

# Hit clustering can improve virtual fragment screening: CDK2 and PARP1 case studies

Alexey A. Zeifman · Victor S. Stroylov · Fedor N. Novikov · Oleg V. Stroganov ·  
Alexandra L. Zakharenko · Svetlana N. Khodyreva · Olga I. Lavrik ·  
Ghermes G. Chilov

Received: 12 May 2011 / Accepted: 12 October 2011 / Published online: 9 November 2011  
© Springer-Verlag 2011

**Abstract** Virtual fragment screening could be a promising alternative to existing experimental screening techniques. However, reliable methods of *in silico* fragment screening are yet to be established and validated. In order to develop such an approach we first checked how successful the existing molecular docking methods can be in predicting fragment binding affinities and poses. Using our Lead Finder docking software the RMSD of the binding energy prediction was observed to be  $1.35 \text{ kcal/mol}^{-1}$  on a set of 26 experimentally characterized fragment inhibitors, and the RMSD of the predicted binding pose from the experimental one was  $<1.5 \text{ \AA}$ . Then, we explored docking of 68 fragments obtained from 39 drug molecules for which co-crystal structures were available from the PDB. It appeared that fragments that participate in oriented non-covalent interactions, such as hydrogen bonds and metal coordination, could be correctly docked in 70–80% of cases suggesting the potential success of rediscovering of corresponding drugs by *in silico* fragment approach. Based on these findings we've developed a virtual fragment screening technique which involved structural filtration of

protein-ligand complexes for specific interactions and subsequent clustering in order to minimize the number of preferable starting fragment candidates. Application of this method led to 2 millimolar-scale fragment PARP1 inhibitors with a new scaffold.

**Keywords** CDK2 · FBDD · Molecular docking · PARP1 · Virtual screening

## Introduction

The fragment approach in drug discovery

The target-based drug discovery (TBDD) is one of the modern paradigms of drug discovery [1].

TBDD begins with the screening of compounds that can bind to a protein target. High-throughput screening (HTS) has become an established method, involving the fast experimental processing of a large number (typically from tens of thousands to millions) of organic compounds with the MW in 300–500 range. The hit ligands, identified through the HTS, have  $K_d$  typically in the micromolar scale and that makes them suitable targets for lead optimization.

The fragment-based drug discovery (FBDD) paradigm is different from the classical TBDD approach. The fragment screening is performed on a smaller library of several thousand smaller organic compounds with MW in the range of 150–300 [2]. Due to their small size, the FBDD hits typically have  $K_d$  in the millimolar range. Once the fragment hits are identified, their binding efficiency is enhanced through growth or linking. Effective inhibitors have reportedly been designed from fragments with  $K_d=1 \text{ mM}$  [3].

The FBDD has two main advantages as compared to the HTS. First, a fragment library is capable of covering a

**Electronic supplementary material** The online version of this article (doi:10.1007/s00894-011-1280-4) contains supplementary material, which is available to authorized users.

A. A. Zeifman · V. S. Stroylov · F. N. Novikov ·  
O. V. Stroganov · G. G. Chilov (✉)  
N. D. Zelinsky Institute Of Organic Chemistry (ZIOC RAS),  
Leninsky avenue, 47,  
119991 Moscow, Russia  
e-mail: ghermes@moltech.ru

A. L. Zakharenko · S. N. Khodyreva · O. I. Lavrik  
Novosibirsk Institute of Chemical Biology and Fundamental  
Medicine, Siberian Division of the Russian Academy of Sciences,  
Prospect Lavrentieva 8,  
Novosibirsk 630090, Russia

much larger percentage of the available chemical space. A thousand of diverse fragment structures from 14 million of structures with  $MW < 160$  is a much more representative selection than a million of HTS structures from the chemical space of  $10^{60}$  possible organic substances [4]. Second, the size of a drug candidate molecule is guided by the Lipinski Rule of Five [5]. During the lead optimization stage the improvement of  $pK_d$  leads at least to a linear increase of MW [6], therefore smaller starting fragments possess a significantly greater capacity for optimization.

Despite these advantages a number of technical difficulties are characteristic for FBDD. The detection of weak but specific binding in the millimolar range requires the use of such sophisticated techniques as SPR [7], high-throughput X-ray crystallography [8], NMR [9], etc. Also, fragment linking is a significant problem by itself. In some cases, fragment linking is outright impossible, for example when the fragment binding sites overlap, or when a linker with the right geometry does not exist.

### In silico FBDD

In silico implementation of FBDD could be a valuable enhancement of the existing experimental routines and is supposed to reduce the scope of wet lab work and bring some structural insights regarding further optimization of fragment hits. However, the in silico FBDD raises a number of additional issues. The generation of fragment libraries is usually performed either through the application of “the rule of three” [10] to a set of synthetically available structures, or through the algorithmic generation of all possible structures that do not exceed a specified molecular weight. In the former case, the generated library may lack diversity, while in the latter the compounds may not be synthetically available. Another problem of the in silico FBDD originates from the fact that the molecular docking scoring functions are typically trained on, and optimized for, larger ligands. Consequently, the use of such scoring functions on fragments leads to errors in scoring and rank-ordering [11].

### Suggested implementation of the in silico FBDD method

Clearly, the core of in silico FBDD should comprise a reliable virtual fragment screening routine, that is a software tool capable of finding fragment hits and predicting their binding mode, which is necessary for further lead

optimization. In this work we evaluate our Lead Finder docking software as an engine of virtual fragment screening and develop a practical protocol of such simulations. In order to validate the developed workflow, we first assess the accuracy of fragment docking with Lead Finder by using a test set of experimentally characterized fragment inhibitors. Subsequently, we hypothesized the feasibility of reconstruction of known drugs by breaking them into fragments, docking such fragments and characterizing the docked fragments of a drug by various parameters. As a result of those experiments, we identified drugs that are more likely to be reconstructed from their fragments. Finally, we developed the method of virtual fragment screening and probed it in a series of practical runs. The developed virtual screening protocol aims at prioritizing fragment hits for further lead optimization and consists of the docking-based virtual screening, structural filtration of docked ligand poses, clustering of the obtained hits and visual inspection of the best hits from each cluster. To validate our virtual screening protocol, we performed a search of novel fragment inhibitors of PARP1 and CDK2 important therapeutic targets, for which a number of clinical drug candidates have been developed recently [12–15].

## Materials and methods

### Ligand preparation


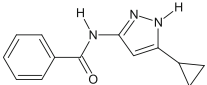
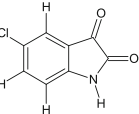
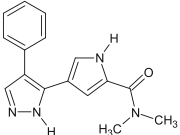
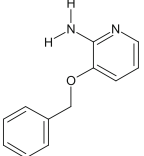
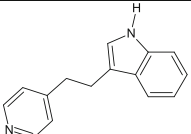
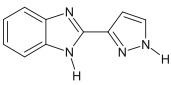

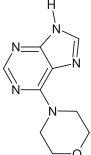
The experimental information on the selected fragment inhibitors, taken from the literature, is provided in Table 1.

The list of drugs with known protein-ligand structure was taken from the EBI database [16]. In this study, we used only the protein targets whose biological activity had been known to be modulated by ligand binding.

