



Published in final edited form as:

*J Chem Inf Model.* 2011 November 28; 51(11): 2904–2914. doi:10.1021/ci200257b.

## Discovery of Novel Checkpoint Kinase 1 Inhibitors by Virtual Screening Based on Multiple Crystal Structures

Yan Li<sup>1</sup>, Dong Joon Kim<sup>1</sup>, Weiya Ma<sup>1</sup>, Ronald A. Lubet<sup>2</sup>, Ann M. Bode<sup>1</sup>, and Zigang Dong<sup>1,\*</sup>

Zigang Dong: zgdong@hi.umn.edu

<sup>1</sup>The Hormel Institute, University of Minnesota, Austin, MN

<sup>2</sup>The National Cancer Institute, Bethesda, MD

### Abstract

Incorporating receptor flexibility is considered crucial for improvement of docking-based virtual screening. With an abundance of crystallographic structures freely available, docking with multiple crystal structures is believed to be a practical approach to cope with protein flexibility. Here we describe a successful application of the docking of multiple structures to discover novel and potent Chk1 inhibitors. Forty-six Chk1 structures were first compared in single structure docking by predicting the binding mode and recovering known ligands. Combinations of different protein structures were then compared by recovery of known ligands and an optimal ensemble of Chk1 structures were selected. The chosen structures were used in the virtual screening of over 60,000 diverse compounds for Chk1 inhibitors. Six novel compounds ranked at the top of the hits list were tested experimentally and two of these compounds inhibited Chk1 activity—the best with an IC<sub>50</sub> value of 9.6 μM. Further study indicated that achieving a better enrichment and identifying more diverse compounds was more likely using multiple structures than using only a single structure even when protein structures were randomly selected. Taking into account conformational energy difference did not help to improve enrichment in the top ranked list.

### Introduction

Molecular docking and receptor-based virtual screening have been an indispensable component within structure-based drug design for hits identification and lead optimization.<sup>1-3</sup> Although ligand flexibility can be handled by a variety of algorithms in current docking implementation, receptor flexibility remains a major outstanding challenge in the practice of docking-based virtual screening because of the high dimensionality of the conformational space and the complexity of energy function.<sup>4</sup> Protein flexibility has long been acknowledged to be often coupled to ligand binding in numerous experimental and theoretical studies. Two kinds of ligand-binding mechanisms have been well discussed<sup>5</sup> and include conformational selection, which assumes that the ligand binds to a pre-existing receptor conformation in an equilibrated ensemble, and induced fit, which presumes that the receptor is induced to the bound conformation when ligand binding. In either case, structural fluctuations of receptors need to be taken into account in docking studies.

Various modeling methods have been developed to incorporate receptor flexibility in molecular docking and virtual screening.<sup>6</sup> Soft docking, which docks ligands to a rigid receptor with a soft scoring function tolerating some steric clashes, has been reported to be worse for identifying known ligands than the hard scoring function when multiple receptor

\*Corresponding Author Footnote Zigang Dong, M.D., Dr. P.H., The Hormel Institute University of Minnesota, 801 16th Ave NE, Austin MN 55912; Telephone 507-437-9600; FAX 507-437-9606.

conformations were used.<sup>7</sup> Several docking programs limit protein flexibility to side chains by exploration of rotamer libraries and make the problem less computationally demanding, but cannot deal with backbone movement or other major structural rearrangements. Another implementation is docking of ligands to multiple receptor conformations, which may either be obtained experimentally by X-ray crystallography<sup>8-12</sup> and NMR spectroscopy<sup>13</sup> or computationally by molecular dynamics,<sup>14</sup> normal mode analysis<sup>15-17</sup> and other techniques.<sup>18</sup>

Checkpoint kinase 1 (Chk1), a serine/threonine kinase, is involved in the S-phase checkpoint and G2 checkpoint, which is a key regulator in the DNA damage-induced signaling pathway.<sup>19</sup> In response to DNA damage, normal cells are arrested at various cell cycle checkpoints (G1/S/G2) to allow DNA to be repaired; however, cancer cells with p53-deficiency, which is common in tumors, are arrested only at the S or G2 checkpoint because p53 is required at the G1 checkpoint.<sup>20</sup> The inhibition of Chk1 is believed to abrogate the remaining checkpoints in cancer cells, consequently leading to cell death due to accumulation of cytotoxicity of chemotherapeutics. Chk1 as a target for chemosensitization and chemoprevention has been reviewed previously.<sup>21-22</sup> Therefore, Chk1 inhibitors are capable of increasing the therapeutic efficacy of anticancer drugs as sensitizing agents. A number of Chk1 inhibitors have been studied during the past decade, and have been reviewed previously.<sup>23-25</sup> Several compounds have been in advanced preclinical or early clinical development.<sup>20, 22</sup> However, a clear need exists for potent and selective Chk1 inhibitors derived from distinct chemotypes, given the unfavorable properties and toxicities of known compounds.

Chk1 is a 54 kDa protein of 476 amino acids comprised of a highly conserved N-terminal kinase domain followed by a linker region and a less conserved C-terminal domain. The first crystal structure of the kinase domain of human Chk1 was reported at 1.7Å resolution in 2000.<sup>26</sup> Now, 43 complex structures of Chk1 and ATP-competitive inhibitors have already been solved as well as 3 complex structures of Chk1 and allosteric inhibitors. The abundance of available Chk1 structures offers the opportunity not only to investigate ligand docking but also to explore the impact of receptor flexibility.

Here we present a successful application of docking-based virtual screening with multiple crystal structures in the discovery of novel Chk1 inhibitors. Although this method was evaluated previously, few successful applications have been reported. A plethora of crystallographic structures of Chk1 were employed in this work targeting both the ATP-binding site and the allosteric site. Each structure first was evaluated by the docking accuracy in prediction of binding mode and the enrichment factor in recovery of known ligands from decoy compounds. Then the optimal ensemble of structures was determined and utilized in virtual screening of more than 60,000 diverse compounds for Chk1 inhibitors. Finally, six compounds were purchased and tested in experiments and two of them showed good inhibitory effects. In addition, the conformational energy difference was calculated with targeted molecular dynamics and whether considering conformational energy could improve the enrichment factor in virtual screening is discussed. We also discuss the advantage of multiple structures over a single structure and the chemical space of known ligands explored by each protein structure.

## Methods

### 1 Receptor preparation

Forty-six crystal structures of Chk1 with different ligands were retrieved from the Protein Data Bank : PDB entries 1NVQ, 1NVR, 1NVS, 1ZLT, 1ZYS, 2AYP, 2BR1, 2BRB, 2BRG, 2BRH, 2BRM, 2BRN, 2BRO, 2C3J, 2C3K, 2C3L, 2CGU, 2CGV, 2CGW, 2CGX, 2E9N,

2E9O, 2E9P, 2E9U, 2E9V, 2GDO, 2GHG, 2HOG, 2HXL, 2HXQ, 2HY0, 2QHM, 2QHN, 2R0U, 2WMQ, 2WMR, 2WMS, 2WMT, 2WMU, 2WMV, 2WMW, 2WMX, 2YWP, 3F9N, 3JVR and 3JVS. Each structure was prepared using the protein preparation workflow in Maestro, and missing atoms and loops were added with Prime in Schrödinger Suite 2009. All water molecules, metal ions and small chemical groups were removed. AMBER03 partial charge was assigned to each receptor atom with Chimera.<sup>27</sup>

## 2 Ligand preparation

Forty-six ligands were extracted from the Chk1 crystal complexes listed above and were prepared with LigPrep. More than 60,000 structurally diverse compounds were downloaded from all-purchasable subset of the publicly accessible ZINC database with the Tanimoto index at 70% cutoff. These compounds are selected because they are commercially available and display some kind of diversity. Two thousand compounds were randomly selected from this dataset to be used in validation of virtual screening. All small molecules, including crystal ligands and compounds from the ZINC database, were submitted to ConfGen<sup>28</sup> to generate ensembles of low-energy conformations in advance. With the random 2,000 compounds, ConfGen output ensembles of conformations for 1,996 compounds. A total of 2,042 small molecules were generated, including 1,996 random compounds and 46 crystal ligands, in the testing set used in validation of virtual screening. All operations were performed within Schrödinger Suite 2009.

