Transplant−Insert−Constrain−Relax−Assemble (TICRA): Protein−Ligand Complex Structure Modeling and Application to Kinases

Siavash Meshkat,* Anthony E. Klon, Jinming Zou, Jeffrey S. Wiseman,[1] and Zenon Konteatis

Ansaris, Four Valley Square, 512 Township Line Rd, Blue Bell, Pennsylvania 19422, United States

Received July 6, 2010

We introduce TICRA (transplant−insert−constrain−relax−assemble), a method for modeling the structure of unknown protein−ligand complexes using the X-ray crystal structures of homologous proteins and ligands with known activity. We present results from modeling the structures of protein kinase−inhibitor complexes using p38 and Lck as examples. These examples show that the TICRA method may be used prospectively to create and refine models for protein kinase−inhibitor complexes with an overall backbone rmsd of less than 0.75 Å for the kinase domain, when compared to published X-ray crystal structures. Further refinement of the models of the kinase domains of p38 and Lck in complex with their cognate ligands from the published crystal structures was able to improve the rmsd's of the model complexes to below 0.5 Å. Our results show that TICRA is a useful approach to the problem of structure-based drug design in cases where little structural information is available for the target proteins and the binding mode of active compounds is unknown.

INTRODUCTION

The accuracy of a protein structure is one of the most critical factors that affect the calculation of binding energies and prediction of binding poses in protein−ligand binding. The current gold standard for protein structure determination is X-ray crystallography. However, in many cases, the complexity, cost, and delays in obtaining X-ray crystal structures have necessitated the exploration of computational alternatives in structure-based design programs. While computational approaches based on first principles have been developed that can predict protein structure from a protein's amino acid sequence, these methods are currently limited in accuracy to relatively small proteins.[1−3] Alternative methods are required because a typical protein target for small molecule therapeutics may have over 200 amino acids. Furthermore, even first-principle computation of a large-scale conformational change can take tremendous computational resources. For example, the reorientation of the protein kinase DFG motif from the in to the out conformation can take several weeks of computation.[4,5] Heuristic computational methods, such as using a reduced-space model of a protein where amino acids are represented by Cα, Cβ, side-chain mass center, and the peptide bond center, have been developed as a consequence.[2] Homology-based methods use the geometry of homologous proteins to model an unknown structure. These methods have been augmented by using energy minimization to add more accuracy to the structure.[6] Other approaches have used molecular dynamics[7] or statistical methods to improve the quality of the homology models.[8] A comprehensive introduction to homology modeling methods has been compiled in a recent survey of commercially available approaches.[9]

We have developed an alternative computational method, TICRA (transplant−insert−constrain−relax−assemble), for the refinement of a protein structure from the available crystallographic data and small molecules with experimentally measured binding affinities to the target. The method is used to refine a homology model or X-ray crystal structure in the presence of substituent fragments derived from known small molecule ligands. We have extended molecular mechanics to include the presence of ligands or their constituent fragments, with interaction constraints, and incorporate torsion-space energy minimization and molecular dynamics in an iterative process of fragment assembly and validation. Our approach relies upon our previously reported fragment-based molecule assembly method, as described in a separate publication.[10] We refer to the specific protocol of molecular dynamics and energy minimization as relaxation. Our method relaxes the protein structure in response to one or more ligands and tests whether active molecules can be assembled from their substituent fragments in the context of the relaxed binding site. It is important to note that the approach we describe is not a high-throughput virtual screening method nor is it analogous to docking. Rather, the utility of this approach lies in the ability to model a conformation for a given protein that is suitable for the design of novel inhibitors. As such, the accurate scoring of the resulting protein/ligand complexes obtained through TICRA is not described because this is a separate computational problem that is not relevant to the task at hand. In this paper, we apply the refinement protocol to protein kinases.

Two major binding modes of inhibitors to protein kinases have been studied extensively.[11] The type I binding mode is similar to how ATP binds to kinases and is characterized by the Asp of the conserved DFG-loop oriented toward the ATP binding site in the "DFG-in" conformation. This binding mode will be referred to as ATP-like. The type II binding mode was first postulated based on the Abl−Gleevec binding complex[12] and was later crystallographically confirmed with the

* To whom correspondence should be addressed. E-mail: smeshkat@Locuspharma.com.

[1] Present address: Pharmatrope Ltd., 324 Croton Road, Wayne, PA 19087.

10.1021/ci100256u
Published on Web 11/30/2010

p38−BIRB796 complex[13] and the Abl−Gleevec complex.[14] It is characterized by a reorganization of the DFG-loop that orients the DFG Asp away from the ATP site, exposing an allosteric pocket distal to the ATP binding site. We refer to this binding mode as allosteric.

As of early 2007, several ATP-like cocrystal structures of Lck were available on the Protein Data Bank, but no allosteric structures were available.[15] This led us to develop a computational costructure of Lck kinase bound with Gleevec, a known inhibitor. Amgen subsequently published two cocrystal structures of Lck with allosteric compounds.[16] This allowed us to evaluate our computational structure and the underlying method by comparing our results with the reported structures from Amgen.

A method for protein−ligand conformation prediction has been proposed which samples multiple poses of the ligand and protein and performs induced-fit docking. The best protein−ligand complexes are selected using a scoring function.[17] An alternative method proposed for the specific case of allosteric protein kinase inhibitors uses a novel pharmacophore-like field to morph the kinase DFG motif into a type II conformation.[18] We will discuss the differences between these approaches and the method we present here. In order to illustrate the application of TICRA to modeling allosteric structures, we compute the allosteric complex of p38 bound to a known inhibitor, using an ATP-like crystal structure and a known allosteric compound reported by Astex[19] as a starting point. The computed structure is compared with the known crystal structure published for the allosteric compound. We also describe the application of TICRA to develop a computational model of Gleevec bound to Lck. The crystal structures reported by Amgen are then compared to this computational model.

METHODS

Transplant−Insert−Constrain−Relax−Assemble (TICRA). In its most general form, the TICRA method involves a sequence of iterations to model a protein−ligand complex. One or more of the steps may be omitted from an iteration in specific cases. These iterations are performed sequentially for each constituent fragment in a ligand until the entire molecule may be placed in the binding site. The final protein−ligand complex is the result of these calculations at the conclusion of this process. Below, we describe the steps involved in each iteration.