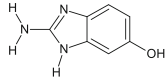
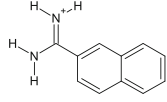
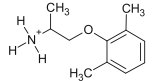
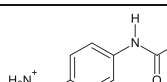
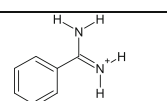
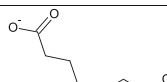
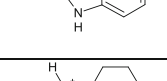
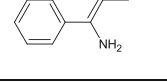
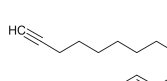
We used ACD/ChemSketch v 12.0 [17] to break down the drug molecules into fragments in accordance with the following rules (**Supporting information**):

- a) the fragments must satisfy “the rule of three” [10];
- b) the breakdown into fragments must not create new ionogenic groups, since the presence of such groups may critically alter the nature of ligand binding;
- c) the breakdown into fragments must split as few bonds as possible so that the reconstruction of the original drug structure is still feasible;

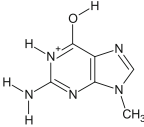
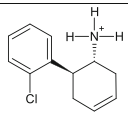
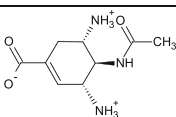
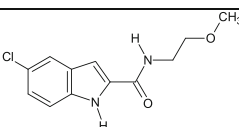
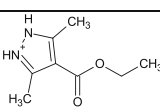
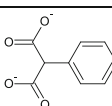
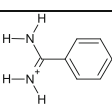
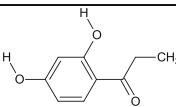
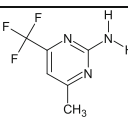
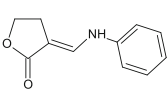
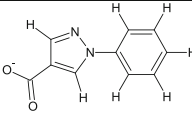
**Table 1** The fragments with experimentally measured data. \* denotes complexes of the target protein that contain an analogue of a docked fragment (all other complexes contain docked fragments as native ligands)

Target	PDB ID	Starting fragment	RMSD, Å	dG <sub>exp</sub> , kcal/mol	dG <sub>calc</sub> , kcal/mol	ΔdG , kcal/mol
CDK2 [12]	2VU3*		0.34	-5.09	-4.53	0.56
CDK2 [28]	1VYZ		0.65	-7.94	-7.70	0.25
CDK2 [29]	2BHE*		0.83	-	-5.25	
ERK2 [30]	2OJG		0.77	-7.69	-7.67	0.03
P38α [31]	1W7H		0.88	-3.94	-6.89	2.95
P38α [29]	1W84		0.58	-6.08	-7.10	1.02
Aurore kinase [32]	2W1D		0.67	-8.24	-5.78	2.46
Proteinkinase B [33]	2UVX		0.29	-5.46	-5.10	0.36
ChK1 [34]	2WMU		0.51	-5.88	-6.67	0.79

**Table 1** (continued)

Target	PDB ID	Starting fragment	RMSD, Å	dG <sub>exp</sub> , kcal/mol	dG <sub>calc</sub> , kcal/mol	ΔdG , kcal/mol
Urokinase [35]	1FV9		0.83	-6.82	-5.44	1.38
Urokinase [36]	1SQO*		0.23	-7.13	-7.63	0.50
Urokinase [37]	2VIN		0.67	-	-7.64	-
PTP1B [38]	1WAX		0.75	-5.55	-7.10	1.55
Coagulation factor Xa [39]	1FJS*		0.21	-5.05	-6.06	1.01
PPAR-1 [40]	3ET0*		0.89	-5.22	-6.62	1.40
Acetylcholinesterase [41]	2CKM*		0.66	-8.49	-7.30	1.19
Acetylcholinesterase [42]	1Q84*		0.71 0.51	- -	-7.99 -6.76	- -
β-Secretase [43, 44]	2OHM		0.58	-4.79	-5.11	0.32

**Table 1** (continued)

Target	PDB ID	Starting fragment	RMSD, Å	dG <sub>exp</sub> , kcal/mol	dG <sub>calc</sub> , kcal/mol	ΔdG , kcal/mol
Dihydroneopterin aldolase [45]	1RRW		0.34	-6.21	-5.95	0.25
Dipeptidylpeptidase [46]	2I78*		1.14	-8.30	-7.00	1.30
Neuraminidase [47]	1F8B*		0.55	-6.15	-6.73	0.58
Glycogenphosphorylase [48]	1EM6*		0.35	-6.69	-7.65	0.96
Phosphodiesterase IV [49]	1Y2B		1.12	-5.76	-4.4	1.36
SRC kinase SH2 domain [50, 51]	1O4P		0.67	-3.55	-7.1	3.57
Tryptase [52]	1A0L*		0.70	-6.35	-6.14	0.22
HSP90 [53]	3EKR*		1.11	-4.09	-5.25	1.16
HSP90 [54]	2QFO		0.36	-6.41	-5.90	0.51
			0.30	-5.22	-5.20	0.02
Prostaglandin D <sub>2</sub> synthase [55]	2VCW		0.69	4.16	-5.59	1.43

- d) the fragments must be relatively stable structures; they must not contain highly active functional groups that would complicate the experimental work;
- e) a drug molecule is broken down into two or more non-overlapping fragments; if such breakdown creates fragments that are too small to be used in molecular docking, only the larger fragments are docked.

The library of fragments for virtual screening was obtained by the application of “the rule of three” [10] to the STK library, donated by Vitas-M Laboratory [18], by using an in-house pybel script [19].

Ligand protonation and 3D-structure optimization were performed with the use of ACD/ChemSketch v 12.0 [17] in the case of mol input files and with the corina [20] for sdf-files.

#### Protein model preparation and molecular docking

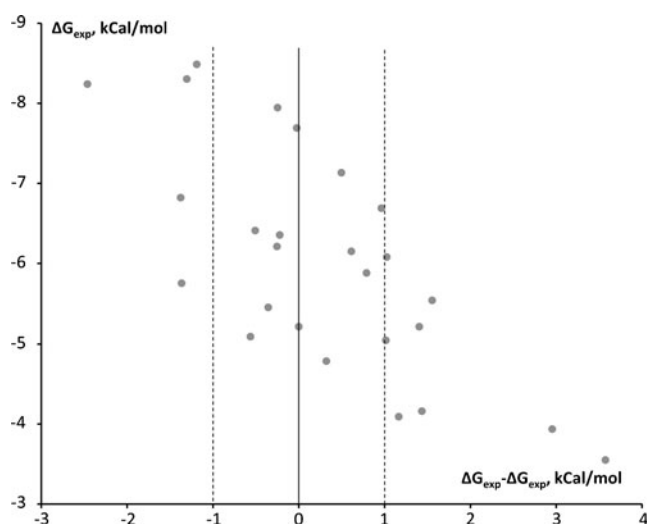
The full-atom models of all proteins used in the current study were prepared from the raw PDB structures by adding hydrogen atoms and assigning ionization states of the amino acids with the Model Builder (*build\_model*) program of the Lead Finder software package v 1.1.13 [21]. Each model was validated by docking its cognate ligand (as extracted from the corresponding PDB structure) and measuring the root mean square deviation (RMSD) of docked ligand pose from its crystallographic position. All of the mentioned models revealed  $\text{RMSD} < 2 \text{ \AA}$  suggesting their applicability for the molecular docking studies.

