The transition between active and inactive conformations of Abl kinase studied by rock climbing and Milestoning

Brajesh Narayan, Arman Fathizadeh, Clark Templeton, Peng He, Shima Arasteh, Ron Elber, Nicolae-Viorel Buchete, Ron M. Levy

PII: S0304-4165(19)30297-1
DOI: https://doi.org/10.1016/j.bbagen.2019.129508
Reference: BBAGEN 129508
To appear in: BBA - General Subjects

Received date: 29 September 2019
Revised date: 11 December 2019
Accepted date: 19 December 2019

Please cite this article as: B. Narayan, A. Fathizadeh, C. Templeton, et al., The transition between active and inactive conformations of Abl kinase studied by rock climbing and Milestoning, BBA - General Subjects(2019), https://doi.org/10.1016/j.bbagen.2019.129508

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019 Published by Elsevier.
The Transition Between Active and Inactive Conformations of Abl Kinase Studied by Rock Climbing and Milestoning

Brajesh Narayan¹, Arman Fathizadeh², Clark Templeton³, Peng He⁵, Shima Arasteh⁵, Ron Elber²,⁴, Nicolae-Viorel Buchete¹, and Ron M. Levy⁵

¹School of Physics, University College Dublin, Belfield, Dublin 4, Ireland

²Oden Institute for Computational Engineering and Sciences, University of Texas at Austin, 201 E. 24th Street, 1 University Station (C0200) Austin, Texas, 78712-1229, USA

³McKetta Department of Chemical Engineering, University of Texas at Austin, 200 E. Dean Keaton St. Stop C0400, Austin, TX, 78712-1589

⁴Department of Chemistry, University of Texas at Austin, 2506 Speedway STOP A5300, Austin, TX 78712-1224, USA

⁵Department of Chemistry, Temple University, 1801 N Broad Street, Philadelphia, PA 19122, USA
**Abstract:**

**Background:** Kinases are a family of enzymes that catalyse the transfer of the $\gamma$-phosphate group from ATP to a protein’s residue. Malfunctioning kinases are involved in many health problems such as cardiovascular diseases, diabetes, and cancer. Kinases transitions between multiple conformations of inactive to active forms attracted considerable interest.

**Method:** A reaction coordinate is computed for the transition between the active to inactive conformation in Abl kinase with a focus on the DFG-in to DFG-out flip. The method of Rock Climbing is used to construct a path locally which is subsequently optimized using a functional of the entire path. The discrete coordinate sets along the reaction path are used in a Milestoning calculation of the free energy landscape and the rate of the transition.

**Results:** The estimated transition times are between a few milliseconds and seconds, consistent with simulations of the kinetics and with indirect experimental data. The activation requires the transient dissociation of the salt bridge between Lys271 and Glu286. The salt bridge reforms once the DFG motif is stabilized by a locked conformation of Phe382. About ten residues are identified that contribute significantly to the process and are included as part of the reaction space.

**Conclusions:** The transition from DFG-in to DFG-out in Abl kinase was simulated using atomic resolution of a fully solvated protein yielding detailed description of the kinetics and the
mechanism of the DFG flip. The results are consistent with other computational methods that simulate the kinetics and with some indirect experimental measurements.

**General significance**: The activation of kinases includes a conformational transition of the DFG motif that is important for enzyme activity but is not accessible to conventional Molecular Dynamics. We propose a detailed mechanism for the transition, at a timescale longer than conventional MD, using a combination of reaction path and Milestoning algorithms. The mechanism includes local structural adjustments near the binding site as well as collective interactions with more remote residues.

Keywords: Reaction Path; DFG motif; activation loop; Molecular Dynamics
1. Introduction

Kinases form one of the largest family of enzymes. In the human genome, there are about 500 predicted protein kinases. They catalyze the transfer of the γ-phosphate group from ATP to the hydroxyl group of a serine, threonine or tyrosine residue, a type of transfer that is found in many biological processes. Malfunctioning kinases are involved in many major human health-related problems such as cardiovascular diseases, diabetes, and cancer. Despite their diversity of function, the structure of their catalytic domain is shared across the kinase family. Roughly, the kinase domains consist of an N-lobe, and a C-lobe connected by a flexible hinge region (Fig. 1). The active sites consist of three conserved structural elements: the activation loop (A-loop), the Asp-Phe-Gly (DFG) motif, and the C helix (which is part of the N-lobe). These conserved structural elements make it challenging to design a drug that would be specific to only one kinase. Nevertheless, such a design is desirable to minimize unwanted, yet likely side effects, given the high structural similarity of members of this family.

Nevertheless, the drug imatinib was found to be selective and inhibit BCR-Abl but not c-Src. Understanding the origin of this selectivity is of significant interest and potential for enhancing drug design efforts. An intriguing proposal explains the selectivity using variation in active site flexibility and binding of the drug to the inactive kinase conformation. However, a recent experiment suggests that the selectivity of imatinib towards Abl-kinase and not Src-kinase is a result of a slow conformational change that occurs after ligand binding. The importance of the DFG flip to selectivity is therefore in doubt. Nevertheless, the FG must change a structure at some step along the reaction to allow the entrance of the inhibitor to the active site.
Only kinases that are able to form DFG-out (inactive) conformation can open up a pocket to facilitate binding to the imatinib. Supporting evidence for a high activation loop conformational flexibility emerged from X-ray crystallography (significant variations were observed in many structures of kinase proteins). Also illustrating diversity are NMR spectroscopy, and molecular simulations.

