J Mol Model (2009) 15:1337–1347
DOI 10.1007/s00894-009-0497-y

ORIGINAL PAPER

Developing novel approaches to improve binding energy estimation and virtual screening: a PARP case study

Fedor N. Novikov · Viktor S. Stroylov · Oleg V. Stroganov · Val Kulkov · Ghermes G. Chilov

Received: 10 February 2009 / Accepted: 5 March 2009 / Published online: 16 April 2009

Abstract Poly-(ADP-ribose)-polymerase (PARP) is a promising anti-cancer target as it plays a crucial role in the cellular reparation and survival mechanisms. However, the development of a robust and cost effective experimental technique to screen PARP inhibitors is still a scientific challenge owing to the difficulties in quantitative detection of the enzyme activity. In this work we demonstrate that the computational chemistry tools including molecular docking and scoring can perform on par with the experimental studies in assessing binding constants and in the recovery of active compounds in virtual screening. Using the recently introduced Lead Finder software we were able to dock a set of 142 well characterized PARP inhibitors and obtain a good correlation between the calculated and experimentally measured binding energies with the *rmsd* of 1.67 kcal mol^{-1}. Additionally, fine-tuning of the energy scaling coefficients within the Lead Finder scoring function has further decreased *rmsd* to the value of 0.88 kcal mol^{-1}. Moreover, we were able to reproduce the selectivity of ligand binding between the two isoforms of the enzyme-PARP1 and PARP2-suggesting that the Lead Finder software can be used to design isoform-selective inhibitors of PARP. An impressive enrichment was obtained in the virtual screening experiment, in which the mentioned set of PARP inhibitors was mixed with a commercial library of 300,000 compounds. We also demonstrate that the virtual screening performance can be significantly improved by an additional structural filtration of the docked ligand poses through detection of the crucial hydrogen bonding interactions with the enzyme.

Keywords Binding energy calculations · PARP1 · PARP2 · Protein-ligand docking · Selectivity modeling · Virtual screening

Electronic supplementary material The online version of this article (doi:10.1007/s00894-009-0497-y) contains supplementary material, which is available to authorized users.

F. N. Novikov · O. V. Stroganov · G. G. Chilov
MolTech Ltd,
119992 Moscow, Leninskie gory, 1/75A, Russia

F. N. Novikov · V. S. Stroylov
Department of Bioengineering and Bioinformatics,
Lomonosov Moscow State University,
119992 Moscow, Leninskie gory, 1/73, Russia

O. V. Stroganov · G. G. Chilov (✉)
Belozersky Institute of Physicochemical Biology,
Lomonosov Moscow State University,
119992 Moscow, Acad. Khokhlov st. 6, Russia
e-mail: Ghermes@moltech.ru

V. Kulkov
BioMolTech Corp,
226 York Mills Rd, Toronto, Ontario M2L 1 L1, Canada

Introduction

Poly-(ADP-ribose) polymerase (PARP) is an enzyme abundantly expressed in the nucleus. It catalyzes poly-(ADP-ribosyl)-ation of various proteins using NAD^+ as an ADP-ribose donor [1]. PARP-catalyzed covalent modification regulates activity of proteins involved in the DNA repair and maintenance of the genome stability, gene transcription, cell proliferation and differentiation, and other processes [2, 3]. Thus, the inhibition of PARP that is intended to suppress reparation and survival mechanisms upon the radio- or chemotherapy-induced cell damage is viewed as a promising strategy in the anti-cancer therapy [4]. Another therapeutic benefit of the PARP inhibition is to rescue a cell from necrosis caused by the depletion of

NAD^+ pool, which takes place during the myocardial infarction and other pathophysiological conditions [5]. Finally, the inhibition of PARP-stimulated transcription of pro-inflammatory genes might have a positive therapeutic effect in the cardiovascular therapy [5].