Docking and virtual screening (VS) of ligands to the prepared models of proteins and binding energy calculations were performed with the Lead Finder v. 1.1.13 software using its default configuration parameters. The energy grid box for ligand docking and virtual screening was set to span 6 Å (the default value) in each direction from the cognate ligand of the corresponding PDB structure. The dG-score produced by Lead Finder was taken as the value of the ligand binding energy. Only the top-ranked poses were used in structural and energy analyses.

#### Structural filtration and rank-ordering of virtual screening results

The structural filtration of the docked complexes obtained in virtual screening experiments was performed by the *structure\_filter* module that is part of the Lead Finder software package v 1.1.13 [21].

The virtual screening results after the structural filtration were ranked by VS/MW (virtual screening score divided by molecular weight) in ascending order, as this brings the list into the closest compliance with Binding Efficiency Index (BEI) [22].



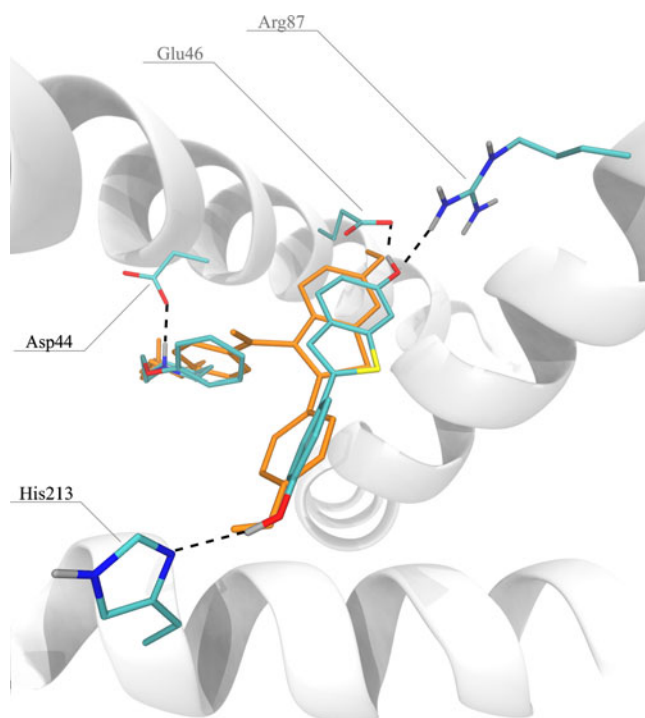
**Fig. 1** The accuracy of binding energy prediction for the fragment inhibitors

#### The clustering of VS results and selection of compounds for experimental verification

We performed the clustering of VS results by a visual analysis in the VMD visualization software [23]. Compounds with specific binding interactions in protein-ligand complexes were placed in the same cluster on the basis of the common scaffold. In the case of CDK2 and PARP1 such specific binding interactions were hydrogen bonds with Leu83 and Gly202 respectively. In many cases these bonds were formed by a heterocyclic moiety of a ligand molecule thus the same heterocycles were placed into one cluster. Non-cyclic groups involved in this bonding were classified according to their chemical nature (e.g., amides, hydroxamates). For experimental verification, compounds from each cluster were selected by an expert.

**Table 2** The influence of different properties of protein-ligand complexes on the success rate of fragment docking

Property	Value	Number of structures	Correctly docked structures, %
Hydrogen bonds	0	15	13
	1	11	64
	>1	42	83
Coordination of metals	–	64	64
	+	4	75
Fraction of occupied binding site, %	15	7	14
	25	30	63
	50	29	75
	75	2	100
Total	–	68	65

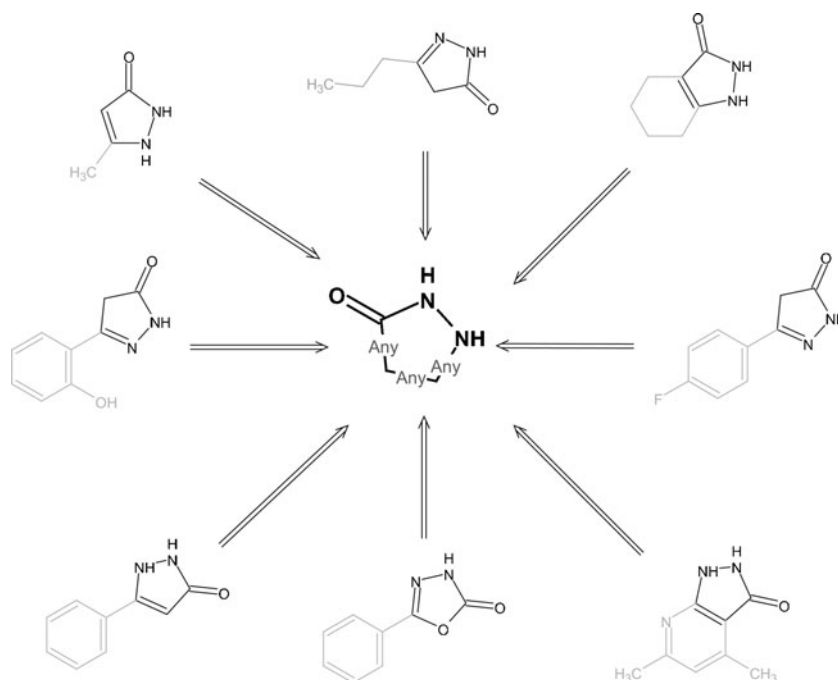


**Fig. 2** Binding of raloxifene (XRay structure, orange) and its fragments (as predicted by docking) with estrogen receptor, PDB ID 2QXS

#### In vitro study of selected potential inhibitors

CDK2 inhibition measurements were carried out by Caliper Discovery Alliances and Services [24]. The inhibition was tested on the human recombinant enzyme in the phosphor-

**Fig. 3** An example of clustering of the virtual screening hits. "Any" denotes a single or double bond

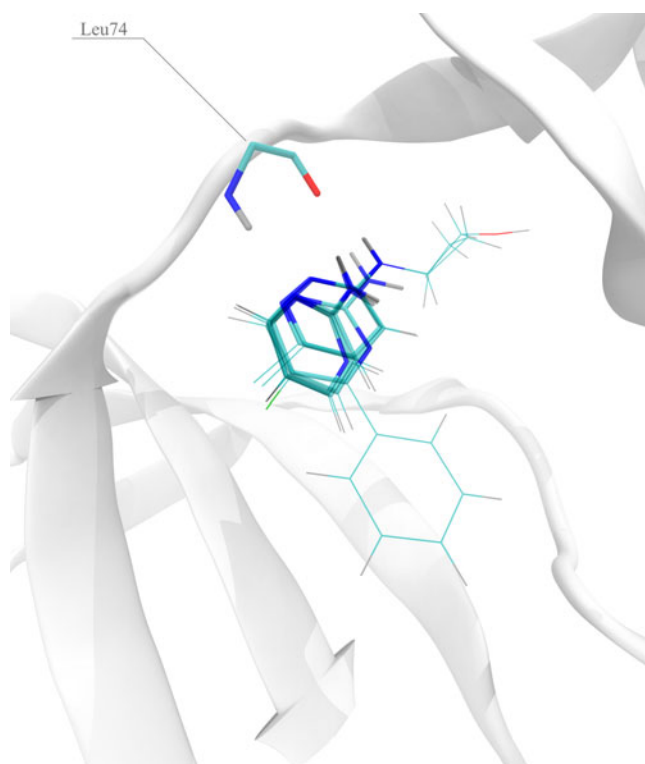


ylation reaction between ATP (35  $\mu\text{M}$ ) and fluorescein-labeled peptide (1.5  $\mu\text{M}$ ). Enzyme activity was estimated as an intensity of the fluorescence of phosphorylated peptide after electrophoretic separation of the reaction mixture. Staurosporin was used as a reference inhibitor ( $\text{IC}_{50}$  6.7 nM).