Transplant. During the transplant step, features of the protein model are modified, or additional features are added. A feature is typically a group of related atoms comprising secondary structure elements, such as an α helix or a loop. This step is often necessary because an initial protein model may have amino acids that are missing from the structure or have been reported with zero occupancy. It may also be used to splice secondary structural elements of interest from another protein structure that exist in different conformations. We use the transplant step here in the context of kinases, where an existing crystal structure for the protein of interest exists in the DFG-in conformation, but the goal is to model the protein−ligand complex with the kinase in the DFG-out conformation. The ligands discussed here are type II kinase inhibitors, which bind to the DFG-out conformation of their respective kinases. The existing DFG-in conformation of these kinases is therefore inappropriate, as it would sterically inhibit the binding of these known inhibitors. Coordinates for secondary structural elements are taken from structurally homologous protein crystal structures with high sequence identity to the protein being modeled. A structural feature may include additional water molecules. Water molecules may be taken either from the coordinates of crystallographically resolved water molecules in the X-ray structure or the computed position of high-affinity water molecules determined using the systematic sampling protocol described in detail elsewhere.[10]

Insert. In the insert step, a known active molecule is placed in the current protein model. The new molecule may be a whole ligand or one or more fragments. To obtain an initial placement of the molecule fragments, a systematic sampling of the fragments is performed in the presence of the current conformation of the protein, with or without additional constraints, to identify low-energy poses of each fragment, using a method described in a separate paper.[20] If no low-energy pose is found in the expected region, we incrementally reduce the protein's van der Waals radius by about 5% and repeat the process, until a low-energy pose of the fragment is found.

Constrain. Two general computational protocols are practiced in the constrain step of the TICRA method, which we refer to as the constraint lifting protocol and the new feature protocol.

Computational constraints are enforced using an energy incentive. For example, to enforce a constraint to keep two atoms at a preferred distance, an energy incentive is attached that is similar to the empirical bond energy. The energy incentive of the constraint is gradually reduced to zero while relaxation is repeated. This is referred to as the *constraint lifting protocol*. The final relaxation step is performed with no enforced constraints to ensure the model naturally maintains the constraints.

Because the insert step may result in fragments of ligand being placed in the binding site with a van der Waals radius of 80−90%, the initial coordinates may not be energetically favorable. In order to preserve key molecular interactions, it may be necessary to add geometric constraints between the positions of the protein and ligand atoms or explicit water molecules. These constraints may consist of distance constraints between two atoms, angular constraints between three atoms, or weights applied to the force field to energetically penalize the movement of atoms in the simulation from their starting coordinates.

A "new feature" refers to a set of atoms that have been transplanted at a new position, as in the case of protein atoms, or inserted, as in the case of a small molecule. The new feature of a protein model added in the transplant step may have the constraints applied to the backbone only or include side chains. The new feature protocol refers to performing simulations in successive stages. In the first stage, the new feature is kept fixed, but allows only the neighboring atoms from the initial model to move during the relax (described below) step. In the second stage, the new feature is kept fixed, but allows all other atoms in the model to move. In the final stage, all atoms are allowed to move freely during the relax step without any restraints.

Relax. A sequence of short time scale molecular dynamics and energy minimization steps is performed in torsion space

for 10−200 ps on the modified protein in the presence of new protein features, fragments, whole molecules, water molecules, and constraints added to the system in the first three steps. This allows the protein and the ligand atoms to rearrange. Unless otherwise specified, the relax protocol consists of energy minimization in Cartesian space, then in torsion space, followed by 100 ps of molecular dynamics in torsion space, followed by an energy minimization in torsion space, and a final minimization in Cartesian space. Molecular dynamics in torsion space was performed using the Imagiro system.[1] Torsion-space simulation is typically faster than Cartesian-space simulation due to fewer degrees of freedom.

Assemble. The assembly step tests the current protein structure against other molecules with known activity. The assembly involves two main stages. The first step is to systematically sample the relaxed protein interactions with the fragments of the expected molecule. This is followed by assembling the fragments into a whole molecule, with or without additional constraints. The linking of fragments is performed using internally developed software, which has been previously reported.[10]

A systematic sampling program is applied to perform the first step to identify the most favorable poses in the binding site and to calculate the interaction energies between the protein and the fragment.[20] The assembly of the fragments into a whole molecule is done using a systematic assembly algorithm.[10] This algorithm considers all the poses in the six-dimensional configuration space (consisting of translations and rotations) of the fragment distributions. By considering only two populations of rigid fragments at a time and using prudent preprocessing and search pruning techniques, the algorithm avoids costly combinatorial and geometric processing.

Force Field. All the computations in this paper were performed using two compatible empirical force fields. For the proteins, Amber 1999 is used.[21] For ligands, its companion force field, Gaff 2003 (General Amber Force Field),[22] is used. AM1-BCC[23] partial charges were calculated for the ligands using QUACPAC.[24] These Cartesian-based force fields are projected into torsion space.[25] The torsion space MD simulations were carried out at 300 K with MM/GBSA continuum solvent model and no counterions, using the Imagiro internal software.[1]

RESULTS

The TICRA Method Can be Used To Generate a Model Allosteric Structure of p38 using the ATP-like Crystal Structure as a Starting Point. We applied the TICRA method to the p38 protein kinase to test whether an accurate allosteric conformation of the protein may be generated using the ATP-like conformation as a starting point. Several structures of p38 are available from the Protein Data Bank,[26] crystallized in the presence or absence of ligands. P38 structures have been reported with cocrystallized ligands in both the ATP and allosteric binding sites, corresponding to the DFG loop made up of residues 168−170 existing in the in and out conformations. In this study, a cocrystal structure of p38 in complex with a ligand in the ATP site is used as the initial protein model for the ATP-like conformation of p38.[27] Residues Ala^{399} through Asp^{410} from the allosteric structure of the Abl tyrosine kinase

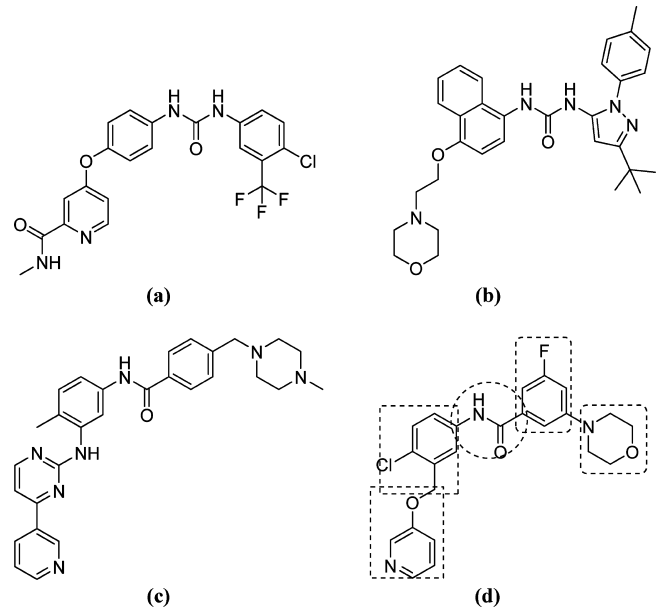

Figure 1. Compounds used for TICRA: (a) BAY-43-9006; (b) BIRB-796; (c) Gleevec; (d) Astex compound 9, from PDB code 1W83.

