Published in final edited form as:

Science. 2015 February 20; 347(6224): 882–886. doi:10.1126/science.aaa1823.

Using ancient protein kinases to unravel a modern cancer drug's mechanism

C. Wilson^{1,*}, R. V. Agafonov^{1,*}, M. Hoemberger¹, S. Kutter¹, A. Zorba¹, J. Halpin¹, V. Buosi¹, R. Otten¹, D. Waterman¹, D. L. Theobald², and D. Kern^{1,†}

¹Howard Hughes Medical Institute and Department of Biochemistry, Brandeis University, Waltham, MA 02452, USA

²Department of Biochemistry, Brandeis University, Waltham, MA 02452, USA

Abstract

Macromolecular function is rooted in energy landscapes, where sequence determines not a single structure but an ensemble of conformations. Hence, evolution modifies a protein's function by altering its energy landscape. Here, we recreate the evolutionary pathway between two modern human oncogenes, Src and Abl, by reconstructing their common ancestors. Our evolutionary reconstruction combined with x-ray structures of the common ancestor and pre-steady-state kinetics reveals a detailed atomistic mechanism for selectivity of the successful cancer drug Gleevec. Gleevec affinity is gained during the evolutionary trajectory toward Abl and lost toward Src, primarily by shifting an induced-fit equilibrium that is also disrupted in the clinical T315I resistance mutation. This work reveals the mechanism of Gleevec specificity while offering insights into how energy landscapes evolve.

The evolution of protein kinases is a key event in the origin of multicellularity (1), which enabled the development of more complex signaling cascades essential for the evolution of higher organisms. The key role of protein kinases in the cell cycle has placed them at the center of cancer drug research. Despite an explosion in diversity in the kinome (2), the catalytic kinase domains have maintained nearly identical structures (2–5). It is therefore surprising that the clinically successful cancer drug Gleevec has such strong selectivity toward Abl versus other tyrosine kinases, including the closely related Src. This is puzzling because the Gleevec-bound structures of Abl and Src are nearly identical, including the N- and C-terminal lobes and the drug-binding pocket (Fig. 1A) (3). Extensive work led to a frequently cited but controversial model where Gleevec selectivity is rooted in a pre-existing equilibrium between two alternative conformations of the DFG-motif (for Asp-Phe-Gly), a

[†]Corresponding author: dkern@brandeis.edu.

^{*}These authors contributed equally to the work.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/347/6224/882/suppl/DC1

Material and Methods

Supplementary Text

Figs. S1 to S12

Tables S1 and S2

References (35–48)

conserved segment of the activation loop (3, 6–12). A number of x-ray structures have revealed the sampling of a Gleevec-binding-competent DFG-out position and a binding-incompetent DFG-in position in free kinases (Figs. 1A and 2A). Recently reported data rule out the predominant role of the DFG-in/out equilibrium (conformational selection) in Gleevec selectivity. It led to a different binding scheme (Scheme 1) that accounts for the 3000-fold difference in Gleevec affinity between Src and Abl (11) due to a global conformational change after drug binding (induced fit, Scheme 1).

However, the molecular mechanism of this selectivity remains unknown. Based on x-ray crystal structures, specific point mutations have been made in attempts to convert Abl to Src-like specificity and vice versa. Despite two decades of efforts, such sequence swaps in modern kinases have failed to illuminate the atomistic determinants of selectivity (3, 6). The differences between Src, Abl, and other homologous kinases have evolved over a billion years from their common ancestor—not via residue swaps from one modern kinase to another. Sequence-swap experiments using modern enzymes have a fundamental shortcoming by neglecting epistasis (the effect of the surrounding amino acid background) (13). However, evolution has already navigated the complex epistatic protein space by producing functional proteins at each stage despite large numbers of accumulated mutations. We therefore examined the evolution of Src and Abl along their phylogenetic branches using ancestral reconstruction to understand differences not only in their equilibrium structures but also in their energy landscapes.

Ancestral reconstruction has provided a way to achieve mechanistic insight into protein function (14–19). Here, we elucidate the basis of modern kinase specificity toward Gleevec with atomic resolution by recapitulating the evolution of the Src and Abl catalytic domain from their last common ancestor. Analysis of the ancestral kinases allows us to track the evolution of the protein energy landscape (20, 21). We define “energy landscape” as a set of free energy and kinetic parameters linking kinetically distinct states that are relevant to biological processes.