Human recombinant PARP1 enzyme was purified as described in [25]. The PARP1 enzyme inhibiting activities of the compounds were assayed as follows. Samples (15  $\mu\text{L}$ ) containing 200 nM purified PARP1 protein, 2 OU/mL DNase I-activated calf thymus DNA, 600  $\mu\text{M}$  NAD<sup>+</sup>, 0.4  $\mu\text{Ci}$  [3H]NAD<sup>+</sup>, 10% DMSO, and various concentrations of test compounds were incubated in sample buffer [50 mM Tris, pH 8.0, 20 mM MgCl<sub>2</sub>, 150 mM NaCl, 7 mM  $\beta$ -mercaptoethanol] at 37  $^{\circ}\text{C}$  for 1 min. Under these conditions, the reaction rate was linear up to 20 min. The reaction was stopped by dropping the 12  $\mu\text{L}$  aliquote on the Whatman 1 paper filters soaked in 5% trichloroacetic acid. Filters were washed three times with 150 mL of 5% trichloroacetic acid, TCA was removed by 90% ethanol and then filters were dried. [3H]ADP-ribose incorporation into the acid insoluble material was quantified using a scintillation counter QuantaSmart in the toluene scintillator.  $\text{IC}_{50}$  values were calculated using Origin Pro 8.0 software by a nonlinear regression analysis.

CDK2 binding assay of Reaction Biology is a widely used one, that's why we did not try validating this assay on reference fragment inhibitors. At the same time less common PARP inhibition assay needed a validation on some known fragment inhibitors. We have chosen two known PARP inhibitors: 4-aminobenzamide ( $\text{IC}_{50}$  9  $\mu\text{M}$  [26]) and quinazoline-2,4-dione ( $\text{IC}_{50}$  8.1  $\mu\text{M}$  at 200  $\mu\text{M}$





**Fig. 4** The predicted 3D structure of fragment inhibitors bound to CDK2. The fragment inhibitors are based on the same scaffold

NAD<sup>+</sup> [66]). Experimentally obtained IC<sub>50</sub> values were 17 and 49 μM (600 μM NAD<sup>+</sup>) respectively, which were in good agreement with the literature data.

## Results and discussion

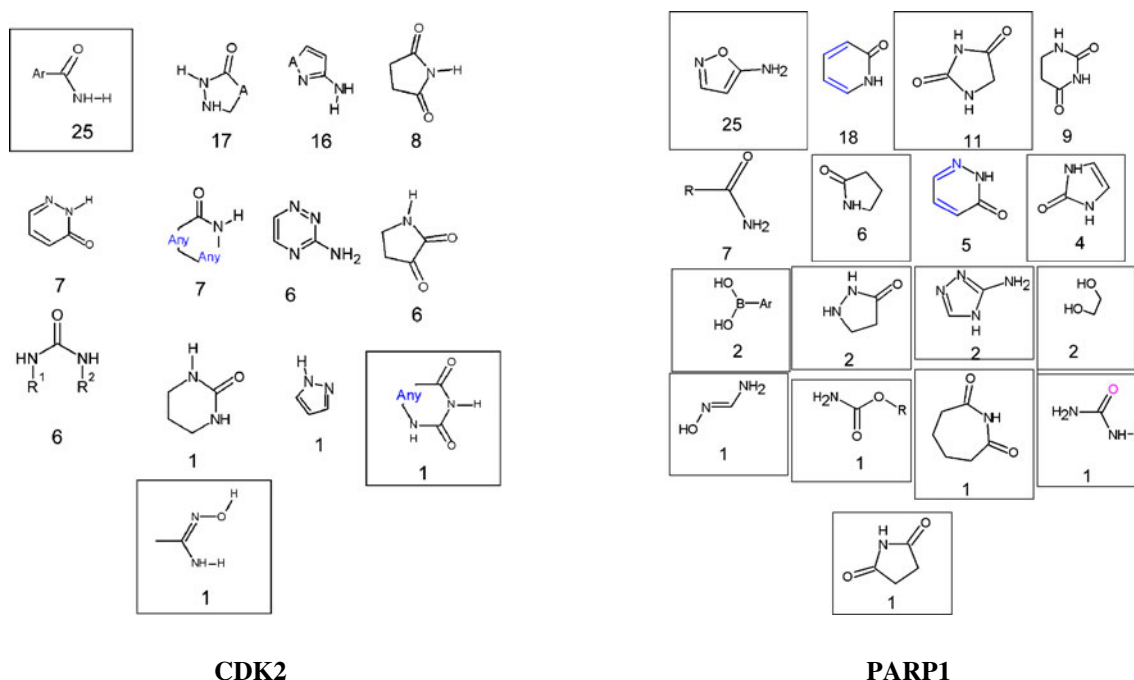
### Docking of fragment inhibitors with LeadFinder

In order to verify the applicability of the molecular docking method to fragment docking, we applied this method to the fragment hits with experimentally observed 3D structures (Table 1). The average RMSD value for the crystallographic protein-ligand structures for such hits did not exceed 1.5 Å.

The mean square error of the binding energy prediction was 1.35 kcal mol<sup>-1</sup>, which is less than one order of magnitude in the K<sub>d</sub> units. In approximately half of the cases the absolute error did not exceed 1 kcal mol<sup>-1</sup>, and only in three cases it was more than 2 kcal mol<sup>-1</sup> (Fig. 1). Therefore, the accuracy of LeadFinder docking predictions was found to be sufficient for adequate prediction of the pose and binding energy for the fragment ligands.