## 3 Docking

DOCK 6.3<sup>29</sup> was utilized to dock the pre-generated conformational ensemble into Chk1 structures. The DMS program was used to generate a molecular surface for each receptor. The SPHGEN algorithm was utilized to create a negative image of the binding pocket, which is complementary to the molecular surface. All the spheres within 10 Å of the ligand were selected for docking. The receptor box delimiting the binding pocket was calculated using SHOWBOX with an additional box size so that it was greater than 20 Å in length, which is large enough to contain all crystal ligands and most of the compounds. Potential grids were calculated for each receptor using a 0.3 Å spacing that fully enclosed the spheres, and were generated by the GRID program in advance for increasing docking speed. Each conformation was treated as rigid in the docking simulation, and the number of ligand poses sampled was set to 500. All docking experiments were carried out using the IBM Blue Gene/L Supercomputer.

## 4 Ensemble of multiple structures

For ensembles with 2, 3, 40, 41, 42, and 43 structures, all possible combinations were enumerated and the enrichment factor (EF) for each combination was calculated. For other ensembles, the structures were randomly selected and duplicate combinations were discarded until 40,000 distinct combinations for each ensemble were generated. The combination with the highest EF was reported as the best solution. All of the analyses were performed with a utility program and scripts developed at our laboratory on the Linux workstation.

The formulae used for the EF calculation is:

$$EF = \frac{q/n}{A/N}$$

Where  $N$  is the total size of the compound library;  $n$  is the number of compounds selected after screening;  $A$  is the total number of known inhibitors; and  $a$  is the number of known inhibitors in the selection.

## 5 Targeted molecular dynamics

The targeted molecular dynamics (TMD) simulation was performed with Amber 11. The Chk1 crystal structure from PDB entry 1IA8<sup>26</sup> was used as the starting structure. Of 46 complex structures, 1ZYS was excluded from TMD simulation because of a residue mutation, and the other 45 structures were used as target structures. Amber03 parameters were applied to describe protein atoms. All structures were subjected to energy minimization first using the steepest descent algorithm in 100 steps and then the conjugate gradient algorithm in 1,000 steps. A time step of 1 fs was used to integrate the equation of motion. The system was started with initial velocities of a given temperature (100 K) and then was held at a constant temperature of 300 K. A restraint force of 15.0 was applied to backbone atoms and the RMSD of backbone atoms was calculated after superposition of all protein atoms. All TMD simulations were performed in the implicit solvent model of Generalized Born (GB) for 100 ps.

## 6 Experiments

All of compounds tested *in vitro* were purchased from InterBioScreen (Moscow, Russia). The active CHK1 human recombinant protein and CHKtide (substrate peptide) for the kinase assay were purchased from Millipore (Temecula, CA).

The kinase assay was performed in accordance with instructions provided by Millipore (Temecula, CA). Briefly, the reaction was carried out in the presence of 10  $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P]ATP with each compound in 40  $\mu$ l of reaction buffer containing 20 mM HEPES (pH 7.4), 10 mM MgCl<sub>2</sub>, 10 mM MnCl<sub>2</sub>, and 1 mM dithiothreitol. After incubation at room temperature for 30 min, the radioactivity was determined by scintillation counter. Each experiment was repeated three times.

## Results

### 1 Chk1 ligands

Most of the known Chk1 inhibitors are competitive with ATP. Out of 46 complex structures published, 43 complexes<sup>30-43</sup> have a ligand occupying the ATP-binding site, and the other 3 complexes<sup>44-45</sup> have a ligand bound to the P-5 substrate binding site, which is about 13 Å<sup>44</sup> from the ATP-binding site and is referred to as an allosteric site (Figure 1). Because the ATP-binding site of kinases is highly conserved in sequence and structure, targeting the ATP site could result in off-target side effects. On the other hand, this might offer an advantage of synergistic effects.<sup>46</sup> The allosteric site is considered to be more structurally distinct than the ATP binding site, and makes possible the discovery and design of highly selective kinase inhibitors. Thus, interest has been growing in the development of kinase inhibitors interacting with the allosteric site, and progress has been reviewed recently.<sup>47</sup>

### 2 Chk1 crystal structures

Forty-six Chk1 complex structures with a crystallographic resolution from 1.70 Å to 3.50 Å were utilized in this work. To explore the diversity of these Chk1 structures, we compared both the backbone conformation and binding site conformation for all structures. To compare the backbone conformation of a pair of structures, one structure was used as the reference and the other structure was superposed onto it, and then the pairwise C- $\alpha$  RMSD was calculated. To compare the binding site, only residues within 5 Å of the co-crystallized

ligand were used in the superposition and then the C- $\alpha$  RMSD of selected residues was calculated. The calculation was performed with scripts in Schrödinger Suite 2009.

**Allosteric site**—In three complexes with allosteric inhibitors, the C- $\alpha$  RMSD of backbone between 3JVR and 3JVS is 5.68 Å, indicating that they have similar global structures. In contrast, the C- $\alpha$  RMSD of the backbone between 3F9N-3JVR and 3F9N-3JVS is 11.75 Å and 9.13 Å respectively, showing that 3F9N has a slightly different structure from the other two (Figure 1). However, the C- $\alpha$  RMSD of the allosteric site is between 0.18 and 0.56 Å for any pair of structures, suggesting that the allosteric site for all three structures is very similar.

**ATP-binding site**—For forty-three complex structures with ATP-competitive inhibitors, the C- $\alpha$  RMSD of backbone varies from 0.13 Å (2CGX-2CGV) to 18.90 Å (2WMS-2GHG) (Figure 2). The structural fluctuation comes mainly from the loop rearrangement in the N-terminal domain (Figure 1). The C- $\alpha$  RMSD of the ATP-binding site is between 0.05 Å (2WMW-2WMX) and 3.72 Å (1NVS-2YWP) (Figure 2). Notably, 61.4% of pairwise structures have a RMSD value over 10.0 Å (Figure 2, left) and 31.0% of pairwise structures have a RMSD value over 3.0 Å (Figure 2, right), indicating that the ensemble of Chk1 structures display conformational diversity in both backbone and binding site.

### 3 Prediction of binding mode

A primary measurement of docking accuracy is the ability to reproduce correctly the experimentally determined binding mode of known ligands. To compare docking performance with different protein structures, the ligands from crystallographic complexes were docked to all Chk1 structures, which is also called cross-docking. For each ligand, the cognate protein structure was utilized as the reference and other structures were superposed onto it. In this process, only the C- $\alpha$  atoms of residues forming the binding site were used. The docked pose of the ligand was moved alongside the protein, resulting in an alignment of the docked pose with respect to the reference ligand. Then the RMSD of heavy atoms was computed between the docked pose with the lowest docking score and the crystal pose of the ligand. The common criterion is that the binding mode of a ligand is considered to be successfully predicted if the RMSD is no greater than 2 Å. Thus, performance of one protein structure was simply evaluated by counting successfully docked ligands. The superposition of structures and calculation of RMSD were performed using scripts in Schrödinger Suite 2009.

**Allosteric site**—Table 1 lists the results from cross-docking for 3 allosteric inhibitors and co-crystallized protein structures. In native docking, that is, re-docking of the ligand into the co-crystallized protein structure, the docked pose is within 2.0 Å of the crystal ligand pose for all 3 ligands. In cross-docking, the protein structure from 3JVR performs better than the other two, with which the predicted binding modes for all three known inhibitors are close (RMSD < 2.0 Å) to the corresponding reference ligand.

**ATP-binding site**—Figure 3 displays the performance of 43 Chk1 structures with ATP-competitive ligands, expressed as a percentage of successfully docked ligands. In native docking, the success rate is 90.7%, and for all ligands, the RMSDs are less than 3.0 Å between the docked pose and the reference ligand. In cross-docking, the success rate of individual structures varies from 16.3% to 60.5%, and the average rate of docking accuracy is 34.9%. Any ligand can be recognized by at least one structure, not necessarily the co-crystallized one. Our results are consistent with previous reports<sup>48-51</sup> and also show a significant decrease in success rates from native docking to cross-docking.

#### 4 Recovery of known ligands

Another measurement to assess docking performance is the ability to identify active compounds from a pool of decoys. To evaluate performance of different Chk1 structures, a small compound library was docked to each protein structure. The compound library was constructed as described in Methods and comprises more than 2,000 compounds, including 46 ligands from crystal complexes. The performance of each structure was evaluated by enrichment factor (EF) at the top 1% of ranked list.

**Allosteric site**—Only three true binders exist in the testing set for the allosteric site. With the structure of 3JVR, all three inhibitors are ranked at the top 1%; with the structure of 3F9N, only the native ligand is ranked at the top 1%; with the structure of 3JVS, none of them is ranked at the top 1%.