Table 1. Structural Templates for p38 and Abl

protein↓	DFG-in (ATP-like)	DFG-out (allosteric)
Abl	2G2F	1OPJ
p38	1A9U	to be computed (known result = 1W83)

bound with Gleevec (1OPJ) were used as a template for the conformation of the activation loop and the conserved DFG motif. Although the a-loop itself does not interact directly with the ligand, its inclusion into the model is necessary because the conformation of this loop is affected by the conformation of the DFG-loop (residues 168−170), which immediately preceeds it. Using the Abl activation loop as the template for the homology model poses a challenging test of our method because Abl and p38 share only 23% sequence identity in the kinase domain and 21% in the activation loop.[28] Gleevec cannot be used for the refinement of the model protein structure because it does not bind to p38. Astex compound 9, shown in Figure 1d, is used for the structural refinement instead. The Astex compound has been cocrystallized with p38 and shows a similar binding mode as Gleevec binding to Abl, with the protein in the allosteric conformation.[19] The Astex compound is reported to have an IC_{50} value of 65 nM against p38. Table 1 summarizes the structures used for the structure refinement and validation study.

The ATP-like conformation of Abl was first aligned to the ATP-like conformation of p38 along the conserved residues of the binding site to obtain an alignment. The internally developed Imagiro program was used to carry out the alignment.[1] Subsequently, the allosteric conformation of Abl was aligned to the ATP-like conformation of Abl obtained above, along the binding site residues (excluding the a- and p-loops).

The backbone of the activation loop (a-loop) from the allosteric conformation of Abl kinase spanning Ala^{399} to Asp^{410} was transplanted into p38-ATP to replace residues Leu^{167} through Glu^{178} of the ATP-like p38 template. The

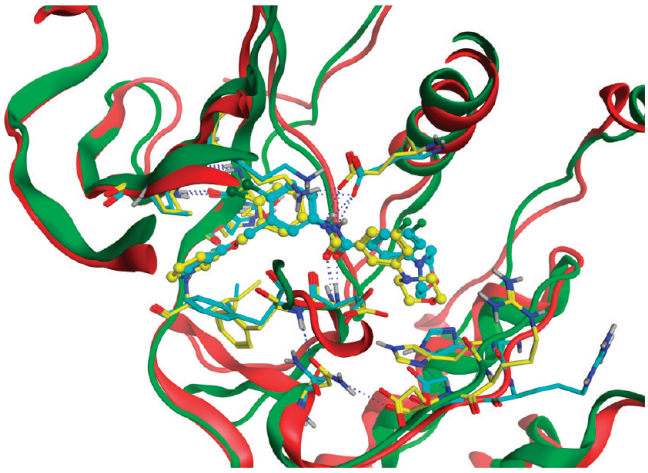

Figure 2. 1W83 crystal structure superimposed with the computed p38 allosteric structure. The ligand is shown in ball-stick. Protein residues shown in stick (yellow, 1W83; cyan, model). Protein backbone shown in cartoon (green, 1W83; red, model). Hydrogen bonds are shown in blue dashed lines. The graphics was generated using MOE.[29]

starting and ending residues were selected to be homologous residues that occupied the same relative space in both ATP-like and allosteric conformations after the structural alignment. The initial conformation of side chain atoms was generated using Imagiro.[1] The protein model was constrained such that only side chains of the a-loop were permitted to move, and the protein model was relaxed. Another round of relaxation was applied with the constraint that all atoms of the a-loop and no others were allowed to move. A third round of relaxation was carried out with the constraint that only atoms in the a-loop and within 5 Å of the a-loop were permitted to move.

Astex compound 9 was then inserted into the allosteric binding site one fragment at a time, and the intermediate protein-fragment models were relaxed. Figure 1d shows the fragment decomposition of compound 9 using dashed contours enclosing each fragment. After each insertion of a fragment, the protein model−ligand complex was first relaxed with the constraint that the position of the fragment atoms were held fixed. This was followed by a subsequent round of relaxation in which the protein-fragment complex had no constraints. Once the ligand had been fully assembled, the complex was relaxed in the presence of Astex compound 9, first with the constraint that the compound was held fixed, followed by relaxation without any constraints on the positions of the ligand atoms. The resulting model for the p38−Astex compound 9 complex is shown in Figure 2.

The TICRA Method Accurately Reproduces the Experimentally Observed Allosteric Structure of p38. Table 2 shows the rmsd of p38 computed structures compared to the 1W83 structure. The most significant differences are highlighted in red. The largest rmsd for the critical residues occurs at the DFG Phe[169] side chain. This may be due to the large degree of flexibility of this residue. In the X-ray structure of 1W83, the average crystallographic *B*-factor of the residues shown in table 2 is 29.06, while the average *B*-factor for the heavy atoms of Phe[169] is 31.76.

The compound's amide group and its surrounding phenyl groups align very well, with rmsd less than 0.7 Å as compared to the same compound in the crystal structure. The

Table 2. Rmsd (Å) of p38 Allosteric Crystal Structure versus the Original p38 ATP-like Crystal Structure and the Computed p38 Allosteric Structure

	1W83 **vs 1A9U**		1W83 **vs p38 allosteric model**	
residues	backbone	side chain	backbone	side chain
ASP-168	1.56	4.63	0.30	1.33
THR-106	0.52	0.47	0.33	0.81
HIS-107	0.36	0.77	0.24	0.65
MET-109	0.43	0.52	0.30	0.87
LYS-53	1.36	1.58	0.39	0.56
ASN-155	0.67	1.55	0.32	1.67
VAL-38	1.70	1.66	0.43	0.35
GLU-71	0.94	1.05	0.65	0.62
PHE-169	6.33	10.42	*0.73*	*4.41*
ALA-51	0.87	0.74	0.33	0.43
mean	**1.47**	2.34	**0.40**	1.17

Table 3. IC_{50} (μM) of Allosteric Compounds against Lck and Abl

protein	BIRB-796	Gleevec	BAY-43-9006
Lck	1.1	0.062	6.9
Abl	1.5	0.0022	0.13

p38 model is very similar to the p38 X-ray structure in the hinge region (residues 104−109), with a backbone rmsd of 0.5 Å. Likewise, the α-C helix region (residues 67−79) has a backbone rmsd of 0.4 Å.