### The reconstruction of drug molecules by fragment docking

A drug can be identified by the fragment method only if the fragments' binding geometry as individual structures matches their binding geometry in the original drug. No drugs designed purely by the fragment method currently exist [4]. Therefore, we attempted to break down a few known drugs into fragments and dock those fragments in order to check the feasibility of reconstruction of known drugs. In our analysis, we used 39 known



**Fig. 5** The found clusters of fragments binding to CDK2 and PARP1 (the number under a fragment indicates the number of drug molecules containing that fragment). A = C(H)<sub>n</sub>, N(H), O; "Any" denotes a single or double bond. Novel scaffolds are depicted in rectangular boxes



**Table 3** The structural motifs, predicted by the fragment screening method, of known CDK2 and PARP1 inhibitors.  $K_i$  or  $IC_{50}$  are provided for certain compounds. A = C(H)<sub>n</sub>, N(H), O; "Any" denotes a single or double bond

Number of structures in a cluster	Core structure	Examples of inhibitors
<b>CDK2</b>		
16		<p><math>IC_{50}</math> 37 nM (S-isomer) [56]</p> <p>48 nM [57]</p>
6		<p><math>K_i</math> 30 nM [58]</p>
17		<p><math>IC_{50}</math> 1-20 nM [59]</p>
7		<p>26 nM [60]</p>
25		-
6		<p>96 nM [61]</p>

**Table 3** (continued)

Number of structures in a cluster	Core structure	Examples of inhibitors
7		<p>3.3 <math>\mu</math>M [62]</p>
6		<p>As hydrazones:</p> <p>30 nM [63]</p>
8		<p>5.6 nM [57]</p>
1		<p>8.2 <math>\mu</math>M [64]</p>
1		<p>1.6 <math>\mu</math>M [65]</p>
1		-
1		-

**Table 3** (continued)

Number of structures in a cluster	Core structure	Examples of inhibitors
PARP		
18		 0.44 $\mu$ M [66]
9		 8.1 $\mu$ M [67]
7		 22 $\mu$ M [66]
5		 12 $\mu$ M [66]

drugs that were broken down into 68 fragments ([Supporting information](#)).

The predicted docked poses were found to be correct in 65% of the cases (44 of 68) ( $\text{RMSD} < 2 \text{ \AA}$ ). We analyzed how the forming of a hydrogen bond, the coordination of metal ion and the degree of occupation of the binding site by the fragment affect the correctness of molecular docking prediction. We found that the formation of more than one hydrogen bond by the fragment increases the correctness of docking predictions to 83% (35 cases out of 42). The absence of hydrogen bond formation reduces the correctness to 13% (two cases out of 15). Also, the larger occupancy of the binding site by the fragment and the coordination with metal ion increase the correctness of docking predictions (Table 2).

The increase of fragment docking accuracy when a fragment participates in directed non-covalent interactions, such as hydrogen bonds and the coordination with metals, can be explained by the fact that these interactions contribute significantly to the overall energy of binding. Furthermore,

directed interactions are significant for the orientation of a fragment structure since the latter has only a few atoms that are capable of forming such interactions.

When a fragment structure occupies a small part of the binding site, several possible docked poses with close binding energies may exist. This appears to be the likely reason for the low accuracy rate of docking predictions for fragments with low degree of occupancy of the binding site. Taken together these binding characteristics can be evaluated for the desired protein with known 3D structure and yield an estimate of the success of developing fragment inhibitors for the corresponding target.

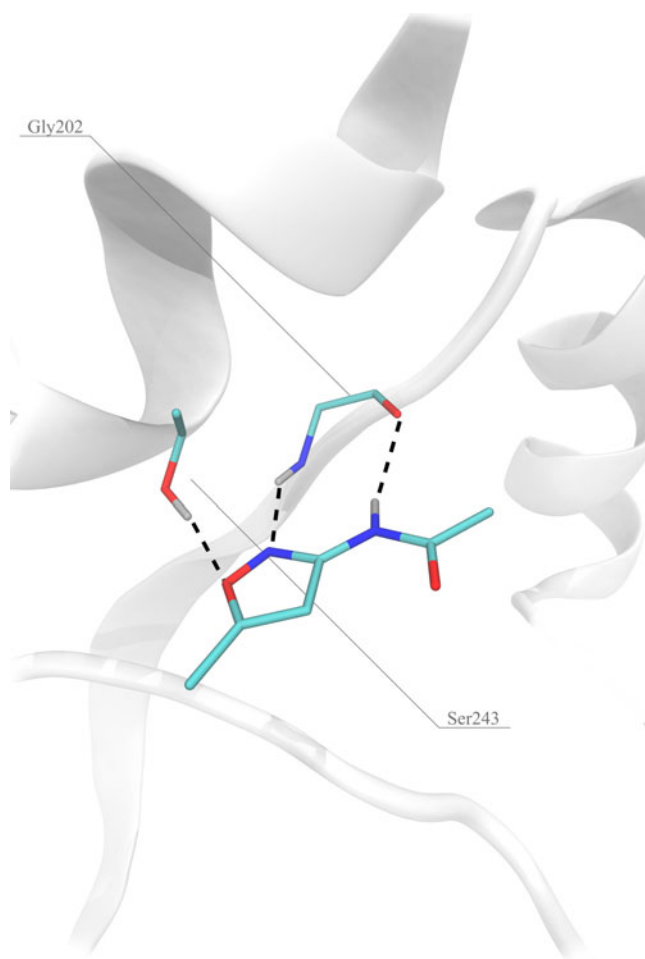
The results presented in (Fig. 1) suggest a reasonable correlation between the experimental data and the predicted binding of drug fragments, obtained via molecular docking, when oriented non-covalent interactions are formed, or when the drug fragment occupies a large part of the binding site. In such cases the full reconstruction of drug molecules from fragments appears to be possible (Fig. 2).

#### The virtual screening method with clustering

Having demonstrated that molecular docking can be successfully applied to predict fragment binding, we attempted to develop a virtual fragment screening method.

**Table 4** The structural motifs, predicted by the fragment screening method, of known CDK2 fragment inhibitors.  $K_i$  or  $\text{IC}_{50}$  are provided for certain compounds. A = C(H)<sub>n</sub>, N(H), O; "Any" denotes a single or double bond

Number of structures in a cluster	Core structure	Examples of known fragments
6		 $\text{IC}_{50}$ 1 mM [12]
16		 $\text{IC}_{50}$ 1.5 $\mu$ M [26]
1		 $\text{IC}_{50}$ 0.185 mM [12]



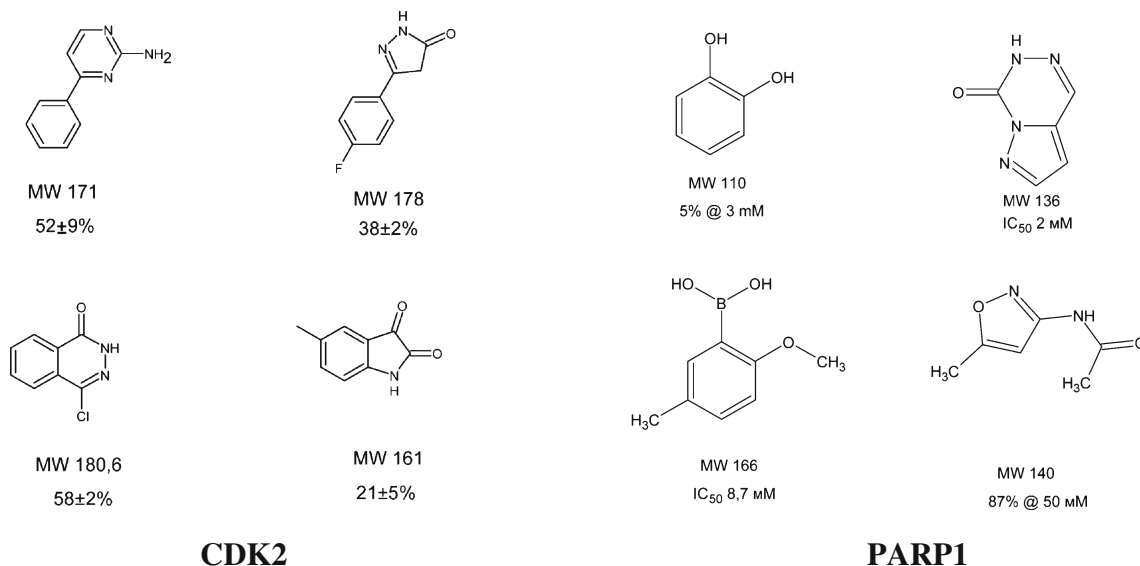
**Fig. 6** The predicted 3D structure of the new fragment inhibitor of PARP1 (87% inhibition at 50 mM). Receptor model was based on PDB ID 1UK0