**ATP-binding site**—Forty-three true binders exist in the testing set for the ATP-binding site, giving a theoretical maximum EF of 47.6 at the top 1%. Figure 4 shows that docking to a single structure can deliver an EF up to 21.4 with two structures (1NVR or 2R0U). In the worst case, none of the known ligands is ranked at the top 1% with two structures (2GDO or 2WMS). The structure giving a high success rate in prediction of binding mode does not necessarily obtain a high EF value in recovery of known ligands, which agrees with previous results.<sup>48-49</sup>

**Multiple structures**—When several structures are combined, obtaining better results is more likely than only using the best single structure (Figure 5). Notably, the EF can rise from 21.4 with the best single structure to 28.5 with only two structures combined. The EF reaches the maximal value (33.2) when six structures are combined, whereas it is reduced to 30.9 when 27 structures are combined and begins to decrease as the number of combined structures increases. When all 43 structures are combined, the EF is only 19.0, which is a little worse than that with the best single structure (21.4). We agree with the so-called anti-cooperative behavior<sup>49</sup> as reported previously, which means the result may degrade with the increasing number of protein structures. As indicated, combining all 43 structures (EF=19.0) performs just a little worse than with the best single structure (EF=21.4), but better than with most of the single structures. Actually only 3 out of 43 structures (7.0%) have an EF greater than or equal to 19.0. This will be further discussed later.

#### 5 Selection of Chk1 structures used in virtual screening

The choice of the most appropriate receptor structure is key for successful virtual screening. The ideal structure should have a high success rate in prediction of binding mode and a high EF in recovery of known ligands. Prediction of binding modes is considered the most successful area in docking and scoring, whereas scoring functions remain too primitive to rank a docking hit list correctly.<sup>52</sup> In our experience, to choose a structure that performs best in both, prediction of binding mode and recovery of known ligands, is difficult. In reality, possible lead compounds are generally chosen according to scoring rank. Therefore, we determined the optimal structures for virtual screening mainly by the enrichment in recovery of known ligands.

**Allosteric site**—In prediction of the binding mode, all three structures successfully reproduce the experimental poses of native ligands, while 3JVR correctly predicts docking poses of all three ligands. In recovery of known ligands, 3F9N and 3JVR rank their native ligands at the top 1%, whereas 3JVR manages to rank all three known ligands at the top 1%. That is, 3JVR succeeds in reproducing the crystal poses of all three ligands and distinguishing them from decoy compounds. Thus, it is the ideal structure for virtual screening and will be used in the following virtual screening for Chk1 inhibitors.



**ATP-binding site**—Finding a small subset of Chk1 structures that can perform the best in both prediction of binding mode and recovery of known ligands is possible. In the practical environment, the results from recovery of known ligands are more intuitive than those from the prediction of binding mode. Therefore, we would select out several structures based on results from recovery of known ligands. Three structures, 1NVR, 2R0U and 2WMX, perform the best or near the best in single structure docking and their combinations perform the best in multiple structures docking. The EF for a single structure at the top 1% is respectively: 1NVR, 21.4; 2R0U, 21.4; 2WMX, 16.6. The EF for the structure ensemble at the top 1% is respectively: 1NVR-2R0U, 28.5; 1NVR-2WMX, 28.5; 1NVR-2R0U-2WMX, 30.9. The three structures are used in the following virtual screening for Chk1 inhibitors.

## 6 Virtual screening and experimental tests

More than 60,000 diverse compounds were retrieved from the ZINC<sup>53</sup> database and prepared as described in Methods. They were each docked to the four chosen Chk1 structures sequentially using Dock 6.3. More than one hundred compounds with the lowest docking score were visually inspected focusing on chemically attractive and novel structures. Finally six compounds were selected to be tested experimentally based on the availability and cost of purchasing (Figure 6). Two of the compounds (chk#2 and chk#4) exhibited an inhibitory effect against Chk1 with IC<sub>50</sub> values of 9.6  $\mu$ M and 49.5  $\mu$ M. The two active compounds and 43 crystal ligands were compared using Canvas v1.4.<sup>54</sup> The similarity was calculated based on fingerprints with Tanimoto index as a metric. For chk#2, the similarity with 43 crystal ligands varies from 0.005 to 0.039; for chk#4, the similarity varies from 0.002 to 0.039. As a comparison, the similarity between 43 crystal ligands ranges from 0.002 to 0.89. Therefore, the two active compounds are different from known inhibitors.

## Discussion

### 1 Advantage of multiple crystal structures over single structure

Figure 7 displays the distribution of enrichment factor for single structure and all possible combinations of 2-structure, 3-structure, 40-structure, 41-structure and 42-structure. The EF is 19.0 when all 43 structures are utilized in docking, and this value is used as a threshold to examine how EF varies with an increasing number of structures. For a single structure, 7.0% of 43 structures obtain an EF no less than 19.0; for ensembles with 2 and 3 structures, 15.5% of 903 combinations and 21.9% of 12,341 combinations, respectively, have an EF equal to or greater than 19.0; for ensembles with 40, 41, 42 and 43 structures, the percentage rises to 96.6%, 98.8%, 100% and 100%, respectively. The probability to achieve an EF no less than 19.0 is rising monotonically with the increasing number of structures combined. This is true even if we don't calculate the percentage of ensembles with 4 – 39 structures because of the huge number of possible combinations. Our results clearly indicate that obtaining a good enrichment at the top ranked list with multiple structures is more likely than with a single structure when structures are chosen randomly. On the other hand, the percentage of single structure or structure combinations, which might have an EF of 0, decreases steadily from 4.7% with single structure to 3.2% with 2 structures and 2.5% with 3 structures. Moreover, the lowest EF is 11.9, 14.2 and 19.0, respectively, for ensembles with 40, 41 and 42 structures, suggesting that the risk of getting the worst enrichment is greatly reduced when multiple structures are used. Therefore, the advantage of multiple structures is to be able to not only outperform the best single structure (Figure 5), but also circumvent the worst result and guarantee a good enrichment if it's not the best.

## 2 Conformational energy corrections

One common criticism of the multiple structures approach is the exclusion of conformational energy difference. Taking into account conformational energy would improve enrichment,<sup>55</sup> binding mode prediction<sup>49</sup> and ligand selectivity profile<sup>12</sup> even with very crude approximations when multiple receptor conformations are applied in docking. Here, in order to evaluate if considering the conformational energy difference can improve docking performance with multiple structures, we calculate the conformational energy difference with targeted molecular dynamics,<sup>56</sup> a computationally demanding method but more precise than previous reports,<sup>49, 55</sup> where the conformation energy difference is simply estimated. The maximal difference of docking score between the top 1% compounds for all structures is 20.6 Kcal/mol (Figure 8). The conformational energy difference ranges from 0.19 to 278.95 Kcal/mol for pairwise structures, and only 27.6% of pairwise structures have an energy difference less than 20 Kcal/mol (Figure 8). This means that, for most of the structures, when multiple structures are combined, the enrichment factor would be no better than that from the best single structure, because conformational energy plays a dominant role in ranking. A total of 238 pairs of structures have an energy difference that is less than 20 Kcal/mol. We examined whether the enrichment factor could be improved for 2-structure ensembles when taking into account conformational energy difference. Our results show that considering conformational energy could hardly improve the enrichment factor in recovery of known ligands. In contrast, the enrichment factor deteriorates in most cases. In a few ensembles, with which the enrichment factor is increased, a large conformational energy difference exists between the two protein structures. Therefore, it performs no better than the better single structure.

## 3 Potential biases in crystal structures

Of the 43 ATP competitive inhibitors, only 22 ligands (51%) are ranked at the top 1% of the screening results in 43 independent docking simulations. The other 21 true inhibitors are falsely classified into a negative set no matter what structure is used with 1% as threshold. The most frequently recognized ligands are from 2E9O, 1ZYS, 2E9V and 2GHG, and they are ranked at the top 1% by 31, 28, 27 and 26 structures, respectively (Figure 9), indicating that the crystal structures show an inclination to identify some ligands more easily than others. The 22 ligands are recognized 187 times totally at the top 1%, and these 4 ligands explain 60% of them. To further explore the chemical diversity of Chk1 ligands, we utilized the hierarchical clustering method in Canvas<sup>54</sup> to classify all ligands from crystal complexes (Figure 10). For the 4 most recognizable ligands, the ligand from 2E9O belongs to cluster 10; the ligands from 1ZYS and 2E9V belong to cluster 5; the ligand from 2GHG belongs to cluster 4. After checking the results from recovery of known ligands, most known ligands ranked at the top 1% belong to cluster 5 no matter which structure is employed. Twelve of 22 recognizable ligands belong to cluster 5, the biggest cluster (19 ligands), and they occur 106 times at the top 1%. However, in cluster 6, which contains 12 ligands and is the second biggest cluster, only three ligands are ranked at the top 1% 8 times. For the 3 structures employed in virtual screening, the ligand from 1NVR belongs to cluster 11; the ligand from 2R0U belongs to cluster 5; the ligand from 2WMX belongs to cluster 6. With 1NVR, the three ligands in cluster 11 are ranked at the top 1% and, with 2WMX, 2 ligands in cluster 6 are ranked at the top 1%, except those easily recognized ligands in cluster 5. The results suggest that it is more likely in virtual screening to identify diverse compounds with multiple structures than with a single structure.