There have been numerous experimental studies to estimate the rate of transition between DFG-in and DFG-out for different members of the kinase family. It was argued that the operating mechanism of the enzyme is influenced by the rate of the DFG flip. The early evidence for this mechanism came from the experimental observations of the significant differences between the rates of binding of inhibitors to the DFG-out and the DFG-in states in P38 kinase. While the DFG-in inhibitors binding is quite fast and within the diffusion-controlled regime, the rate for DFG-out inhibitors were orders of magnitude slower. This observation suggests that DFG-in states are more populated in equilibrium. Kinetic may also play a role. If the rate of transition between the protein conformations is fast, the drug can always find a ready conformation to bind, and the ratio of the unbound population: [DFG-in]/[DFG-out] remains the same. If the rate is slow a shift in the conformation of the unbound protein will be observed and the binding of the inhibitors to one of the states may slow down if saturation of proteins bound to ligands is reached.

In an NMR study, it was shown that the binding of different inhibitors can influence the dynamics of the DFG motif in Abl-kinase. A wide range of timescales was suggested for the dynamics of different residues in presence of different inhibitors. However, determining the accurate time scale of the DFG flip remains challenging due to the lack of signal for some key
residues from DFG motif and activation loop. In another study, probing the kinetics of the DFG-flip in NMR measurements for P38 kinase, the observed experimental line-shape is very broad and indicates an intermediate time scale on the NMR time scale (milliseconds). In a combined experimental and computational paper, the time scale of imatinib binding, under a variety of perturbations, was estimated to be in the range of ten milliseconds. A recent study combines NMR and tryptophan fluorescence on imatinib binding to Abl-kinase and observes two time-scales for the binding event. One time-scale is at, or below a few milliseconds and a second time is of 100 milliseconds to seconds. The process with a slower time scale was identified as a large conformational transition. The shorter time scale is assigned to a local binding event. If we assume that the local and fast event of imatinib binding to Abl kinase is associated with the DFG flip, we can obtain an indirect estimate of the time scale ~2ms, as we illustrate in the discussion.

Summarizing experimental results of particular relevance to the present study, these studies shared the following observations (i) DFG-in state is more populated than DFG-out state in Abl kinases for unprotonated Asp381 and (ii) Some indirect evidence is available that the timescale for the flip DFG-in to DFG-out in kinases is in the millisecond time scale.

Computational studies are in general agreement for the slightly larger stability of the DFG-in active conformation compared to DFG-out state in Abl kinase with Asp 381 unprotonated. However, the study of kinetics is more restricted due to limitations on conventional MD. A recent study of Abl kinase, using MD and the Markov State Models (MSM) suggest a time scale for the transition of milliseconds. However, the statistics of transitions was small. This estimate is consistent with conventional MD simulations that were not able to sample the DFG-in and DFG-out transition on hundreds of nanoseconds simulation time without an assisting mutation.
The kinetic of the DFG-in to DFG-out transition is an important component of the Abl kinase activity. It is therefore of interest to further investigate the mechanism of the transition and quantify the time scale of the process.

In the present manuscript we combine a reaction path algorithm (Rock Climbing\textsuperscript{18}) with the theory and algorithm of Milestoning to study kinetics of the DFG flip and compute the mechanism and the rate of the transition in Abl kinase.

2. Material and Methods

2.1 Choice of reactant and product structures

The reaction path approach that we use requires as input the conformations of the two end states, the reactant and the product.\textsuperscript{18} Given the richness of crystallographic structures of kinases and specifically of Abl kinases the choice of the end structures requires discussion.

The “classical DFG-out” cluster of structures in the PDB (with \textasciitilde200 structures) is the most prevalent inactive state observed among kinase structures with DFG-out, in which the activation loop (A-loop) is fully folded/closed.\textsuperscript{19} The second most populated cluster of DFG-out inactive structures is a cluster we refer to as “DFG-out minimally perturbed”, and 2G2F is a member of this cluster. The root-mean-square deviation (RMSD) of the activation loop over all DFG-out inactive structures with respect to 2F4J which is a representative of an active kinase with the A-loop extended in the active conformation, indicates that the A-loop in the DFG\textsubscript{out}-A-loop\textsubscript{minimally-perturbed} cluster differs by 1-5 Å from 2F4J, in contrast to the large range variation of between 11-19 Å in the classical DFG-out cluster (Ref structure: 2F4J chain A). Representatives of the
DFG_{out}-A\text{-}loop_{\text{minimally-perturbed}} cluster have been observed for 11 kinase families, and Ab1 (with ~10 unique pdbs) is the most observed family among them.

The Protein Data Bank (PDB) structures of the Ab1 kinase with accession codes of 2F4J\textsuperscript{20} and 2G2F\textsuperscript{21} were used as active (reactant) and inactive (product) conformations, respectively. We picked two conformations that are not profoundly different with the exception of the activation loop, making it possible for us to focus on the transition of the DFG motif. For example, there is no significant shift of the $\alphaC$ helix between the two states. In previous studies it was also shown that the inactivated set of structures is highly flexible (see for instance\textsuperscript{8, 19}). A significant number of structures are sampled with conventional Molecular Dynamics (MD) in the flexible state and it is likely that a single pathway calculation assisted with unbiased trajectory sampling of Milestoning (section 2.3) will probe the transition network.