Seventy-six modern-day sequences spanning the cytosolic tyrosine kinase family (Src/Abl/Tec families) were used in a Bayesian phylogenetic analysis with receptor tyrosine kinases as the out-group (Fig. 1B). Because the quality of the ancestral reconstruction strongly depends on the alignment, we estimated the tree and alignments simultaneously. The most probable sequences were inferred for four key ancestral proteins between modern Src and Abl and their last common ancestor (Fig. 1B and figs. S1 and S2), and their corresponding proteins were expressed, purified, and characterized.

We denote the reconstructed protein corresponding to the last common ancestor of Src and Abl as ANC-AS. Similarly, on the lineage leading from ANC-AS to modern Abl, ANC-A1 represents the common ancestor between humans and colonial choanoflagellates, and ANC-A2 corresponds to the common ancestor between humans and *Caenorhabditis elegans*. On the lineage leading to modern Src, ANC-S1 is the last common ancestor between humans and colonial choanoflagellates/sponges. Despite the fact that ANC-AS differs by 96 residues from any modern cytosolic tyrosine kinase, all ancestral kinases reconstructed here are fully active and thermostable (Fig. 1C and fig. S3). We evaluated the specificity of Gleevec

toward the ancestral kinases by measuring inhibition (Fig. 1D) and dissociation constants (fig. S4). The inhibition of ANC-AS is intermediate between modern Src and Abl. Gleevec affinity increases gradually toward Abl along the evolutionary pathway, whereas it drastically decreases toward Src.

Src and Abl differ by 146 amino acids, and experiments with the modern kinases could not identify the subset of residues responsible for the changes in dynamics (11). Because our reconstructed kinase ancestors have intermediate Gleevec affinities, we can now explore the evolution of these dynamical changes. To this end, we characterized the changes in the energy landscape from ANC-AS to modern Src and Abl by comparing the kinetics of Gleevec binding. All ancestors follow the same kinetic scheme as modern Src and Abl (Scheme 1), but with differences in individual conformational steps. As detailed in (11) and (22), the double exponential binding kinetics (Fig. 1E and fig. S5) reflect the faster physical binding step (identified by the linear dependence of the observed rate on Gleevec concentration) (Fig. 1G), followed by the slower induced-fit step, with the observed rate approaching a maximum at Gleevec saturation (Fig. 1H).

The gradual change in these kinetic parameters (k_{fast} and k_{slow}) from the weak binders to the tight binders is clearly visible, and the physical off rates (k_{off}), identified by the intercept in Fig. 1G, remain similarly fast. In contrast, the observed overall dissociation of the inhibitor-enzyme complex is extremely slow for ancestors ANC-AS, ANC-A1, and ANC-A2, and much faster for ANC-S1 (Fig. 1F), revealing that the rate-limiting step in Gleevec release is a conformational change before dissociation ($E^*.I \rightarrow E.I$) (Scheme 1 and Fig. 1F). To summarize (Fig. 2), we detect a systematic shift in the conformational equilibrium from $E^*.I$ to $E.I$ when traversing the evolutionary tree from Abl to Src, caused by a gradual decrease in the forward rate ($k_{\text{conf}+}$) and a more dramatic increase in the reverse rate ($k_{\text{conf}-}$) (Fig. 2C). This conformational step, independently validated previously by a direct visualization of the $E.I$ and $E^*.I$ conformers by nuclear magnetic resonance (NMR) on the enzyme-drug complex (11), accounts for the major difference in binding energy between different ancestors and modern Src and Abl, whereas changes in the drug's binding/dissociation step are nearly negligible (Fig. 2).