## 4 Characteristics in the optimal ensemble of Chk1 structures

Previous results also raise another common question as to whether a simple and reliable strategy exists to choose the optimal ensemble of receptor structures according to available properties of receptors and ligands. Several efforts have been made to identify the most



enriching structure without running a small-scale virtual screening. Selecting significant structures based on docking results from prediction of binding mode seems unsuccessful because of absence of correlation between good docked poses and good scores or ranks.<sup>48-49</sup> Interestingly, the volume of the binding site that the co-crystallized ligand occupies is considered a possible reason to explain the performance of different crystal structures.<sup>50</sup> Five strategies to construct ensembles have been compared in a recent study.<sup>11</sup>

Here we examine the protein structures in the optimal ensemble and corresponding co-crystallized ligands to look for properties that can be utilized as selection criterion. The backbone C- $\alpha$  RMSD for pairwise structures in the optimal ensemble is respectively: 1NVR-2R0U, 7.2 Å; 1NVR-2WMX, 15.7 Å; and 2R0U-2WMX, 14.7 Å. Although 2R0U and 2WMX have a different backbone conformation, the C- $\alpha$  RMSD of the binding site is less than 0.3 Å suggesting that they have a similar binding site. The C- $\alpha$  RMSD of the binding site for other pairwise structures is between 2.2 and 3.1 Å. As mentioned earlier, the ligand from 1NVR, 2R0U and 2WMX belongs to cluster 11, 5 and 6, respectively. In summary, the optimal ensemble of Chk1 structures can be described by three criteria: 1) they have different backbone conformation; 2) their cognate ligands belong to different chemical spaces (clusters); and 3) the ligands come from the three biggest clusters. However, determining the optimal ensemble based on protein backbone comparison and ligand clustering is not sufficient because a great number of ensembles meet the three requirements. According to our results, with active ligands available, the approach to evaluate protein structures by the recovery of known ligands is more appropriate for selection of the optimal ensemble applied in virtual screening than a simple strategy based on properties of receptors and ligands, though this approach needs more computer time.

## 5 Computation time

One disadvantage of multiple structures docking is that the required computation time increases linearly with the number of structures, which makes it a computationally demanding process. However, now with the emergence of powerful and parallel supercomputers and clusters, multiple structures docking can be performed daily in drug discovery. In this work, all docking simulations were performed on an IBM Blue Gene/L supercomputer. In predicting the binding mode, less than 17 minutes with 128 nodes was required to dock one Chk1 structure to the 46 known ligands, and about 3 hours with 512 nodes were required to dock 46 Chk1 structures to 46 known ligands. In the recovery of known ligands, less than 15 minutes with 512 nodes was needed to dock one Chk1 structure to the testing set of over 2,000 compounds. When 46 Chk1 structures were docked to the testing set, less than 6 hours were needed because 1,024 nodes were used. In virtual screening of compounds extracted from ZINC database, about 20 hours were needed with 512 nodes to dock over 60,000 compounds into one Chk1 structure.

## Conclusion

In this work, we examined the performance of a number of Chk1 crystal structures in the prediction of binding mode and recovery of known ligands. An optimal ensemble of Chk1 structures were selected mainly based on the results from recovery of known ligands and used for the successful discovery of novel Chk1 inhibitors. Furthermore, we compared the performance of multiple structures versus a single structure in the recovery of known ligands when structures were selected randomly. We found that obtaining a good enrichment is more likely with multiple structures than with a single structure. Identifying more diverse ligands is also more likely with multiple structures than with a single structure according to our ligand clustering results. We found that the chosen protein structures displayed diverse conformations in backbone and binding site, and the co-crystallized ligands belonged to different clusters. However, to select a small subset of protein structures according to the

available information of receptors and ligands was still challenging. Evaluating protein structures by the recovery of known inhibitors is a reliable approach to determine the optimal ensemble of structures. Although this process was computationally expensive, increasing the power of supercomputers and clusters would make it less time-consuming. We demonstrated that the evaluation process could be finished in 10 hours using the IBM Blue Gene/L supercomputer. Finally, the conformational energy difference was calculated by targeted molecular dynamics and our results indicated that this could not improve the enrichment.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This work was supported by The Hormel Foundation, National Institutes of Health Grant R37 CA081064 and NCI Contract Number HHSN-261200533001C - NO1-CN-53301.

## References

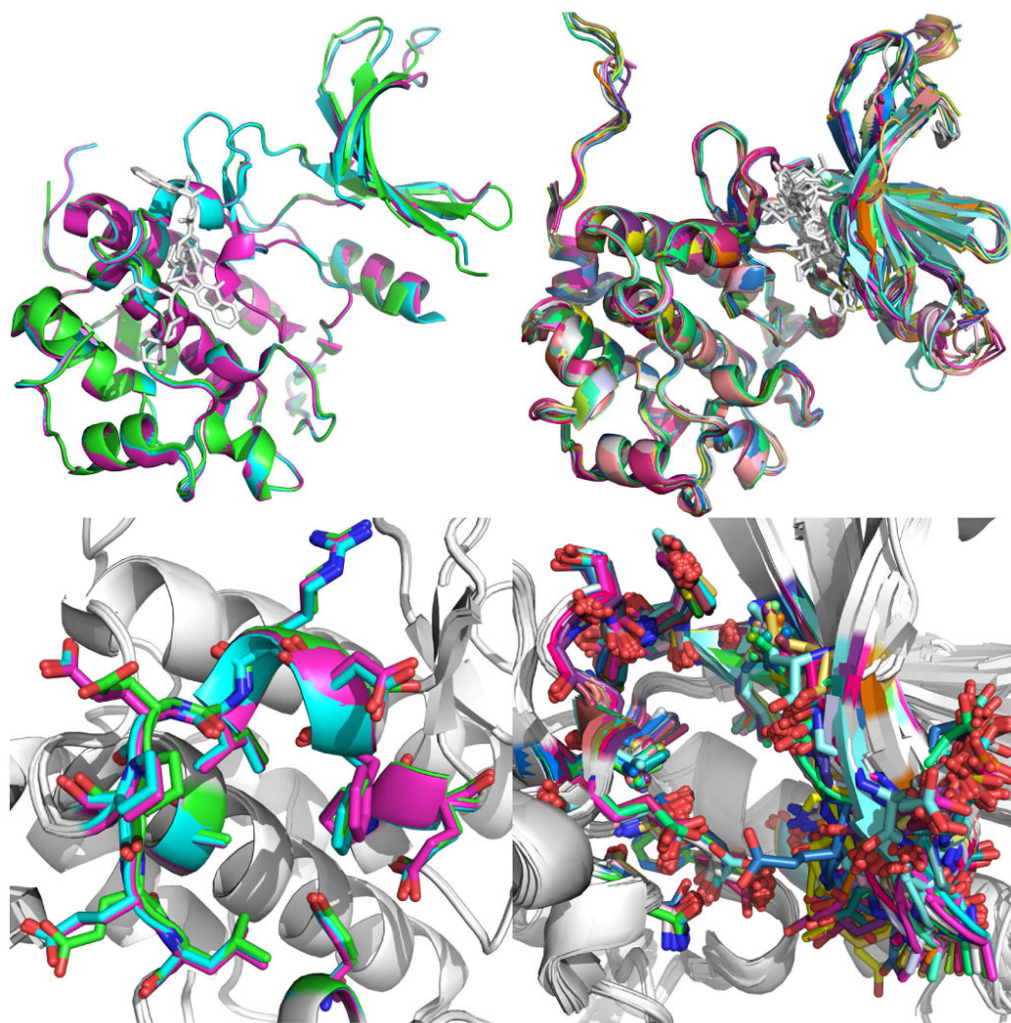
1. Jorgensen WL. The many roles of computation in drug discovery. *Science*. 2004; 303:1813–1818. [PubMed: 15031495]
2. Shoichet BK. Virtual screening of chemical libraries. *Nature*. 2004; 432:862–865. [PubMed: 15602552]
3. Mobley DL, Dill KA. Binding of Small-Molecule Ligands to Proteins: “What You See” Is Not Always “What You Get”. *Structure*. 2009; 17:489–498. [PubMed: 19368882]
4. Totrov M, Abagyan R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr Opin Struct Biol*. 2008; 18:178–184. [PubMed: 18302984]
5. Grant BJ, Gorfe AA, McCammon JA. Large conformational changes in proteins: signaling and other functions. *Curr Opin Struct Biol*. 2010; 20:142–147. [PubMed: 20060708]
6. Teodoro ML, Kavraki LE. Conformational flexibility models for the receptor in structure based drug design. *Curr Pharm Design*. 2003; 9:1635–1648.
7. Ferrari AM, Wei BQQ, Costantino L, Shoichet BK. Soft docking and multiple receptor conformations in virtual screening. *J Med Chem*. 2004; 47:5076–5084. [PubMed: 15456251]
8. Yoon S, Welsh WJ. Identification of a minimal subset of receptor conformations for improved multiple conformation docking and two-step scoring. *J Chem Inf Comput Sci*. 2004; 44:88–96. [PubMed: 14741014]
9. Rao S, Sanschagrin PC, Greenwood JR, Repasky MP, Sherman W, Farid R. Improving database enrichment through ensemble docking. *J Comput-Aided Mol Des*. 2008; 22:621–627. [PubMed: 18253700]
10. Rueda M, Bottegoni G, Abagyan R. Recipes for the Selection of Experimental Protein Conformations for Virtual Screening. *J Chem Inf Model*. 2010; 50:186–193. [PubMed: 20000587]
11. Craig IR, Essex JW, Spiegel K. Ensemble Docking into Multiple Crystallographically Derived Protein Structures: An Evaluation Based on the Statistical Analysis of Enrichments. *J Chem Inf Model*. 2010; 50:511–524. [PubMed: 20222690]
12. Park SJ, Kufareva I, Abagyan R. Improved docking, screening and selectivity prediction for small molecule nuclear receptor modulators using conformational ensembles. *J Comput-Aided Mol Des*. 2010; 24:459–471. [PubMed: 20455005]
13. Bolstad ESD, Anderson AC. In pursuit of virtual lead optimization: The role of the receptor structure and ensembles in accurate docking. *Proteins*. 2008; 73:566–580. [PubMed: 18473360]
14. Cheng LS, Amaro RE, Xu D, Li WW, Arzberger PW, McCammon JA. Ensemble-based virtual screening reveals potential novel antiviral compounds for avian influenza neuraminidase. *J Med Chem*. 2008; 51:3878–3894. [PubMed: 18558668]

