

Consensus Induced Fit Docking (cIFD): methodology, validation, and application to the discovery of novel Crm1 inhibitors

Ori Kalid · Dora Toledo Warshaviak ·
Sharon Shechter · Woody Sherman ·
Sharon Shacham

Received: 21 June 2012 / Accepted: 20 September 2012 / Published online: 30 September 2012
© Springer Science+Business Media Dordrecht 2012

Abstract We present the Consensus Induced Fit Docking (cIFD) approach for adapting a protein binding site to accommodate multiple diverse ligands for virtual screening. This novel approach results in a single binding site structure that can bind diverse chemotypes and is thus highly useful for efficient structure-based virtual screening. We first describe the cIFD method and its validation on three targets that were previously shown to be challenging for docking programs (COX-2, estrogen receptor, and HIV reverse transcriptase). We then demonstrate the application of cIFD to the challenging discovery of irreversible Crm1 inhibitors. We report the identification of 33 novel Crm1 inhibitors, which resulted from the testing of 402 purchased compounds selected from a screening set containing 261,680 compounds. This corresponds to a hit rate of 8.2 %. The novel Crm1 inhibitors reveal diverse chemical structures, validating the utility of the cIFD method in a real-world drug discovery project. This approach offers a pragmatic way to implicitly account for protein flexibility without the additional computational costs of ensemble

docking or including full protein flexibility during virtual screening.

Keywords Docking · Virtual screening · Structure-based drug design · Induced Fit Docking · Receptor flexibility · Crm1 · Exportin-1 · Nuclear transport

Introduction

It is well known that proteins are inherently flexible and different ligands can bind to distinctive protein conformations [1, 2]. While the degree of flexibility varies greatly among proteins, it is generally agreed that accounting for protein flexibility is important in structure-based drug design [3, 4]. Most docking algorithms rely on a rigid protein structure, which in many cases is sufficient to find some active compounds [5–9]. However, a binding site refined around one ligand, either experimentally or computationally, may not be adequate to retrieve a wide range of diverse actives in virtual screening. The binding site shape and amino acid orientations may not be suitable for accommodating significantly different chemotypes. Even highly related compounds may fail to dock well, especially in spatially constrained binding sites where small differences in the ligands may result in clashes with the protein that cannot be alleviated within the rigid receptor framework. Simply softening the potential energy function to allow for steric overlap is an inadequate strategy, as sometimes even small clashes are important for binding affinity and selectivity discrimination [10].

There are multiple ways to account for protein flexibility in docking. The most straightforward approach conceptually is to include explicit protein sampling during docking. This approach has been successfully applied in a number of cases to predict poses for a small number of ligands [11, 12].

Ori Kalid, Dora Toledo Warshaviak: equal contributors.

Electronic supplementary material The online version of this article (doi:10.1007/s10822-012-9611-9) contains supplementary material, which is available to authorized users.

O. Kalid (✉) · S. Shechter · S. Shacham
Karyopharm Therapeutics, 2 Mercer Road, Natick,
MA 01760, USA
e-mail: ori@karyopharm.com; orikalid@gmail.com

O. Kalid
Ori Kalid Consulting, Hahotnim 40, 37013 Pardes-Hanna, Israel

D. Toledo Warshaviak · W. Sherman
Schrodinger Inc., 120 West 45th Street, New York,
NY 10036, USA

Unfortunately, full protein flexibility makes sampling of the protein–ligand complex computationally impractical for virtual screening. Another approach is to use an ensemble of protein structures, derived either through experiment or simulation, and dock to each one individually. This approach has been successfully applied to virtual screening and has been shown to produce improved retrieval of actives compared with screening a single rigid receptor structure [13–15]. A variety of approaches have been taken to generate receptor ensembles. In some cases crystal structures have been used [15, 16], although the choice of structures is not always straightforward [17]. If multiple crystal structures are not available, it is possible to generate ensembles using sampling approaches such as molecular dynamics, Monte Carlo, [18] or low mode analysis [19, 20]. However, as for with crystal structures, the choice of which receptors structures from the simulation to use for the ensemble is not straightforward [21]. Recent progress has been made on the selection of receptor structures for virtual screening ensembles using a method based on binding site shape clustering, which was demonstrated to work on crystal structures and snapshots from molecular dynamics simulations [22, 23].

However, while ensemble docking is much more computationally tractable than the explicit protein sampling approach, it still requires docking to each receptor in the ensemble; thus, an ensemble of five receptors would take five times longer than a single rigid receptor screen. Hybrid approaches have also been proposed, where flexibility of the receptor is partially accounted for by using a restricted conformational space, such as a selected set of side chains or normal modes of the receptor [20, 24, 25]. These hybrid approaches are promising but still require significantly more computational resources than rigid receptor docking and often require user knowledge about the protein degrees of freedom to consider.

One possible solution to implicitly account for protein flexibility while maintaining the computational efficiency of screening a single structure has recently been proposed [26]. This approach utilizes a protocol in which the protein binding site is preprocessed before virtual screening by optimizing it in the presence of several bound active compounds simultaneously, thus generating a single binding site conformation that can accommodate different ligand classes. In this protocol, based on the Locally Enhanced Sampling (LES) concept [27], protein–ligand interactions are scaled asymmetrically, such that all inter-ligand interactions are annulled to allow spatial overlap while each ligand “feels” the full force exerted by the protein and the protein “feels” the average force exerted by all ligands. This process has been successfully applied in structure-based screening of a variety of GPCRs including mGluR5 and class-A peptide receptors [26].

In this work we propose a new method, called Consensus Induced Fit Docking (cIFD), which combines Induced Fit Docking (IFD) [28] of multiple ligands for preliminary binding mode determination followed by receptor optimization in the presence of a “hybrid” ligand that combines selected poses of the IFD-docked ligands. We first describe the cIFD methodology and perform a retrospective analysis on three targets [Cyclooxygenase-2 (COX-2), estrogen receptor (ER), and human immunodeficiency virus reverse transcriptase (HIV-rt)] demonstrating the potential benefits of using cIFD. We then describe a successful prospective application of cIFD in an active drug discovery project to find covalent protein–protein interaction (PPI) inhibitors blocking chromosome region maintenance 1 protein/exportin 1/Xpo1 (Crm1) binding to its cargo proteins.

Crm1 is a key nuclear exporter protein responsible for shuttling a large number of proteins, including tumor suppressors such as p53, pRB, FOXO and APC/ β -catenin, growth regulatory proteins such as p21CIP1, p27Kip1 and NF- κ B/I- κ B and chemotherapeutic targets such as DNA topoisomerases I and IIA and Bcr-ABL. Crm1 cargo proteins carry leucine-rich nuclear export signals (NESs), through which they associate with a shallow binding groove on the surface of Crm1. These are 10–15 residue long amino-acid stretches containing regularly spaced hydrophobic anchors that form combined α -helical extended or entirely extended tertiary structures [29]. Crm1 is a validated molecular target for treatment of cancer [30, 31] and is attractive due to its effect on multiple growth suppressive signaling pathways. A number of Crm1 inhibitors have been reported in the literature including the structurally related natural toxins Leptomycin B (LMB), Anguinomycin, Ratjadones (RATs), Goniotalamin and synthetic analogs [31–34], and synthetic chalcones [35], maleimides [35], halomethyl(ethyl)ketones [35], *N*-azolyacrylates [36], Karyopharm compounds [37], and most recently the pyrrole-2,5-dione CBS9106 [38], all of which bind covalently to Cys528, which is located in the NES-binding groove of human Crm1.