TICRA Can Be Used To Generate an Allosteric Binding Conformation of Lck in Complex with Gleevec from the Crystal Structure of the ATP-like Conformation of Lck. Lck is a nonreceptor cytoplasmic tyrosine kinase of the Src family that binds three well-known compounds (BIRB-796, Gleevec, and BAY-43-9006; binding affinities shown in Table 3) that typically bind to kinases in an allosteric binding mode. While Gleevec binding mode with most kinases is allosteric, there is a crystal structure where it binds to Syk entirely in the ATP pocket.[30] Because Lck and Abl have a very similar binding profile for allosteric compounds, as shown in Table 3, we hypothesize that Gleevec binds to Lck in an allosteric fashion analogous to its observed binding mode with Abl. In addition to having a similar activity profile with the three allosteric compounds shown in Table 3 and comparable binding affinities reported for Gleevec, the Lck and Abl activation loops are of the same length. The sequence alignment of the 24 residue a-loop is colored blue in Table 4.

Table 5 depicts the relationship between three known structures of Abl and Lck and the unknown structure we would like to compute. Of the known structures for Abl and Lck, we chose three that have a good degree of structural similarity with each other, using the rmsd as an indicator. To improve the alignment, the Abl ATP-like conformation was first aligned to the Lck ATP-like conformation, using eight conserved binding site residues. Subsequently, the Abl allosteric conformation was aligned to the Abl ATP-like conformation obtained above, using the binding site residues (excluding the a- and p-loops). The alignment was done along eight conserved and key residues of these two kinases. The rmsd of these two alignments in shown in Table 6, with the eight aligned residues shaded. This resulted in an adequate alignment of Abl-allosteric to Lck-ATP, as shown in Table 7 and Figure 3. Near the binding site, the main differences

Table 4. Residue Alignment of Lck and Abl near the a-Loop (blue)

```
                                380       390       400       410
                                |....*....|....*....|....*....|....*....|
Lck: 1QPC 370 ILVSDTLSCKIADFGlarliednextareGAkFPIKWTAPE 410
Abl: 1OPJ 388 CLVGENHLVKVADFGlsrlmtgdtytahaGAkFPIKWTAPE 428
```

Table 5. Structural templates for Lck and Abl

Protein	ATP-bound	Gleevec-bound
Abl	2G2F	1OPJ[31]
Lck	1QPC[32]	unknown

between the Lck ATP-like conformation and the Abl allosteric conformation are in the DFG conformation and the p-loop (also known as the "glycine-rich loop").

In particular, rmsd values of the hinge (Lck Thr^{316} to Gly^{322} and Abl Thr^{334} to Gly^{340}), the α-C helix (Lck Pro^{282} to Leu^{285} and Abl Val^{298} to Ile^{312}), and the residues prior to the DFG motif (Lck Cys^{378} to Ala^{381} and Abl Val^{396} to Ala^{399}) are within 0.5 Å.

The backbone position of the 24-residue activation loop of the Abl allosteric structure (*a-loop*) was transplanted into the Lck ATP-like structure to assemble an Lck allosteric model. In addition, eight water molecules within 6 Å of the inhibitor were extracted from the Abl allosteric structure and inserted into the Lck allosteric model. We also considered the case where no crystallographic water molecules were inserted into the structure. The protein model was constrained such that only the side chains of the a-loop were allowed to relax in torsion space. This was followed by another round of relaxation in torsion space in which the a-loop backbone and side chain atoms were allowed to move. In the final round of relaxation, all of the atoms in the a-loop and the protein atoms within 5 Å were allowed to move freely. Gleevec was then assembled from its constituent fragments whose binding modes were simulated in the resulting structure. Gleevec was constrained and the rest of the protein was relaxed in torsion space. This was followed by a final round of relaxation where all protein and ligand atoms were permitted to move freely.

The resulting structure of the allosteric model for Lck is shown in Figure 4 in complex with Gleevec. The allosteric structure of the Lck−Gleevec complex is superimposed on the ATP-bound structure of Lck in Figure 5. The calculated rmsd values of the key residues resulting from the structural superposition are shown in Table 8. The main differences between the two structures lie in the a-loop, including the conserved DFG-motif, and the α-C helix. In particular, Phe^{383} shows the greatest variability in the a-loop due to the great difference in conformation between the DFG-in and DFG-out states. Glu^{288} in the α-C helix also shows a great difference in the calculated rmsd values. The location of the main-chain atoms in this residue are more varied due to the fact that the entire α-C helix is translated outward

Table 6. Alignment of the Abl ATP-like Conformation to the Lck ATP-like Conformation and the Ab-Allosteric Conformation[a]

lck-1QPC Residue	abl-2G2F Residue	Backbone	Sidechain	abl-2G2F Residue	abl-1OPJ Residue	Backbone	Sidechain
THR-316	THR-315	0.21	0.24	THR-315	THR-334	0.10	0.27
GLU-317	GLU-316	0.49	0.99	GLU-316	GLU-335	0.12	0.30
MET-319	MET-318	0.35	1.51	MET-318	MET-337	0.50	1.41
ALA-271	ALA-269	0.17	0.24	ALA-269	ALA-288	0.27	0.28
ASN-369	ASN-368	0.60	0.82	ASN-368	ASN-387	0.37	0.57
VAL-259	VAL-256	0.63	1.68	VAL-256	VAL-275	0.57	1.87
LYS-273	LYS-271	0.27	1.61	LYS-271	LYS-290	0.32	1.70
ALA-271	ALA-269	0.17	0.24	ALA-269	ALA-288	0.27	0.28
GLU-288	GLU-286	2.25	3.30	GLU-286	GLU-305	0.75	2.97
ASP-382	ASP-381	3.26	7.68	ASP-381	ASP-400	4.64	7.85
PHE-383	PHE-382	2.68	8.66	PHE-382	PHE-401	5.80	8.52
HIS-362	HIS-361	1.89	1.65	HIS-361	HIS-380	0.83	1.34
ARG-363	ARG-362	3.26	6.16	ARG-362	ARG-381	2.06	6.40
ASP-364	ASP-363	1.64	2.19	ASP-363	ASP-382	0.99	0.89

[a] Shaded regions mark residues used for the alignment.