15. Cavasotto CN, Kovacs JA, Abagyan RA. Representing receptor flexibility in ligand docking through relevant normal modes. *J Am Chem Soc.* 2005; 127:9632–9640. [PubMed: 15984891]
16. Rueda M, Bottegoni G, Abagyan R. Consistent Improvement of Cross-Docking Results Using Binding Site Ensembles Generated with Elastic Network Normal Modes. *J Chem Inf Model.* 2009; 49:716–725. [PubMed: 19434904]
17. Sperandio O, Mouawad L, Pinto E, Villoutreix BO, Perahia D, Miteva MA. How to choose relevant multiple receptor conformations for virtual screening: a test case of Cdk2 and normal mode analysis. *Eur Biophys J Biophys Lett.* 2010; 39:1365–1372.
18. Amaro RE, Li WW. Emerging Methods for Ensemble-Based Virtual Screening. *Curr Top Med Chem.* 2010; 10:3–13. [PubMed: 19929833]
19. Bartek J, Lukas J. Chk1 and Chk2 kinases in checkpoint control and cancer. *Cancer Cell.* 2003; 3:421–429. [PubMed: 12781359]
20. Bucher N, Britten CD. G2 checkpoint abrogation and checkpoint kinase-1 targeting in the treatment of cancer. *Brit J Cancer.* 2008; 98:523–528. [PubMed: 18231106]
21. Zhou BBS, Bartek J. Targeting the checkpoint kinases: Chemosensitization versus chemoprotection. *Nat Rev Cancer.* 2004; 4:216–225. [PubMed: 14993903]
22. Tse AN, Carvajal R, Schwartz GK. Targeting checkpoint kinase 1 in cancer therapeutics. *Clin Cancer Res.* 2007; 13:1955–1960. [PubMed: 17404075]
23. Tao ZF, Lin NH. Chk1 inhibitors for novel cancer treatment. *Anticancer Agents Med Chem.* 2006; 6:377–388. [PubMed: 16842237]
24. Janetka JW, Ashwell S. Checkpoint kinase inhibitors: a review of the patent literature. *Expert Opin Ther Pat.* 2009; 19:165–197. [PubMed: 19441917]
25. Gullotta F, De Marinis E, Ascenzi P, di Masi A. Targeting the DNA Double Strand Breaks Repair for Cancer Therapy. *Curr Med Chem.* 2010; 17:2017–2048. [PubMed: 20423312]
26. Chen P, Luo C, Deng YL, Ryan K, Register J, Margosiak S, Tempczyk-Russell A, Nguyen B, Myers P, Lundgren K, Kan CC, O'Connor PM. The 1.7 angstrom crystal structure of human cell cycle checkpoint kinase Chk1: Implications for Chk1 regulation. *Cell.* 2000; 100:681–692. [PubMed: 10761933]
27. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF chimera - A visualization system for exploratory research and analysis. *J Comput Chem.* 2004; 25:1605–1612. [PubMed: 15264254]
28. Watts KS, Dalal P, Murphy RB, Sherman W, Friesner RA, Shelley JC. ConfGen: A Conformational Search Method for Efficient Generation of Bioactive Conformers. *J Chem Inf Model.* 2010; 50:534–546. [PubMed: 20373803]
29. Moustakas DT, Lang PT, Pegg S, Pettersen E, Kuntz ID, Brooijmans N, Rizzo RC. Development and validation of a modular, extensible docking program: DOCK 5. *J Comput-Aided Mol Des.* 2006; 20:601–619. [PubMed: 17149653]
30. Zhao B, Bower MJ, McDevitt PJ, Zhao HZ, Davis ST, Johanson KO, Green SM, Concha NO, Zhou BBS. Structural basis for Chk1 inhibition by UCN-01. *J Biol Chem.* 2002; 277:46609–46615. [PubMed: 12244092]
31. Foloppe N, Fisher LM, Howes R, Kierstan P, Potter A, Robertson AGS, Surgenor AE. Structure-based design of novel Chk1 inhibitors: Insights into hydrogen bonding and protein-ligand affinity. *J Med Chem.* 2005; 48:4332–4345. [PubMed: 15974586]
32. Foloppe N, Fisher LM, Francis G, Howes R, Kierstan P, Potter A. Identification of a buried pocket for potent and selective inhibition of Chk1: Prediction and verification. *Bioorgan Med Chem.* 2006; 14:1792–1804.
33. Foloppe N, Fisher LM, Howes R, Potter A, Robertson AGS, Surgenor AE. Identification of chemically diverse Chk1 inhibitors by receptor-based virtual screening. *Bioorgan Med Chem.* 2006; 14:4792–4802.
34. Fraley ME, Steen JT, Brnardic EJ, Arrington KL, Spencer KL, Hanney BA, Kim Y, Hartman GD, Stirdivant SM, Drakas BA, Rickert K, Walsh ES, Hamilton K, Buser CA, Hardwick J, Tao WK, Beck SC, Mao XZ, Lobell RB, Sepp-Lorenzino L, Yan YW, Ikuta M, Munshi SK, Kuo LC, Kreatsoulas C. 3-(Indol-2-yl)indazoles as Chk1 kinase inhibitors: Optimization of potency and

