


Evaluation of residue variability in a conformation-specific context and during evolutionary sequence reconstruction narrows drug resistance selection in Abl1 tyrosine kinase

Felipe A. M. Otsuka^{1,2} | Sinisa Bjelic¹ 

¹Department of Chemistry and Biomedical Sciences, Linnaeus University, Kalmar, Sweden

²Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

Correspondence

Sinisa Bjelic, Department of Chemistry and Biomedical Sciences, Linnaeus University, Kalmar, Sweden.
Email: sinisa.bjelic@lnu.se

Funding information

Fundação de Amparo à Pesquisa do Estado de São Paulo, Grant/Award Numbers: 2018/18537-4, 2019/25955-0; Vetenskapsrådet, Grant/Award Number: SNIC 2021/5-401; Conselho Nacional de Desenvolvimento Científico; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior; Swedish Research Council

Review Editor: John Kuriyan

Abstract

Diseases with readily available therapies may eventually prevail against the specific treatment by the acquisition of resistance. The constitutively active Abl1 tyrosine kinase known to cause chronic myeloid leukemia is an example, where patients may experience relapse after small inhibitor drug treatment. Mutations in the Abl1 tyrosine kinase domain (Abl1-KD) are a critical source of resistance and their emergence depends on the conformational states that have been observed experimentally: the inactive state, the active state, and the intermediate inactive state that resembles Src kinase. Understanding how resistant positions and amino acid identities are determined by selection pressure during drug treatment is necessary to improve future drug development or treatment decisions. We carry out *in silico* site-saturation mutagenesis over the Abl1-KD structure in a conformational context to evaluate the *in situ* and conformational stability energy upon mutation. Out of the 11 studied resistant positions, we determined that 7 of the resistant mutations favored the active conformation of Abl1-KD with respect to the inactive state. When, instead, the sequence optimization was modeled simultaneously at resistant positions, we recovered five known resistant mutations in the active conformation. These results suggested that the Abl1 resistance mechanism targeted substitutions that favored the active conformation. Further sequence variability, explored by ancestral reconstruction in Abl1-KD, showed that neutral genetic drift, with respect to amino acid variability, was specifically diminished in the resistant positions. Since resistant mutations are susceptible to chance with a certain probability of fixation, combining methodologies outlined here may narrow and limit the available sequence space for resistance to emerge, resulting in more robust therapeutic treatments over time.

KEYWORDS

Abl1, ancestral reconstruction, conformational energy, kinase, resistant positions

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

1 | INTRODUCTION

Acquisition of drug resistance is a major strategy that diseases, such as cancer and pathogens, use as an escape route against drug therapies. During the treatment, the medical intervention kills the susceptible population, while the resistant one continues to propagate serving as seeds for resistant genes that explore a new landscape of fitness mutations.^{1,2} It is notorious in the field of targeted drug therapy, where the treatment of, for example, cancer has positively impacted survival rate in patients.³ However, the development of new drugs more robust toward resistance mutations is inefficient due to the limitations of treatment period versus the time needed for escape mutations to appear. Better approaches are therefore needed that take into the account possible resistance mutations during the initial stages of drug development.⁴

Loss of control mechanisms during cell growth may lead to cancer, as in chronic myeloid leukemia (CML), when the constitutively active BCR–Abl1 oncogene is expressed in the stem cells of bone marrow.^{5,6} First line therapy against CML uses small molecule inhibitors as imatinib, dasatinib, nilotinib, and bosutinib that prevent ATP binding to the kinase domain (KD) of BCR–Abl1, leading to growth impairment followed by cell death.⁷ The Abl1-KD structure has been experimentally determined in three distinct conformational states (Figure 1a), that is, inactive (PDB ID 2hyy), Src-like (PDB ID 2g1t), and active (PDB ID 2g2i).^{8,9} These states have N- and C-terminal lobes that share an interface with a highly mobile activation loop (Figure 1a, Figure S1) that flips

~180° between inactive and active conformations.¹⁰ Each inhibitor binds specifically to a given conformation of the Abl1-KD, however, distinct genetic mutations located in the Abl1-KD have been clinically found and correlated with the relapse.¹¹ Due to the dynamic nature of Abl1 and its ability to sample several biological relevant conformational states, drug resistance may emerge by synergy or consequence of many escape routes like in situ steric hindrance, reshaping of the conformational energy landscape, increase in free energy of inhibitor binding, and altered dynamics.^{6,12–18}

To expand our understanding of how drug resistance emerges in relation to conformational energy landscape and evolution, two approaches were investigated in parallel. Our initial hypothesis assumed that the resistance mutations may be accessible to certain extent through neutral genetic drift, which would be confirmed by calculations on Abl1-KD. Our second assumption was that the sequence variability in the resistant positions in the ancestors should be more pronounced as they are not yet specialized as the current sequences, especially with respect to those positions not directly in contact with a substrate. However, it is important to note that these sites may still be under the selective pressure to keep a kinase in a state that is less active and responsive to activation upon mutation.^{19,20} To start, we evaluated the mutational energy landscape of Abl1-KD in three known meta-stable states: inactive, Src-like, and active.⁹ In each conformation, every position was sampled by 20 common amino acids and its residue and total energy were calculated using Rosetta macromolecular modeling program.²¹

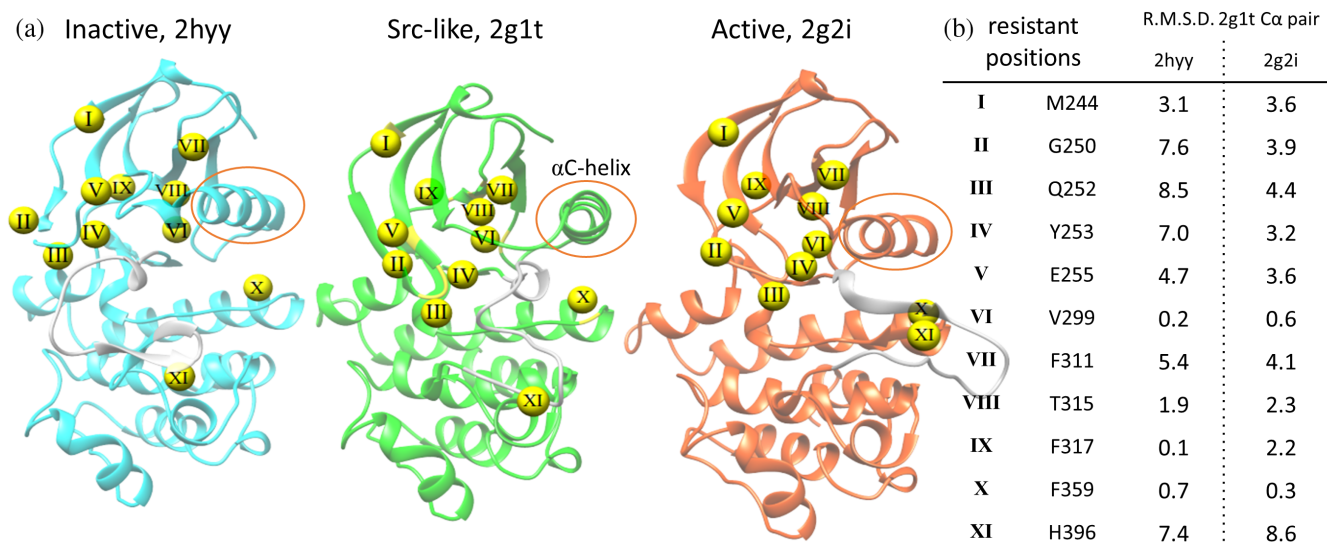


FIGURE 1 (a) Conformational states of Abl1-KD: inactive (cyan), Src-like (green), and active (red) structure with positions of residues subject to resistance mutations showed as yellow spheres (PDB ID 2hyy, 2g1t, and 2g2i, respectively). The activation loop (gray) contains the H396 mutation (XI); α C-helix is marked with the orange ellipse. (b) Native residues at the resistant positions; the C α R.M.S.D. (Å) for 2hyy and 2g2i conformation relative to 2g1t structure as reference

Here, we specifically tracked 11 positions in Abl1-KD known to give rise to clinically relevant resistant mutations M244V, G250E, Q252H, Y253H, E255(K,M,V), V299L, F311I, T315(I,M), F317L, F359(V,C), H396(P,R) (Figure 1b), and compound (double) substitutions that arise concurrently with the gatekeeper mutation, T315(I,M). Substitutions in these positions are found in clinical CML-patients before and during treatment^{22,23} and they confer resistance to a certain kinase inhibitor, for example, imatinib-resistant (E255, Y253, T315),^{24,25} dasatinib-resistant (V299, T315, F317),²⁶ nilotinib-resistant (M244, Y253, E255, T315, F359)^{6,27} in addition to compound mutants with the pan-resistant T315I that confer resistance against the highly potent Abl1 inhibitor, ponatinib.^{6,11,14,28} Next, we investigate the evolutionary history of the Abl1-KD to determine the residue variability at resistant positions in modeled ancestrals. This was accomplished by inferring the maximum likelihood sequence of ancestors and extracting their residue probabilities per position. Expression and characterization of four ancestral sequences benchmarked our reconstruction to support the theoretical investigation.