Table 7. Calculated Rmsd Values for Selected Active Site Residues Resulting from the Alignment of the Abl Allosteric Conformation to the Lck ATP-like Conformation[a]

lck-1QPC Residue	abl-1OPJ Residue	Backbone	Sidechain
THR-316	THR-334	0.21	0.38
GLU-317	GLU-335	0.59	1.06
MET-319	MET-337	0.77	0.84
ALA-271	ALA-288	0.22	0.20
ASN-369	ASN-387	0.63	0.64
VAL-259	VAL-275	0.59	0.82
LYS-273	LYS-290	0.24	1.27
ALA-271	ALA-288	0.22	0.20
GLU-288	GLU-305	1.84	1.88
ASP-382	ASP-400	2.00	3.21
PHE-383	PHE-401	6.81	9.39
HIS-362	HIS-380	2.05	1.70
ARG-363	ARG-381	2.04	3.77
ASP-364	ASP-382	1.54	1.87

[a] Shaded regions mark the residues used for the alignment.

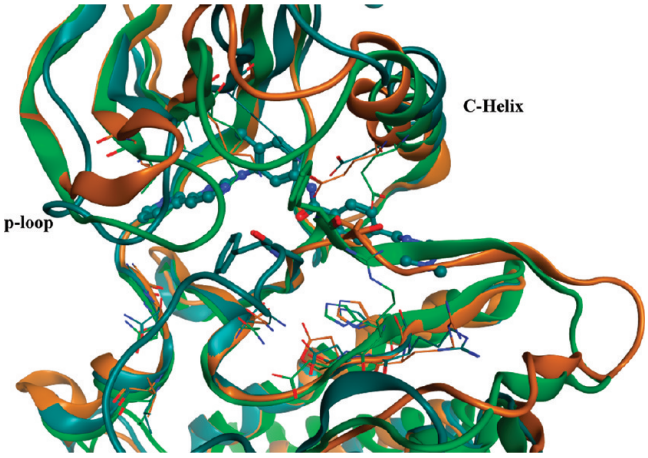

Figure 3. Structural alignment of the Lck ATP-like (orange, PDB ID 1QPC), the Abl ATP-like (green, PDB ID 2G2F), and the Abl allosteric (cyan, PDB ID 1OPJ) crystal structures. 1OPJ ligand shown in ball stick. DFG Phenylalanine shown in stick. The graphics were generated in MOE.[29]

from the active site in the allosteric model relative to the ATP-like crystal structure. In addition, the side-chain oxygen atoms of Glu[288] participate in an intermolecular hydrogen-bonding interaction between Gleevec and Lck, resulting in a different torsion conformation for the side chain of this residue.

Changes to the p-loop were not explicitly modeled using the TICRA method because it was possible to assemble Gleevec in the allosteric model with the transplanted a-loop and inserted crystallographic water molecules alone. With longer duration torsion space molecular dynamics simulations, it is plausible that the p-loop will close in on the binding site in a manner similar to the conformation observed in the Abl−Gleevec crystal structure. However, the putative differences in the p-loop are not anticipated to have a significant impact on the design of allosteric compounds, because most allosteric compounds do not make strong interactions with this loop. Indeed, the successful application of TICRA to develop a model for the Lck−Gleevec complex suggests that a fragment-based drug design approach could be readily applied to the search for novel type II inhibitors that bind to the allosteric conformation of Lck.

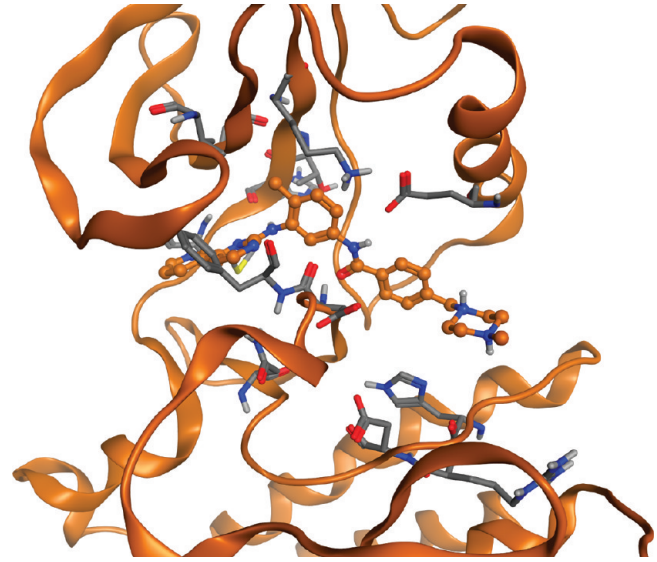

Figure 4. Lck allosteric computed structure in complex with Gleevec. The graphics were generated in MOE.[29]

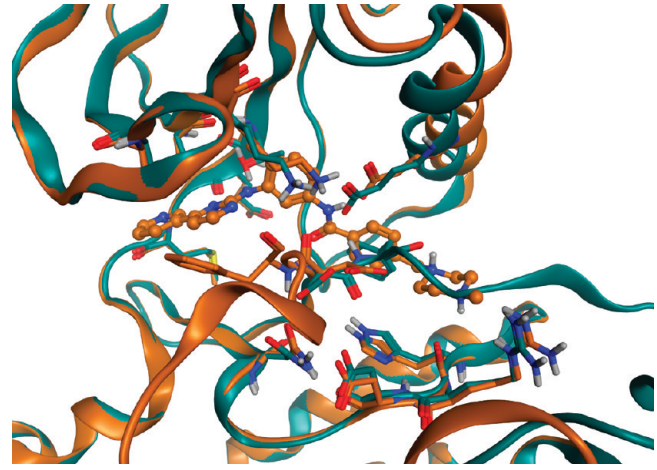

Figure 5. Lck allosteric model (orange) versus Lck ATP-bound crystal structure (cyan). The graphics were generated in MOE.[29]

The Lck Allosteric Model Computed Using TICRA Prospectively Closely Matches the Published Crystal Structures. As mentioned earlier in the paper, two crystal structures were subsequently published by Amgen and provided the first opportunity we had to validate our computed structures against an experimental one. Table 9 shows the rmsd of critical residues of these structures, 2G08 and 2FOV, after alignment along these residues.