- selectivity via substitution at C6. *Bioorg Med Chem Lett.* 2006; 16:6049–6053. [PubMed: 16978863]
35. Huang SE, Garbaccio RM, Fraley ME, Steen J, Kreatsoulas C, Hartman G, Stirdivant S, Drakas B, Rickert K, Walsh E, Hamilton K, Buser CA, Hardwick J, Mao XZ, Abrams M, Beck S, Tao WK, Lobell R, Sepp-Lorenzino L, Yan YW, Ikuta M, Murphy JZ, Sardana V, Munshi S, Kuo L, Reilly M, Mahan E. Development of 6-substituted indolylquinolinones as potent Chk1 kinase inhibitors. *Bioorg Med Chem Lett.* 2006; 16:5907–5912. [PubMed: 16990002]
  36. Li GQ, Hasvold LA, Tao ZF, Wang GT, Gwaltney SL, Patel J, Kovar P, Credo RB, Chen ZH, Zhang HY, Park C, Sham HL, Sowin T, Rosenberg SH, Lin NH. Synthesis and biological evaluation of 1-(2,4,5-trisubstituted phenyl)-3-(5-cyanopyrazin-2-yl)ureas as potent Chk1 kinase inhibitors. *Bioorg Med Chem Lett.* 2006; 16:2293–2298. [PubMed: 16446090]
  37. Lin NH, Xia P, Kovar P, Park C, Chen ZH, Zhang HY, Rosenberg SH, Sham HL. Synthesis and biological evaluation of 3-ethylidene-1,3-dihydro-indol-2-ones as novel checkpoint 1 inhibitors. *Bioorg Med Chem Lett.* 2006; 16:421–426. [PubMed: 16242328]
  38. Ni ZJ, Barsanti P, Brammeier N, Diebes A, Poon DJ, Ng S, Pecchi S, Pfister K, Renhowe PA, Ramurthy S, Wagman AS, Bussiere DE, Le V, Zhou Y, Jansen JM, Ma S, Gesner TG. 4-(aminoalkylamino)-3-benzimidazole-quinolinones as potent CHK-1 inhibitors. *Bioorg Med Chem Lett.* 2006; 16:3121–3124. [PubMed: 16603354]
  39. Zhu GD, Gandhi VB, Gong JC, Luo Y, Liu XS, Shi Y, Guan R, Magnone SR, Klinghofer V, Johnson EF, Bouska J, Shoemaker A, Oleksijew A, Jarvis K, Park C, De Jong R, Oltersdorf T, Li Q, Rosenberg SH, Giranda VL. Discovery and SAR of oxindole-pyridine-based protein kinase B/Akt inhibitors for treating cancers. *Bioorg Med Chem Lett.* 2006; 16:3424–3429. [PubMed: 16644221]
  40. Brnardic EJ, Garbaccio RM, Fraley ME, Tasber ES, Steen JT, Arrington KL, Dudkin VY, Hartman GD, Stirdivant SM, Drakas BA, Rickert K, Walsh ES, Hamilton K, Buser CA, Hardwick J, Tao WK, Beck SC, Mao XZ, Lobell RB, Sepp-Lorenzino L, Yan YW, Ikuta M, Munshi SK, Kuo LC, Kreatsoulas C. Optimization of a pyrazoloquinolinone class of Chk1 kinase inhibitors. *Bioorg Med Chem Lett.* 2007; 17:5989–5994. [PubMed: 17804227]
  41. Garbaccio RM, Huang S, Tasber ES, Fraley ME, Yan YW, Munshi S, Ikuta M, Kuo L, Kreatsoulas C, Stirdivant S, Drakas B, Rickert K, Walsh ES, Hamilton KA, Buser CA, Hardwick J, Mao XZ, Beck SC, Abrams MT, Tao WK, Lobell R, Sepp-Lorenzino L, Hartman GD. Synthesis and evaluation of substituted benzoisoquinolinones as potent inhibitors of Chk1 kinase. *Bioorg Med Chem Lett.* 2007; 17:6280–6285. [PubMed: 17900896]
  42. Tong YS, Claiborne A, Stewart KD, Park C, Kovar P, Chen ZH, Credo RB, Gu WZ, Gwaltney SL, Judge RA, Zhang HY, Rosenberg SH, Sham HL, Sowin TJ, Lin NH. Discovery of 1,4-dihydroindeno[1,2-c]pyrazoles as a novel class of potent and selective checkpoint kinase 1 inhibitors. *Bioorgan Med Chem.* 2007; 15:2759–2767.
  43. Matthews TP, Klair S, Burns S, Boxall K, Cherry M, Fisher M, Westwood IM, Walton MI, McHardy T, Cheung KMJ, Van Montfort R, Williams D, Aherne GW, Garrett MD, Reader J, Collins I. Identification of Inhibitors of Checkpoint Kinase 1 through Template Screening. *J Med Chem.* 2009; 52:4810–4819. [PubMed: 19572549]
  44. Converso A, Hartingh T, Garbaccio RM, Tasber E, Rickert K, Fraley ME, Yan YW, Kreatsoulas C, Stirdivant S, Drakas B, Walsh ES, Hamilton K, Buser CA, Mao XZ, Abrams MT, Beck SC, Tao WK, Lobell R, Sepp-Lorenzino L, Zugay-Murphy J, Sardana V, Munshi SK, Jezequel-Sur SM, Zuck PD, Hartman GD. Development of thioquinazolinones, allosteric Chk1 kinase inhibitors. *Bioorg Med Chem Lett.* 2009; 19:1240–1244. [PubMed: 19155174]
  45. Vanderpool D, Johnson TO, Ping C, Bergqvist S, Alton G, Phonephaly S, Rui E, Luo C, Deng YL, Grant S, Quenzer T, Margosiak S, Register J, Brown E, Ermolieff J. Characterization of the CHK1 Allosteric Inhibitor Binding Site. *Biochemistry.* 2009; 48:9823–9830. [PubMed: 19743875]
  46. Morphy R. Selectively Nonselective Kinase Inhibition: Striking the Right Balance. *J Med Chem.* 2010; 53:1413–1437. [PubMed: 20166671]
  47. Garuti L, Roberti M, Bottegoni G. Non-ATP Competitive Protein Kinase Inhibitors. *Curr Med Chem.* 2010; 17:2804–2821. [PubMed: 20586715]
  48. Cavasotto CN, Abagyan RA. Protein flexibility in ligand docking and virtual screening to protein kinases. *J Mol Biol.* 2004; 337:209–225. [PubMed: 15001363]

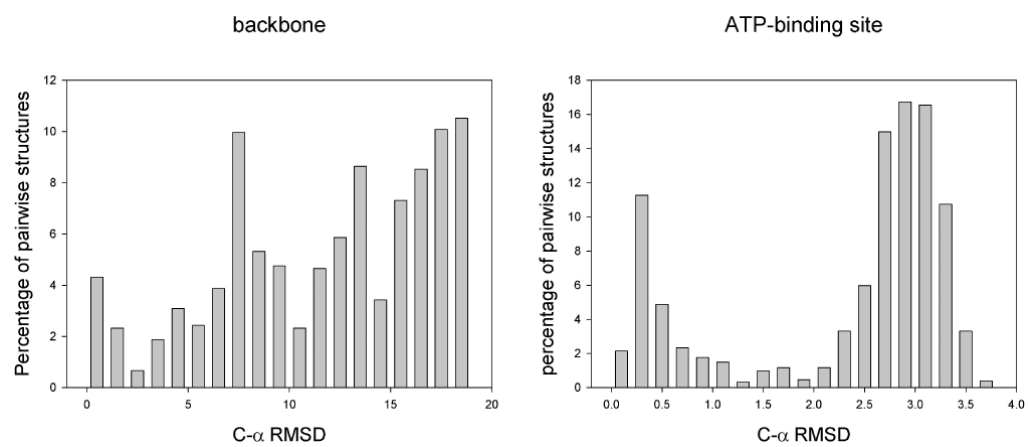
49. Barril X, Morley SD. Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. *J Med Chem.* 2005; 48:4432–4443. [PubMed: 15974595]
50. Thomas MP, McInnes C, Fischer PM. Protein structures in virtual screening: A case study with CDK2. *J Med Chem.* 2006; 49:92–104. [PubMed: 16392795]
51. Verdonk ML, Mortenson PN, Hall RJ, Hartshorn MJ, Murray CW. Protein-Ligand Docking against Non-Native Protein Conformers. *J Chem Inf Model.* 2008; 48:2214–2225. [PubMed: 18954138]
52. Leach AR, Shoichet BK, Peishoff CE. Prediction of Protein-Ligand Interactions. Docking and Scoring: Successes and Gaps. *J Med Chem.* 2006; 49:5851–5855. [PubMed: 17004700]
53. Irwin JJ, Shoichet BK. ZINC - A free database of commercially available compounds for virtual screening. *J Chem Inf Model.* 2005; 45:177–182. [PubMed: 15667143]
54. Duan JX, Dixon SL, Lowrie JF, Sherman W. Analysis and comparison of 2D fingerprints: Insights into database screening performance using eight fingerprint methods. *J Mol Graph Model.* 2010; 29:157–170. [PubMed: 20579912]
55. Wei BQ, Weaver LH, Ferrari AM, Matthews BW, Shoichet BK. Testing a flexible-receptor docking algorithm in a model binding site. *J Mol Biol.* 2004; 337:1161–1182. [PubMed: 15046985]
56. Schlitter J, Engels M, Kruger P. Targeted Molecular-Dynamics - a New Approach for Searching Pathways of Conformational Transitions. *J Mol Graphics.* 1994; 12:84–89.