The structures were solvated separately with TIP3P water molecules and salt concentration of 0.15M (NaCl). The systems consist of ~45,000 atoms. All the simulations have been conducted with the NAMD program\textsuperscript{22} and the CHARMM36\textsuperscript{23} forcefield has been used. Periodic boundary conditions were used, and the system was minimized using conjugate gradient algorithm for 10,000 steps. Equilibration followed in the NPT ensemble with Nose-Hoover Langevin piston pressure control for 5 ns at pressure of 1 atm and temperature of 310 K.\textsuperscript{24-25} Then the system was equilibrated in the NVT ensemble at 310 K using Langevin thermostat for additional 10 ns. Water molecules were kept rigid with the SETTLE algorithm\textsuperscript{26} and all other bonds with hydrogen atoms were kept fixed with the SHAKE algorithm.\textsuperscript{27} The cut off distance for non-
bonded interactions was 12 Å and the Particle Meshed Ewald method was used to sum the electrostatic interactions.\textsuperscript{28} The timestep was 1 fs. The final configurations of the equilibrated structures were used as the reactant and product states for pathway generation, with root mean square distances (RMSD) of the equilibrated structures relative to the initial structure were 0.9 Å and 1.1 Å, respectively.

### 2.2. Generation and optimization of pathway

Examining the active and inactive conformations (Fig. 1) we realize that the major differences between the two conformations are concentrated at the activation loop (from residue number 380 to 400).
Therefore, the reaction space or the coarse variables that guide the reaction path calculations were selected from this region. First, the backbones of the two structures were aligned for a best overlap for the entire structures excluding the activation loop. Then 24 atoms along the loop were selected to represent the coarse space. The selected 24 atoms include both alpha carbon and atoms from the side chains of the residues with the highest RMSD values between the reactant
and product. These atoms are listed in Table 1. The actual number of degrees of freedom in the coarse space is smaller than $3^{24} = 72$ since bond lengths and bond angles do not vary significantly in the calculations and may be considered fixed. We define an active torsion in the reaction space if at least one atom from the coarse space is included in the definition of the torsion, and the torsion changes along the reaction coordinate by at least 60 degree. The number of such torsions that have significant contribution to the coarse space is 22. The reaction space that we considered is therefore quite large compared to other studies in the field.

**Table 1.** List of the 24 atoms used to define the coarse space in the calculations of the pathway. The final 12 atoms that are used in the Milestoning calculations are indicated in red. See text for more details about the selection.

<table>
<thead>
<tr>
<th>Val379(CA)</th>
<th>Val379(CB)</th>
<th>Ala380(CA)</th>
<th>Ala380(CB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp381(CA)</td>
<td>Asp381(CG)</td>
<td>Phe382(CA)</td>
<td>Phe382(CG)</td>
</tr>
<tr>
<td>Gly383(CA)</td>
<td>Leu384(CA)</td>
<td>Leu384(CG)</td>
<td>Ser385(CA)</td>
</tr>
<tr>
<td>Arg386(CA)</td>
<td>Arg386(CZ)</td>
<td>Leu387(CA)</td>
<td>Leu387(CG)</td>
</tr>
<tr>
<td>Met388(CA)</td>
<td>Met388(CE)</td>
<td>Thr389(CA)</td>
<td>Thr389(CB)</td>
</tr>
<tr>
<td>Asp391(CA)</td>
<td>Tyr393(CA)</td>
<td>His396(CA)</td>
<td>Ala397(CA)</td>
</tr>
</tbody>
</table>

The method of Milestoning (see section 2.3) is used to compute kinetic and thermodynamic observables. As a first step in a Milestoning calculation we require a rough sample of the space linking the reactant and product. The number of intermediate configurations that we use as a sample varies depending on the system characteristics. It is between a few tens to several thousand structures. These configurations form centers of Voronoi cells and their interfaces are used as Milestones, either in the Markovian Milestoning approach, or in other variants of
Milestoning.\textsuperscript{31} The choice of the Voronoi cells impact the rate of convergence of the calculation but should not impact the final results if the system is close to equilibrium, or if iterations of the exact Milestoning approach are used.\textsuperscript{32}

One way of generating centers of Voronoi cells that cover the relevant space is by a reaction path calculation. We have extensive experience in generating reaction coordinates in complex systems, starting with the study of a conformational transition in myoglobin\textsuperscript{33} and continuing to a number of other complex systems.\textsuperscript{34-35} In these approaches, and variants of them \textsuperscript{36-37}, a guess is generated for the path and then optimized. Unfortunately, the end result may be biased by the initial guess, especially on rough energy landscapes in which multiple pathways exist. This problem was discussed in Ref. \textsuperscript{16} in the context of kinase conformational transitions. The activation loop is highly flexible and transitions in a space of several dimensions. Nevertheless, our current interest is focused on the DFG transition, which is spatially small. The transition is also activated, and is expected to be a rare event. Determining the pathway is challenging for conventional MD, but it is more straightforward to compute with reaction path calculations.

Recently we introduced a new method for computing reaction coordinate in complex systems that does not require an initial guess (Rock Climbing\textsuperscript{18}). The local optimization of the path were carried with implicit solvent, using a Generalized Born method\textsuperscript{38} with ion concentration of 0.15M and solvent dielectric of 78.5 while the global optimization was done in explicit solvent. During the pathway generation only two regions of the protein are allowed to move. The first region is of residues 376 to 405 that includes the activation loop. The second region is of
residues 278 to 299 that contains the $\alpha$C helix. The backbone of the rest of the structure was restrained with harmonic potentials to their initial positions during the pathway calculations.

We divide the path calculation into two steps. First, using a local and greedy algorithm we generate a pathway from the coordinates of the reactant to the product as follows: The generation of the pathway starts with adding a displacement vector, $\delta$, along the vector connecting reactant to product. By providing information about the end points we made the process global. However, we do not provide an initial guess for the entire path and the path generation follows a local procedure. To begin with, each atom of the coarse space of the reactant is shifted by $\delta = 0.25 \, \text{Å}$ toward the product (equilibrated inactive kinase). Then harmonic restraints with force constant of 2000 kcal/mol $\text{Å}^2$ were applied on the selected atoms of the coarse space and the rest of the system was minimized for 500 steps followed by 5000 steps of MD at 310 K. Finally, the harmonic restraints were released, and the system was minimized for 50 steps.