The 2G08 and 2FOV structures were aligned to the Lck allosteric model using 11 residues shown in Table 10. As an example, a superposition of 2OG8 and our original structure is shown in Figure 6. The rmsd of the Lck allosteric model versus the aligned 2OG8 and 2OFV is summarized in Table 10. The mean calculated rmsd value of the backbone atoms for these residues was 0.66 Å. When the eight structural waters were omitted from the model, little differ-

Table 8. Calculated Rmsd Values for the Structural Alignment of the Abl Allosteric to the Lck ATP-like[a]

Lck Residue	Backbone	Sidechain
THR-316	0.23	0.31
GLU-317	0.32	0.26
MET-319	0.27	0.28
ALA-271	0.22	0.49
ASN-369	0.57	0.61
VAL-259	0.62	0.82
LYS-273	0.59	0.85
ALA-271	0.22	0.49
GLU-288	1.13	1.10
ASP-382	1.98	3.19
PHE-383	7.00	12.28
HIS-362	1.57	1.49
ARG-363	1.58	1.47
ASP-364	1.41	1.48

[a] residues used in the alignment are shaded.

Table 9. Calculated Rmsd Values of 2OG8 versus 2OFV

	residue	backbone	side chain
α-C	GLU-288	0.75	0.74
hinge	THR-316	0.18	0.38
	GLU-317	0.28	0.25
	MET-319	0.66	1.36
DFG	ASP-382	0.37	1.47
	PHE-383	0.75	1.90
β-3	LYS-273	0.95	0.94
	ALA-271	0.26	0.38
β-2	VAL-259	1.63	1.58
C-lobe HRD	ARG-363	0.46	0.91
	ASP-364	0.62	0.93
	mean	0.63	0.98

ence was observed in the calculated rmsd values between the final model and the published crystal structures. The Lck model is very similar to the 2G08 and 2OFV X-ray structures in the hinge and α-C helix regions of the binding site. For example, in the hinge region it has a backbone rmsd of 0.8 Å, versus the 2G08 structure. Likewise, the α-C helix region (residues 67−79) has a backbone rmsd of 1.3 Å.

Because Gleevec is sufficiently different than the 2OG8 and 2OFV compounds, we minimized our protein structure in the presence of the 2OG8 and 2OFV ligands. A superposition of 2OG8 and our minimized structure is shown in Figure 7. The rms deviation of results are summarized in Table 11. The mean calculated rmsd of the backbone atoms in these residues was 0.46 Å.

CONCLUSIONS

We have been able to use the TICRA protocol to create computational complexes of two protein kinases given an initial protein structure from X-ray crystallography and active compounds with known (or presumed) binding modes. The resulting models are within the accuracy of known experimental methods as evidenced by results for the Lck allosteric model, which was obtained prior to any available experimental data.

Table 10. Rmsd of Lck-Allosteric versus 2OG8 and 2OFV

		rmsd vs 2OG8		rmsd vs 2OFV	
	residue	backbone	side chain	backbone	side chain
α-C	GLU-288	1.39	1.47	0.70	0.88
hinge	THR-316	0.14	0.27	0.31	0.58
	GLU-317	0.39	0.84	0.58	0.89
	MET-319	0.43	1.11	0.65	1.87
DFG	ASP-382	1.00	1.58	0.68	1.77
	PHE-383	0.75	3.94	0.93	3.58
β-3	LYS-273	0.95	1.24	0.65	1.08
	ALA-271	0.28	0.50	0.20	0.46
β-2	VAL-259	1.23	1.30	1.86	1.82
C-lobe HRD	ARG-363	0.39	1.55	0.80	1.52
	ASP-364	0.27	0.93	0.53	0.59
	mean	0.66	1.34	0.72	1.37
	mean (w/o water)	0.6	1.2	0.7	1.2

There are several differences between our approach and that of Sherman,[17] which samples multiple poses of the ligand and multiple poses of the protein and performs induced-fit docking using a scoring function. Our initial ligand conformation is computed using a fragment-based systematic sampling and scoring followed by molecule assembly. We allow the protein and ligand atoms to move in torsion space in addition to Cartesian space while carrying out molecular mechanics calculations during the relaxation part of TICRA. We also allow explicit enforcement of specific intermolecular interactions through the judicious use of constraints. Furthermore, we employ homology techniques for transplanting features. As a result of these differences, we are able to achieve sub-angstrom rmsd value as compared to known crystal structures.

The advantage of using a fragment-based molecule assembly over traditional flexible ligand docking is that the

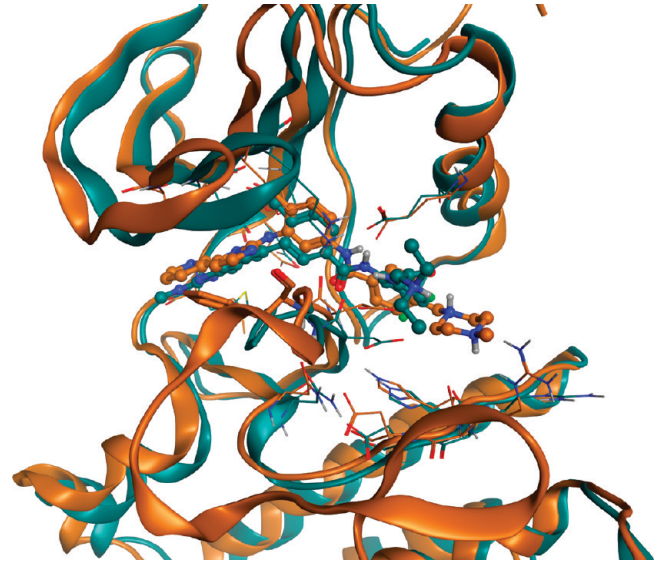

Figure 6. 2OG8 crystal structure (cyan) overlaid with Lck allosteric model (orange). The graphics were generated in MOE.[29]