Our data reveal the evolution of the induced-fit step that is essential for Gleevec selectivity. It additionally provides experimental estimates of the relative populations of the DFG-in and -out conformations and reinforces that this equilibrium plays only a minor role in Gleevec affinity (Fig. 2, A and D). In the past, quantification of the equilibrium between these two alternative states has proven elusive, despite direct observation of both states in crystal structures (3). This new opportunity arises from the time-resolved detection of the binding step. The relative amplitude of the fast binding step (Fig. 2A, bottom) reflects the propensity to populate the DFG-out conformation ($p_{\text{DFG-out}}$). As apparent from Fig. 2A, one can indeed “watch” this flip in population from mainly being in a DFG-in state for modern Src and ANC-S1 to increasing DFG-out population in ANC-AS as an intermediate, and to even higher DFG-out populations for ANC-A1, ANC-A2, and Abl (large amplitudes). The DFG-out population is also an intrinsic component of the observed binding rate constant, $k_{\text{on}}^{\text{obs}} = k_{\text{on}} \times p_{\text{DFG-out}}$. Notably, the increase in $p_{\text{DFG-out}}$ measured from the amplitudes (Fig. 2A) is mirrored in the gradual increase in $k_{\text{on}}^{\text{obs}}$ (Fig. 2B), implying that the true k_{on}

rate constants are very similar. The populations of DFG-out in ANC-S1 and Src are too small to allow a quantitative analysis of the fast binding step (Fig. 2, A and D). The overall equilibrium dissociation constant K_d agrees well with the K_d calculated from all microscopic rate constants (figs. S4 and S6) (11, 22), which corroborates the kinetic scheme and the accuracy of the fitted values.

During the evolution of the energy landscape from the last common ancestor, ANC-AS, to the modern tight-binding Abl and the weak-binding Src, the major contribution to increased affinity arose from an induced-fit mechanism (up to 5 kcal/mol) (Fig. 2E) with a minor contribution from the pre-existing DFG-in/out flip in the free enzymes (Fig. 2D) (11) arising from a depletion of the binding competent DFG-out conformation for the weaker binders. The actual binding/unbinding step, which is commonly used in structure-guided rational drug design (e.g., docking analyses), is very similar between the weak and strong binders.

What are the sequence differences responsible for the two major changes in the energy landscape, the DFG loop equilibrium, and the $E.I \leftrightarrow E^*.I$ equilibrium? The ancestral reconstruction narrows down the regions responsible for these changes. Modern Abl and Src differ at 146 amino acids, yet only 70 differences separate ANC-AS and ANC-A2 and only 42 differences separate ANC-AS and ANC-S1 (fig. S2). These sequence changes are located throughout the protein, in agreement with globally distributed NMR chemical shift changes upon Gleevec binding (11).

X-ray crystal structures of ANC-AS bound to phosphomethylphosphonic acid adenylate ester (AMPPCP) (fig. S7) and Gleevec (Fig. 3) illustrate the structural consequences of sequence evolution. As expected, the overall structure of ANC-AS is highly similar to modern Src and Abl with subtle differences in the P-loop, C-helix and $\beta 4$ - $\beta 5$ loop (Fig. 3C and figs. S7 and S8). Ancestral reconstruction identified a subset of 70 mutations potentially responsible for the dramatic shift of the $E.I \leftrightarrow E^*.I$ conformational equilibrium between ANC-AS and ANC-A2, but not all of them are necessarily important for the increased affinity. To pinpoint the essential differences, we analyzed the ANC-AS-Gleevec structure and divided these 70 mutations into four groups (fig. S9). Constructs containing subgroups of mutations were then tested for activity and Gleevec binding (fig. S9). Remarkably, changing only 15 amino acids in the core of the ANC-AS N-terminal lobe to the Abl sequence [named AS(+15)] drastically increased Gleevec affinity to a level similar to Abl (Fig. 3B and fig. S10). Therefore, a small subset of mutations located only in the N-terminal lobe are responsible for the majority of the change in the $E.I \leftrightarrow E^*.I$ equilibrium (fig. S11).

With the importance of these 15 residues clearly established, we can attempt to rationalize the changes in the energy landscape at an atomistic level using the ANC-AS x-ray structures. Most of these 15 amino acids are distant from the drug-binding pocket and are part of a hydrogen-bonding network in both the AMPPCP- and Gleevec-bound conformations in ANC-AS and Src. In contrast, the corresponding residues in AS(+15) and Abl prohibit such hydrogen-bonding networks (Fig. 3, D and E). We hypothesize that the lack of these hydrogen bonds allow the P-loop (another conserved element of kinases) (Fig. 1A) to close over Gleevec in a kinked conformation (Fig. 3C), whereas in Src and ANC-AS, the identified hydrogen bonds prohibit such a conformational change. A stabilizing role of

the N-lobe hydrogen bond networks for the P-loop is consistent with the clear P-loop electron density in Src (3DQW) and ANC-AS (4UEU) bound to nucleotide, in contrast to missing P-loop electron density in the corresponding Abl structures (2G2I).