The last configuration is used to generate a new displacement vector, $\delta$, toward the product configuration. The displacement is added to the current coordinate set and the relaxation process described above follows. The process is repeated until the product is reached. The value of $\delta$ is tuned during the process to allow for faster convergence or better accuracy. We obtained 850 structures interpolating between the reactant and product, with an average displacement length of about 0.1 Å. Out of these 850 structures, 50 structures were selected that are approximately equidistant in the coarse space. These structures are used in the next step of path refinement.
In the second step, the path is globally optimized. All the 50 structures including the reactant and product, are solvated and equilibrated for 1ns in the NPT ensemble using Nose-Hoover Langevin piston\textsuperscript{24-25} followed by 5 ns in the NVT at 310 K. The coarse variables are restrained to their corresponding positions along the path using harmonic force constants of 150 kcal/mol. After equilibration of the bath coordinates (coordinates that are not included in the coarse space), we refine the pathway. Each configuration along the pathway in the coarse space is represented by the vector $x_i$. The following target function is optimized:

$$T = \frac{1}{2} \sum_{i=1}^{n} [ ( |\nabla U(x_i)| + |\nabla U(x_{i-1})|) \Delta l_{i-1,i} + k(\Delta l_{i-1,i} - \langle \Delta l \rangle)^2 ] + \sum_{i=2}^{n} H(\theta_{i,i-1,i-2} - \theta_0)k'(\theta_{i,i-1,i-2} - \theta_0)^2$$

The heavy side function is denoted by $H(y)$.

$$H(y) = \begin{cases} 
    0 & y < 0 \\
    1 & y \geq 0 
\end{cases}$$

The norm of the force in coarse space, $|\nabla U(x_i)|$, is an average over all coordinates of the bath, i.e. the coordinates that are not included in the coarse space. The first term in Eq. (1) is a discrete approximation to a functional of the path:

$$S[x(l)] = \int_{x(0)}^{x(L)} |\nabla U(x(l))| dl$$

The path that minimizes $S[x(l)]$ is the Steepest Descent Path (SDP) in the free energy landscape of the coarse variables, $x$.\textsuperscript{39}
Eq. (1) is a discrete version of Eq. (2) in which we must ensure that the configurations are distributed uniformly along the path. Therefore, we added the second term. The norm of the distance vector between the structures $x_i$ and $x_{i-1}$, is $\Delta l_{i-1,i}$ and $<\Delta l>$ is the average distance over all the neighbor distances of the path ($\langle l \rangle = \frac{1}{n} \sum_{i=1}^{n} l_{i, i-1}$). Using $<\Delta l>$ allows the path to gradually expand or shrink in every iteration, if required, and keeping the points equally spaced along the pathway. The force constant $k$ is 200 kcal/molÅ$^2$.

Another difference between the discrete and continuous paths is that the discrete path can make sharp turns of high curvature that reduce the value of the discrete functional. The third term prevents large path curvatures. It is applied only when the angle, $\theta_{i,i+i}$, between three sequential configurations of the path is larger than a threshold value, $\cos(\theta_0)$.

$$\left(\cos\left(\theta_{i,i,i+1}\right) = \frac{(x_{i+1} - x_i) \cdot (x_i - x_{i-1})}{\left|(x_{i+1} - x_i) \cdot (x_i - x_{i-1})\right|}\right).$$

Here we use $\theta_0 = 60^\circ$ and $k' = 5$ kcal/mol$\cdot$deg$^2$.

With the above definition of the target function of the path, the optimization is conducted in iterations. In every iteration, the 48 intermediate structures are simulated for 1ns to compute the average force. Harmonic restraints with force constants of 150 kcal/molÅ$^2$ are applied on the atoms of the coarse space during the 1ns simulations and the mean forces at each structure are computed. The coordinates of the coarse variables of the path are then adjusted by a small step (typically 0.01-0.03Å) to minimize the function of Eq. (1) and to move toward the minimum free energy path. This procedure is repeated one hundred times. Convergence is assumed when the
norm of the force gradient does not vary significantly. Only the force component parallel to the reaction path remains when the SDP is reached (Fig. 2).

Figure 2. The reduction in the norm of force, averaged over the entire reaction path, as a function of iteration number. The norm of the force drops rapidly in the first ten minimization steps and then decays more gradually. It seems to stabilize at around 100 iterations. The final gradient is around 0.35 kcal/mol Å\(^{-1}\). It is not zero since the norm of the force along the reaction coordinate is included.

After the calculations of the optimized path were complete, we found that some of the atoms included in the coarse space were hardly moving. We therefore adjusted the coarse space to include only coordinates that were modified significantly along the optimal reaction pathway.
This reduces the number of coarse variables that define the coarse space from 72 to 36. The final list of atoms that determine the coarse space is given in Table 1.