Results

Consensus Induced Fit Docking was developed to improve the enrichment and diversity of active compounds in structure-based virtual screening while minimizing additional computational costs. In internal studies at Karyopharm (data not shown), we frequently observed that rigid receptor virtual screening calculations were failing to retrieve known active compounds due to small rearrangements needed in the protein. To overcome this, we developed a method that would generate one receptor conformation that could bind

multiple diverse ligands that were not docking properly to a single crystal structure. The method, called Consensus Induced Fit Docking (cIFD), involves an initial generation of a receptor-ligand complex for multiple ligands, followed by binding site optimization around a hybrid compound frozen in space. Initial testing showed that while the resulting binding sites were often highly similar to the original ones, docking accuracy was improved, providing superior binding mode consistency for diverse chemotypes. Preliminary screening experiments utilizing cIFD structures resulted in improved enrichment rates. Encouraged by these results, we performed retrospective validation of the protocol on additional targets and applied the protocol to the structure-based discovery of Crm1 inhibitors, as described below.

cIFD retrospective validation

To test the benefit and applicability of the cIFD procedure, we ran calculations on COX-2, ER, and HIV-rt, which were previously shown to be challenging targets for docking (see “Methods”) [39]. Results are compared with rigid receptor docking using a single crystal structure and ensemble docking using the individual IFD structures. As seen in Fig. 1, the cIFD results typically fall between the single structure rigid-receptor docking and the ensemble docking results. This encouraging outcome is possibly expected, given that our hope was for cIFD to improve on rigid receptor docking while knowing that full ensemble docking offers a more realistic approximation of the true ensemble of receptor states. In addition to the favorable enrichments, the cIFD computational times are equivalent to single-structure rigid-receptor docking, since docking calculations scale linearly with the number of structures used. Ensemble docking, on the other hand, took approximately five times more computational resources to complete. It is interesting to note that for each target there is at least one IFD conformation that performs significantly better than the crystal structure (Figure S1), although determining a priori which structure to use for virtual screening presents a significant challenge for the field, as noted in previous work [17].

The BEDROC enrichment, which uses a Boltzmann weighting to favor actives that score well but still accounts for the entire ROC curve, shows that cIFD performs 25 % better on average than docking to the crystal structure [0.22 vs. 0.17 using BEDROC($\alpha = 20$)]. In addition, the method was as good as or better than the ensemble docking approach for both COX-2 and ER. It is also interesting to note that cIFD performs better or as good as the crystal structure or ensemble docking when looking at the enrichment in the top 10 % of the database (EF10%). On the other hand, while the EF1% values are comparable between the crystal structure and cIFD for both COX-2 and ER, they deteriorated for HIV-rt. The improved performance for EF10% can be understood

directly from the method, which generates a structure that should be able to accommodate more of the active compounds but possibly not fit any single active compounds as well as the ideal receptor structure for that compound. Given that, very early enrichment may be diminished with the cIFD approach but overall retrieval of active compounds should be relatively high because the receptor has been adapted to bind multiple active ligands. It is also worth noting that the improvements in enrichment using a cIFD model are based on specific rearrangements in the protein that allow binding of actives that would not fit into the rigid crystal binding site otherwise. This is different than softened-potential docking where the van der Waals radii are reduced for receptor and/or ligand atoms. We observed that the enrichment values, especially EF1%, deteriorate in softened-potential docking whereas they improve with cIFD docking (Figure S1).

The cIFD results presented above use a fully automated protocol with no user or experimental input in determining the structures to use for docking. The only input needed is a starting crystal structure and a set of active ligands. The method then combines the best poses for each ligand from IFD (i.e. lowest energy structures) to be used in the cIFD calculations. While a fully automated method is useful, in many cases there is substantial experimental biophysical data suggesting what ligand binding modes might be correct even in the absence of crystal structures. Using literature data to eliminate improbable poses, enrichments can be improved over the default cIFD protocol. In the cases of ER and HIV-rt, the poses with the lowest IFD scores agreed well with the known binding modes of similar actives. However, for COX-2 the top scoring IFD pose for one active ligand [ligand 1 (Figure S2)] does not extend toward the selectivity pocket formed by residues Leu352, Ser353, Tyr355, Phe518, and Val523 [40]. Taking an alternative pose for ligand 1 where the fluorophenyl group binds to the selectivity pocket greatly improves EF1% values using cIFD, with EF1% enrichments going from 4.3 to 7.7. The binding mode of the other four actives predicted by using the lowest IFD score agreed with the biochemical data.

The above results for cIFD assume that only a single initial crystal structure is available and crystal structures are not known for any of the active molecules of interest. While that might be a realistic scenario very early in a project, many projects have multiple crystal structures that could be used to reduce the potential for incorrect poses inherent to IFD predictions. Indeed, using crystal structures of known actives in cIFD, if available, improves the average EF1% value calculated for the three targets from 8.8 to 13.0 (Figure S3), with the largest improvement coming from ER.

COX-2 provides an excellent example for the dependence of screening results on the choice of X-ray structure. While