We note that the difference in P-loop conformation for kinase-Gleevec structures has been discussed previously as a potential basis for differential affinity (23). However, a sequence swap of the two P-loop differences placing Abl residues into Src, F278Y, and Q275G failed to increase Src's affinity toward Gleevec (3). Our data suggest that an additional mutation, N341, is needed for stabilization of the kinked P-loop, enabling a hydrogen bond between Y272 in the P-loop and N341 in the D-helix (Fig. 3F). This kinked P-loop conformation results in additional favorable interaction between the kinase and the drug; however, they are only possible in the absence of the restricting hydrogen bonds in the N-lobe identified above (Fig. 3E).

A long-standing problem in molecular biology is how to establish the sequence determinants for specificity within protein families. As a modern anthropogenic creation, Gleevec could not have provided evolutionary pressure for the divergence of the Src and Abl kinase families. However, the ancestral kinases deliver a deeper understanding of the molecular mechanism underlying the impressive selectivity of a modern cancer drug. Surprisingly, Gleevec takes advantage of the evolutionary differences in the Src and Abl energy landscapes. In addition, Gleevec binding serves as an experimental readout for the natural evolution of the DFG in/out equilibrium, which is widely considered to be a key element for differential regulation in the protein kinase kingdom, although the corresponding mechanism has been elusive (3, 4, 6–8). We hypothesize that the gradual evolution of both the DFG in/out equilibrium and kinase plasticity responsible for differences in the induced-fit step was driven by evolutionary pressure for differential regulation.

Currently, natural evolutionary selection is in play in the development of Gleevec resistance. During the therapeutic use of Gleevec in chronic myelogenous leukemia patients, a number of clinically relevant resistance mutations have evolved, including the most common Abl(T315I) mutation (24) (Fig. 4). This mutation drastically decreases the affinity for Gleevec (K_d of $12 \pm 5 \mu\text{M}$ at 25°C) and has been called the “gatekeeper” mutation because of the hypothesis that the Ile residue obstructs binding due to steric hindrance (25, 26). Surprisingly, we find that the binding step is unaltered by the T315I mutation but that the subsequent induced-fit step is severely hampered (Fig. 4). As described before, this latter step of conformational dynamics after drug binding is the key for high affinity in the wild-type protein, and it is the same step that is compromised under the evolutionary pressure of Gleevec treatment of cancer cells.

Previous ancestral reconstructions investigated highly conserved protein families that remain relatively unchanged in function and sequence over a vast period of time (up to 4 billion years) (27–30) and metazoan lineages (within the past 600 million years) with large functional divergence caused by a small number of mutations (31, 32). Our system differed in the number of residues involved and focused mainly on revealing the atomistic mechanism of a modern cancer drug for modern kinases. The observed gradual change in energy landscape from the common ancestor to modern kinases, and the mechanism for the

resistance mutant that evolved under natural pressure, advocate that altering conformational dynamics, and hence energy landscapes, may be a crucial driving force in evolution.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the Howard Hughes Medical Institute (HHMI), the Office of Basic Energy Sciences, Catalysis Science Program, U.S. Dept. of Energy, award DE-FG02-05ER15699, and NIH (GM100966-01) to D.K and NIH (GM096053 and GM094468) to D.L.T. R.O. is an HHMI Fellow of the Damon Runyon Cancer Research Foundation, DRG-2114-12. We thank the Advanced Light Source (ALS), Berkeley, CA, USA, for access to beamline BL8.2.2. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences, and the HHMI. The ALS is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under contract DE-AC02-05CH11231. D.K. is the inventor on a patent applied for by Brandeis University that describes a biophysical platform for drug development based on energy landscapes. The Protein Data Bank accession codes are 4CSV (ANC-AS.Gleevec) and 4UEU (ANC-AS.AMPPCP).