However, we encountered some difficulties in the development of an in silico FBDD method.

One of the difficulties was that the existing scoring functions do not typically do a good job in distinguishing active fragments from the inactive ones. One possible method to substantially increase efficiency of virtual screening is the structural filtration approach [27] that is based on the selection of docked protein-ligand complexes by the presence of interactions that are characteristic for inhibitors/binders of the given protein structure, such as hydrogen bonds and metal interactions. For example, in all known 3D complexes of CDK2 and PARP-1 with their inhibitors a hydrogen bond between a ligand and main-chain of residues Leu83 and Gly202, correspondingly, is formed. Thus it is reasonable to expect that formation of the mentioned hydrogen bonds is a prerequisite for a ligand to be a true inhibitor of these enzymes. Therefore, only those ligands that formed hydrogen bonds with Leu83 and Gly202, were selected by structural filtration procedure as potential inhibitors of CDK2 and PARP1 correspondingly.

The virtual screening method does not allow to reliably separate binding energies of structurally similar fragments. As a general rule, larger structures have larger VS score values. The use of VS/MW characteristic to rank-order ligands makes it possible to identify small ligands that are more promising for further optimization. Also, the VS/MW characteristic is closely related to the Binding Efficiency Index (BEI) that is widely used in rank-ordering fragment ligands in experimental studies [22].

Since our experimental library of compounds had a large number of similar structures, in many cases different only by a substituent, we obtained a number of similar-looking hits after the rank-ordering and structural filtration operations. Their VS



**Fig. 7** The results of experimental studies of binding of the CDK2 inhibitors (the molar mass and the inhibition rate are provided at 0.5 mM concentration of the inhibitor) and the PARP1 inhibitors (IC<sub>50</sub> or the inhibition rate at a selected concentration, see Supporting information)

scores do not provide a good basis for further selection, since their binding is similar and their VS score differences are not material. At this point, further selection is possible through clustering of similar structures. The core of a cluster is a small structural element with few degrees of freedom, forming specific interactions with the protein target — for example, in the case of CDK2 there are two specific hydrogen bonds with Leu83, Fig. 4. Each structural filtration operation requires such specific interactions with the protein target. An example set of substances forming a cluster is presented in Fig. 3. This example suggests that all substances belonging to the cluster bind similarly to the target (Fig. 4). Therefore, virtually any compound from a cluster can serve as a basis for further optimization. This consideration can be used to prioritize hits once their clustered by their 3D pattern of binding. For example it is reasonable to drain clusters from compounds with toxicophoric groups, compounds that seem to be hardly functionalized during further optimization, or simply fragments with relatively high molecular weight. Finally if there is a known list of off-target proteins only compounds with affordable predicted selectivity can remain. Also, the clustering facilitates the identification of novel inhibitors outside of the patent protection space. It appears that the clusters with cores that are not present in the patented inhibitors may be the ones to attract the greatest interest of practical researchers.

In summary, our method of virtual fragment screening requires the following steps to be taken (the size of our starting library was about 600,000 substances):

1. The generation of the fragment library: the application of the rule of three to the library of synthetically available compounds (in our case, it brought about 20,000 compounds).
2. Virtual screening of the fragment library.
3. The structural filtration of the docked fragments (in our case, it resulted in the selection of about 1000 compounds)
4. Rank-ordering by VS/MW (we obtained about 100 compounds).
5. The clustering of the best-suited fragment hits (~20 clusters).

Validation of the fragment screening method on the CDK2 and PARP1 targets

As a result of the application of the rule of three to our VitasM library, we selected 20,152 fragment compounds for virtual screening.

Based on the analysis of binding of known inhibitors, we selected two hydrogen bonds, with L83 for CDK2 and G202 for PARP1, as structural filters.

Rank-ordering by VS/MW of the compounds that passed the structural filtration stage, selection of 100 best compounds and their clustering resulted in 13 clusters for CDK2 and 18 clusters for PARP1 (Fig. 5).

In order to verify the correctness of our method, we evaluated the cores of found clusters to see how often those cores are present in the known CDK2 or PARP1 inhibitors. It turned out that ten out of the 13 and four of the 18 cores we found were present in the known CDK2 and PARP1 inhibitors, respectively, and some cores are present in the known fragment inhibitors of CDK2 (Tables 3 and 4).

In order to verify binding of the found fragments we selected one candidate molecule from each cluster of the potential fragment CDK2 and PARP1 inhibitors and experimentally measured binding for the most promising molecules. All four available fragment CDK2 inhibitors and three out of four available PARP1 inhibitors were found to be binding in the millimolar range (Fig. 6, 7). Notably, two fragment PARP1 inhibitors are based on new cores (namely, arylboronic acid and 3-aminoisoxazole core) that have not been characterized earlier. They appear to be promising candidates for further optimization.

## Conclusions

We have shown that the current docking approaches could be quite accurate in predicting poses and binding energies of fragment inhibitors. Based on these observations we have designed a virtual fragment screening technique that in addition to docking involves structural filtration of docked ligand poses by specific interactions present in protein-ligand complexes, and clustering of the selected fragment hits. The application of this method allowed us to identify ten out of 14 known scaffolds of CDK2 inhibitors and two PARP1 inhibitors with new scaffolds that were experimentally proven to be active binders. We believe that this method can facilitate the earlier stages of FBDD. We see the full automation of the fragment clustering stage and the development of growth and linking algorithms as natural extensions of the proposed method that will enable the effective application of this technique in the fragment-based search of potent inhibitors.

**Acknowledgments** We thank Sukhanova M. V. for PARP1 isolation and Val Kulkov (BioMolTech Corp) for revising the manuscript.