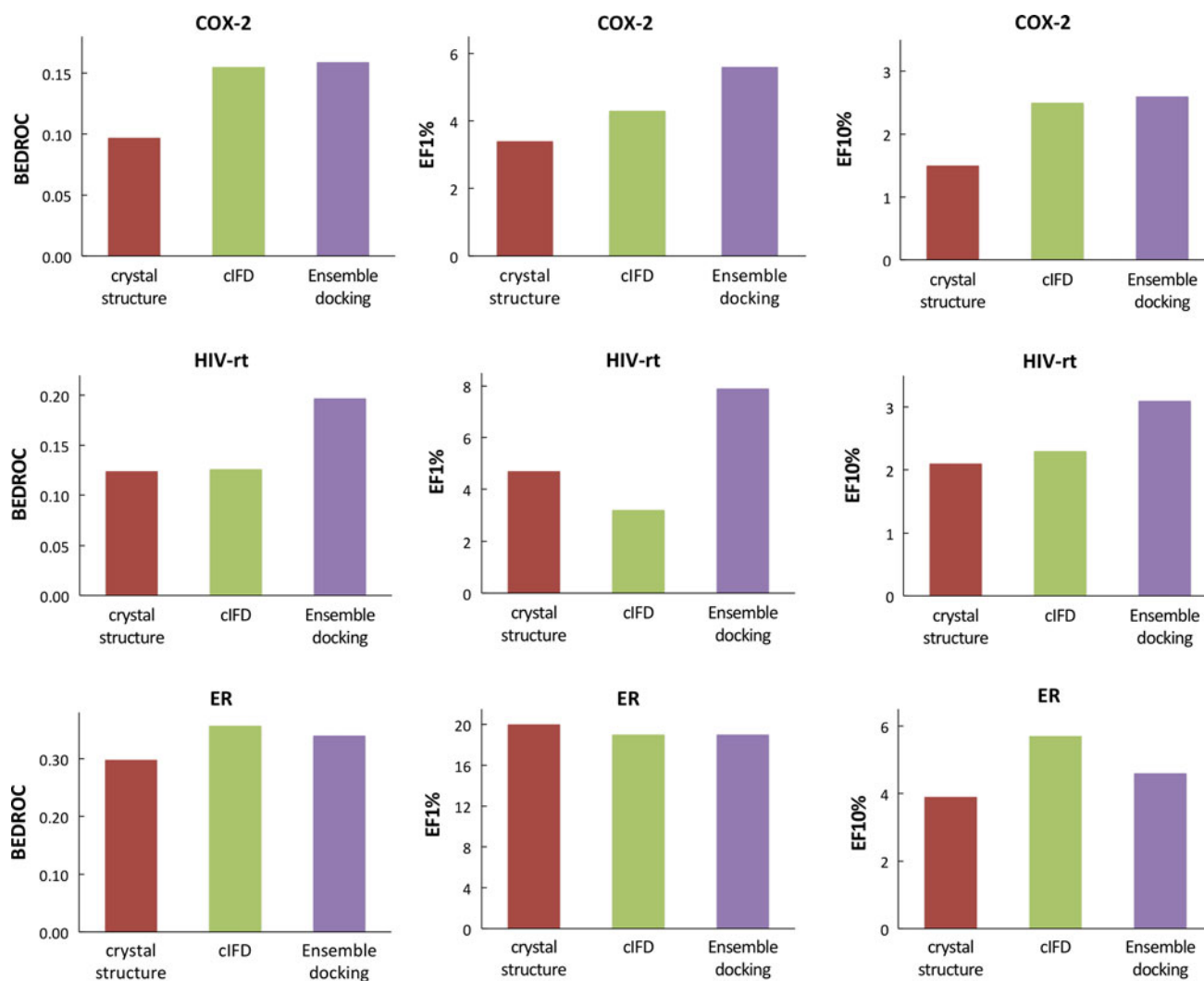


Fig. 1 Enrichment values for retrospective validation of the cIFD method. For each graph, enrichment values are shown for docking to different structures (red crystal, green cIFD, purple ensemble docking to all 5 IFD structures). Enrichment values shown include BEDROC

with $\alpha = 20$ (left column), enrichment in the top 1 % of the database (middle column), and enrichment in the top 10 % of the database (right column) for each of the targets studied (COX-2 top, HIV-rt middle, and ER bottom)

1CVU produces relatively low enrichment ($EF1\% = 3.4$), using alternative structures, e.g. 3LN1, results in significantly improved enrichment ($EF1\% = 17.0$ for 3LN1 rigid receptor docking). These structures are different in that one is a COX-2 complex with its substrate arachidonic acid (PDB ID 1CVU [41]) and the other is a complex with the inhibitor celecoxib, a non-steroidal anti-inflammatory drug (PDB ID 3LN1 [42]). In the case of the superior 3LN1 template, $EF1\%$ is reduced in cIFD compared to the crystal structure ($EF1\% = 17.0$ for 3LN1 crystal structure and $EF1\% = 14.0$ for cIFD using 3LN1 as the template). However, diverse actives with different binding modes were retrieved that could not be retrieved in rigid receptor docking at top 1 % of the screening library.

For ER, the early enrichment values ($EF1\%$) were comparable for all three methods described (single crystal structure, cIFD, and ensemble docking). However, the number of unique scaffolds as determined by the scaffold decomposition tool in the cheminformatics package Canvas [43, 44] is higher for cIFD and ensemble docking (21 for single crystal structure, 27 for cIFD, and 27 for ensemble docking), highlighting the value of cIFD in being able to produce results on par with ensemble docking while being significantly faster. Furthermore, the $EF10\%$ values for cIFD are higher than either crystal structure or ensemble docking.

Finally, HIV-rt is the only case in which enrichment did not improve significantly using cIFD. Many of the non-

nucleoside reverse transcriptase inhibitors (NNRTIs) bind in different modes to the flexible HIV-rt allosteric binding site [45, 46]. Alignment of HIV-rt crystal structures in complex with different NNRTIs shows the plasticity of HIV-rt allosteric binding site (see Figure S4B). As seen in the figure, the loops in the binding site change conformation to adapt to various NNRTIs. This case presents a limitation of the cIFD method, which works best when ligands bind in ways that are not mutually exclusive (e.g. COX-2 ligands shown in Figure S4A), as opposed to cases with multiple binding modes and large-scale flexibility that would preclude the simultaneous modeling of multiple diverse ligands binding to a single protein structure. This limitation is exemplified by a target like P38 MAP kinase, in which type I and type II ligands bind to a DFG-in and DFG-out conformation, respectively [47]. In such a case, the two binding sites cannot exist simultaneously because inducing the binding site to accommodate one class of ligands explicitly excludes the other binding site from forming. In such cases, ensemble docking should produce better results and has been shown to be successful for P38 [13].

This retrospective analysis demonstrates that cIFD is capable of producing a single receptor structure that can efficiently retrieve diverse active compounds. In the sections below, we describe the application of cIFD in a prospective drug discovery project to screen for Crm1 inhibitors.

Structure-based discovery of irreversible Crm1 inhibitors

The primary objective of the Crm1 project, performed at Karyopharm Therapeutics (KPT), was to discover novel Crm1 inhibitors by structure-based screening utilizing the NES-bound crystal structure of Crm1 available at the time. Initial testing of Glide rigid receptor docking protocols revealed that not all of the known Crm1 inhibitors could be docked correctly into the NES-bound crystal structure binding site, as judged by shape complementarity, ability to mimic NES hydrophobic interactions (Fig. 2), and the ability to position a thiol reactive warhead within ~ 4 Å of the Cys528 sulfur atom. Compounds tested included inhibitors reported by Kau et al. [35] and *N*-azolylacrylate analogs generated at Karyopharm. Our hypothesis was that small receptor rearrangements were needed to accommodate all of the actives. However, due to the requirements at Karyopharm for a computationally efficient virtual screening method, it was a principal objective to generate a single receptor structure that could be used for rigid receptor docking. Therefore, we used cIFD to generate a new conformation

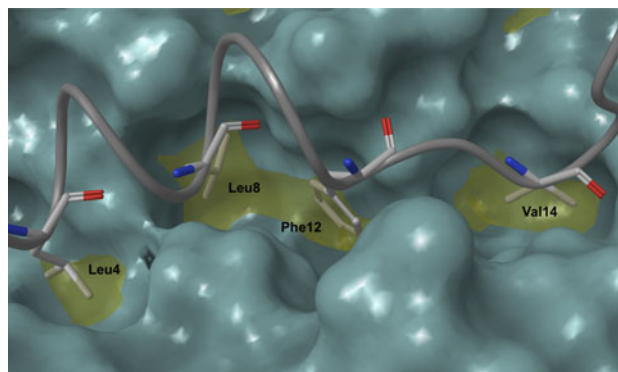


Fig. 2 Snurportin NES (gray tube and sticks) bound to Crm1 (blue surface) in the 3GJX X-ray structure [49]. Four main hydrophobic anchors are clearly observed. Yellow regions correspond to hydrophobic pockets identified by SiteMap

of the NES binding site that would enable improved chemotype coverage.