### 2.3 Milestoning

Milestoning is a versatile theory and algorithm to compute thermodynamics and kinetics of complex systems using a large number of short trajectories. It was discussed extensively in the literature\(^\text{32, 40-41}\) and a review article is available.\(^\text{42}\) We therefore described it below only briefly. In the first step of a Milestoning application we provide a sample of configurations from the coarse space between the reactants and products (Fig. 3). In the present study the sample consists of the configurations along the reaction coordinate, \(\{x\}\), described in the previous section. Each configuration is called an anchor and is a center of a Voronoi cell in coarse space. The boundaries between the Voronoi cells (say cells \(i\) and \(j\)) are called milestones \(M_{ij}\). Milestone \(M_{ij}\) is the set of points with equal minimal distances to anchors \(i\) and \(j\) and larger distances from all other anchors. For brevity we index milestones also by a single Greek letter, e.g., \(M\).
Fig 3. A schematic representation of the discretization of the coarse space following the transition pathway. R and P represent the reactant and product states, respectively. The black line shows the reaction pathway. The red dots are the anchors, and the blue lines are the milestones. Every milestone is numbered by its corresponding anchors. For example, milestone (j,k) is the boundary between cells j and k. The green arrows show 4 unbiased trajectories initiated from milestone (i,j). The trajectories are terminated when they hit any other milestone for the first time. Re-crossing the original milestone does not lead to trajectory termination.

In the second step of Milestoning we sample configurations from the Boltzmann distribution using MD simulations constrained to the milestones. We generate configurations from the conditional probability density

$$p(X|M_{ij}) = \exp\left( -bH(X) \right) I_{X \in M_{ij}}$$

where $X$ denotes the phase points constrained to milestone $M_{ij}$.

In the third step we launch unbiased MD trajectories from the sampled configurations at the milestone generated in the second step. We terminate these trajectories when they hit for the first time a milestone different from the milestone they started from. We recorded the identities of the milestones and the time of termination. In one variant of Milestoning, which is not used here,
(exact Milestoning\textsuperscript{32}) we also retain the terminating phase space configuration at the milestone.

In the fourth step we use the information gathered in the first step to compute two functions \( K \), the probability that a trajectory initiated in milestone \( a \) will terminate at milestone \( b \) and \( t \), the lifetime of milestone \( a \). Let \( n \) be the number of trajectories initiated at milestone \( a \). Let \( \nabla \) be the number of trajectories that were initiated at milestone \( a \) and were terminated at milestone \( b \). We estimate the transition probability also called the kernel as \( K_{ab} = \frac{n_{ab}}{n_{a}} \) and the lifetime \( t_{\alpha} = \frac{1}{n_{\alpha}} \sum_{l=1}^{n_{\alpha}} t_{l} \), where \( l \) is the index of the trajectory and the time length of trajectory \( l \) is \( t_{l} \).

In the final and fifth step we compute the thermodynamic and kinetic observables. The stationary flux of trajectories through a milestone is the eigenvector, \( q \), of the matrix \( K \) with an eigenvalue of one: \( q^{T}K = q^{T} \) where we used bold face for vectors and matrices. The free energy of a milestone \( a \) is given by \( F_{a} = k_{B}T \log[q_{a}] \) and the mean first passage time (MFPT, \( \langle \ldots \rangle \)) is given by \( \langle \ldots \rangle = p_{0}(I - K')^{-1}t \). \( p_{0} \) is the vector of the initial distribution and \( I \) is the identity matrix. The matrix \( K' \) is an adjusted \( K \) matrix in which absorbing boundaries are placed at the product state.\textsuperscript{32}

More computational details on the implementation of the Milestoning method to the kinase problem are given below.
Step 1: We use 50 structures equally distributed along the path to define the anchors. Sequential anchors are separated by a distance of 0.5 Å along the reaction coordinate (RMSD of the coarse variables) which is a typical value for a Milestoning calculation.

Step 2: Since the anchors are sampled along a one-dimensional reaction coordinate, it is suggestive to place the initial milestones between sequential anchors. We therefore initiate trajectory sampling for 49 milestones between each of two consecutive anchors along the path. We conduct 1ns MD simulation at a constant temperature while two restraints were applied to keep the sampling trajectory restrained to the milestone. The first restraint is a harmonic term $k(d_i-d_j)^2=0$ to keep the distances $d_i$ and $d_j$ (in coarse space) between the current configuration and the two anchors $i$ and $j$ equal. The restraining force constant is 2000 kcal/molÅ$^2$. A second set of half-harmonic restraints are $k'(d_m-d_l)^2$ when $d_m<d_l$, $l=i,j$ and zero otherwise with $k'=1000$ and $m$ being any milestone other than $i$ or $j$. This restraint prevents the system from getting closer to any other milestone, $m$. 100 samples were kept from the final 0.5ns of the restrained simulation at each milestone.

Step 3: We release all the restraints and conduct unbiased short MD trajectories at the NVE ensemble starting from the configurations at each milestone that were sampled in step 2. The trajectories are terminated when they hit a milestone different from the one that they were initiated at. The typical length of each of the unbiased trajectories is 10 ps. The trajectories could either reach one of the initial 49 milestones we started from or they may reach a new milestone. For example, in Fig. 3 one of the trajectories initiated from the milestone $(ij)$ reaches a new milestone $(j,k)$ that connects non-sequential milestones along the path. We therefore add the new
milestone to the list and sample configurations at the newly discovered milestones. Finally, we 
launch unbiased trajectories from these new samples. The process is continued until we do not 
visit new milestones, or a connected network is observed in which the MFPT value for the 
transition between reactant and product converges to a finite value. The final number of 
milestones we ended up with in the current project was 171 (Fig. 4).

Fig 4. Representation of all the 171 milestones considered for computing the transition matrix. 
Every point corresponds to one milestone. The red points are the initial milestones between the 
consecutive anchors along the transition path. The blue points represent the milestones 
discovered during the analysis of the free trajectories that are used to enrich the sampling of the 
pathways.