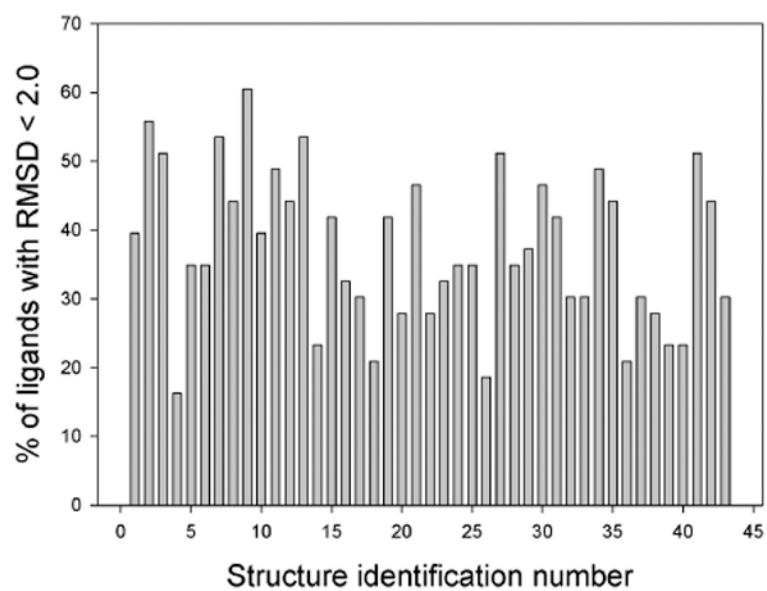


**Figure 1.**

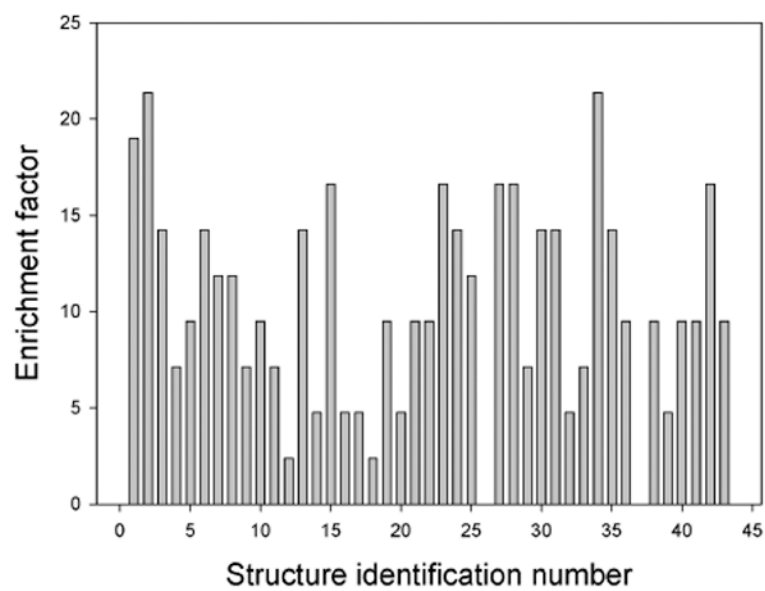
Upper left: Superposition of three Chk1 structures with allosteric inhibitors. 3JVR is used as the reference structure. 3F9N is colored green; 3JVR is colored cyan; and 3JVS is colored magenta; ligands indicating the allosteric site are colored white. Upper right: Superposition of 43 Chk1 structures with ATP-competitive inhibitors. 1NVQ is used as the reference structure, and ligands indicating the ATP-binding site are colored white. Lower left: Superposition of the allosteric sites of three Chk1 structures. Residues forming the allosteric sites are shown. Carbons of 3F9N are colored green; carbons of 3JVR are colored cyan; carbons of 3JVS are colored magenta. Lower right: Superposition of the ATP-binding sites of 43 Chk1 structures. Residues forming the ATP-binding sites are shown.



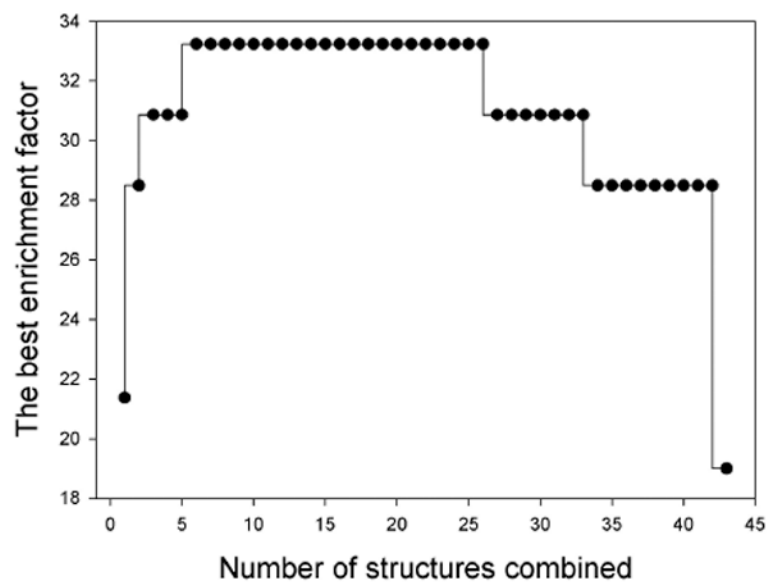
**Figure 2.** Left: Distribution of C- $\alpha$  RMSD of backbone for 43 Chk1 structures with ATP-competitive inhibitors. Right: Distribution of C- $\alpha$  RMSD of the ATP-binding site for 43 Chk1 structures.



**Figure 3.** Docking accuracy of single structure in prediction of binding mode for 43 Chk1 structures with ATP-competitive inhibitors.

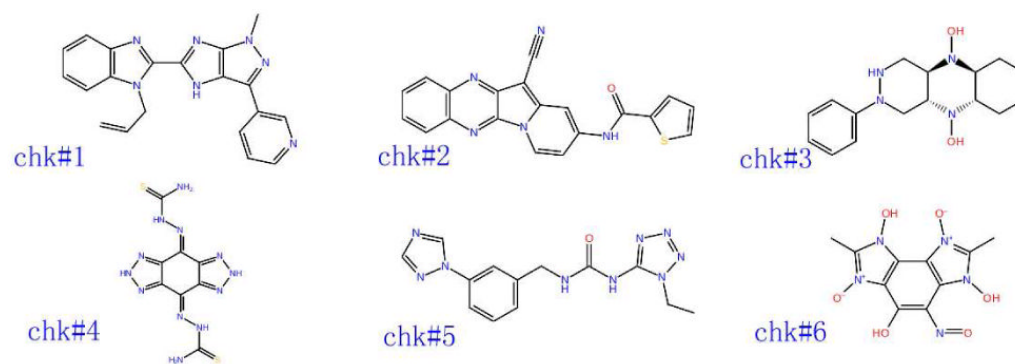


**Figure 4.** Enrichment factor of single structure in recovery of known ligands for 43 Chk1 structures with ATP-competitive inhibitors.