Consensus Induced Fit Docking was performed with four representative compounds including three *N*-azolylacrylates and compound 521996 from Kau et al. While the modeled structure is highly similar to the NES-bound crystal structure, there are two notable differences that affect compound binding. One difference is a rotation of Glu529 toward solvent (Fig. 3), which strongly affects the electrostatic properties of the binding site. The other is a domino effect of conformational changes in which Met545 makes way for the bound small molecule inhibitors (original movement spotted in IFD of Kau et al. compound 521996) consequently pushing Met583 away from the NES binding site (Fig. 3). Strikingly, this conformational change

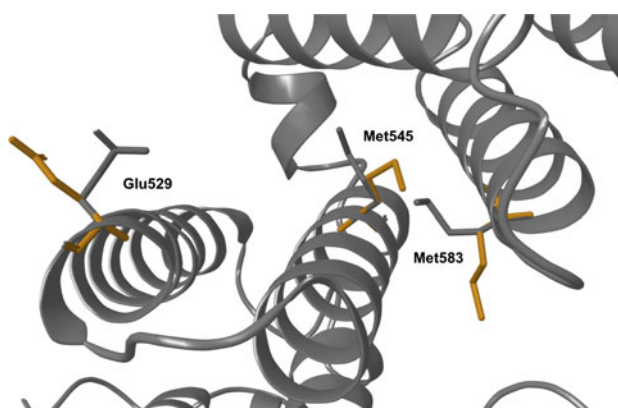


Fig. 3 Conformational changes in the Crm1 cIFD model. Gray ribbons and sticks NES-bound Crm1 crystal structure. Gold sticks cIFD model side chains. In the cIFD structure E529 assumes a more solvent exposed conformation and a switching motion is observed for M545 and M583, which move synchronously away from the NES binding groove to make way for bound small molecule inhibitors

was later validated by experimental co-crystal structures with bound KPT compounds [48].

The quality of the cIFD structure was evaluated by re-docking the known covalent inhibitors using both constrained (thiol-reactive warhead required to approach Cys528 thiol within ~ 4 Å) and unconstrained Glide docking (Methods). In general, improved binding modes (in better accordance with the criteria mentioned above, comprising our binding hypothesis) were obtained with the cIFD structure coupled with constrained docking (Methods), as exemplified for compound 521996 and CBS59106 in Figs. 4 and 5, respectively. The analysis of CBS9106, which is reported to be a reversible covalent binder [50] was performed in hindsight since the structure of this compound was only recently published.

The known inhibitors included in our analysis pose a significant challenge to the docking software since binding seems to be guided mainly by hydrophobic interactions and reactivity, of which only the former is recognized by Glide and in itself does not constitute a sharp enough signal. As a result, Glide Scores are relatively poor (mostly > -7.0) and binding modes are not consistent between related molecules. CBS59106 stands out in clearly forming a hydrogen bond with Lys568. It is possible that inclusion of this compound in cIFD modeling would have resulted in a slightly different model structure and would have affected the results of the virtual screen described below. Notably, the recently published structure of CRM1 bound to Karyopharm compound KPT-251 [48] provided support for the dominance of

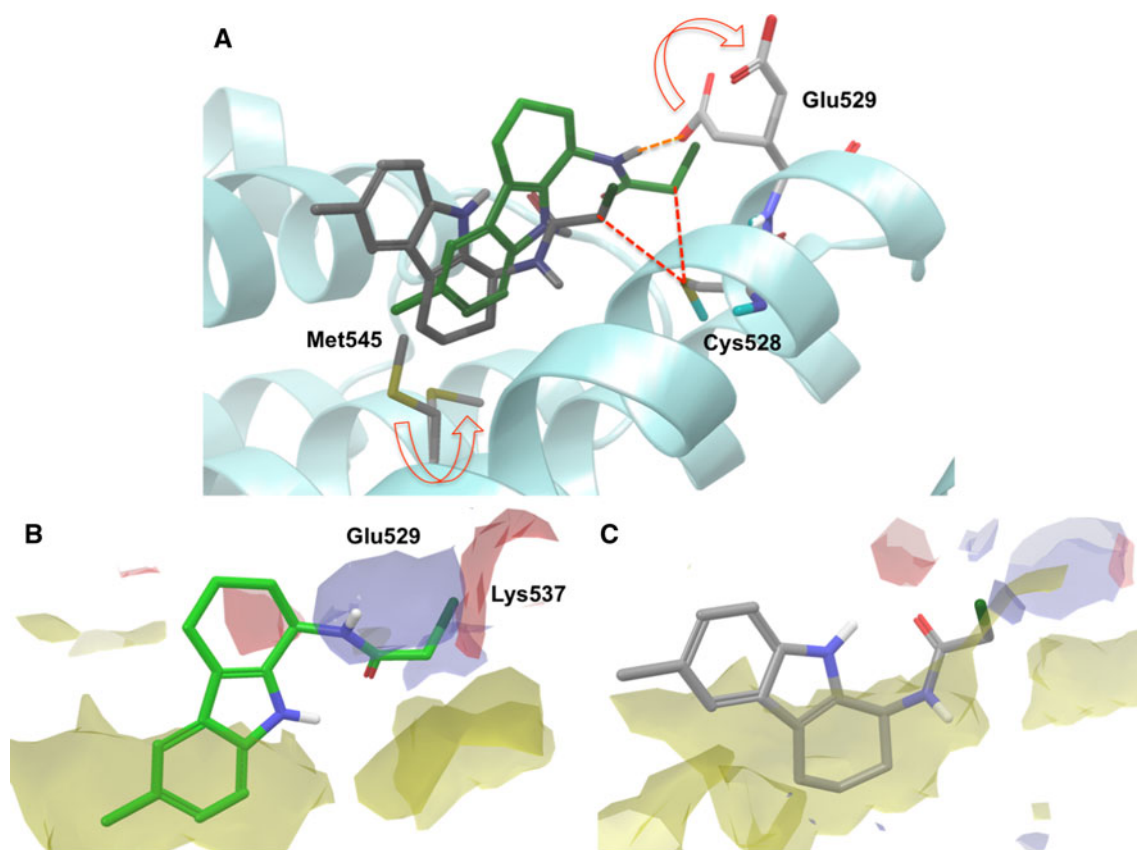


Fig. 4 Predicted binding modes of compound 521996 in the 3GJX Crm1 crystal structure and cIFD model (constrained docking). **a** Binding modes are shown superimposed. *Green carbons* 521996 docked to crystal structure. *Gray carbons* 521996 docked to cIFD model. Both binding modes position the reactive halo-methyl carbon within ~ 3.3 Å of the Cys528 thiol (*dashed red lines*), however binding site occupancy by Met545 coupled to electrostatic attraction by Glu529 in the X-ray structure lead to a shallower binding and hydrogen bond formation (*dashed orange line*) with Glu529. *Red arrows* highlight rotamer changes in cIFD structure compared to

3GJX. **b, c** show 521996 docked to X-ray and model structures, respectively, along with SiteMap [51, 52] projection of binding site properties. *Yellow* hydrophobic region, *Blue* hydrogen bond donor region, *Red* hydrogen bond acceptor region. The binding mode in the model structure is dominated by hydrophobic interactions and is in better agreement with our NES-mimetic binding hypothesis discussed above. These are significantly reduced in the X-ray binding mode, which also involves electrostatic interactions between amide and Glu529 (*blue patch*) and between terminal chlorine and Lys537 (*red patch*). Glide score = -5.4 for the X-ray, and -6.1 for the model