## References

1. Sams-Dodd F (2005) Target-based drug discovery: next term is something wrong? *Drug Disc Today* 10:139–147
2. Chessari G, Woodhead AJ (2009) From fragment to clinical candidate: a historical perspective. *Drug Discov Today* 14:668–675
3. Hubbard RE (2008) Fragment approaches in structure-based drug discovery. *J Synchrotron Radiat* 15:227–230
4. Warr WA (2009) Fragment-based drug discovery. *J Computer-Aided Mod Des* 23:453–458

5. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Del Rev* 32:3–25
6. Hajduk PJ, Greer JA (2007) Decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Disc* 6(3):211–219
7. Huber W (2005) A new strategy for improved secondary screening and lead optimization using high-resolution SPR characterization of compound-target interaction. *J Mol Recognition* 18:273–281
8. Hartshorn MJ, Murray CW, Cleasby A, Frederickson M, Tickle IJ, Jhoti H (2005) Fragment-based lead discovery using X-ray crystallography. *J Med Chem* 48:403–413
9. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996) Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* 274:1531–1534
10. Congreve M, Carr R, Murray C, Jhoti H (2003) A ‘Rule of Three’ for fragment-based drug discovery? *Drug Discov Today* 8:876–877
11. Warren GL et al (2006) A Critical Assessment of Docking Programs and Scoring Functions. *J Med Chem* 49:5912–5931
12. Wyatt PG, Woodhead AJ, Berdini V, Boulstridge JA, Carr MG, Cross DM, Davis DJ, Devine LA, Early TR, Feltell RE et al (2008) Identification of N-(4-piperidinyl)-4-(2, 6-dichlorobenzoylamino)-1H-pyrazole-3-carboxamide (AT7519), a novel cyclin dependent kinase inhibitor using fragment-based X-ray crystallography and structure based drug design. *J Med Chem* 51:4986–4999
13. Lapenna S, Giordano A (2009) Cell cycle kinases as therapeutic agents for cancer. *Nature Reviews Drug Discovery* 8:547–566 <http://www.clinicaltrials.gov/ct2/show/NCT00540358>
14. NIH (2011) A Phase 2 Trial of Standard Chemotherapy, With or Without BSI-201, in Patients With Triple Negative Metastatic Breast Cancer. <http://www.clinicaltrials.gov/ct2/show/NCT00679783>
15. NIH (2011) Phase II Study of AZD2281 in Patients With Known BRCA Mutation Status or Recurrent High Grade Ovarian Cancer or Patients With Known BRCA Mutation Status/ Triple Neg Breast Cancer
16. EMBL-EBI. Index of approved drugs and nutraceuticals in the PDB [http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/drugport/GetPage.pl?template=drugindex\\_pdb.html](http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/drugport/GetPage.pl?template=drugindex_pdb.html)
17. Advanced Chemistry Development (2011) <http://acdlabs.com/home>
18. Vitas-M Laboratory Ltd. <http://vitasmlab.com>
19. Noel M, O’Boyle CM, Geoffrey RH (2008) Pybel: a Python wrapper for the OpenBabel cheminformatics toolkit. *Chem Central J* 2:5
20. Life Science Center. <http://www.csc.fi/english/research/software/corina>
21. Stroganov OV, Novikov FN, Stroylov VS, Kulkov V, Chilov GG (2008) Lead Finder: An Approach To Improve Accuracy of Protein-Ligand Docking, Binding Energy Estimation, and Virtual Screening. *J Chem Inf Model* 48:2371–2385
22. Abad-Zapatero C, Metz JT (2005) Ligand efficiency indices as guideposts for drug discovery. *Drug Discov Today* 10:464–469
23. Humphrey W, Dalke A, Schulten K (1996) VMD - Visual Molecular Dynamics *J Mol Graph* 14:33–38. <http://www.ks.uiuc.edu/Research/vmd>
24. Caliper Life Sciences <http://www.caliperls.com/products/contract-research/in-vitro/kinases/cdk2cyclina-h.htm>
25. Sukhanova MV, Khodyreva SN, Lavrik OI (2004) Poly(ADP-ribose) polymerase-1 inhibits strand-displacement synthesis of DNA catalyzed by DNA polymerase beta. *Biochemistry (Mosc)* 69:558–568
26. Wen-Ting Z et al (2009) Design, Synthesis, and Cytoprotective Effect of 2-Aminothiazole Analogues as Potent Poly(ADP-Ribose) Polymerase-1 Inhibitors. *J Med Chem* 52:718–725
27. Novikov FN, Stroylov VS, Stroganov OV, Chilov GG (2009) Improving performance of docking-based virtual screening by structural filtration. *J Mol Model* 16, published online
28. Pevarello P et al (2004) 3- Aminopyrazole Inhibitors of CDK2/ Cyclin A as Antitumor Agents. 1. Lead Finding. *J Med Chem* 47:3367–3380
29. Congreve MS, Davis DJ, Devine L, Granata C, O’Reilly M, Wyatt PG, Jhoti H (2003) Detection of Ligands from a Dynamic Combinatorial Library by X-ray Crystallography. *Angew Chem Int Ed* 42:4479–4482
30. Aronov AM et al (2007) Flipped Out: Structure-Guided Design of Selective Pyrazolopyrrole ERK Inhibitors. *J Med Chem* 50:1280–1287
31. Gill AL et al (2005) Identification of Novel p38 MAP Kinase Inhibitors Using Fragment-Based Lead Generation. *J Med Chem* 48:414–426
32. Howard S, Berdini V, Boulstridge JA, Carr MG, Cross DM, Curry J, Devine LA, Early TR, Fazal L, Gill AL et al (2009) Fragment-based discovery of the pyrazol-4-yl urea (AT9283), a multi-targeted kinase inhibitor with potent aurora kinase activity. *J Med Chem* 52:379–388
33. Caldwell JJ et al (2008) Identification of 4-(4-aminopiperidin-1-yl)-7Hpyrrolo[ 2,3-d]pyrimidines as selective inhibitors of protein kinase B through fragment elaboration. *J Med Chem* 51:2147–2157
34. Reader J et al (2008) Identification and structure-guided optimization of novel inhibitors of checkpoint kinase 1 (Chk1) through combined biochemical and crystallographic screening. *AACR Annual Meeting April 12–16, 2008, San Diego, CA*
35. Hajduk PJ, Boyd S, Nettesheim D, Nienaber V, Severin J, Smith R, Davidson D, Rockway T, Fesik SW (2000) Identification of Novel Inhibitors of Urokinase via NMR-Based Screening. *J Med Chem* 43:3862–3866
36. Wendt MD et al (2004) Identification of novel binding interactions in the development of potent, selective 2-naphthamide inhibitors of urokinase: synthesis, structural analysis, and SAR of N-phenyl amide 6-substitution. *J Med Chem* 47:303–324
37. Frederickson M et al (2008) Fragment-based discovery of mexiletine derivatives as orally bioavailable inhibitors of urokinase-type plasminogen activator. *J Med Chem* 51:183–186
38. Hartshorn MJ, Murray CW, Cleasby A, Frederickson M, Tickle IJ, Jhoti H (2007) Fragment-Based Lead Discovery Using X-ray Crystallography. *J Med Chem* 48:403–413
39. Agnelli G, Haas S, Ginsberg JS, Krueger KA, Dmitrienko A, Brandt JT (2007) A phase II study of the oral factor Xa inhibitor LY517717 for the prevention of venous thromboembolism after hip or knee replacement. *J Thromb Haemost* 5:746–753
40. Artis DR et al (2009) Scaffold-based discovery of indeglitazar, a PPAR pan-active anti-diabetic agent. *Proc Natl Acad Sci USA* 106:262–267
41. Pang YP, Quiram P, Jelacic T, Hong F, Brimijoin S (1996) Highly potent, selective, and low cost bis-tetrahydroaminacrine inhibitors of acetylcholinesterase Steps toward novel drugs for treating Alzheimer’s disease. *J Biol Chem* 271:23646–23649
42. Lewis WG, Green LG, Grynspan F, Radi Z, Carlier PR, Taylor P, Finn MG, Sharpless KB (2002) Click chemistry in situ: acetylcholinesterase as a reaction vessel for the selective assembly of a femtomolar inhibitor from an array of building blocks. *Angew Chem Int Edn Eng* 41:1053–1057
43. Murray CW, Callaghan O, Chessari G, Cleasby A, Congreve M, Frederickson M, Hartshorn MJ, McMenamin R, Patel S, Wallis N (2007) Application of Fragment Screening by X-ray Crystallography to b-Secretase. *J Med Chem* 50:1116–1123
44. Congreve M, Aharony D, Albert J, Callaghan O, Campbell J, Carr RA, Chessari G, Cowan S, Edwards PD, Frederickson M, McMenamin R, Murray CW, Patel S, Wallis N (2007) Application