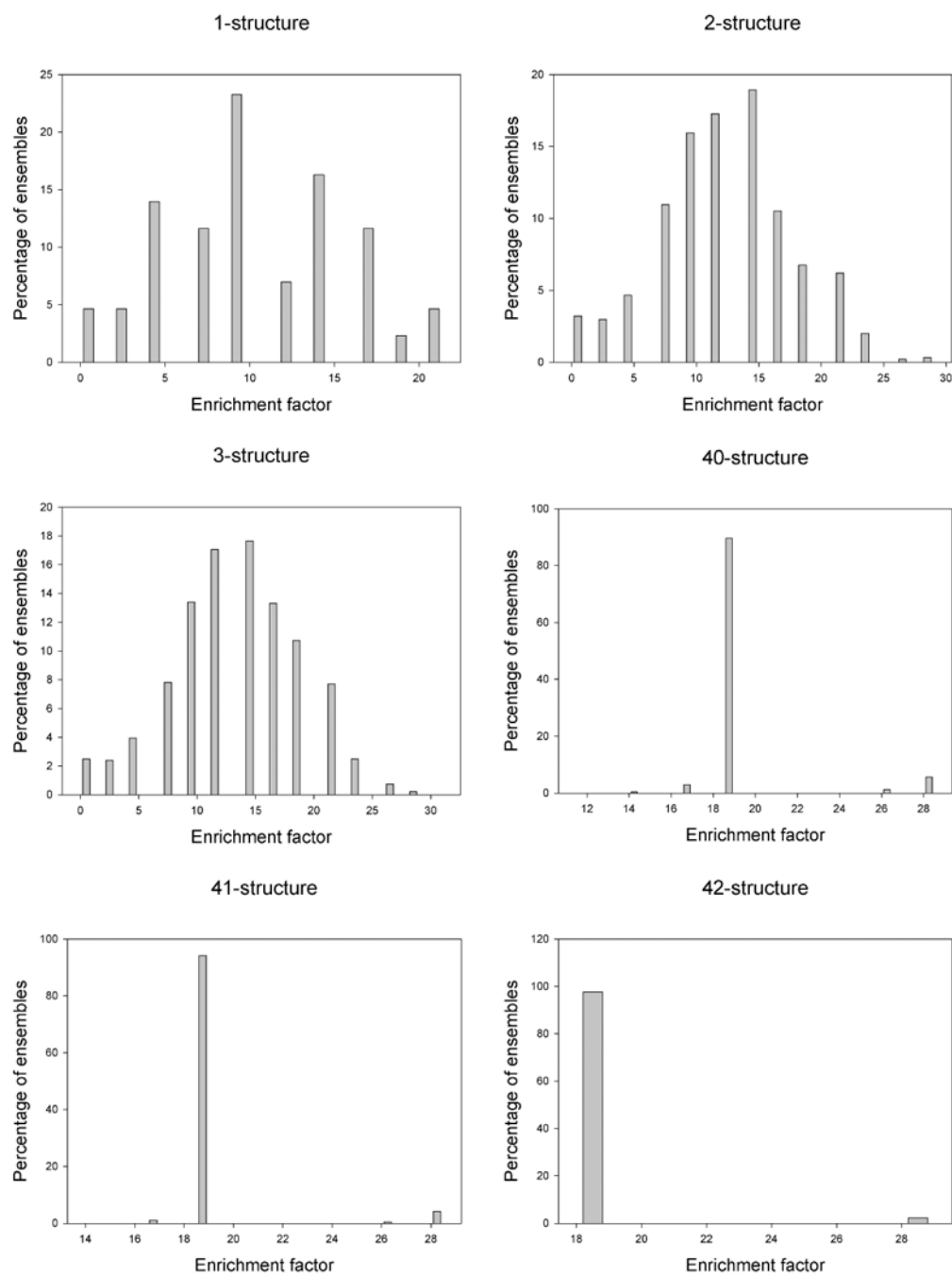


**Figure 5.**  
The best enrichment factor achieved with an increasing number of structures.

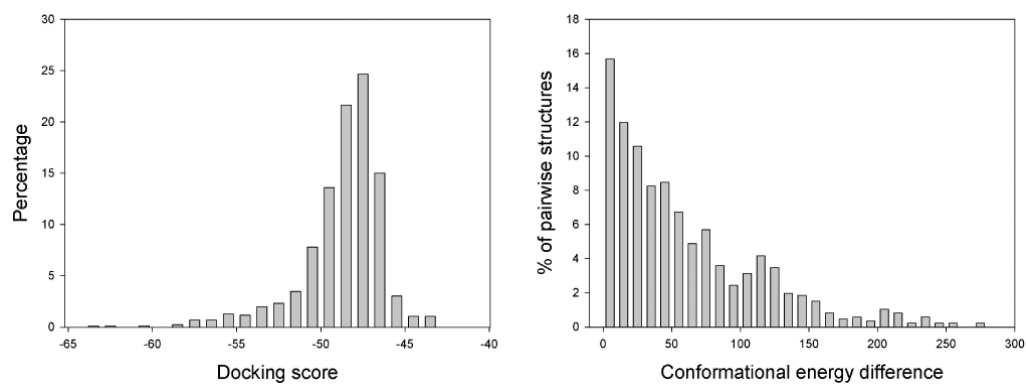




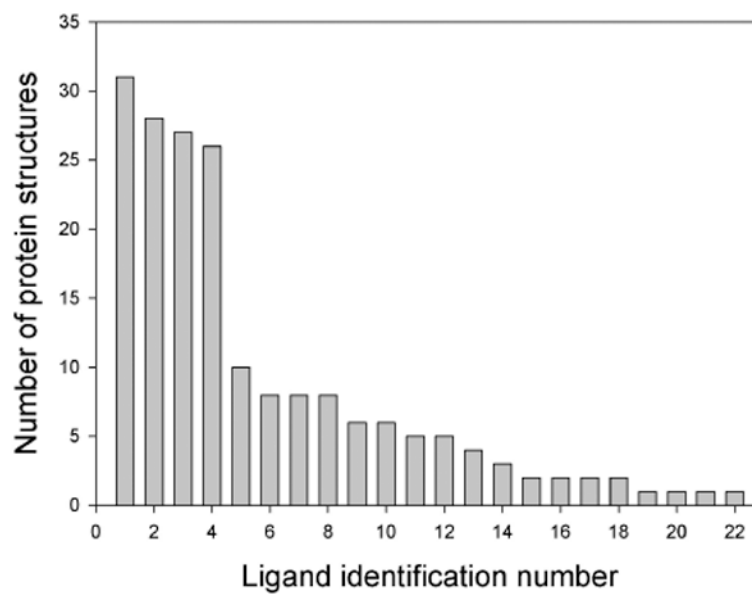
**Figure 6.**  
2D structures of six compounds tested in experiments. Non-polar hydrogens are hidden.



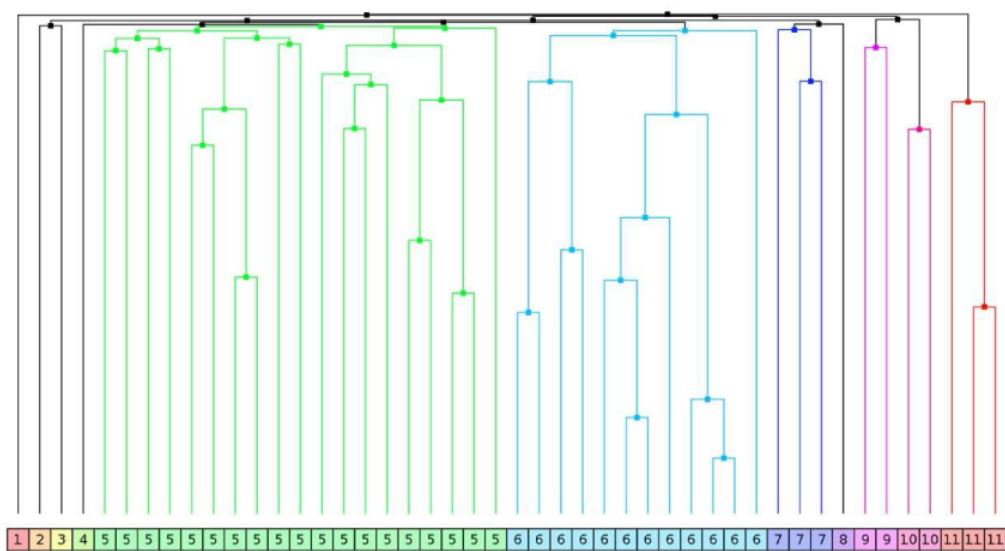
**Figure 7.** Distribution of enrichment factors when 1, 2, 3, 40, 41, and 42 structures are combined in recovery of known ligands.



**Figure 8.** Left: Distribution of docking score of the top 1% ranked list for 43 Chk1 structures. Right: Distribution of conformational energy difference of pairwise structures obtained from targeted molecular dynamics.



**Figure 9.**  
The number of Chk1 structures by which a ligand can be ranked at the top 1%.



**Figure 10.**  
Hierarchical clustering of 46 Chk1 ligands.



**Table 1**

RMSDs of three allosteric inhibitors in prediction of binding mode. RMSDs less than 2.0 Å are colored light red; RMSDs between 2.0 and 3.0 Å are colored olive green.

|               |     | <i>Protein structure (PDB entry)</i> |             |             |
|---------------|-----|--------------------------------------|-------------|-------------|
|               |     | <b>3F9N</b>                          | <b>3JVR</b> | <b>3JVS</b> |
|               | 38M | 0.77                                 | 1.12        | 8.89        |
| <i>Ligand</i> | AGX | 8.09                                 | 0.25        | 0.38        |
|               | AGY | 2.80                                 | 1.68        | 1.67        |