Once PARP had been identified as an important therapeutic target, a number of HTS techniques to recover its inhibitors were developed. The original methods employed the count of ^{32}P included into the poly-(ADP-ribose) produced by PARP. Those methods were quite laborious, expensive and hazardous [6]. Later, a less hazardous technique employing ^{3}H-labeled NAD^+ was developed [7] and a variety of other isotope-free methods appeared, including: the application of monoclonal antibodies against the poly-(ADP-ribose) [8]; use of biotinylated NAD^+ to produce the biotinylated poly-(ADP-ribose) imaged by biotin [9, 10]; the quantitative determination of the chemically modified NAD^+ by fluorescence [11]; a technique based on the observation that PARP activity retards the growth rate of yeast [12]. However, those techniques were still laborious and expensive compared to other HTS assays, in which a direct detection of the reaction products was possible (e.g., in case of proteases). With this in mind, the development of a robust *in silico* method to model binding of PARP with its inhibitors would benefit the discovery of novel hits and lead compounds targeting PARP.

So far, numerous works on modeling the PARP inhibition by means of molecular docking and QSAR approaches have been published [13, 14]. However, these works have been confined to modeling the structure and the binding energy of PARP-inhibitor complexes, while the ability to distinguish between the true binders and the decoy ligands, which is crucial for practical research, has not been attempted. In addition, the cited works exploited relatively small sets of the known PARP inhibitors, which hampered extrapolation toward the new scaffolds.

In this study we evaluate the performance of the recently introduced Lead Finder docking software [15] in predicting binding energies and poses of the PARP inhibitors. Lead Finder is a novel docking software that combines the genetic algorithm search with local optimization procedures and smart exploitation of the knowledge generated during docking process to perform global minimum search on the multidimensional surface describing free energy of protein-ligand binding. Lead Finder introduces three scoring functions specifically optimized for (a) accurate binding energy predictions; (b) correct energy-ranking of docked ligand poses; and (c) correct rank-ordering of active and inactive compounds in virtual screening experiments.

A set of 142 well characterized compounds with known binding potencies was used for benchmarking. We also evaluate the ability of Lead Finder to distinguish the true PARP binders from the decoy ligands in a virtual screening experiment, in which the mentioned set of compounds was mixed with an actual library of 300,000 ligands. In addition, we demonstrate how the performance of Lead Finder in docking and scoring can be further improved. First, adjusting the energy scaling coefficients of Lead Finder's internal scoring function lowers the root mean square deviation (*rmsd*) of the binding energy prediction from 1.67 kcal mol^{-1} to 0.88 kcal mol^{-1}. Second, the application of an additional structural filter to the docked ligand poses that identifies the formation of conservative hydrogen bonds between the ligand and the enzyme has resulted into a 10-fold increase of the enrichment factor. Both techniques-the adjustment of the energy scaling coefficients and the structural filtration of the docked ligand poses-can be easily exploited by a Lead Finder user. Finally, we evaluate if the current accuracy of the binding energy prediction is good enough to model the selectivity of ligand binding between the two PARP isoforms-PARP1 and PARP2, where the latter structure was developed by means of homology modeling.

Experimental section

PARP inhibitors

A set of experimentally characterized PARP inhibitors was retrieved from the recent literature sources to evaluate the efficiency of the PARP inhibition modeling. In total, 142 structures, comprising 11 different scaffolds, were selected from the original publications [13, 14, 16–19] (Fig. 1). The physicochemical characteristics (cLogP, charge, number of hydrogen bond donors and acceptors) of the selected PARP inhibitors are provided in the Supplementary Material. For the current molecular modeling purposes, the 3D-coordinates of inhibitors were obtained with ACD/ChemSketch Freeware 9.0. The ligand protonation states were manually checked and corrected by an expert. The final set of PARP inhibitors with 3D-coordinates and their protonation states used in this study is available for download [20].

Preparation of protein structure models

The full-atom models of PARP1 were prepared from the raw PDB structures by adding hydrogen atoms and assigning ionization states of the amino acids using the Model Builder (*build_model.exe)* module of the Lead Finder software package. The optimization of amino acid side chains (provided in the Results and discussion section) was performed using the covalent docking option of Lead Finder.