REFERENCES AND NOTES

1. Richter DJ, King N. *Annu Rev Genet.* 2013; 47:509–537. [PubMed: 24050174]
2. Manning G, Plowman GD, Hunter T, Sudarsanam S. *Trends Biochem Sci.* 2002; 27:514–520. [PubMed: 12368087]
3. Seeliger MA, et al. *Structure.* 2007; 15:299–311. [PubMed: 17355866]
4. Taylor SS, Kornev AP. *Trends Biochem Sci.* 2011; 36:65–77. [PubMed: 20971646]
5. Kornev AP, Taylor SS. *Biochim Biophys Acta.* 2010; 1804:440–444. [PubMed: 19879387]
6. Lin YL, Meng Y, Jiang W, Roux B. *Proc Natl Acad Sci USA.* 2013; 110:1664–1669. [PubMed: 23319661]
7. Aleksandrov A, Simonson T. *J Biol Chem.* 2010; 285:13807–13815. [PubMed: 20200154]
8. Lovera S, et al. *J Am Chem Soc.* 2012; 134:2496–2499. [PubMed: 22280319]
9. Cowan-Jacob SW, et al. *Structure.* 2005; 13:861–871. [PubMed: 15939018]
10. Dar AC, Shokat KM. *Annu Rev Biochem.* 2011; 80:769–795. [PubMed: 21548788]
11. Agafonov RV, Wilson C, Otten R, Buosi V, Kern D. *Nat Struct Mol Biol.* 2014; 21:848–853. [PubMed: 25218445]
12. Shan Y, et al. *Proc Natl Acad Sci USA.* 2009; 106:139–144. [PubMed: 19109437]
13. DePristo MA, Weinreich DM, Hartl DL. *Nat Rev Genet.* 2005; 6:678–687. [PubMed: 16074985]
14. Harms MJ, Thornton JW. *Nat Rev Genet.* 2013; 14:559–571. [PubMed: 23864121]
15. Pauling L, Zuckerkandl E. *Acta Chem Scand.* 1963; 17:S9–S16.
16. Dean AM, Thornton JW. *Nat Rev Genet.* 2007; 8:675–688. [PubMed: 17703238]
17. Liberles, DA. *Ancestral Sequence Reconstruction.* Oxford Univ. Press; Oxford: 2007.
18. Williams PD, Pollock DD, Blackburne BP, Goldstein RA. *PLOS Comput Biol.* 2006; 2:e69. [PubMed: 16789817]
19. Krishnan NM, Seligmann H, Stewart CB, De Koning AP, Pollock DD. *Mol Biol Evol.* 2004; 21:1871–1883. [PubMed: 15229290]
20. Hunter T. *Curr Opin Cell Biol.* 2009; 21:140–146. [PubMed: 19269802]
21. Eckhart W, Hutchinson MA, Hunter T. *Cell.* 1979; 18:925–933. [PubMed: 229973]
22. Materials and methods are available as supplementary materials on *Science* Online.
23. Seeliger MA, et al. *Cancer Res.* 2009; 69:2384–2392. [PubMed: 19276351]
24. Gorre ME, et al. *Science.* 2001; 293:876–880. [PubMed: 11423618]
25. Modugno M, et al. *Cancer Res.* 2007; 67:7987–7990. [PubMed: 17804707]

26. Daub H, Specht K, Ullrich A. *Nat Rev Drug Discov.* 2004; 3:1001–1010. [PubMed: 15573099]
27. Ingles-Prieto A, et al. *Structure.* 2013; 21:1690–1697. [PubMed: 23932589]
28. Perez-Jimenez R, et al. *Nat Struct Mol Biol.* 2011; 18:592–596. [PubMed: 21460845]
29. Gaucher EA, Govindarajan S, Ganesh OK. *Nature.* 2008; 451:704–707. [PubMed: 18256669]
30. Gaucher EA, Thomson JM, Burgan MF, Benner SA. *Nature.* 2003; 425:285–288. [PubMed: 13679914]
31. Harms MJ, Thornton JW. *Curr Opin Struct Biol.* 2010; 20:360–366. [PubMed: 20413295]
32. Field SF, Matz MV. *Mol Biol Evol.* 2010; 27:225–233. [PubMed: 19793832]
33. Nagar B, et al. *Cancer Res.* 2002; 62:4236–4243. [PubMed: 12154025]
34. Redelings BD, Suchard MA. *Syst Biol.* 2005; 54:401–418. [PubMed: 16012107]