3. Results

3.1. Free energy, pathway, and committor
Fig. 5 shows the free energy landscape of all the milestones in two dimensions where the dimensions are the anchor indices. To achieve a better qualitative understanding of plausible transition paths between the active and inactive states, we determine optimal pathways in the milestones space. Every milestone is a node in a network and we assign weights to the network's edges as rate coefficients for transitions between milestones. Rate coefficients are computed from the transition matrix $K$ and average lifetimes of the milestone, $t$. The rate coefficient for a transition between a milestone pair $(i,j)$ is given by

$$k(i,j) = \frac{K_{ij}}{t_i} \quad (3)$$

It was shown that choosing rate coefficients between the milestones in this way leads to a Master Equation where the milestones are the states. It was also shown that the exact MFPT is obtained with this Master Equation (but not higher moments of the first passage time).\textsuperscript{43} Using the rate coefficients, global maximum weight paths (GMWP) from reactant to product and vice-versa, were obtained from the network, shown in Fig 4, using recursive Dijkstra's algorithm\textsuperscript{44} (Fig. 5). In other words, these are the pathways on the network with the fastest rates. 1D free energy profiles are shown in Fig. 6 for the two GMWPs between the active and inactive states. According to these plots, the free energy landscape consists of multiple barriers with a maximum height of $\sim 15.7 \pm 2.5$ kcal/mol with respect to the active state. Also, the free energy of the inactive state is $2.8\pm 2.0$ kcal/mol higher than the active state, suggesting that the active state is more stable.

The local minima and maxima that are observed along the paths follow bond rotations of different loop segments. These rotations are roughly independent, and they are observed at
different positions along the reaction coordinate. For example, the rotation of the DFG motif contributes to the first barrier along the pathway near position 13 in Fig 6a. The local minima between positions 16 and 24 correspond to states where the rotation of DFG and residue 385 are complete. Finally, the rotations of Leu387 and Met388 contribute to the free energy hump from positions 24 to 31.

**Fig. 5.** Free energy of pairs of anchors as computed from the Milestoning theory. The energy values are in Kcal/mol. The two paths with maximum flux from the reactant to the product are
lines in red and in magenta. Note that the significantly off-diagonal “jumps” on the surface are a consequence of long-range connection between milestones that are not in sequence along the reaction pathway (see also Fig. 4).

**Fig. 6.** Two optimal free energy profiles along the two max-flux pathways from active to inactive state. In panel a, the milestones are numbered from 1 to 41 for the corresponding points along the red path shown in figure 5 starting from active state, and from 1 to 32 for the magenta path in figure 5, starting from the inactive state for panel b.

Note that Fig. 6 is plotted as a function of the index of the milestones, which is different from the anchor number. This is because we consider nodes and edges on a network that are labelled separately. The mapping between anchor pairs and milestones can be found in Fig. 5. The mapping is not trivial since the path is not monotonic in the index of the anchor and is monotonic in the milestone index.
The committor function is another useful quantity that can be calculated directly from Milestoning. The committor function, $C$, at every milestone, is the probability that a complete trajectory initiated at that milestone will reach the product before reactant. Interesting milestones are those with values close to $C\approx 0.5$ that can serve as a definition of the transition state.

According to Fig. 7 this occurs near milestones (21,22), (22,23), and (23,24). Fig 8 shows the conformations of the first residues of the activation loop for anchor 23, and the active and inactive states. At anchor 23 residue Arg386 already moved to its final state while Asp381 did not change its configuration significantly. The sidechains of Phe382 from the DFG switch and of Leu384 are rotated approximately half of the way through. Their significant motions at the transition state and their intermediate structures suggest them as a bottleneck for the transition.

**Fig. 7.** Color-coding the committor function at every milestone. The committor of a milestone is the probability of a complete trajectory initiated at that milestone to reach the product before the reactant state.
Fig. 8. A stick model of residues 381 to 386 for active (yellow), inactive (red), and a sample configuration at anchor 23 where the committor value is near 0.5 (blue). Note that Arg386 already reached its final position at the transition state, while Asp381 did not change its configuration significantly. The residue Phe382 is found at half of the way of the transition.

Although the DFG-in and extended conformations of the A-loop are essential elements for a kinase to be active, active conformations are also associated with other features such as $\beta_3$\textsubscript{Lys271} to $\alpha$-chelix\textsubscript{Glu286} salt-bridge. The orientation of Lys interacting with both Glu and the DFG-Asp with the assistance of ions provides a proper network for coordinating the ATP phosphate in active kinase structures. Detailed structural investigation of the DFG\textsubscript{out}-A-loop\textsubscript{minimally-perturbed} structures indicates that this salt-bridge is not always maintained due to differences in the orientation of the DFG-Phe side chain. There are three possibilities: 1) DFG-Phe is located between the Lys and Glu, in a way that Glu is free to interact with the HRD-Arg or DFG-Gly amide groups. Therefore, the salt-bridge is broken. 2) The DFG-Phe points to the back-pocket and the salt-bridge is maintained. 3) The DFG-Phe points out to the solvent and the salt-bridge is maintained. Abl kinase structures are observed in all three groups, 2G2F chain B belongs to the second group for which the salt bridge is present in the inactive state. For the structures studied
in this paper, this salt bridge exists at both, active and inactive states. With the reaction coordinates between the two states at hand, we are able to probe the status of this salt bridge at different steps. Our results show that the rotation of the DFG residues and more specifically, Phe382, requires breakage of this salt bridge. This was pointed out already in Ref. 8 The salt bridge breaks between anchors 22 to 28 along the reaction pathways, which the committor analysis suggests to be close to C~0.5. Fig 9 shows a sample configuration from anchor 27 in which the distance between the charged groups of Lys271 and Glu286 reaches ~7 Angstroms. When the Phe382 reaches its final destination, the salt bridge reforms.
Fig. 9. Changes in the salt bridge between Lys271 and Glu286 for inactive (A), an intermediate state (B), and active (C) states. The salt bridge exists in the active state and inactive states but during the transition from active to inactive state, the salt bridge breaks. The intermediate state shown is anchor 27. The DFG residues are shown in red, blue, and yellow for inactive, intermediate, and active states, respectively.
3.2. Non-Markovianity and MFPT