- of Fragment Screening by X-ray Crystallography to the Discovery of Aminopyridines as Inhibitors of  $\beta$ -Secretase. *J Med Chem* 50:1124–1132
45. Sanders WJ et al (2004) Discovery of Potent Inhibitors of Dihydroneopterin Aldolase using CrystaLEAD High-Throughput X-ray Crystallographic Screening and Structure-Directed Lead Optimization. *J Med Chem* 47:1709–1718
46. Pei Z et al (2006) Discovery of ((4R,5S)-5-Amino-4-(2,4,5-trifluorophenyl)-cyclohex-1-enyl)-(3-(trifluoromethyl)-5,6-dihydro-[1,2,4]triazolo[4,3-a]pyrazin-7(8H)-yl)methanone (ABT-341), a Highly Potent, Selective, Orally Efficacious, and Safe Dipeptidyl Peptidase IV Inhibitor for the Treatment of Type 2 Diabetes. *J Med Chem* 49:6439–6442
47. Hochgürtel M, Kroth H, Piecha D, Hofmann MW, Nicolau C, Krause S, Schaaf O, Sonnenmoser G, Eliseev AV (2002) Target-induced formation of neuraminidase inhibitors from in vitro virtual combinatorial libraries. *Proc Natl Acad Sci USA* 99:3382–3387
48. Rath VL, Ammirati M, Danley DE, Ekstrom JL, Gibbs EM, Hynes TR, Mathiowetz AM, McPherson RK, Olson TV, Treadway JL, Hoover DJ (2000) Human liver glycogen phosphorylase inhibitors bind at a new allosteric site. *Chem Biol* 7:677–682
49. Card GL et al (2005) A family of phosphodiesterase inhibitors discovered by cocrystallography and scaffold-based drug design. *Nat Biotechnol* 23:201–207
50. Lesuisse D, Lange G, Deprez P, Bénard D, Schoot B, Delettre G, Marquette J-P, Broto P, Jean-Baptiste V, Bichet P, Sarubbi E, Mandine E (2002) SAR and X-ray. A new approach combining fragment-based screening and rational drug design: application to the discovery of nanomolar inhibitors of Src SH2. *J Med Chem* 45:2379–2387
51. Lange G, Lesuisse D, Deprez P, Schoot B, Loenze P, Bénard D, Marquette JP, Broto P, Sarubbi E, Mandine E (2003) Requirements for specific binding of low affinity inhibitor fragments to the SH2 domain of (pp 60)Src are identical to those for high affinity binding to full length inhibitors. *J Med Chem* 46:5184–5195
52. Burgess LE, Newhouse BJ, Ibrahim P, Rizzi J, Kashem MA, Hartman A, Brandhuber BJ, Wright CD, Thomson DS, Vigers GP, Koch K (1999) Potent and selective nonpeptidic inhibitors of human lung tryptase. *Proc Natl Acad Sci USA* 96:8348–8352
53. Dymock BW et al (2005) Novel, potent small-molecule inhibitors of the molecular chaperone Hsp90 discovered through structure-based design. *J Med Chem* 48:4212–4215
54. Huth JR et al (2007) Discovery and design of novel HSP90 inhibitors using multiple fragment-based design strategies. *Chem Biol Drug Des* 70:1–12
55. Hohwy M et al (2008) Novel prostaglandin D synthase inhibitors generated by fragment-based drug design. *J Med Chem* 51:2178–2186
56. Pevarello P et al (2005) 3-Aminopyrazole Inhibitors of CDK2/Cyclin A as Antitumor Agents. 2. Lead Optimization. *J Med Chem* 48:2944–2956
57. Kim KS et al (2002) Discovery of Aminothiazole Inhibitors of Cyclin-Dependent Kinase 2: Synthesis, X-ray Crystallographic Analysis, and Biological Activities. *J Med Chem* 45:3905–3927
58. The Binding Database. <http://www.bindingdb.org>
59. Green J, Arnost MJ, Pierce A (2003) Patent WO 2003011287 A1 2003
60. Hoelder S et al (2004) Pyridazinone derivatives as cdk2-inhibitors. Patent WO 2004046130
61. Ikuta M et al (2001) Crystallographic Approach to Identification of Cyclin-dependent Kinase 4 (CDK4)-specific Inhibitors by Using CDK4 Mimic CDK2 Protein. *J Biol Chem* 276:27548–27554
62. James ET et al (2002) Photochemical preparation of a pyridone containing tetracycle: A jak protein kinase inhibitor. *Bioorg Med Chem Let* 12:1219–1223
63. Congreve MS, Davis DJ, Devine L, Granata C, O'Reilly M, Wyatt PG, Jhoti H (2003) Detection of Ligands from a Dynamic Combinatorial Library by X-ray Crystallography. *Angew Chem Int Edn* 42:4479–4482
64. Schlappach A et al (2008) Pyrrolo-pyrimidones: A novel class of MK2 inhibitors with potent cellular activity. *Bioorg Med Chem Let* 18:6142–6146
65. Furet P, Meyer T, Strauss A, Raccuglia S, Rondeau J-M (2002) Structure-based design and protein X-ray analysis of a protein kinase inhibitor. *Bioorg Med Chem Let* 12:221–224
66. Griffin RJ, Srinivasan S, Bowman K, Calvert AH, Curtin NJ, Newell DR, Pemberton LC, Golding BT (1998) Resistance modifying agents. 5. Synthesis and biological properties of quinazolinone inhibitors of the DNA repair enzyme poly(ADPribose) polymerase (PARP). *J Med Chem* 41:5247–5256
67. Banasik M, Komura H, Shimoyama M, Ueda K (1992) Specific inhibitors of poly(ADP-ribose) synthetase and mono(ADP-ribosyl) transferase. *J Biol Chem* 267:1569–1575