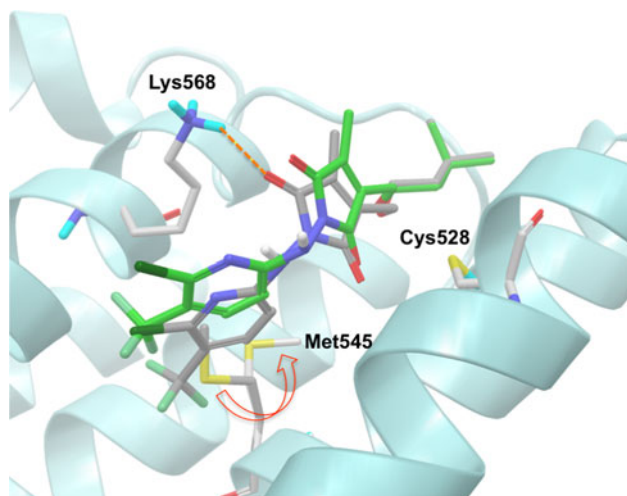


Fig. 5 Improved binding mode of CBS9106 in the cIFD structure. *Green carbons* CBS9106 docked to X-ray structure. *Gray carbons* CBS9106 docked to cIFD model, binding more deeply and forming a hydrogen bond with Lys568 (*dashed orange line*). The distance of the maleimide beta carbon from the Cys528 thiol is 3.8 and 3.7 Å for the X-ray and model structures, respectively. Respective Glide scores are -6.0 and -7.0

hydrophobic interactions as well as the binding modes predicted for this class of compounds (results not shown).

Subsequently, the cIFD model was used for structure-based screening (see “[Methods](#)” for details). A screening library of ~ 250 K potential covalent inhibitors, all containing thiol reactive groups (e.g. α,β -unsaturated ketones, halomethylketones, nitriles etc.), was prepared. The library was docked to the Crm1 model structure using constrained Glide docking and compounds with adequately positioned warheads were subjected to rescoring followed by a knowledge based enrichment guided filtering procedure. In this procedure, a series of structure-based and ligand-based filters were applied as described in “[Methods](#)”, reducing library size to $\sim 3,400$ compounds (1.3 %) while retaining 60 % of the known actives discussed above, corresponding to an enrichment factor equal to 46. These were subsequently clustered based on molecular similarity and 232 diverse compounds were selected and purchased for testing in a Rev-GFP localization assay.

The current approach suffers from three main limitations: (1) The majority of known inhibitors discussed above bind Crm1 without forming hydrogen bonds or salt bridge interactions and thus Glide scoring is mostly limited to hydrophobic and weak polar terms, which may not be sufficient for activity discrimination; (2) There is currently no computational method that would have enabled efficient evaluation of the actual reactivity of the diverse warheads included in the screening library; (3) The assay measures

Table 1 Results of in vitro screening of compounds selected with protocol 1

Activity range (IC_{50})	No. of compounds
30–100 μ M	6
10–30 μ M	9
1–10 μ M	1
<1 μ M	1
Total tested in vitro	232
% Active	7.3 %

Table 2 Results of in vitro screening of compounds selected with protocol 2

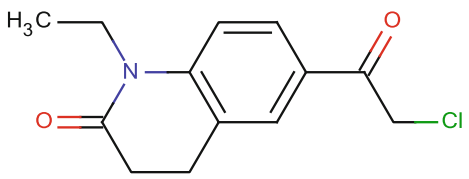
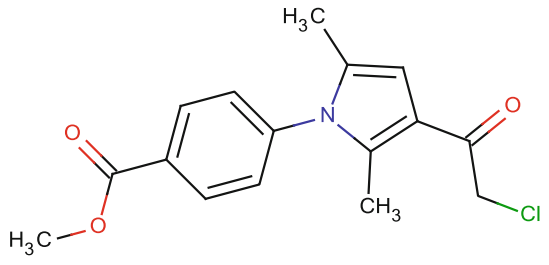
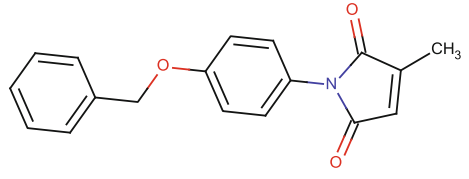
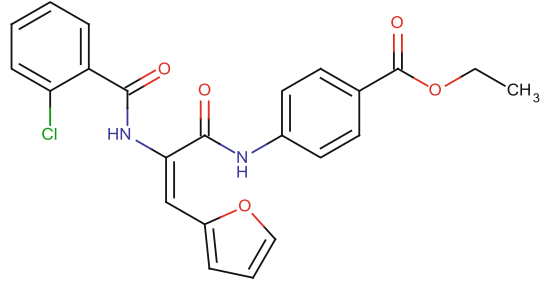
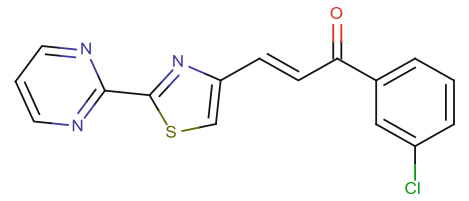
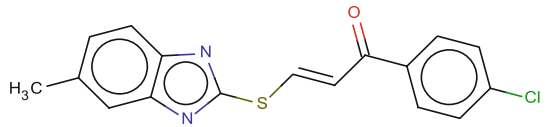
Activity range (IC_{50})	No. of compounds
30–100 μ M	8
10–30 μ M	7
1–10 μ M	1
<1 μ M	0
Total tested in vitro	170
% Active	9.4 %

cell-based functionality rather than direct binding and does not directly reflect binding affinity.

Despite these limitations, 17 of the tested compounds were found to inhibit Crm1 activity with an IC_{50} under 100 μ M (Table 1), corresponding to a hit rate of 7.3 %. While this hit rate represents a successful applications of the methodology, we were concerned that the knowledge based filtering procedure was introducing a bias that was limiting the chemical diversity of the hits obtained and possibly also reducing the hit rate (lower diversity leads to a selection of fewer representatives).

Therefore, a rescreen was performed using a smaller library containing only 11,680 compounds (Methods) and a blind filtering procedure was performed, in which a simple GlideScore filter was applied ($GlideScore \leq -6.0$) to the docked compounds. The remaining 3,053 compounds (26 %) were clustered (Methods) and a set of 170 diverse compounds were selected and purchased for testing. In this case, a hit rate of 9.4 % was obtained with a similar distribution of activities albeit with larger chemical diversity (Table 2). The aggregate hit rate when combining the two screens is 8.2 %. The results of this comparison are not conclusive and may benefit from a larger screen using the blind filtering process. Examples for hits obtained in the two screening projects are shown in Table 3.

Table 3 Examples of compounds active against Crm1 retrieved in the virtual screen using the cIFD structure

Structure	Vendor	Catalogue number	RevGFP IC50 (μM)
	Otava	1087055	2
	Otava	129720256	13
	Chembridge	5106795	10
	Specs	AL-281/14557040	13
	Enamine	T6657853	9
	Maybridge	MFCD00204680	6

Conclusions

In this work, we presented a new method to generate a receptor structure conformation that would improve virtual screening enrichments and boost the retrieval of diverse ligands. The method, called Consensus Induced Fit Docking (cIFD), involves an initial generation of a ligand-receptor

complex for several ligands (via crystal structure or Induced-Fit Docking calculations) followed by a Prime side chain refinement and minimization of the protein atoms around a hybrid ligand. The protein “feels” the force of all of the ligands but the ligands do not interact with each other. The primary advantage of the cIFD method is the ability to indirectly account for some amount of protein flexibility while not

adding to the computational costs of rigid receptor docking to a single target. Although the cIFD structure is unlikely to be a physically accurate representation for the binding of any single ligand, it provides a useful model structure that can help retrieve diverse ligands that might not be able to bind the same co-crystallized receptor conformation.

