. Mechanistic insight into alk receptor tyrosine kinase in human cancer biology. Nat Rev Cancer. 2013;13(10):685-700.
3. Lemmon MA, Schlessinger J. Cell signaling by receptor tyrosine kinases. Cell. 2010;141(7):1117-34.
4. McClendon CL, Kornev AP, Gilson MK, Taylor SS. Dynamic architecture of a protein kinase. Proc Natl Acad Sci U S A. 2014;111(43):E4623-31.
5. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419.
6. George RE, Sanda T, Hanna M, Frohling S, Luther W, 2nd, Zhang J, et al. Activating mutations in alk provide a therapeutic target in neuroblastoma. Nature. 2008;455(7215):975-8.

7. Hopkins AL, Groom CR. The druggable genome. Nat Rev Drug Discov. 2002;1(9):727-30.
8. Rodig SJ, Shapiro GI. Crizotinib, a small-molecule dual inhibitor of the c-met and alk receptor tyrosine kinases. Curr Opin Investig Drugs. 2010;11(12):1477-90.
9. Katayama R. Therapeutic strategies and mechanisms of drug resistance in anaplastic lymphoma kinase (alk)-rearranged lung cancer. Pharmacol Ther. 2017;177:1-8.
10. Roskoski R, Jr. Anaplastic lymphoma kinase (alk) inhibitors in the treatment of alk-driven lung cancers. Pharmacol Res. 2017;117:343-56.
11. Friboulet L, Li N, Katayama R, Lee CC, Gainor JF, Crystal AS, et al. The alk inhibitor ceritinib overcomes crizotinib resistance in non-small cell lung cancer. Cancer Discov. 2014;4(6):662-73.
12. De Pas T, Pala L, Catania C, Conforti F. Molecular and clinical features of second-generation anaplastic lymphoma kinase inhibitors ceritinib. Future Oncol. 2017.
13. Johnson TW, Richardson PF, Bailey S, Brooun A, Burke BJ, Collins MR, et al. Discovery of (10r)-7-amino-12-fluoro-2,10,16-trimethyl-15-oxo-10,15,16,17-tetrahydro-2h-8,4-(m etheno)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. J Med Chem. 2014;57(11):4720-44.
14. Menichincheri M, Ardini E, Magnaghi P, Avanzi N, Banfi P, Bossi R, et al. Discovery of entrectinib: A new 3-aminoindazole as a potent anaplastic lymphoma kinase (alk), c-ros oncogene 1 kinase (ros1), and pan-tropomyosin receptor kinases (pan-trks) inhibitor. J Med Chem. 2016;59(7):3392-408.
15. Bursulaya BD, Totrov M, Abagyan R, Brooks CL, 3rd. Comparative study of several algorithms for flexible ligand docking. J Comput Aided Mol Des. 2003;17(11):755-63.
16. Konc J, Lesnik S, Janezic D. Modeling enzyme-ligand binding in drug discovery. J Cheminform. 2015;7(1):48.
17. Lee CC, Jia Y, Li N, Sun X, Ng K, Ambing E, et al. Crystal structure of the alk (anaplastic lymphoma kinase) catalytic domain. Biochem J. 2010;430(3):425-37.
18. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. Nucleic Acids Res. 2000;28(1):235-42.
19. Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM. Mustang: A multiple structural alignment algorithm. Proteins. 2006;64(3):559-74.
20. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. Cosmic: Somatic cancer genetics at high-resolution. Nucleic Acids Res. 2017;45(D1):D777-D83.
21. Patterson SE, Liu R, Statz CM, Durkin D, Lakshminarayana A, Mockus SM. The clinical trial landscape in oncology and connectivity of somatic mutational profiles to targeted therapies. Hum Genomics. 2016;10:4.
22. Epstein LF, Chen H, Emkey R, Whittington DA. The r1275q neuroblastoma mutant and certain atp-competitive inhibitors stabilize alternative activation loop conformations of anaplastic lymphoma kinase. J Biol Chem. 2012;287(44):37447-57.
23. Huang Q, Johnson TW, Bailey S, Brooun A, Bunker KD, Burke BJ, et al. Design of potent and selective inhibitors to overcome clinical anaplastic lymphoma kinase mutations resistant to crizotinib. J Med Chem. 2014;57(4):1170-87.
24. Shaw AT, Friboulet L, Leshchiner I, Gainor JF, Bergqvist S, Brooun A, et al. Resensitization to crizotinib by the lorlatinib alk resistance mutation l1198f. N Engl J Med. 2016;374(1):54-61.
25. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, et al. Drugbank 4.0: Shedding new light on drug metabolism. Nucleic Acids Res. 2014;42(Database issue):D1091-7.
26. Guex N, Peitsch MC. Swiss-model and the swiss-pdbviewer: An environment for comparative protein modeling. Electrophoresis. 1997;18(15):2714-23.
27. Trott O, Olson AJ. Autodock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31(2):455-61.