One of the advantages of the Milestoning method is that the description of the dynamics is not required to be Markovian. The assumption of Markovianity fails in numerous biological processes. As an illustration, Fig. 10 shows the distribution of the lifetimes for a few milestones. If the kinetics is Markovian, then the probability distribution of the transition time between two states (in our case milestones) must follow an exponential behaviour of the type:

\[ P(t) = Ae^{-t/\langle t \rangle} \]  \hspace{1cm} (4)

Where \( P(t) \) is the distribution of the lifetime, \( A \) is a constant, and \( \langle t \rangle \) is the average lifetime of a milestone. In Fig. 10, we first calculated the average lifetime from the distribution and then fitted an exponential curve to the computed histograms. The \( R^2 \) values are shown in every window. The distributions deviate significantly from a Markovian behaviour. A deviation from a single exponential behaviour for the binding of imatinib was pointed out experimentally.\(^4\)
**Fig. 10.** Distribution of lifetime for (a) milestone (21,22), (b) milestone (22,23), (c) milestone (24,25) and (d) milestone (40,41).

Finally, the MFPT for the transition can be calculated from Milestoning by sampling transition matrices and lifetimes from their known distributions. The sampling procedure was discussed in details in reference 46. **Fig. 11** shows the distribution of the MFPTs obtained from 1,000 sample transition matrices and lifetimes. The averaged MFPTs for forward and backward processes are $1.0 \pm 0.9$ s and $0.2 \pm 0.1$ s, respectively. These values are obtained by ignoring the bins containing values less than 5% of the bin with the maximum population.

It can be seen that the mean transition time for going from active to inactive state is slightly higher than the reverse process since the active state is located approximately 2.8 kcal/mol lower than the inactive state.
4. Discussions

In the present study, we investigate computationally the kinetics of the DFG-in to DFG-out transition. There are two main observables that we discuss below. The first is the time scale for the process and the second is the mechanism or the structural features of the reaction. Since significant variations were observed in the measured and computed rates for different kinases, there is a significant uncertainty in the comparisons.

Fig. 11. Distributions of MFPT for transition from active to inactive states (top) and the reverse process (bottom). The insets show the corresponding distributions for 1/MFPT which are estimates of the rate coefficients consistent with the simulation data and the error analysis. We quote the mean values of the rate coefficients for the forward and backward transition.
4.1 The time-scale for the DFG-in to DFG-out transition in Abl-kinase.

Milestoning simulations provide an estimate of the Mean First Passage Time without assuming the existence or the need to identify a bottleneck or a transition state. The only assumption invoked in the present version of Milestoning is that the system remains close to equilibrium.\textsuperscript{42} Even though we have used a set of coarse variables to guide the calculations, the calculations are exact provided that the equilibrium assumption is satisfied and the coarse variables are able to differentiate between reactants and products. This theory is different from other approaches in the field to study kinetics\textsuperscript{8} and it is interesting to compare its results to those of related calculations and experiments.

Experimentally, there are a few indirect observations about the rate of the DFG flip that we discuss below.\textsuperscript{4, 9} One observation is an NMR measurement of the dynamics of p38 kinase.\textsuperscript{6} The NMR spectrum of the DFG motif was found to be very broad suggesting an intermediate NMR time scale (milliseconds). Fast processes in NMR are characterized by a single averaged peak while multiple peaks of different conformations correspond to a slow process by NMR scale.

Combined NMR and flow experiments on Abl kinase provide an alternative picture.\textsuperscript{4} The first is fast and was associated with the local binding of imatinib, and the second, a slower process. The second process was interpreted as an induced fit that follows the binding event. The experiment does not provide a direct structural information on the DFG dynamics during those events, however, a reasonable assumption is that the fast process includes the DFG flip. The measurements that include the insertion of the inhibitor are significantly different from our study that focuses only on the DFG flip. We therefore declined to analyze our results in the context of reference \textsuperscript{4}. 
From computational point of view, a Markov State Model was used to estimate the transition rate of the DFG flip at milliseconds. The model was based on samples of significantly shorter MD trajectories (similar to Milestoning). A small number of transitions was detected making the uncertainty of the longer time scale significant, as noted by the authors. The estimates of the kinetics that use unbiased trajectories, MSM and Milestoning, provide answers close to each other and close to the experimental finding.

4.2 The structural features of the reaction pathways from DFG-in to DFG-out in Abl kinase

In this section we discuss in more details the structural features of the transition from DFG-in to DFG-out. In Fig. 12 we show the residue root mean square difference (RMSD) between the reactant, and several structures along the reaction pathway. We include a comparison of the reactant and of the product to anchor 27, which is near a committor value of 0.5 (Fig. 7).