The results indicated that mutations in resistant positions could reshape the conformational energy landscape by favoring either the active conformation or the inactive conformation, suggesting that resistance mechanism in these cases may not entirely rely on conformational stabilization. The ancestral reconstruction revealed that resistance positions were less sampled over the course of natural selection, if compared to positions where neutral genetic drift had taken place. Overall, the resistant mutations were predominantly selected to favor the active state and their positions were subjected to restrictions based on evolutionary constraints. This agrees with the correlation that increases in higher specific activity confers a potential advantage during growth and selection upon treatment.

2 | RESULTS

2.1 | Evaluation of the state-specific conformational energy

The native state of a protein is described by the average of an ensemble of biologically relevant meta-stable states or conformations which are sampled according to their relative free energy. To build a more detailed understanding of the accessible Abl1-KD conformational energy landscape, we determined, for the three most relevant conformations: active, Src-like, and inactive, the energy distribution for the native sequence using Rosetta structural refinement protocol.^{29,30} The total energy

distribution of 9,240 trajectories for each of the three main conformations (Figure 2a) showed that the Src-like structure has a substantially lower energy followed by the inactive and then active states. The inactive conformation was about 4.4 ± 2.2 Rosetta energy units (REU) lower than the active one (*t*-test, *p*-value < .05) and the Src-like structure was lower by additional 46.0 ± 2.3 REU compared to the inactive state (*t* test, *p*-value < .05). Despite the statistically significant difference in total energy for native sequences, the individual wild type residues have energy interactions that correlate well across positions (Figure 2b). Src-like and active conformations scored the poorest per residue energy correlation ($R^2 = .66$) and had the most dispersed energy at resistant positions (Figure 2b), while Src-like and inactive conformations correlated better ($R^2 = .77$). The substantial contribution to the energetic stabilization in Src-like structure compared to the inactive and active states (under cutoff of -4 REU, Figure 2c) came from seven residues: Y253 (resistant position), D276, M278, G383, L387, T389, and D421. Although these residues are situated in regions critical for ATP binding and kinase regulation (P-loop, α C-loop, and the activation loop; Figure 2d),¹⁰ their lower energies in the Src-like structure are explained by the many contacts surrounding them (Figure S3). Other explanation for the lower energy of the Src-like state may be a necessary compensation to offset destabilizing interactions upon solvent rearrangement (to solubilize the α C-helix, Figure 1a) as well as the possible entropic cost involved in the rearrangement of water molecules (not explicitly modeled). Because our method strongly preferred the Src-like state, we focused the evaluation between the inactive and active state where the “small” difference (4.4 ± 2.2 REU) was ideal to compare the effects of single mutations. The mutations change the total energy between 0 and 6 REU, allowing fair comparison between inactive and active state. These findings indicated that the native Abl1-KD is energetically stable with respect to the enthalpy of the system in nonactive states and suggested that resistant positions in the modeled structures have a substantial energy to impact the transition from inactive to Src-like conformations towards the active state.

2.2 | Computational site-saturation mutagenesis for each amino acid position in Abl1-KD

Here we evaluated how single mutations affect the relative conformational energy landscape of Abl1-KD to categorize positions and residues that may have the strongest impact on the equilibrium balance between the

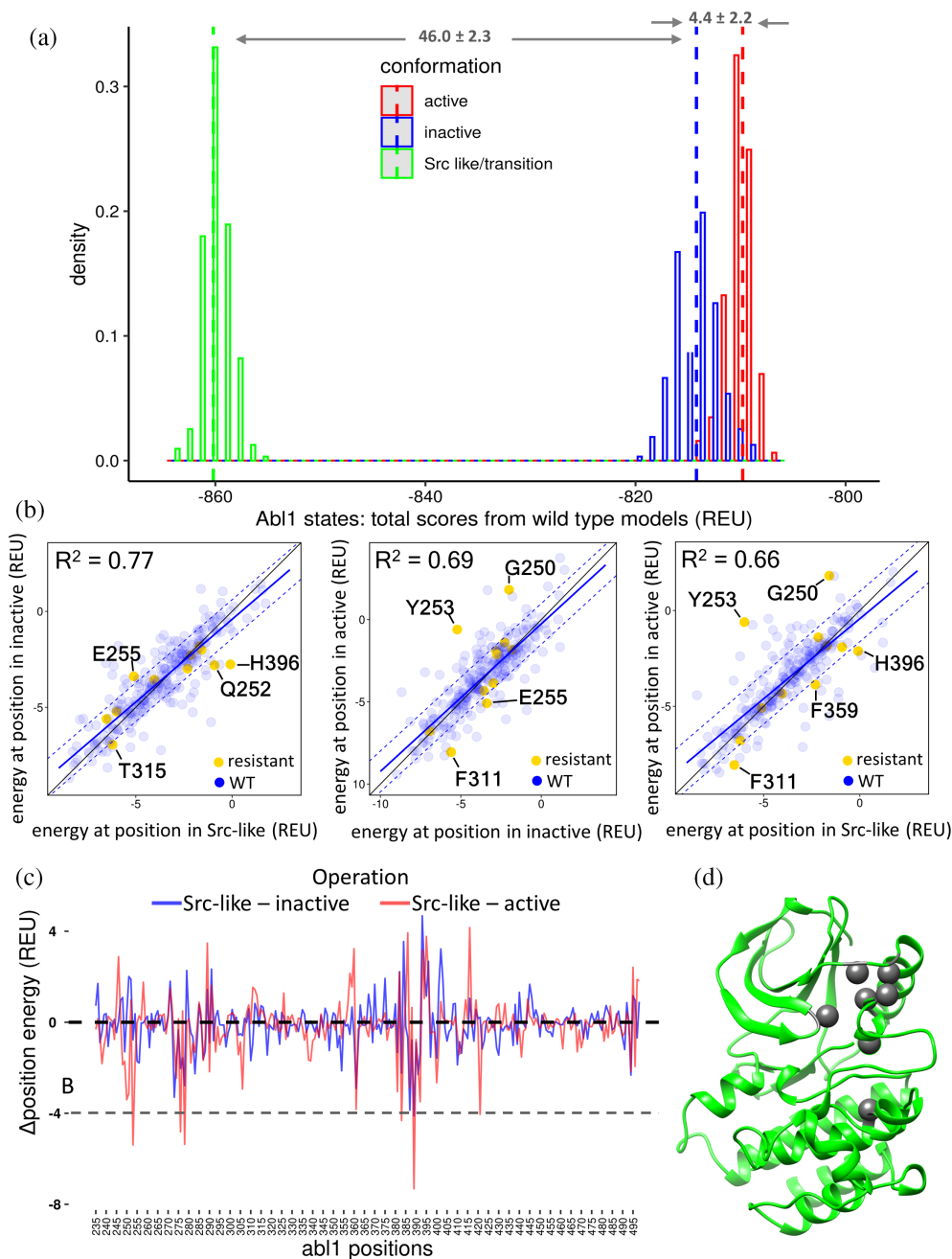


FIGURE 2 Total energy of native Abl1-KD and residue energy pair comparison between conformations. (a) Histogram of the total energy; dashed lines are the mean for the Src-like conformation (green; -860 ± 1 REU), inactive conformation (blue; -814 ± 2 REU) and active conformation (red; -810 ± 1 REU). (b) Correlation plots for each pair of residue energies between conformations; solid line, $x = y$; blue line is the best linear fit with standard error of estimate as dashed lines; yellow dots are resistant positions analyzed in this study.