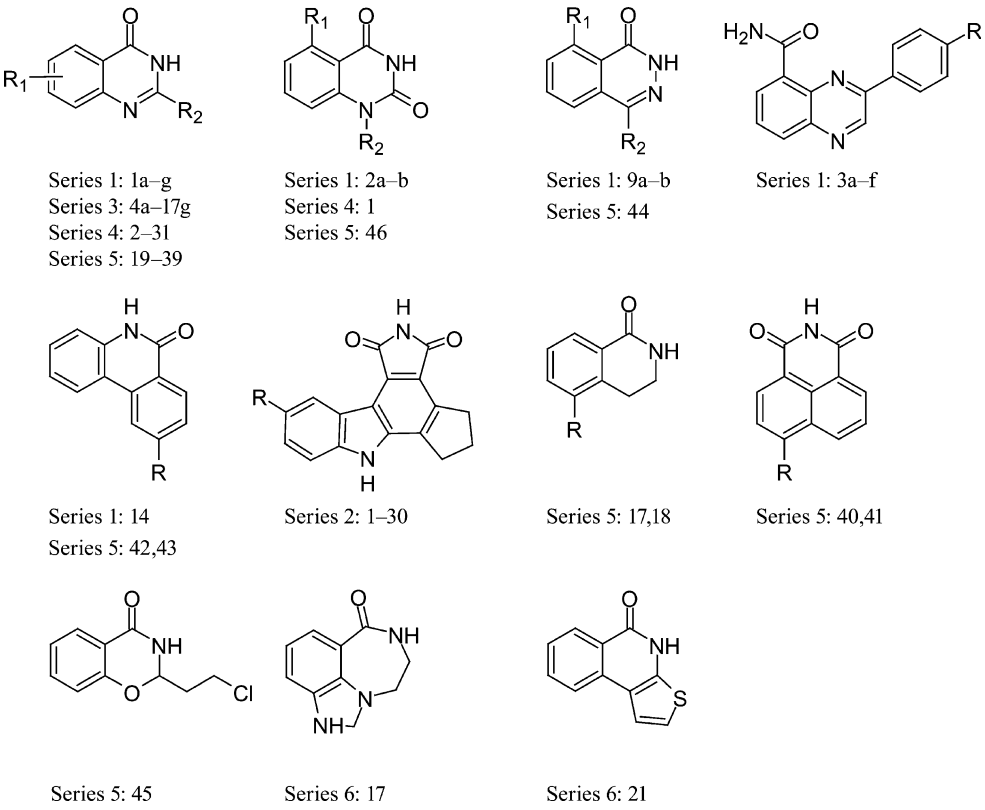

Fig. 1 PARP inhibitors used in the current modeling study. Ligand numeration corresponds to the original works, except for some ligands that were dropped off due to the lack of robustly measure binding constants (see the Supplementary Material for details)

The full-atom model of PARP2 isoform was built by homology modeling with PARP1 using the default options of the Modeller software. The addition of hydrogen atoms and the optimization of amino acid side chains for PARP2 model were performed in the same way as described above for PARP1.

Docking and binding energy calculations

Ligand docking and binding energy calculations were performed with the Lead Finder v. 1.1.10 software under its default configuration parameters. The reference ligand for mapping the inhibitor binding site was taken from the PDB structure 1efy. The size of the grid box for ligand docking was set to extend 6 Å (the default value) in each direction from the reference ligand. The dG-score produced by Lead Finder was taken as the predicted value of the ligand binding energy. Only the top-ranked poses were used for structural and energy analyses.

To evaluate the accuracy of energy calculations, the experimental values of binding energy were derived from the IC50 values cited in the original publications. For this purpose, IC50 values were first converted into the inhibition constants (K_i) according to the formula of competitive inhibition:

$$K_i = IC50 \big/ \left(1 + [NAD^+] \big/ K_m^{NAD}\right) \quad (1)$$

where $[NAD^+]$ is the concentration of NAD^+ used in a particular experiment, and K_m^{NAD} is its Michaelis constant (50 μM according to the literature data [21]). The values of $[NAD^+]$ used in particular experimental works were carefully extracted from the corresponding literature sources (as specified in the Supplementary Material). The derived K_i values were further transformed into the free energies of binding according to the well-known formula:

$$\Delta G = RT\ln(K_i) \quad (2)$$

The resulting experimental free energies were then compared with the dG-score values predicted by Lead Finder.