Fig. 1. Reconstructing ancestors of the cytosolic tyrosine kinase family to probe the energy landscape of Gleevec selectivity

(A) Structures of Abl (33) and Src (3) bound to Gleevec. (B) Phylogenetic tree constructed with Bali-Phy (34). Reconstructed nodes (stars) and human Abl and Src are marked with the colors used throughout the manuscript. For the full tree, including reconstruction posteriors, see fig. S1. (C) Kinase activity measured by phosphorylating the target peptide. (D) Gleevec inhibition constants (K_i) at 25°C. Kinetics of Gleevec binding (E) and dissociation (F) were measured by stopped-flow fluorescence at 5°C. (E) Mixing 50 nM of kinase with Gleevec displays double-exponential kinetics with the fast phase reporting on the binding step (G) and the slow step monitoring the induced fit (H). (F) Rate of dissociation, measured by dilution of the kinase-Gleevec complex, is dominated by $E^*.I$ to $E.I$ transition (Scheme 1), whereas k_{off} is much faster [intercept in G)]. Uncertainties in all figures are $\pm$ SEM from three experiments.

Fig. 2. Evolution of the free-energy landscape in tyrosine kinases based on data in Fig. 1 (A) Evolution of the DFG-in/DFG-out equilibrium, (B) Gleevec binding step, and (C) induced fit step. (A) The gradual population shift between DFG-out (blue, 4CSV) and DFG-in (pink, 4UEU, top) is reflected in the differences in amplitude of the fast phase (bottom). (B) The k_{on}^{obs} , the product of the true k_{on} and the population of DFG-out, increases from ANC-AS to Abl in parallel with the increase in the DFG-out population seen in (A). The measured microscopic k_{off} 's for Gleevec are equivalent. (C) For the induced-fit step, a gradual decrease in the forward rate constant (k_{conf+} , top) and a drastic increase in the reverse rate constant (k_{conf-} , bottom) from Abl via the common ancestor to ANC-S1 and Src are apparent. (D and E) Free-energy contributions of conformational selection plus binding (D) and the induced-fit step (E) to the overall binding energy.

Fig. 3. Atomistic mechanism for Gleevec selectivity

(A) Mutations between ANC-AS and AS(+15) are mapped on the ANC-AS structure bound to Gleevec (4CSV) as black spheres. (B) Gleevec inhibition constants (K_i) at 25°C. (C) Structural comparison of Src(2OIQ), ANC-AS(4CSV), and Abl(1OPJ) bound to Gleevec highlighting the different P-loop conformations. (D) Ten out of the 15 identified mutations that are visible in all three x-ray structures are listed and (E) shown in the corresponding structures, illustrating how mutations in AS(+15) disrupt the hydrogen bond network (dotted lines) that are present in weak binders Src and ANC-AS. (F) Abl.Gleevec structure close-up showing the additional stabilizing interactions between the P-loop and the D-helix.

Fig. 4. Mechanism of Gleevec-evolved resistance in Abl(T315I)

(A and B) Mixing 50 nM of kinase with Gleevec displays double-exponential kinetics (fig. S12) with the fast-phase reporting on the binding step (A) and the slow step monitoring the induced fit (B). (C) Rate of dissociation measured by rapid dilution of the kinase-Gleevec complex is dominated by $\text{E}^* \cdot \text{I}$ to $\text{E} \cdot \text{I}$ conversion. (D) Site of mutation T315I is plotted onto the x-ray structure of Abl bound to Gleevec (gray) (1OPJ). (E) Individual rate constants for binding and induced-fit step and (F) the corresponding free-energy contributions.

Scheme 1. Proposed Gleevec binding scheme to human Src and Abl (II) and the ancestors (see the supplementary materials)

E and E.I correspond to free and inhibitor-bound kinase; E*.I corresponds to inhibitor-bound kinase in a distinct conformational state; DFG-in and DFG-out subscripts specify the conformation of the DFG loop (Fig. 2A).