(c) Residue energy difference between Src-like and inactive (blue) and Src-like and active (red); a dashed grey line cutoff the most stabilizing residues in the Src-like compared to active and inactive structure. (d) Gray spheres highlight in the Src-like structure for which residues the energy difference is most negative compared to inactive and active conformations (c.f. in Figure S3). REU, Rosetta energy units

conformations. All 20 amino acids per position were sampled in the Abl1-KD with Rosetta in the three conformations outlined in Figure 1. Residue energies were converged, $R^2 = .99$, while total energy showed lower correlation, $R^2 = .78$, according to the independent test runs (Figure S2). Comparisons with total energy between mutants can thus give general trends, as it was only partially converged due to sampling limitation, while the residues energies were accurate enough to evaluate the mutational effect in situ. Therefore, it is more relevant to focus on investigating the relative differences in residue energies.

Next, the total energy for all single mutants and their individual residue energies were compared among the

conformational states, to identify the structural influences upon mutation. A weak linear tendency between different conformations was observed for the total energy (Figure S4): $R^2 = .45$ Src-like vs. inactive, $R^2 = .39$ inactive vs. active, and $R^2 = .38$ Src-like vs. active, indicating the possibility to shift conformational ensemble upon mutation. Interestingly, mutants at resistant positions had an even lower correlation of total energy between conformations (Figure S4), where the correlation between the inactive and active states was affected the most ($R^2 = .08$).

The correlation of residue energies between Src-like and active conformations resulted in the most divergent energy per position ($R^2 = .48$, Figure S5B), while between

inactive and active (Figure 3a, $R^2 = .66$), mutations kept similar energetic correlation as seen with wild type residues (Figure 2b). Apart from the overall behavior of the energy-specific conformational states for single residue substitutions, we assessed residue substitution effects for the resistant positions (Figure 3b, Figure S5). Overall, the correlation of residue energy was similar between the conformations with a slightly lower correlation between inactive and active conformation (Figure 3b, $R^2 = .53$). We note also that the gatekeeper position (T315) was substantially stable (energetically favorable in terms of calculated Rosetta energy units with respect to the compared state), as both native and resistant residue (energy for the corresponding amino acid is less than -5.0 REU) in all three conformational states. A cluster of six resistant mutants, G250E, Q252H, E255M, E255V, F359C, and F359V were close to destabilizing (residue energy ~ 0 REU) in the inactive conformation compared to both active and Src-like states (Figure 3B, Figure S5A).

2.3 | Effect of residue substitution on resistant positions

To investigate mutational susceptibility of the resistant positions, we plotted the total energy, residue energy, and relative conformation-specific energy of each mutant (Figure 4 and Figure S6). The difference in conformational (total) energy for these positions showed that Src-like state was still more stable than the other two states (Figure S6C,E). Change in residue energy after resistant

mutation, were more stable in the Src-like compared to the inactive state for residues G250E, Q252H, E255V, T315(I,M), F359(C,V), and H396R; comparison between Src-like to the active state showed residue energy in T315 (I,M), F317L, F359V and H396R to be more stable in the Src-like conformation while residues G250E, Y253H, V299L, and H396P became more stable in the active conformation (Figure S6D,F). Despite the fact that resistant residue energy of the gatekeeper is more stable in the Src-like conformation, the total energy of mutants T315I,M remained close to the native (Figure S6A,C,E). More importantly, the relative residue energy between inactive and active states for resistant mutants (as seen in the Figure 4f, red arrows) revealed that the resistant substitutions G250E, Q252H, Y253H, E255V, V299L, F359 (C,V), and H396P became more favorable in the active conformation. Because the Src-like structure is so low in energy, evaluation will focus on the relative energy between inactive and active state, otherwise, mutants would need to greatly stabilize the active state relative to Src-like state to be meaningful, additionally the canonical inactive and active structures have been historically used to design kinase inhibitors^{8,25,28,31,32} and therefore are more appealing to understand inhibition resistance in Abl1-KD.

The change in residue energy (or relative residue energy) between conformations showed that we can identify in situ energy changes that reshape the energy landscape of the states. The mutations at the gatekeeper position did not produce a substantial change in the total and residue energy in either of the inactive and active

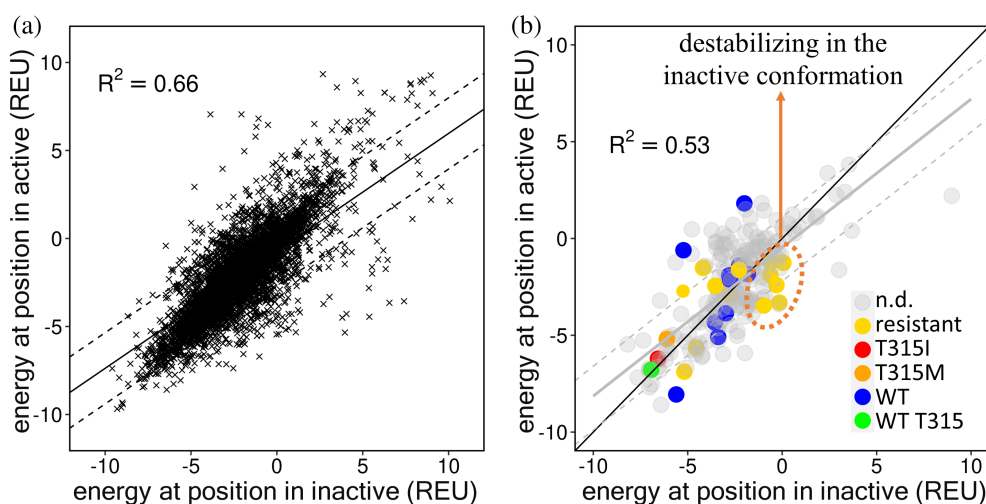


FIGURE 3 Correlation plot of sampled 20 amino acid residue energies per position in inactive and active Abl1-KD calculated for all positions in the structure (a) and plotted only for resistant positions (b). Dots in blue are energies from wild type residues in those positions; in yellow are the studied resistant mutations; in green, orange and red are T315, T315M, and T315I, respectively. Grey denotes experimentally not determined (n.d.) mutations to be resistant. Reference line ($x = y$) was drawn only for the right box (b). All plots show the best linear fit with standard error of estimate as dashed lines (p -value $< .05$)