We first validated the method in a retrospective study of three targets (COX-2, ER, and HIV-rt). These three targets were chosen because they were previously shown to be particularly challenging for docking programs, possibly due to the inability of a single receptor structure to dock diverse ligands. We showed that the method consistently performed better than using a single rigid crystal structure as the target for docking. In addition, the method was able to achieve results on par with ensemble docking, which combines results from separate docking calculations to different protein conformations and takes significantly longer than cIFD. HIV-rt is the only case in which enrichment did not improve significantly using cIFD. This is mainly due to the fact that many of the non-nucleoside reverse transcriptase inhibitors (NNRTIs) bind in distinct modes to the flexible HIV-rt allosteric binding site. The cIFD method works best when ligands adopt similar binding modes as opposed to poses involving large-scale protein rearrangements, which would compromise modeling of a single protein conformation simultaneously bound to multiple diverse ligands.

We then applied cIFD to an active drug discovery project in pursuit of finding novel covalent inhibitors of Crm1. Our analysis of the model structure suggests that cIFD improves docking results of known inhibitors by facilitating receptor movements required for small molecule binding. Application of cIFD in two separate screens yielded a total of 33 covalent protein–protein interaction inhibitors with measured affinity of at least 100 μM . Analysis of a recently reported Crm1 inhibitor suggests that as new ligands displaying novel interaction patterns are revealed, cIFD modeling may be revisited and updated models could be used in future screening campaigns.

While the results from this study are encouraging, more work is needed to establish the value of cIFD in virtual screening campaigns. First, it will be necessary to screen a larger number of targets from diverse protein classes. Internal results from ongoing drug discovery programs (data not shown) suggest great utility in the discovery of Type-I kinase inhibitors. Next, various aspects of the protocol could be explored in more detail to determine whether systematic improvements can be realized. For example, the choice of the initial receptor structure is likely to be important and for our studies on COX-2 we saw that starting with a structure containing a potent inhibitor produced better enrichments than a substrate-bound structure. Also, the number of ligands to use in the initial cIFD refinement was not explored in this work. We may find that more or less ligands are needed to obtain

good results, depending on the system and the amount of receptor movement that is needed. Finally, the method is not capable of dealing with simultaneous receptor movements that are mutually incompatible. For example, in kinases it would not be possible to generate a single cIFD structure that could bind both DFG-in (type I) and DFG-out (type II) inhibitors because the movement of the activation loop to accommodate ligands from one class prohibits the binding of the other class. While the current framework of cIFD is not capable of handling systems like this, we aim to develop a strategy to detect such systems in advance. Then, cIFD could be performed on each state and the results of screening to each cIFD structure could be merged using ensemble docking techniques. The above issues are the aim of our current research and will be addressed in future publications.

Methods

Target validation set

COX-2, ER and HIV-rt were chosen as the targets because it has previously been shown that these targets presented challenges for multiple docking programs [39]. The PDB codes used for these targets are 1CVU (COX-2), 3ERT (ER), and 1EP4 (HIV-rt). The proteins were prepared with the Protein Preparation Wizard in Maestro [53]. In short, this included assignment of bond orders for ligands, addition of hydrogen atoms, optimization of the hydrogen bonding network, and a restrained minimization. All default options were used.

Ligand validation set

Active ligands were retrieved from the literature and prepared with LigPrep [54]. For each target, the ligands subjected to hierarchical clustering using Canvas [43, 44] using radial fingerprints [55] with Tanimoto similarity and complete linkage. A clustering level of five was chosen as a reasonable number of compounds for the cIFD procedure. This value was not varied, so it is possible results could be improved with more or less compounds. The tightest binding compound from each of cluster was retained for cIFD calculations (Figure S2).

Database compounds

The database compounds were taken from the MDDR, as described in McGaughey et al. [39]. The initial database of approximately 129,000 compounds was clustered using the Butina algorithm [56] with a similarity cutoff of 0.7 using the Dice similarity metric and atom pair descriptors. The centroid was chosen as the representative structure

from each cluster. Molecules with molecular weight greater than 500 Da were removed, resulting in 28,038 compounds among which there were 234 actives for COX-2, 54 actives for ER and 127 actives for HIV-rt.

cIFD protocol

IFD [28] calculations were performed on each target using the five ligands selected for each, as described above. The best complex for each ligand (as defined by the lowest IFD Score) was selected and the ligand poses from each were merged into a single structure. There are many choices for the receptor to use for the cIFD refinement with the merged ligands (initial crystal or one of the IFD structures). We use the IFD structure with the best ligand efficiency (GlideScore divided by MW), since that should represent a complex where most of the ligand is making productive interactions with the receptor. The other four ligand poses are merged into this structure and the resulting complex is refined with Prime [57]. In the refinement, side chains within 5.0 Å from any of the merged ligand atoms were identified and fully minimized while keeping the merged ligand frozen in space.

Virtual screening and enrichment calculations

Docking calculations were performed with the SP mode of Glide [6, 58]. Enrichment values were computed with the *enrichment.py* script available from the Schrödinger Script Center (www.schrodinger.com/scriptcenter). We focused primarily on EF1%, EF10%, and BEDROC($\alpha = 20$) [59]. We also looked at the diversity-weighted enrichment factors DEF1% and DEF10% to see whether the retrieved actives were diverse in addition to looking simply at the number of actives, as described previously [60]. For the studies here, the DEF results showed the same qualitative trends as the EF values and therefore are not reported. In addition to the cIFD calculations, we also performed full ensemble docking, where a separate docking calculation was run on each of the IFD structures. The results from each individual calculation were merged and the top pose for each ligand was selected based on the GlideScore. Finally, docking was also performed on the prepared initial crystal structure to ensure that the cIFD procedure offered an advantage over standard rigid receptor docking.

Crm1 modeling and screening

Binding site optimization in the presence of merged ligand

Ligands were superimposed in the Crm1 X-ray binding site (3GJX) and merged into a single hybrid ligand structure. This hybrid structure was excluded from the selection of residues for Prime side chain refinement, thus keeping it fixed in space.

Preparation of compound libraries for screening

First screen: Drug like collection were obtained from Asinex (www.asinex.com), Maybridge (www.maybridge.com), Bionet (www.keyorganics.co.uk), Specs (www.specs.net), Chembridge (www.chembridge.com), ChemDiv (www.chemdiv.com), and Enamine (www.enamine.net). Compounds were prepared using the Virtual Screening Workflow (VSW) ligand preparation tab in Maestro. “Regularize input geometries” was applied and ionization states and tautomers were determined by the ionizer at a pH 7.4. Compounds were subsequently filtered using the following chemical property ranges: $250 \leq MW < 600$, $RB < 10$, $HBA \leq 10$, $HBD \leq 5$. In the second screen, only collections from Maybridge (www.maybridge.com), Specs (www.specs.net), and Otava (<http://www.otavachemicals.com>) were included and were prepared as in the first screen.

Warhead filtering

Extraction of compounds containing thiol-reactive chemical warheads was performed using the Ligand Filtering utility in Maestro [53]. SMARTS patterns describing chemical warhead were defined manually.