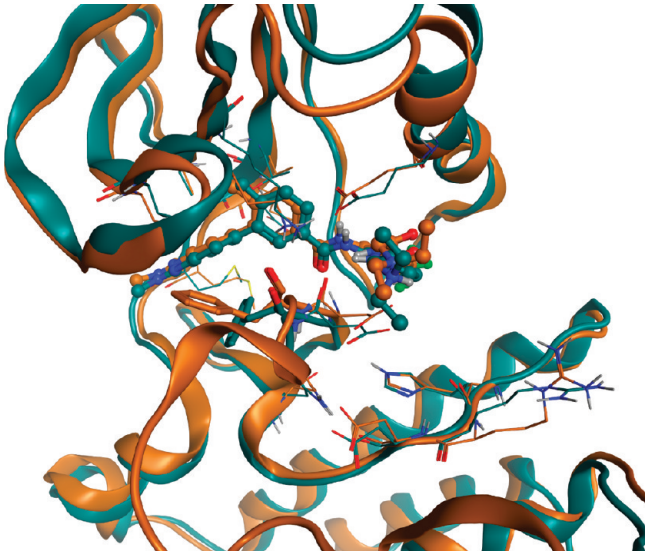

Figure 7. 2OG8 crystal structure (cyan) overlaid with minimized Lck allosteric model and the 2G08 ligand (orange). The graphics were generated in MOE.[29]

Table 11. Rmsd of Minimized Lck-Allosteric versus 2OG8 and 2OFV

		rmsd vs 2OG8		rmsd vs 2OFV	
	residue	backbone	side chain	backbone	side chain
α-C	GLU-288	0.60	0.74	0.37	0.60
hinge	THR-316	0.19	0.34	0.32	0.68
	GLU-317	0.43	0.95	0.71	1.15
	MET-319	0.24	1.11	0.79	1.94
DFG	ASP-382	0.51	0.77	0.77	1.81
	PHE-383	0.72	3.97	0.96	3.32
β-3	LYS-273	0.58	0.94	0.52	0.86
	ALA-271	0.37	0.40	0.27	0.40
β-2	VAL-259	0.47	0.38	0.95	1.15
C-lobe HRD	ARG-363	0.58	1.81	0.82	1.96
	ASP-364	0.37	0.86	0.53	0.65
	mean	0.46	1.05	0.64	1.32

fragment-based method can proceed in an incremental fashion, one fragment at a time. If the addition of a fragment to an intermediate protein conformation is not successful, we have the opportunity to provide a specific remedy. By contrast, failure of ligand docking does not provide any specific hints on how to fix the intermediate protein structure. The purpose of TICRA is to generate protein models that are suitable for prospective structure-based drug design where there is a known small molecule binder but incomplete structural information for the protein. This approach is distinct from the docking of a series of known inhibitors or virtual compounds and scoring the resulting poses. In a docking study, the objective is to rank a series of molecules by their predicted potential to interact with the protein using a scoring function in an attempt to prospectively identify active compounds. In our TICRA method, we use the scoring of individual fragment poses using the computed free energies from systematic sampling to iteratively assemble a known binder in the target site of the protein,[10] allowing the surrounding protein atoms to gradually relax around the growing structure. After using the TICRA method to develop a suitable protein model, we typically simulate several thousand fragments against the model using our proprietary fragment simulation technology. We calculate the interaction energies for each fragment ensemble with the protein, and the highest scoring poses are used for the design of novel compounds.[33]

For modeling the DFG-out conformation of the activation loop in kinases, the Kufareva approach[18] provides an elegant alternative to our transplant step. While their pharmacophore-based approach may require more effort to set up, it may actually speed up our relaxation step, by avoiding unnecessary initial clashes. Their approach is complementary and can be integrated into our general framework.

The TICRA method described here is a valuable tool for the development of high-quality protein models in cases where cocrystal structures do not exist, and enables the application of structure-based drug design approaches to these previously unavailable protein targets.

ACKNOWLEDGMENT

The authors thank Paolo Carnevali, Matthew Clark, Eric Springman, George Talbot, and Haridasan Namboodiri for their help, review, and support.

Supporting Information Available: Tables of additional rmsd data. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Carnevali, P.; Toth, G.; Toubassi, G.; Meshkat, S. Fast protein structure prediction using Monte Carlo simulations with modal moves. *J. Am. Chem. Soc.* **2003**, *125*, 14244–14245.
(2) Kmiecik, S.; Kolinski, A. Characterization of protein-folding pathways by reduced-space modeling. *Proc. Natl. Acad. Sci. U.S/A.* **2007**, *104*, 12330–12335.
(3) Zhang, Y. Progress and challenges in protein structure prediction. *Curr. Opin. Struct. Biol.* **2008**, *18*, 342–348.
(4) Frembgen-Kesner, T.; Elcock, A. Computational sampling of a cryptic drug binding site in a protein receptor: Explicit solvent molecular dynamics and inhibitor docking to p38 MAP kinase. *J. Mol. Biol.* **2006**, *359*, 202–214.
(5) Shan, Y.; Seeliger, M. A.; Eastwood, M. P.; Frank, F.; Xu, H.; Jensen, M. Ø.; Dror, R. O.; Kuriyan, J.; Shaw, D. E. A conserved protonation-dependent switch controls drug binding in the Abl kinase. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 139–144.
(6) Cardozo, T.; Totrov, M. M.; Abagyan, R. A. Homology modeling by the ICM method. *Proteins* **1995**, *23*, 403–414.
(7) Lee, T.; Ma, W.; Zhang, X.; Giles, F.; Cortes, J.; Kantarjian, H.; Albitar, M. BCR-ABL alternative splicing as a common mechanism for imatinib resistance: Evidence from molecular dynamics simulations. *Mol. Cancer Ther.* **2008**, *12*, 3834–3841.
(8) Subramanian, J.; Sharma, S.; B-Rao, C. Modeling and selection of flexible proteins for structure-based drug design: Backbone and side chain movements in p38 MAPK. *ChemMedChem.* **2008**, *3*, 336–344.
(9) Nayeem, A.; Sitkoff, D.; Krystek, S., Jr. A comparative study of available software for high-accuracy homology modeling: From sequence alignments to structural models. *Protein Sci.* **2006**, *15*, 808–824.
(10) Clark, M.; Meshkat, S.; Talbot, G.; Carnevali, P.; Wiseman, J. Fragment-based computation of binding free energies by systematic sampling. *J. Chem. Inf. Model.* **2009**, *49*, 1901–1913.
(11) Okram, B.; Nagle, A.; Adrian, F. J.; Lee, C.; Ren, P.; Wang, X.; Sim, T.; Xie, Y.; Wang, X.; Xia, G.; Spraggon, G.; Warmuth, M.; Liu, Y.; Gray, N. S. A general strategy for creating "inactive conformations" Abl inhibitors. *Chem. Biol.* **2006**, *12*, 779–786.
(12) Schindler, T.; Bornmann, W.; Pellicena, P.; Miller, W. T.; Clarkson, B.; Kuriyan, J. Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* **2000**, *289*, 1938–1942.
(13) Pargellis, C.; Tong, L.; Churchill, L.; Cirillo, P. F.; Gilmore, T.; Graham, A. G.; Grob, P. M.; Hickey, E. R.; Moss, N.; Pav, S.; Regan, J. Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat. Struct. Biol.* **2002**, *9*, 268–272.
(14) Nagar, B.; Hantschel, O.; Young, M. A.; Scheffzek, K.; Veach, D.; Bornmann, W.; Clarkson, B.; Superti-Furga, G.; Kuriyan, J. Structural