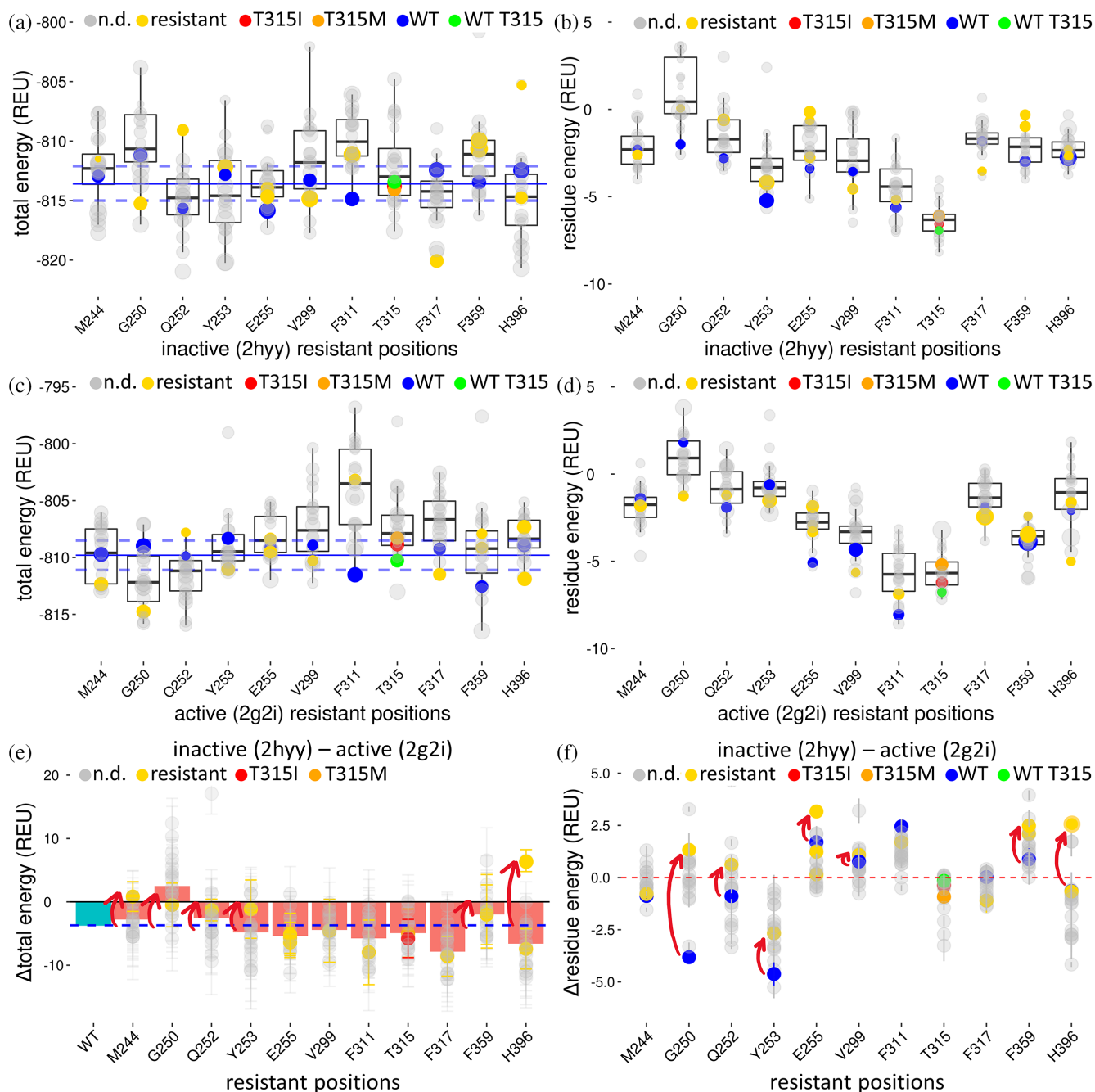


FIGURE 4 Site saturation mutagenesis for total (left graphs) and residue (right graphs) energies for resistant positions of Abl1-KD. Total and residue energy, respectively, of inactive (a,b), and active (c,d) states. Relative total energy (e) and relative per residue energy (f) in resistant positions between inactive and active states. Red arrows show the change in energy that stabilize the active conformation after mutation to a resistant residue. Red bars (e) are the relative energy average between the 20 amino acid substitutions per position and blue lines are the average energy of inactive and active with native sequence (same in Figure 2a). Dots in blue, mean total energy of native structure (left graphs) or residue energy from wild type residue in that position (right graphs). In yellow are from resistant mutations, in orange and red are T315M and T315I, respectively; grey indicates experimentally not determined (n.d.) whether substitutions are resistant. Size of the dots is proportional to the standard deviation

states, both wild type and its resistant substitutions (T315I and T315M) are highly stable (< -5 REU; Figure 4a–d). The effects of substitutions on Thr³¹⁵ residue, are probably best explained in the context of

inhibitor (imatinib) either due to direct contacts or by induce fit events.^{8,12,33,34} Otherwise, to see the substitution effects on the inactive conformation, a structure has to be chosen in which such destabilizing effects have the

possibility to occur. The modeled inactive state here (PDB ID [2hyy](#)) is originally the holo-form with imatinib. In this complex, the residue sidechain of Thr³¹⁵ points towards the inhibitor; removing the inhibitor leaves a space that can be filled with different residue side chains, quite favorably according to our calculations. We also tested the effect of two compound mutants w.r.t. the gatekeeper mutation: G250E/T315I and Y253H/T315I. These inverted the relative total energy in favor to the active conformation (Figure S7C,D) and in a greater extent than the compound mutant G250E/V299L.

Evaluation of amino acid changes in resistant positions that shifted the conformational equilibrium towards the active state, that is, with the most pronounced residue energy difference between the active conformation in comparison to the inactive (excluding native or resistant identities studied in this work) were M244P, G250I, Q252P, Y253K, V299Y, and F359W (Figure S8A); and the mutants that contribute most to the total energy of the active conformation in comparison to the inactive were G250V, Q252P, Y253K, E255W, V299R, F311W, T315H, F317Y, and F359W (Figure S8B).

Together the results show that known resistant positions greatly impact and reshape the relative conformational energy of Abl1-KD. One reason may be from inherent instability of a given conformation, but the overall trend in Figure 3 showed that sampled mutations in resistant positions are about 53% energetically correlated between conformations, thus roughly half of all possible mutations in these positions will have different residue stability according to the state and can impact the total conformational energy. The known resistant mutation that produced the highest stabilization to the active conformation was the H396P (total energy difference = $6.5 \pm$

1.7 REU, residue energy difference = 2.5 ± 0.5 REU, Figure 4a,d-f) and agrees with previous structural observations where this variant makes the active conformation more available by destabilizing the inactive conformation.^{33,34} Substitution that does not change its stability regarding a residue energy in different conformation, such as in M244V (Figure 4f), still may lead to changes in relative total energy between the conformations (Figure 4e) and highlights a possible allosteric stabilization over the structure. Resistant mutations in the gatekeeper position (T315) did not compromise conformational stability and this position was found to be quite tolerable for substitutions (Figure 4f).

2.4 | Designing Abl1-KD in resistant positions

The refinement of the different Abl1-KD conformations was converged for residue interaction energies, while total energy showed potential for better convergence (Figure S2). To thoroughly sample the total conformational energy of mutants in resistant positions we carried out more extensive Rosetta sampling to characterize the optimal sequence for each conformation by allowing design to all 20 residues in the 11 investigated resistant positions. The energetically most favorable residues that were sampled for each conformation are presented in Table 1. The energetically lowest active conformations sampled preferentially five identical resistant residues during the sequence optimization process and five chemically similar ones. In contrast, in the inactive conformation the designed sequence corresponded to only a single identical resistant residue and four chemically similar ones.

TABLE 1 Sequence optimization of Abl1-KD resistant positions w.r.t. studied conformations

WT	Resistant mutation	Design inactive 2hyy	Design Src-like 2g1t	Design active 2g2i
M244	M244V	<i>M244I</i>	<i>M244F</i>	<i>M244P</i>
G250	G250E	G250N	WT	G250E , G250N
Q252	Q252H	<i>Q252K</i>	WT	<i>Q252K</i> , Q252D, WT
Y253	Y253H	Y253L	WT	Y253D
E255	E255V, E255M, E255K	<i>E255A</i>	WT	WT , E255Q, E255K
V299	V299L	V299L	WT	V299Q
F311	F311I	<i>F311L</i>	WT	<i>F311M</i>
T315	T315I, T315M	WT	T315M	<i>T315L</i> , T315M
F317	F317L	F317Y	WT	<i>F317M</i>
F359	F359V, F359C	F359K	WT	F359V
H396	H396P, H396R	H396N	H396S	H396P

Note: Residues in italic are chemically similar to the known resistant residue; bold marks the recovered resistant residues. Ten lowest energy sequences were analyzed.

The data confirms that resistant mutations are more likely to stabilize the active conformation than the inactive, that is, that the kinase native sequence has been evolved to be minimally active, but ready to be activated. The designed sequences for the Src-like structure sampled eight identical wild type residues converging to a minimum energy, meaning that current native sequence was already near its lowest energy sequence. Average total energy of designed sequences after the optimization was $E_{\text{inactive}} = -819.0 \pm 1.7$ REU, $E_{\text{Src-like}} = -869.4 \pm 0.6$ REU, and $E_{\text{active}} = -820.4 \pm 1.5$ REU. These energies were lower compared to the energy of each conformation with native Abl1 sequence shown in Figure 2. With this approach, the resistant mutation T315M in the gatekeeper was sampled in the Src-like and active conformations, while the inactive conformation retained the native threonine. Moreover, positions G250, F359, E255, F359, and H396 were designed to known resistant mutants in the active conformation (Table 1). Thus, it seems likely that compound mutations or combination of them in resistant positions are steering the conformational ensemble to stabilize the active conformation. The exceptions were positions Y253 and V299 that did not sample either chemically similar resistant residue or wild type residue in the active conformation. The resistant mutation V299L was identified in the inactive design and not in the active state during the sequence optimization step (Table 1), which indicates a resistance mechanism where imatinib binding may be destabilized in the inactive state. Also, because the mutant and compounds clones with V299L are selected after dasatinib treatment,^{35,36} the kinase inhibitor that binds the Abl1-KD in the active conformation,^{37,38} V299L could sterically hinder dasatinib binding. The V299L mutation may therefore work through a combination of resistance mechanisms where conformational preference has a minor contribution (in accordance with our site-saturation calculations, Figure 4f).