Glide docking

Glide docking was performed with the “expanded sampling” option. Constrained docking was performed with a 3.5 Å Glide positional constraint centered at the Cys528 SG. Several alternative radii were tested and 3.5 Å was found to produce the superior binding modes for the majority of the known inhibitors evaluated. During screening, 33 SMARTS patterns corresponding to different types of chemical warheads were allowed to match this constraint.

Re-Scoring

Ranges were determined for the following scores based on results obtained for docked known inhibitors: Glide Score [6, 58], XScore [61], Phase Shape similarity [62], MW, and ClogP o/w, QlogS, FISA, and 2D-PISA calculated with QikProp [63]. The Phase Shape similarity was based on one of the Karyopharm lead compounds. These ranges were used to filter the screening library.

Clustering

Compounds were clustered in Canvas using the linear fingerprints [64] and hierarchical clustering with default parameters. Clusters were collected at a Tanimoto cutoff of 0.6 following evaluation of several alternative cutoff values.

Rev-GFP assay

U2OS cells were cultured in McCoy's 5A medium (Invitrogen) supplemented with 10 % heat-inactivated fetal bovine serum (Invitrogen) and 50 µg/ml penicillin/streptomycin (Invitrogen). Stable expression of Rev-GFP (pRev(1.4)-GFP+PKI, Wolff, 1997) was maintained in 200 µg/ml geneticin. U2OS cells were plated in 96-well plate (15,000 cells/well) and left overnight to attach. Cells were treated with serial diluted (started at 10 µM; 1:3 dilution) screening compounds for 4 h to assure steady state Rev-GFP localization. The cells were collected, washed with PBS (Invitrogen), and fixed with 3 % paraformaldehyde solution (3 % w/v paraformaldehyde and 2 % w/v sucrose in 1X PBS) for at least 15 min at room temperature. Nuclei of fixed cells were stained with DAPI (Invitrogen) in PBS for at least 10 min at room temperature. The U2OS cells were imaged using a Nikon fluorescent microscope at 10X magnification. A monochrome camera was used to capture GFP and DAPI images (1 of each per well). Using the Nikon Imaging Software—Elements for capture and analysis, the DAPI image was used to create a threshold of intensity for all wells. The parameter of this threshold was the outline of the nucleus of each cell stained with DAPI. This intensity of the GFP was measured and recorded along with the area for each cell in all images per plate. Each cell was scored by dividing the GFP intensity by total nuclear area. Cells with a ratio (GFP intensity/nuclear area) above a user-defined threshold were scored as positive nuclei. The number of GFP positive nuclei was divided by total number of cells giving the percentage of cells with nuclear Rev-GFP. Three separate wells were analyzed for each concentration of the IC50 curves. XLFit model 205 was used to calculate IC50 curves.

References

1. Tsou CL (1993) *Science* 262(5132):380
2. Teague SJ (2003) *Nat Rev Drug Disc* 2(7):527
3. Carlson HA, McCammon JA (2000) *Mol Pharmacol* 57(2):213
4. Carlson HA (2002) *Curr Opin Chem Biol* 6(4):447
5. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT (2006) *J Med Chem* 49(21):6177
6. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL (2004) *J Med Chem* 47(7):1750
7. Shoichet BK (2004) *Nature* 432(7019):862
8. Shen J, Tan C, Zhang Y, Li X, Li W, Huang J, Shen X, Tang Y (2010) *J Med Chem* 53(14):5361
9. Podvinec M, Lim SP, Schmidt T, Scarsi M, Wen D, Sonntag L-S, Sanschagrin P, Shenkin PS, Schwede T (2010) *J Med Chem* 53(4):1483
10. Huggins DJ, Sherman W, Tidor B (2012) *J Med Chem* 55(4):1424
11. Shan Y, Kim E, Eastwood MP, Dror RO, Seeliger MA, Shaw DE (2011) *J Am Chem Soc* 133(24):9181
12. Gervasio FL, Laio A, Parrinello M (2005) *J Am Chem Soc* 127(8):2600
13. Sherman W, Beard HS, Farid R (2006) *Chem Biol Drug Des* 67(1):83
14. Huang SY, Zou X (2007) *Proteins: Struct, Funct, Bioinf* 66(2):399
15. Rao S, Sanschagrin PC, Greenwood JR, Repasky MP, Sherman W, Farid R (2008) *J Comput Aided Mol Des* 22(9):621
16. Barril X, Morley SD (2005) *J Med Chem* 48(13):4432
17. Craig IR, Essex JW, Spiegel K (2010) *J Chem Inf Model* 50(4):511
18. Bouzida D, Rejto PA, Arthurs S, Colson AB, Freer ST, Gehlhaar DK, Larson V, Luty BA, Rose PW, Verkhivker GM (1999) *Int J Quantum Chem* 72(1):73
19. Rueda M, Bottegoni G, Abagyan R (2009) *J Chem Inf Model* 49(3):716
20. Cavasotto CN, Kovacs JA, Abagyan RA (2005) *J Am Chem Soc* 127(26):9632
21. Nichols SE, Baron R, Ivetac A, McCammon JA (2011) *J Chem Inf Model* 51(6):1439
22. Osguthorpe DJ, Sherman W, Hagler AT (2012) *J Phys Chem B* 116(23):6952
23. Osguthorpe DJ, Sherman W, Hagler AT (2012) *Chem Biol Drug Des* 80(2):182
24. Corbeil CR, Englebienne P, Yannopoulos CG, Chan L, Das SK, Bilimoria D, L'Heureux L, Moitessier N (2008) *J Chem Inf Model* 48(4):902
25. Cavasotto CN, Abagyan RA (2004) *J Mol Biol* 337(1):209
26. Sela I, Golan G, Strajbl M, Rivenzon-Segal D, Bar-Haim S, Bloch I, Inbal B, Shitrit A, Ben-Zeev E, Fichman M, Markus Y, Marantz Y, Senderowitz H, Kalid O (2010) *Curr Top Med Chem* 10(6):638
27. Roitberg A, Elber R (1991) *J Chem Phys* 95:9277
28. Sherman W, Day T, Jacobson MP, Friesner RA, Farid R (2006) *J Med Chem* 49(2):534
29. Kutay U, Güttinger S (2005) *Trends Cell Biol* 15(3):121
30. Turner JG, Sullivan DM (2008) *Curr Med Chem* 15(26):2648
31. Mutka SC, Yang WQ, Dong SD, Ward SL, Craig DA, Timmermans PBMWM, Murli S (2009) *Cancer Res* 69(2):510
32. Köster M, Lykke-Andersen S, Elnakady YA, Gerth K, Washausen P, Höfle G, Sasse F, Kjems J, Hauser H (2003) *Exp Cell Res* 286(2):321
33. Meissner T, Krause E, Vinkemeier U (2004) *FEBS Lett* 576(1):27
34. Bonazzi S, Eidam O, Güttinger S, Wach J-Y, Zemp I, Kutay U, Gademann K (2010) *J Am Chem Soc* 132(4):1432
35. Kau TR, Schroeder F, Ramaswamy S, Wojciechowski CL, Zhao JJ, Roberts TM, Clardy J, Sellers WR, Silver PA (2003) *Cancer Cell* 4(6):463
36. Van Neck T, Pannecouque C, Vanstreels E, Stevens M, Dehaen W, Daelemans D (2008) *Bioorg Med Chem* 16(21):9487
37. Shacham S, Kauffman M, Sandanayaka VP, Shechter S, US 2011/0275607 A1 (2011) Nuclear transport modulators and uses thereof. Google Patents
38. Sakakibara K, Saito N, Sato T, Suzuki A, Hasegawa Y, Friedman JM, Kufe DW, VonHoff DD, Iwami T, Kawabe T (2011) *Blood* 118(14):3922
39. McGaughey GB, Sheridan RP, Bayly CI, Culbertson JC, Kreatsoulas C, Lindsley S, Maiorov V, Truchon J-F, Cornell WD (2007) *J Chem Inf Model* 47(4):1504
40. Kurumbail RG, Stevens AM, Gierse JK, McDonald JJ, Stegeman RA, Pak JY, Gildehaus D, iyashiro JM, Penning TD, Seibert K, Isakson PC, Stallings WC (1996) *Nature* 384(6610):644
41. Kiefer JR, Pawlitz JL, Moreland KT, Stegeman RA, Hood WF, Gierse JK, Stevens AM, Goodwin DC, Rowlinson SW, Marnett LJ, Stallings WC, Kurumbail RG (2000) *Nature* 405(6782):97