We overlap optimally all the heavy atoms (atoms that are not hydrogens) of the entire structures and compute the heavy-atom RMSD between the individual residues. We plot the residue RMSD as a function of the residue number. For comparison we also plot the B factor extracted from the PDB coordinates of the reactant (PDB 2F4J) and the product (PDB 2G2F). The B factors make it possible for us to identify residues that are flexible, and/or are present in multiple static conformational states. In contrast, residues that contribute to the spatial progress of the reaction may have displacements not detected by the B factors which are coupled to the DFG flip. Several of the peaks at residues F382, L384, R386, M388 and A397 are part of the A loop.
Fig. 12: Fluctuations and systematic drifts of residues in Abl-kinase. Top panel reports the B factors of the reactant and product structures as a function of the residue index to identify flexible domains. In the lower panel we compare the structure of Anchor 27 with the reactant and product using room mean square difference (RMSD) between all the heavy atoms of the residues in the protein. The two pink arrows point to Glu286 and Lys271 that forms a blocking salt bridge. Note that the transition state differs about equally from the reactant and from the product structures. There are several spikes at Phe382, Leu384, Arg386, Met388 and Ala397 that belongs to the A loop and are included in the set of coarse variables.

We mark with pink arrows the locations of the salt bridge residues: Glu286 shows significant deviation from both, the reactant and the product, but Lys271 is closer to the position of the product state.

There are several residues that interact strongly with the DFG motif. Two of them form the salt bridge that we already discussed. The side chain of Phe401 is highly flexible. It flips between rotational states several times along the reaction coordinate. It is not coupled to the reaction coordinate. Therefore, the sharp spike we observe in Fig. 13 has no impact on the progress of the reaction. Another important residue is Met290 that was discussed extensively in reference 9. Flips of the DFG were observed using conventional MD only after the mutation of the Met290
residue to alanine, reducing the barrier to enter the binding site. To appreciate the coupling between different residues and Phe382 and the order of events during the transition we plot the distance between the Phe382 ring (the CZ atom) and the edges of the side chains of the other residues (Glu286 min(Oe1, OE2), Met290 SD, and Lys271 NZ) in Fig 14. The distance between Phe382 and Glu286 shows significant fluctuations between 6 and 8Å but not overall drift. The distance between Phe382 and Lys271 is decreasing near anchors 20 to 32, which is close to the committor value of 0.5 (Fig. 7) and hence to a transition state. The motions of Met290 come late in the process and significant displacement is observed after the system is leaving the transition state and continue forward. Hence, the Met 290 transition is a late event in the DFG flip.
The distances between atom CZ of Phe382 to atoms representing the ends of the sidechain of residues Met290, Glu286, and Lys271 are shown. For Glu286, the minimum distance to both oxygens OE1 and OE2 was measured. Lys271 and Glu286 form the salt bridge. Around anchor 30, the Phe382 reaches as close as possible to the salt bridge, (Lys 271) breaks it and then goes away from this residue. The distance to Glu286 remains roughly a constant throughout the transition.

4. Conclusions

The activation loop (A-loop) in kinases attracted considerable experimental and theoretical attention, being highly flexible and able to adopt multiple conformations, which hinders its accurate computational investigation with commonly used sampling methods. Here we investigated in detail a transition pathway in Abl kinase from its active to its inactive form. While we investigated the transition of the entire A-loop, which includes multiple rotational events, we focus our analysis on the kinetics of the activated transition of the DFG motif from a DFG-in to a DFG-out state. Quantifying accurately the kinetics of this transition is, however, difficult from both experimental and computational viewpoints. The long transition time (estimated by NMR\(^6\)) makes the direct atomistic simulations of these rare events challenging. Enhanced sampling techniques such as meta dynamics\(^{10}\) and the string method\(^{16}\) were applied to the system and were used to primarily investigate its equilibrium conformations. Simulations with Milestoning, \(^{42}\) suggest that the time scales for the DFG flip are milliseconds to seconds. The extensive information that we gather from the reaction coordinate calculations as well as from the short Milestoning trajectories, allows us to propose a detailed molecular mechanism for the events of the reaction and their coupling to different residues.
We also note that the calculations reported in this paper are conducted with unprotonated Asp381. Previous simulations highlighted the importance of the protonation state of this residue for the conformational transition.\(^9\) Our study will enable future work to examine the impact of the protonation state on the transition pathway.

One should keep in mind the significant conceptual and sampling challenges that calculation of kinetics in the Abl kinase molecular system poses. We cannot be sure that our set of coarse variables is complete, and we are uncertain if all significant transition pathways were sampled. It is expected that the impact of missing alternative pathways will be to reduce the estimated time scale. Nevertheless, the overall agreement of this calculation and experimental estimates is encouraging and it opens the way for quantitative comparisons between simulations and experimental measurements of long-time events in complex and activated biomolecular systems.

5. Acknowledgements

RE acknowledges support from the NIH GM59796 and a Welch grant F-1896. BN and NVB acknowledge the financial support received from Thomas Preston Ph.D. Scholarship Fund and from the European Union’s Horizon 2020 research and innovation programme “NanoinformaTIX” (topic NMBP-14-2018, grant No 814426). RL acknowledges support from the NIH GM-132090 grant.
Declaration of interests
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

6. References


**Graphical abstract:** Changes in the salt bridge between Lys271 and Glu286 for inactive (A), an intermediate state (B), and active (C) states.
Highlights

- The DFG flip in Abl kinase is simulated in atomic details at the millisecond time scale using the method of Milestoning.
- The active state is about 2 kcal/mol lower than the inactive state.
- The mechanism of the transition includes a network of states. The network cannot be captured quantitatively by a single reaction coordinate.
- The salt bridge, Lys 271 and Glu 286, forms the major barrier and is transiently broken during the transition process. A secondary and a later barrier is presented by Met 388.
- Elements of the transition are diffusive, showing non-Markovian and non-activated behavior.