2.5 | Evolution of Abl1-KD and ancestral sequence design

Since the data indicated that the investigated resistant positions were to a certain degree sensitive to sequence optimization, we investigated what impact evolution had on those resistant positions. The evolutionary aspect was assessed by building a phylogenetic tree and making ancestral sequence reconstruction (ASR) of Abl1-KD ancestors starting from anc0 (the common ancestor of genes Abl1 and Abl2) to anc9 (the common ancestor of Abl, Src, Tek, Csk, and PKA; Figure S9). Each ancestor was derived based on amino acid probabilities for each

position (stored in the format of position specific scoring matrix, PSSM, which contains the probability of an amino acid type for each position in the sequence). For instance, highly conserved structural or functional residues will have a high probability for that amino acid over the course of evolution. The identity for the inferred sequences compared to the current native Abl1-KD sequence varies from ~90% identical for close relatives to ~50% identity for the most distant nodes and we could also compare the inferred variant in this study, Abl-Src common ancestor anc6, to a previously published inferred sequence (Figure 5a,b).¹² The sequences of anc6 variants were 66% identical to Abl1 compared over the length of 264 amino acids (Figure 5a).

To evaluate the variability of residues in ancestral nodes and assess their possible mutagenesis potential during selection, we divided the sequence into three groups: 11 functionally conserved positions, 11 resistant positions, and the remaining ones. For each group we evaluated the residue variability per position (Figure 5c). As expected, the catalytic and functionally critical residues (for instance, DFG and HRD motif) were fixed over time. Importantly, the resistant positions were sampled at a lower rate than the remaining positions. First after anc8 node there was a tendency even for the resistant positions to drift on the same level as the nonconserved positions.

To further explore the ancestral sequence variability of Abl1-KD we sequenced optimized ancestors applying Rosetta design strategy. Specifically, we took into the account the specific conformations combined with the residue variability inferred from phylogenetic PSSM. Thus, we obtained ancestors with two designed sequences, for the inactive and active conformations. The consensus sequence was retrieved from 10 lowest energy structures from each conformation, a strategy that allowed us to merge the conformational landscape encoded with their respective sequences. The final designed consensus ancestral sequences (PhyML vs. Rosetta) were compared to the current Abl1-KD (Figure 5a,b and Figure S10) where sequence identity and similarity to Abl1-KD decreased in designed ancestors similarly as inferred by the maximum likelihood method. With this framework, relevant physical information was incorporated with a desirable degree of sequence conservation found during evolution.

To test the coherence of the design and reconstruction approach, we resurrected and expressed four ancestral genes for experimental validation: two nodes inferred by phylogenetics (anc6 PhyML and anc8 PhyML) and their sequence optimized variants (anc6 Rosetta and anc8 Rosetta). All resurrected ancestors were soluble (Figure S11) and were tested experimentally for activity

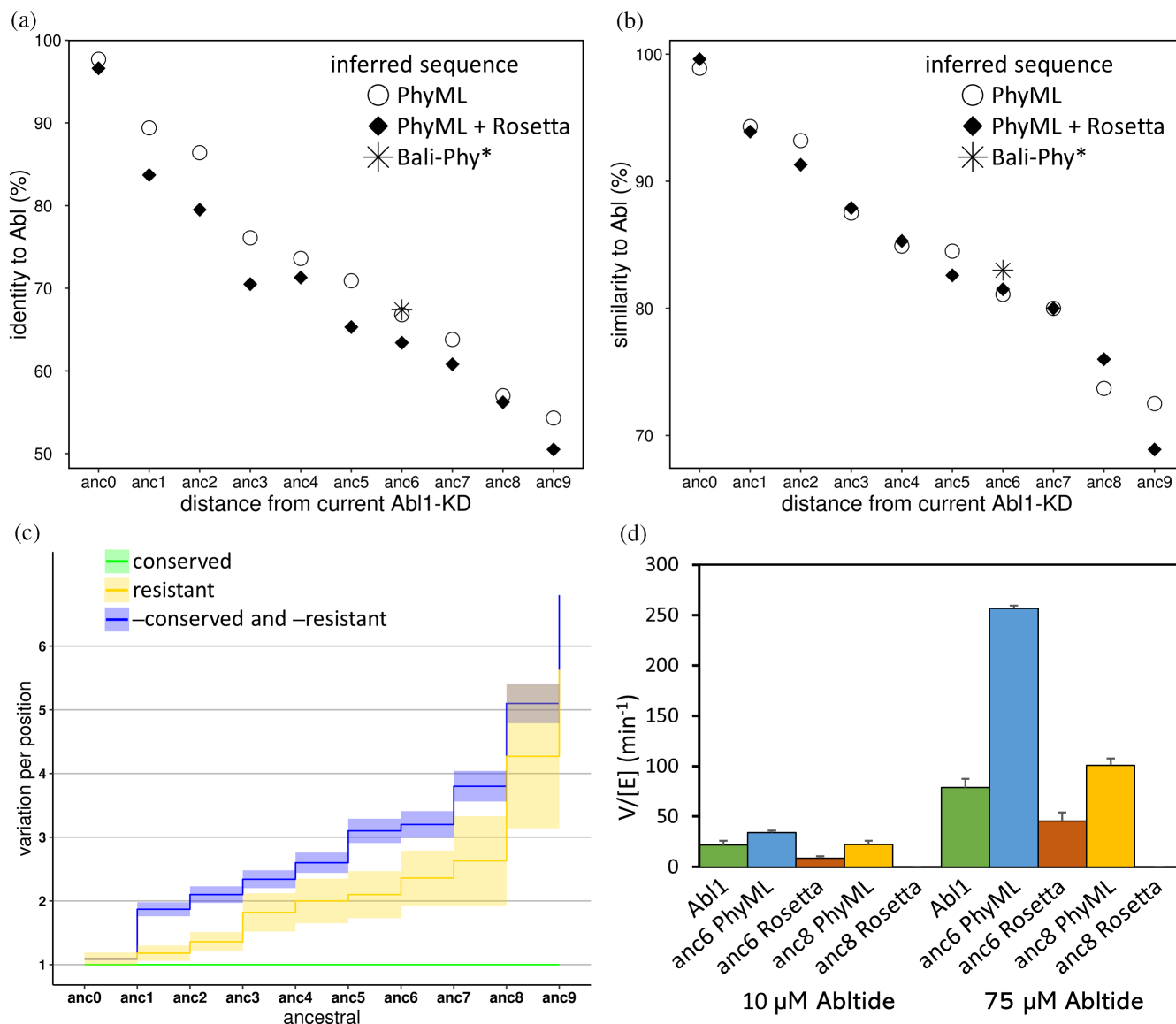


FIGURE 5 Sequence comparison and kinase activity of Abl1-KD ancestors. Percent sequence identity (a) and similarity (b) compared to native Abl1-KD sequence as reference; sequences were ASR inferred by phylogenetics using PhyML or after optimization with Rosetta (PhyML + Rosetta); the sequence of the common Abl1Src ancestor (PDB ID 4csv) was added for comparison in anc6 labeled with “Bali-Phy*.” (c) Variability of residues per position in ancestors with positions clustered in three groups: functional (green), resistant (yellow), and the remaining ones (blue); shaded deviation, standard error of mean (SEM). (d) Catalytic activity assay of kinases in four replicates shown as mean \pm SD. Only anc6 was significantly more active at 75 μ M Abltide concentration than the native Abl1, while anc8 Rosetta had no measurable activity at tested concentrations

by measuring the phosphorylation rate of Abltide substrate at two different substrate concentrations, 10 μ M and 75 μ M, respectively. The Abl1 activity was on parity with previous measurements where k_{cat} and K_M with Abltide were determined to 66 min^{-1} and 17 μ M, respectively.³⁹ Previous validation of ancestors using Bayesian phylogenetic analysis in Wilson et al.¹² indicated that a last common ancestor of Src and Abl would have similar catalytic rates to current Abl1-KD. In our study, the inferred common ancestor from Src and Abl, anc6

PhyML, is about three times more active than the current Abl1-KD (Figure 5d) at the tested substrate concentrations, indicating that ML based ancestral reconstruction successfully captured functionality in kinases and suggests that current Abl1 evolved toward less active kinase. The designed anc6 Rosetta clearly impaired the gain in activity, highlighting the dilemma between function preservation and structure-based sequence optimization for overall stability. This became even more apparent with the anc8 PhyML that retained similar degree of activity

compared to the native Abl1-KD, but its designed variant (anc8 Rosetta) was inactive at the assayed substrate concentrations (Figure 5d). Therefore, ML-based phylogenetic algorithms seem to be better at generating functional proteins than the outlined design strategy that optimized the total energy of the protein without including any functionally relevant information, for example, modeled substrates or reaction transition states.