42. Wang JL, Limburg D, Graneto MJ, Springer J, Hamper JRB, Liao S, Pawlitz JL, Kurumbail RG, Maziasz T, Talley JJ, Kiefer JR, Carter J (2010) *Bioorg Med Chem Lett* 20(23):7159
43. Canvas v1.4. (2011) Schrödinger Inc., Portland
44. Sastry M, Lowrie JF, Dixon SL, Sherman W (2010) *J Chem Inf Model* 50(5):771
45. Das K, Lewi PJ, Hughes SH, Arnold E (2005) *Prog Biophys Mol Biol* 88(2):209
46. Hopkins AL, Ren J, Milton J, Hazen RJ, Chan JH, Stuart DI, Stammers DK (2004) *J Med Chem* 47(24):5912
47. Pargellis C, Tong L, Churchill L, Cirillo PF, Gilmore T, Graham AG, Grob PM, Hickey ER, Moss N, Pav S (2002) *Nat Struct Mol Biol* 9(4):268
48. Etchin J, Sun Q, Kentsis A, Farmer A, Zhang Z, Sanda T, Mansour M, Barcelo C, McCauley D, Kauffman M (2012) *Leukemia*. doi:10.1038/leu.2012.219
49. Monecke T, Güttler T, Neumann P, Dickmanns A, Görlich D, Ficner R (2009) *Science* 324(5930):1087
50. Khanna IK, Weier RM, Yu Y, Xu XD, Koszyk FJ, Collins PW, Koboldt CM, Veenhuizen AW, Perkins WE, Casler JJ, Masferrer JL, Zhang YY, Gregory SA, Seibert K, Isakson PC (1997) *J Med Chem* 40(11):1634
51. Halgren T (2007) *Chem Biol Drug Des* 69(2):146
52. Halgren T (2009) *J Chem Inf Model* 49:377
53. Maestro v9.2. (2011) Schrödinger, Inc., Portland
54. LigPrep v2.4. (2010) Schrödinger Inc., Portland
55. Rogers D, Hahn M (2010) *J Chem Inf Model* 50(5):742
56. Butina D (1999) *J Chem Inf Comput Sci* 39(4):747
57. Prime v3.0. (2011) Schrödinger Inc., Portland
58. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) *J Med Chem* 47(7):1739
59. Truchon J-F, Bayly CI (2007) *J Chem Inf Model* 47(2):488
60. Salam NK, Nuti R, Sherman W (2009) *J Chem Inf Model* 49(10):2356
61. Wang R, Lai L, Wang S (2002) *J Comput Aided Mol Des* 16(1):11
62. Sastry M, Dixon S, Sherman W (2011) *J Chem Inf Model* 51(10):2455
63. QikProp v2.0. (2011) Schrödinger Inc., Portland
64. Daylight Chemical Information Systems (2008) Aliso Viejo
65. Portevin B, Tordjman C, Pastoureau P, Bonnet J, De Nanteuil G (2000) *J Med Chem* 43(24):4582
66. Janusz JM, Young PA, Ridgeway JM, Scherz MW, Enzweiler K, Wu LI, Gan L, Chen J, Kellstein DE, Green SA, Tulich JL, Rosario-Jansen T, Magrisso IJ, Wehmeyer KR, Kühlenbeck DL, Eichhold TH, Dobson RLM (1998) *J Med Chem* 41(18):3515
67. Huang H-C, Li JJ, Garland DJ, Chamberlain TS, Reinhard EJ, Manning RE, Seibert K, Koboldt CM, Gregory SA, Anderson GD, Veenhuizen AW, Zhang Y, Perkins WE, Burton EG, Cogburn JN, Isakson PC, Reitz DB (1996) *J Med Chem* 39(1):253
68. Motakis D, Parniak MA (2002) *Antimicrob Agents Chemother* 46(6):1851
69. Tanaka H, Takashima H, Ubasawa M, Sekiya K, Inouye N, Baba M, Shigeta S, Walker RT, De Clercq E, Miyasaka T (1995) *J Med Chem* 38(15):2860
70. Ludovici DW, Kavash RW, Kukla MJ, Ho CY, Ye H, De Corte BL, Andries K, de Béthune M-P, Azijn H, Pauwels R, Moereels HEL, Heeres J, Koymans LMH, de Jonge MR, Van Aken KJA, Daeyaert FFD, Lewi PJ, Das K, Arnold E, Janssen PAJ (2001) *Bioorg Med Chem Lett* 11(17):2229
71. Himmel DM, Das K, Clark AD, Hughes SH, Benjahad A, Oumouch S, Guillemont J, Coupa S, Poncelet A, Csoka I, Meyer C, Andries K, Nguyen CH, Grierson DS, Arnold E (2005) *J Med Chem* 48(24):7582
72. Sun J, Meyers MJ, Fink BE, Rajendran R, Katzenellenbogen JA, Katzenellenbogen BS (1999) *Endocrinology* 140(2):800
73. Gangloff M, Ruff M, Eiler S, Duclaud S, Wurtz JM, Moras D (2001) *J Biol Chem* 276(18):15059
74. Blizzard TA, DiNinno F, Morgan IJ, Chen HY, Wu JY, Kim S, Chan W, Birzin ET, Yang YT, Pai LY, Fitzgerald PMD, Sharma N, Li Y, Zhang Z, Hayes EC, DaSilva CA, Tang W, Rohrer SP, Schaeffer JM, Hammond ML (2005) *Bioorg Med Chem Lett* 15(1):107
75. Dykstra KD, Guo L, Birzin ET, Chan W, Yang YT, Hayes EC, DaSilva CA, Pai L-Y, Mosley RT, Kraker B, Fitzgerald PMD, DiNinno F, Rohrer SP, Schaeffer JM, Hammond ML (2007) *Bioorg Med Chem Lett* 17(8):2322
76. Richardson TI, Frank SA, Wang M, Clarke CA, Jones SA, Ying B-P, Kohlman DT, Wallace OB, Shepherd TA, Dally RD, Palkowitz AD, Geiser AG, Bryant HU, Henck JW, Cohen IR, Rudmann DG, McCann DJ, Coutant DE, Oldham SW, Hummel CW, Fong KC, Hinklin R, Lewis G, Tian H, Dodge JA (2007) *Bioorg Med Chem Lett* 17(13):3544