basis for the autoinhibition of c-Abl tyrosine kinase. *Cell* **2003**, *112*, 859–871.

(15) Zhu, X.; Kim, J. L.; Newcomb, J. R.; Rose, P. E.; Stover, D. R.; Toledo, L. M.; Zhao, H.; Morgenstern, K. A. Structural analysis of the lymphocyte-specific kinase Lck in complex with non-selective and Src family selective kinase inhibitors. *Structure* **1999**, *7*, 651–661.

(16) DiMauro, E. F.; Newcomb, J.; Nunes, J. J.; Bemis, J. E.; Boucher, C.; Buchanan, J. L.; Buckner, W. H.; Cee, V. J.; Chai, L.; Deak, H. L.; Epstein, L. F.; Faust, T.; Gallant, P.; Geuns-Meyer, S. D.; Gore, A.; Gu, Y.; Henkle, B.; Hodous, B. L.; Hsieh, F.; Huang, X.; Kim, J. L.; Lee, J. H.; Martin, M. W.; Masse, C. E.; McGowan, D. C.; Metz, D.; Mohn, D.; Morgenstern, K. A.; Oliveira-dos-Santos, A.; Patel, V. F.; Powers, D.; Rose, P. E.; Schneider, S.; Tomlinson, S. A.; Tudor, Y.; Turci, S. M.; Welcher, A. A.; White, R. D.; Zhao, H.; Zhu, L.; Zhu, X. Discovery of aminoquinazolines as potent, orally bioavailable inhibitors of Lck: Synthesis, SAR, and in vivo anti-inflammatory activity. *J. Med. Chem.* **2006**, *49*, 5671–5686.

(17) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **2006**, *49*, 534–553.

(18) Kufareva, I.; Abagyan, R. Type-II kinase inhibitor docking, screening, and profiling using modified structures of active kinase states. *J. Med. Chem.* **2008**, *51*, 7921–7932.

(19) Gill, A. L.; Frederickson, M.; Cleasby, A.; Woodhead, S. J.; Carr, M. G.; Woodhead, A. J.; Walker, M. T.; Congreve, M. S.; Devine, L. A.; Tisi, D.; O'Reilly, M.; Seavers, L. C. A.; Davis, D. J.; Curry, J.; Anthony, R.; Padova, A.; Murray, C. W.; Carr, R. A. E.; Jhoti, H. Identification of novel p38alpha MAP kinase inhibitors using fragment-based lead generation. *J. Med. Chem.* **2005**, *48*, 414−426.

(20) Carnevali, P.; Toth, G.; Meshkat; S. Method, system and computer program product for identifying binding conformations of chemical fragments and biological molecules. US Patent 2007/0016374, Jan 18, 2007.

(21) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules. *J. Comput. Chem.* **2000**, *21*, 1049–1074.

(22) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(23) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.

(24) OpenEye Software QUACPAC. http://www.eyesopen.com/quacpac (accessed Aug 30, 2010).

(25) Katritch, V.; Totrov, M.; Abagyan, R. ICFF. A new method to incorporate implicit flexibility into an internal coordinate force field. *J. Comput. Chem.* **2003**, *24*, 254–265.

(26) Berman, H.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.; Weissig, H.; Shindyalov, I.; Bourne, P. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(27) Wang, Z.; Canagarajah, B. J.; Boehm, J. C.; Kassisa, S.; Cobb, M. H.; Young, P. R.; Abdel-Meguid, S.; Adams, J. L.; Goldsmith, E. J. Structural basis of inhibitor selectivity in MAP kinases. *Structure* **1998**, *6*, 1117–1128.

(28) NCB I BLAST Program. http://www.ncbi.nlm.nih.gov/blast (accessed Aug 30, 2010).

(29) Chemical Computing Group MOE. http://www.chemcomp.com/ (accessed Aug 30, 2010).

(30) Atwell, S.; Adams, J. M.; Badger, J.; Buchanan, M. D.; Feil, I. K.; Froning, K. J.; Gao, X.; Hendle, J.; Keegan, K.; Leon, B. C.; Müller-Dieckmann, H. J.; Nienaber, V. L.; Noland, B. W.; Post, K.; Rajashankar, K. R.; Ramos, A.; Russell, M.; Burley, S. K.; Buchanan, S. G. A novel mode of Gleevec binding is revealed by the structure of spleen tyrosine kinase. *J. Biol. Chem.* **2004**, *279*, 55827–55832.

(31) Nagar, B.; Hantschel, O.; Young, M. A.; Scheffzek, K.; Veach, D.; Bornmann, W.; Clarkson, B.; Superti-Furga, G.; Kuriyan, J. Structural basis for the autoinhibition of c-Abl tyrosine kinase. *Cell* **2003**, *112*, 859–871.

(32) Zhu, X.; Kim, J. L.; Newcomb, J. R.; Rose, P. E.; Stover, D. R.; Toledo, L. M.; Zhao, H.; Morgenstern, K. A. Structural analysis of the lymphocyte-specific kinase Lck in complex with non-selective and Src family selective kinase inhibitors. *Structure* **1999**, *6*, 651–61.

(33) Clark, M.; Meshkat, S.; Wiseman, J. Grand canonical free-energy calculations of protein-ligand binding. *J. Chem. Inf. Model.* **2009**, *49*, 934–943.

CI100256U