3 | DISCUSSION

Our results showed that the active conformation as the least stable compared to the Src-like and the inactive states, which agreed with a tight kinase regulation keeping the structural ensemble shifted away from the active state. Src-like state was, however, found to be more stable than the inactive conformation (Figure 2), which is on the contrary to what would be expected from a transient conformation between the active and inactive states.^{40,41} Although it has been evaluated and hypothesized that the Src-like conformation is critical for regulation due to the flip of the DFG motif and possibly connected with the resistance mechanism of some resistant mutations.⁹ The low energy of the Src-like conformation may also be necessary to compensate for the entropic penalty when the α C-helix (Figure 1a) rotates outwards from the N-terminal domain. Our calculations do not take into the account the free energy of the whole system where the contribution of the water molecules is probably a significant factor. The data showed moreover that the active state of Abl1-KD was less stable than the inactive state. Overall, the conformational energy landscape was shifted towards the inactive states. This also agreed with the available experimental and molecular dynamic studies where the inactive state of Abl1 and other kinases homologues were more frequently observed.^{11,16,39,42,43}

The high throughput mutational scan presented an overview of the energetic variability, that is, the residue plasticity, for Abl1-KD conformational states and tendencies for energy correlation between them. Resistant positions were impacted differently in each conformation after evaluation of residue energy for each amino acid at every position. Evaluation of the energetic variability in a position given a structure may show clues about the mutational susceptibility at that position. The relative total energy and the difference in residue energy allowed us to conclude that single resistant mutations M244V, G250E, Q252H, Y253H, E255V, F359VC, and H396P energetically favored the active state (Figure 4e,f). Available NMR experimental data of Abl1-KD showed that mutants Y253H, E255V, F359V, and H396P were responsible for the shifts in the conformational assemble towards the active state.³³ Indeed we show that their

residue and conformational total energies (except the E255V total energy) were more stable in the active conformation (Figure 4e,f), although the same work found the active conformation to be the most frequent state in the solution. Other studies demonstrated an increase in the measured catalytic efficiency for G250E and Y253H mutants^{17,39} which can be explained by the stabilization of the active state (Figure 4c–f) leading to drug resistance.

The same methodology was applied to the relative conformational energy of the Src-like state and may help to delineate more susceptible positions. For instance, the resistant mutant H396R that has unchanged residue energy compared to the native H396 (Figure 4f) but the total energy favored the inactive state (Figure 4f). However, H396R total (−863.7 REU) and residue energy (−1.9 REU) stabilized the Src-like conformation (Figure S6A,B), necessitating the evaluation of the intermediate Src-like state in study of resistance. In support of future drug development efforts, mutants that stabilize the active conformation and to date not experimentally validated as resistant variants (Figure S8), could be tested to minimize possible emergence of drug resistance.

The Abl1-KD shares structural homology with numerous other kinases from diverse organisms differing mainly in substrate specificity, that is, the protein target they phosphorylate.⁴⁴ Since mutations may arise in structurally similar positions and sample similar residues in homologs, there is a possibility that resistant residues in kinases may be predicted based on evolutionary analysis. Thus we decided to ancestrally reconstruct Abl1-KD using homologue sequences chosen based on a hypothesis-driven argument where current mammalian TK regulatory mechanism was specialized since premetazoan age.^{45–48} Indeed, Abl ancestors from the premetazoan times had been biochemically characterized, showing unique set of enzymatic behavior that is distinct from the current Abl1.^{12,45} The evolutionary variability was benchmarked here by resurrecting ancestors with specific conformations and probabilities based on PSSM along the ancestral nodes (Figure 5 and Figure S9). Our ancestral reconstruction showed that the resistant positions in Abl1-KD are more constrained during its evolutionary course (Figure 5, Figure S10). However, to a certain degree, protein design methodology applied here did not account for function as expected and ended up abolishing the enzyme activity as observed with anc8 Rosetta. To further quantify and test these assumptions it would be necessary to perform kinetic experiments of optimized Abl1-KD ancestors for each conformation and to quantitatively evaluate the trade-off between function and stability.

Patients with CML may develop resistance against drugs that were specially developed to target Abl1-KD⁸

and any new drugs may be subjected to reduced efficacy due to resistant mutations.^{4,6} Thus, there is a need to counter drug resistance occurring during CML targeted therapy and during any disease where the risk of resistance is significant. We sought to understand Abl1 resistance susceptibility with two approaches, by systematically evaluating conformational energy plasticity due to all possible substitutions in the isolated Abl1-KD and along its evolutionary trajectory. These approaches were hypothesized previously, where resistant mutations propagate allosteric activation effect in the kinase domain making it more active,^{15,19} because substitutions in mildly conserved positions that stabilize the inactive conformation are involved in resistance after abrupt selective pressure. It has been shown that homologue kinases in different cancers share conserved resistant residues in similar positions after treatment,⁴⁹ therefore, evaluation of residue variability over specific kinase ancestors can be used as strategy to find hotspots positions to test a surmountable number of single or compound mutant variants for in vitro analysis with new drugs. Analysis of ancestral sequences sampled by kinases during evolution followed by evaluation of mutational effects on the conformational energy landscape as outlined here, may be a viable methodology to predict possible drug resistance in kinases where it poses a significant risk to cancer patients.

4 | CONCLUSION

High selection pressure during cancer drug treatment may evolve the sequence space towards regions that are not normally accessible. In an effort to explore alternative methods for the prediction of resistant sites and their residue identities, we carried out sequence sampling at each position to identify susceptible sites and how they respond to the environment with respect to the three different conformational states. Abl1-KD has a conformational plasticity such that even a single mutation can reshape energetics of opposing functional states, that is, inactive versus active, to confer adequate fitness. This property is not exclusive to resistant mutations only, many other substitutions can reshape the conformational energy in resistant positions but do not lead to resistance. This results in a great sequence variability and combination of mutations that can modify the conformational landscape. We demonstrate that the methods here can to certain degree recover the occurring resistance mutations. Ancestral reconstruction analysis revealed that resistant positions are distinguishable from random neutral drift over the natural selection course, making it possible to further investigate mutants in positions that are mildly conserved with residues that stabilize the active

conformation. Taken together, the methodology may narrow down the predicted sequence space accessible for resistant variants in the case of Abl1.

5 | METHODS

5.1 | Abl1-KD structure models

The Abl1-KD structures, with PDB IDs 2hyy, 2g1t, and 2g2i, were trimmed N- and C-terminally to start at residue W235 and end at residue Q498. Structures 2hyy and 2g2i had their missing atoms modeled by Rosetta. Global, structural fitting was made with Chimera using Needleman-Wunsch algorithm with BLOSUM-62 matrix to calculate R.M.S.D. between C-alfa atoms.

5.2 | Structural refinement of native and mutated Abl1-KD

RosettaScripts²⁹ application part of the Rosetta⁵⁰ protein modeling software was used for design, applying the resfile methodology to specify a specific amino acid residue at a certain position in the sequence. PackRotamersMover and FastRelax³⁰ were used with five iterations (c.f. Rosetta example files in Supplementary Information). All 20 amino acids were sampled per position running 35 trajectories each, thus a total of 184,800 mutated and native sequence models were calculated per conformation. Accordingly, the combination of mutations for compound variants was evaluated. To compare the total energy and residue energy for the site-saturation for different conformations, each variant or modified residue had its mean energy and standard deviation calculated from the five lowest values. To analyze data for native conformations, all sampled native sequences were used from modeling (9,240 models per conformation). For each native residue we took the mean energy and standard deviation from the five lowest values. In total, 264 total mean energies for each conformation were analyzed. To evaluate the minimum convergence of our calculations, two independent sets of 15 trajectories were determined for the 2hyy structure using the same FastRelax³⁰ protocol. As above, the total energy and individual residue energy average from the five lowest values were correlated to assess convergence.

5.3 | Sequence optimization of Abl1-KD

To design Abl1-KD in the inactive, Src-like, and active conformation, a similar FastRelax protocol was used (script in Supplementary Information), but instead of

changing one amino acid per position, only the 11 resistant positions were allowed to change to 20 canonical amino acids during the same calculation trajectory. In total, 750 trajectories per conformation were calculated. The 10 lowest energy sequences of each conformation were aligned, and their sequence identity was analyzed. Residues in resistant positions were annotated.

5.4 | Ancestral reconstruction of Abl1-KD and ancestral design

The Abl1-KD sequence (UniProtKB—P00519) was the query to find homologue tyrosine kinases (TK) in the Uniprot database⁵¹ that are SH3-SH2-TK-family proteins restricted to the subgroup of Abl1, Src, Tec, Csk, and two outgroups of protein kinase A (PKA)⁵² and fibroblast growth factor receptor, yielding a total of 41 sequences. Sequences were aligned in GUIDANCE2⁵³ web-server with default MAFFT algorithm followed by phylogenetic tree inference using the software MEGA-X⁵⁴ version 10.1.7 with the Maximum Likelihood method and substitution model parameters LG + G + I with 1,000 bootstraps replicates. Ancestral sequences were inferred using PhyML⁵⁵ version 3.3 (c.f. Supplementary Information for command line arguments). The position specific scoring matrices (PSSMs), extracted with a python script from ancestral reconstruction, were used to determine the residue variability per position.

To infer the residue variability during ancestral reconstruction, the PSSM of each ancestor node was used. The number of residues with probability greater than zero was counted. Three groups of positions were made: (i) a conserved set of 11 functional positions in ancestral sequences that do not change over time (these retained only a single amino acid identity, consequently counted as 1); (ii) the resistant positions, and (iii) a set without functional or resistant positions included. Mean and standard error of the mean of all counted residue variabilities per position for a given group were plotted for all ancestral nodes.

Conformation-specific structures of ancestral sequences were built with Modeller.⁵⁶ Both inactive (PDB ID [2hyy](#)) and active (PDB ID [2g2i](#)) conformations were used as templates. The models were additionally sequence optimized utilizing RosettaScripts with FastRelax and SeqprofConsensus movers. Sequence design was allowed in those positions where the probability was less than 100% and sampling residues with greater than 0% probability according to the inferred PSSMs. Thus, even residues with small probability to occupy a position (less than 5%) had the chance to be incorporated in the final sequence if its energy was favorable. In total, 500 trajectories were simulated. At this stage, each ancestral node had distinct sequences that were optimized in

either the inactive or the active conformation. The 10 lowest total energy sequences per conformation were aligned and the consensus sequence was extracted as the final Rosetta calculated ancestral sequence for a given node. Graphical workflow of this process is outlined in Scheme S1.

5.5 | Expression, purification, and enzymatic assay of Abl1-KD ancestors

We selected the native Abl1-KD (UniProtKB—P00519, residues 229–511) and two ancestor nodes, anc6 and anc8, for heterologous expression in *E. coli* to validate our phylogenetic analysis. Each ancestor node had two corresponding sequences, one inferred by PhyML and the other optimized by Rosetta. Each sequence of interest was synthesized (GenScript) and cloned into pET29b(+) containing a C-terminal hexa-histidine affinity tag. Abl1-KD and ancestors were co-transformed with the YopH phosphatase in pCDFDuet-1b plasmid,⁵⁷ to antagonize Abl1-KD activity and minimize cytotoxicity of *E. coli* BL21(DE3) Tuner cells. Protein expression was induced with 1 mM isopropyl- β -D-thiogalactoside (IPTG) under shaking for 16 h at 20°C in Luria-Bertani (LB) medium with kanamycin and streptomycin (50 μ g/mL each). Cells were harvested by centrifugation (2,300 \times g, 20°C, 20 min), resuspended in Lysis buffer (50 mM HEPES, pH 7.4, 500 mM NaCl, 10 mM imidazole, 20% glycerol, 1% Triton X-100, 5 mM 2-mercaptoethanol) and lysed by sonication followed by centrifugation (30 min centrifugation at 11,770 \times g) to separate cell debris. Supernatant was applied to Co²⁺ affinity column (HisPur™ Cobalt Resin, Thermo Scientific™), washed 3 times with binding buffer (50 mM HEPES pH 7.4, 500 mM NaCl, 10 mM imidazole, 5% glycerol, 5 mM 2-mercaptoethanol), and eluted with elution buffer (binding buffer with 300 mM imidazole). Traces of protein impurities and YopH phosphatase were further removed by ion exchange chromatography using an anion exchange column (Resource Q or HiTrap Q HP) using a linear gradient by mixing buffer A (20 mM Tris-HCl pH 8.0, 5% glycerol, 1 mM DTT, 50 mM NaCl) and buffer B (buffer A + 1 M NaCl). The enzymatic activities of the purified kinases were determined using the ADP-Glo Kinase Assay (Promega Biotech AB), following the manufacturer's manual. The kinases and substrates were diluted in kinase buffer (40 mM Tris pH 7.5, 20 mM MgCl₂, 0.1 mg/mL BSA, 50 μ M DTT). The assay was performed in 20 μ L of kinase buffer containing 3 ng of purified enzyme, ultra-pure 30 μ M ATP, and 10 μ M or 75 μ M Abltide (SignalChem). Reactions in four replicates were done in an opaque white flat bottom half area 96-well polystyrene assay plates (Corning) at 25°C for 1 h and luminescence was measured using TECAN

SPARK 10 M plate reader with an integration time of 1 s. Reaction velocities were obtained by conversion rates (μmol of ADP min^{-1} determined from an ADP calibration curve of ATP–ADP conversion) divided by enzyme concentration measured with Qubit™ (Thermo).

AUTHOR CONTRIBUTIONS

Felipe A. M. Otsuka: Conceptualization (equal); data curation (equal); formal analysis (equal); software (equal); writing – original draft (equal). **Sinisa Bjelic:** Conceptualization (equal); methodology (equal); software (equal); supervision (equal); writing – review and editing (equal).

ACKNOWLEDGMENTS

The computations were enabled by resources provided by SNIC at Lunarc partially funded by the Swedish Research Council through grant agreement no. SNIC 2021/5-401. F.A.M.O. is grateful for the financial support from FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo; grants, 2018/18537-4 and 2019/25955-0), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico). We are grateful to Prof. Ran Friedman for insightful discussions and to Prof. Ingemar André for providing the python script to extract PSSMs.

ORCID

Sinisa Bjelic  <https://orcid.org/0000-0002-9300-614X>

REFERENCES

- Vasan N, Baselga J, Hyman DM. A view on drug resistance in cancer. *Nature*. 2019;575:299–309.
- Merrikh H, Kohli RM. Targeting evolution to inhibit antibiotic resistance. *FEBS J*. 2020;287:4341–4353.
- Baudino T. Targeted cancer therapy: The next generation of cancer treatment. *Curr Drug Discov Technol*. 2015;12:3–20.
- Friedman R. Computational studies of protein–drug binding affinity changes upon mutations in the drug target. *WIREs Comput Mol Sci*. 2021;12:e1563.
- Ren R. Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia. *Nat Rev Cancer*. 2005;5:172–183.
- Braun TP, Eide CA, Druker BJ. Response and resistance to BCR-ABL1-targeted therapies. *Cancer Cell*. 2020;37:530–542.
- Hochhaus A, Baccarani M, Silver RT, et al. European LeukemiaNet 2020 recommendations for treating chronic myeloid leukemia. *Leukemia*. 2020;34:966–984.
- Cowan-Jacob SW, Fendrich G, Floersheimer A, et al. Structural biology contributions to the discovery of drugs to treat chronic myelogenous leukaemia. *Acta Crystallogr D Biol Crystallogr*. 2007;63:80–93.
- Levinson NM, Kuchment O, Shen K, et al. A Src-like inactive conformation in the Abl tyrosine kinase domain. *PLoS Biol*. 2006;4:753–767.
- Taylor SS, Keshwani MM, Steichen JM, Kornev AP. Evolution of the eukaryotic protein kinases as dynamic molecular switches. *Philos Trans R Soc B: Biol Sci*. 2012;367:2517–2528.
- Zabriskie MS, Eide CA, Tantravahi SK, et al. BCR-ABL1 compound mutations combining key kinase domain positions confer clinical resistance to Ponatinib in Ph chromosome-positive leukemia. *Cancer Cell*. 2014;26:428–442.
- Wilson C, Agafonov RV, Hoemberger M, et al. Using ancient protein kinases to unravel a modern cancer drug's mechanism. *Science*. 2015;347:882–886.
- Azam M, Seeliger MA, Gray NS, Kuriyan J, Daley GQ. Activation of tyrosine kinases by mutation of the gatekeeper threonine. *Nat Struct Mol Biol*. 2008;15:1109–1118.
- Shah NP, Sawyers CL. Mechanisms of resistance to STI571 in Philadelphia chromosome-associated leukemias. *Oncogene*. 2003;22:7389–7395.
- Dixit A, Verkhivker GM. The energy landscape analysis of cancer mutations in protein kinases. *PLoS ONE*. 2011;6:e26071.
- Lovera S, Morando M, Pucheta-Martinez E, Martinez-Torrecuadrada JL, Saladino G, Gervasio FL. Towards a molecular understanding of the link between Imatinib resistance and kinase conformational dynamics. *PLoS Comput Biol*. 2015;11:e1004578.
- Hoemberger M, Pitsawong W, Kern D. Cumulative mechanism of several major imatinib-resistant mutations in Abl kinase. *Proc Natl Acad Sci U S A*. 2020;117:19221–19227.
- Dávila-Rodríguez MJ, Freire TS, Lindahl E, Caracelli I, Zukerman-Schpector J, Friedman R. Is breaking of a hydrogen bond enough to lead to drug resistance? *Chem Commun*. 2020;56:6727–6730.
- Foda ZH, Shan Y, Kim ET, Shaw DE, Seeliger MA. A dynamically coupled allosteric network underlies binding cooperativity in Src kinase. *Nat Commun*. 2015;6:5939.
- Bhullar KS, Lagarón NO, McGowan EM, et al. Kinase-targeted cancer therapies: Progress, challenges and future directions. *Mol Cancer*. 2018;17:48.
- Alford RF, Leaver-Fay A, Jeliakov JR, et al. The Rosetta all-atom energy function for macromolecular modeling and design. *J Chem Theory Comput*. 2017;13:3031–3048.
- Iqbal Z, Aleem A, Iqbal M, et al. Sensitive detection of pre-existing BCR-ABL kinase domain mutations in CD34+ cells of newly diagnosed chronic-phase chronic myeloid leukemia patients is associated with Imatinib resistance: Implications in the post-Imatinib era. *PLoS ONE*. 2013;8:e55717.
- Kang H-Y, Hwang J-Y, Kim S-H, Goh H, Kim M, Kim D-W. Comparison of allele specific oligonucleotide-polymerase chain reaction and direct sequencing for high throughput screening of ABL kinase domain mutations in chronic myeloid leukemia resistant to imatinib. *Haematologica*. 2006;91:659–662.
- Gorre ME, Mohammed M, Ellwood K, et al. Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. *Science*. 2001;293:876–880.
- Weisberg E, Manley PW, Breitenstein W, et al. Characterization of AMN107, a selective inhibitor of native and mutant Bcr-Abl. *Cancer Cell*. 2005;7:129–141.
- Shah NP, Skaggs BJ, Branford S, et al. Sequential ABL kinase inhibitor therapy selects for compound drug-resistant BCR-ABL mutations with altered oncogenic potency. *J Clin Investig*. 2007;117:2562–2569.
- Bradeen HA, Eide CA, O'Hare T, et al. Comparison of imatinib mesylate, dasatinib (BMS-354825), and nilotinib (AMN107) in

- an N-ethyl-N-nitrosourea (ENU)-based mutagenesis screen: High efficacy of drug combinations. *Blood*. 2006;108:2332–2338.
28. O'Hare T, Shakespeare WC, Zhu X, et al. AP24534, a pan-BCR-ABL inhibitor for chronic myeloid leukemia, potently inhibits the T315I mutant and overcomes mutation-based resistance. *Cancer Cell*. 2009;16:401–412.
 29. Fleishman SJ, Leaver-Fay A, Corn JE, et al. Rosettascripts: A scripting language interface to the Rosetta macromolecular modeling suite. *PLoS ONE*. 2011;6:e20161.
 30. Khatib F, Cooper S, Tyka MD, et al. Algorithm discovery by protein folding game players. *Proc Natl Acad Sci U S A*. 2011;108:18949–18953.
 31. Levinson NM, Boxer SG. Structural and spectroscopic analysis of the kinase inhibitor Bosutinib and an isomer of Bosutinib binding to the Abl tyrosine kinase domain. *PLoS ONE*. 2012;7:e29828.
 32. Tokarski JS, Newitt JA, Chang CYJ, et al. The structure of Dasatinib (BMS-354825) bound to activated ABL kinase domain elucidates its inhibitory activity against Imatinib-resistant ABL mutants. *Cancer Res*. 2006;66:5790–5797.
 33. Xie T, Saleh T, Rossi P, Kalodimos CG. Conformational states dynamically populated by a kinase determine its function. *Science*. 2020;370:eabc2754.
 34. Young MA, Shah NP, Chao LH, et al. Structure of the kinase domain of an Imatinib-resistant Abl mutant in complex with the Aurora kinase inhibitor VX-680. *Cancer Res*. 2006;66:1007–1014.
 35. Khorashad JS, Kelley TW, Szankasi P, et al. BCR-ABL1 compound mutations in tyrosine kinase inhibitor-resistant CML: Frequency and clonal relationships. *Blood*. 2013;121:489–498.
 36. O'Hare T, Walters DK, Stoffregen EP, et al. *In vitro* activity of Bcr-Abl inhibitors AMN107 and BMS-354825 against clinically relevant Imatinib-resistant Abl kinase domain mutants. *Cancer Res*. 2005;65:4500–4505.
 37. Laurini E, Posocco P, Fermeglia M, Gibbons DL, Quintás-Cardama A, Prichard S. Through the open door: Preferential binding of dasatinib to the active form of BCR-ABL unveiled by *in silico* experiments. *Mol Oncol*. 2013;7:968–975.
 38. Vajpai N, Strauss A, Fendrich G, et al. Solution conformations and dynamics of ABL kinase-inhibitor complexes determined by NMR substantiate the different binding modes of Imatinib/Nilotinib and Dasatinib. *J Biol Chem*. 2008;283:18292–18302.
 39. Georgoulia PS, Todde G, Bjelic S, Friedman R. The catalytic activity of Abl1 single and compound mutations: Implications for the mechanism of drug resistance mutations in chronic myeloid leukaemia. *Biochim Biophys Acta – Gen Subj*. 2019;1863:732–741.
 40. Oruganti B, Friedman R. Activation of Abl1 kinase explored using well-tempered metadynamics simulations on an essential dynamics sampled path. *J Chem Theory Comput*. 2021;17:7260–7270.
 41. Narayan B, Fathizadeh A, Templeton C, et al. The transition between active and inactive conformations of Abl kinase studied by rock climbing and milestoning. *Biochim Biophys Acta (BBA) – Gen Subj*. 2020;1864:129508.
 42. Todde G, Friedman R. Activation and inactivation of the FLT3 kinase: Pathway intermediates and the free energy of transition. *J Phys Chem B*. 2019;123:5385–5394.
 43. Yun CH, Mengwasser KE, Toms AV, et al. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc Natl Acad Sci U S A*. 2008;105:2070–2075.
 44. Sugiyama N, Imamura H, Ishihama Y. Large-scale discovery of substrates of the human kinome. *Sci Rep*. 2019;9:10503.
 45. Aleem SU, Craddock BP, Miller WT. Constitutive activity in an ancestral form of Abl tyrosine kinase. *PLoS ONE*. 2015;10:e0131062.
 46. Hoffmeyer TT, Burkhardt P. Choanoflagellate models—*Monoiga brevicollis* and *Salpingoeca rosetta*. *Curr Opin Genet Dev*. 2016;39:42–47.
 47. Ruiz-Trillo I, Roger AJ, Burger G, Gray MW, Lang BF. A phylogenomic investigation into the origin of Metazoa. *Mol Biol Evol*. 2008;25:664–672.
 48. Carr M, Leadbeater BSC, Hassan R, Nelson M, Baldauf SL. Molecular phylogeny of choanoflagellates, the sister group to Metazoa. *Proc Natl Acad Sci U S A*. 2008;105:16641–16646.
 49. Friedman R. Drug resistance missense mutations in cancer are subject to evolutionary constraints. *PLoS ONE*. 2013;8:e82059.
 50. Leaver-Fay A, Tyka M, Lewis SM, et al. ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol*. 2011;487:545–574.
 51. Bateman A, Martin MJ, Orchard S, et al. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2021;49:D480–D489.
 52. Barker WC, Dayhoff MO. Viral src gene products are related to the catalytic chain of mammalian cAMP-dependent protein kinase. *Proc Natl Acad Sci U S A*. 1982;79:2836–2839.
 53. Sela I, Ashkenazy H, Katoh K, Pupko T. GUIDANCE2: Accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res*. 2015;43:W7–W14.
 54. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018;35:1547–1549.
 55. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59:307–321.
 56. Eswar N, Webb B, Marti-Renom MA, et al. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinformatics*. 2006;15:5–6.
 57. Seeliger MA, Young M, Henderson MN, et al. High yield bacterial expression of active c-Abl and c-Src tyrosine kinases. *Protein Sci*. 2005;14:3135–3139.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Otsuka FAM, Bjelic S. Evaluation of residue variability in a conformation-specific context and during evolutionary sequence reconstruction narrows drug resistance selection in Abl1 tyrosine kinase. *Protein Science*. 2022;31(7):e4354. <https://doi.org/10.1002/pro.4354>