Virtual screening

A set of 300,000 ligands comprising the STK library of Vitas-M Laboratory [22] was used as a decoy ligand set to benchmark the performance of Lead Finder in virtual screening. Virtual screening consisted of docking 142 known inhibitors and 300,000 decoy ligands into the PARP active site in the screening regime of Lead Finder and rank-ordering ligands according to their VS-score values. For the quantitative characterization of virtual screening efficiency, the true positive rate (TP) plot and the enrichment factor (EF) plot were used, as suggested in the recent methodological study [23]. The TP plot depicts the percentage of the recovered active compounds (the true positive rate) as a function of the percentage of the recovered decoy ligands (the false positive rate). The EF at a certain stage of

screening equals the ratio of the achieved TP to the TP obtained by random ligand pick. The EF plot depicts EF as a function of TP. The commonly used discreet indicators of EF are EF20, EF40 and EF70, which denote the enrichment factors at 20%, 40% and 70% of recovered active ligands correspondingly.

Structural filtration of docked ligand poses

For each top-scoring ligand pose obtained in virtual screening the formation of two particular hydrogen bonds, one with the main chain carbonyl and one with the amide groups of residue G769, was checked. This was done automatically using the *structure_filter.exe* module of Lead Finder. This module takes a protein structure file and a list of protein residues for which it checks the formation of hydrogen bonds with potential inhibitors from an *sdf* output file produced by Lead Finder. The formation of a hydrogen bond was deemed to occur when the distance between the hydrogen bond donor and acceptor was found to be less than 3.5 Å and the angle was greater than 150°. For the practical purpose of virtual screening, the application of an additional structural filter resulted in the following: the list of ligands ranked by their VS-score was reorganized by moving the ligands that did not form the hydrogen bonds of interest to the end of the list.

Results and discussion

Preparation of PARP1 structure model

It is well known that the preparation of an adequate full-atom protein structure model is one of the key success factors in modeling protein-ligand interactions. Thus, the development of a high quality model of PARP1 structure was the first objective of our study. A number of PARP1 structures, co-crystallized with different inhibitors, were available from the PDB, namely: 3 structures of the human enzyme and 7 structures of the chicken one. Since all human structures of PARP1 were almost identical, especially in the active site region, we picked only one of them, 1uk1, for conversion to a full-atom model for the reason that it contained the largest bound inhibitor. We also decided to consider an additional PARP1 model prepared from the chicken structure, 1efy, because the active site residues of both enzymes were practically identical while the crystallographic resolution of 1efy (2.2 Å) was notably better than that of 1uk1 (3.0 Å).

In the next step, we wanted to evaluate both protein structure models for their predictive strength in order to select the best model. We decided to evaluate models by docking the set of 142 inhibitors with known binding potencies to both PARP1 models and counting ligands that could be correctly accommodated in the enzyme active site. Since no crystallographic data were available for a direct comparison of docked poses to the crystallographic ligand poses, we had to define what would constitute a *correct position* of a ligand. First, we noticed that all known PARP1 inhibitors had a common structural feature-a pair of adjacent hydrogen bond donor and acceptor atoms that form correlated hydrogen bonds with the main chain carbonyl and amide groups of G769, a conserved residue among PARP isoforms and PARPs from different sources. So, the formation of those hydrogen bonds (Fig. 2) was the first indicator of a correct ligand placement. The second criterion was an appropriate placement of ligand's tail, which is usually attached to the pharmacophore group of nanomolar binders, in a special gorge in the PARP active site (Fig. 2) that can be seen in the available PDB structures. Those were the two structural criteria we used to determine if a particular ligand was docked correctly to a particular model of PARP1. Each ligand was docked independently 20 times in order to determine the probability of finding a correct ligand pose.

Docking to the 1uk1 model revealed that 21 out of 142 ligands were docked incorrectly with the probability of 1.0 (e.g. in each of the 20 independent docking attempts). At the same time, 26 ligands were incorrectly docked with the probability of 1.0 to the 1efy. It appeared that almost all Series 2 inhibitors (Fig. 1) were docked incorrectly to the 1uk1 model. An analysis of the incorrectly docked ligands revealed that those inhibitors were probably too large to accommodate their pharmacophore group in the PARP1 active site even under the soft Lennard-Jones potential implemented in the Lead Finder program. On the contrary, the ligand's pharmacophore group was placed almost

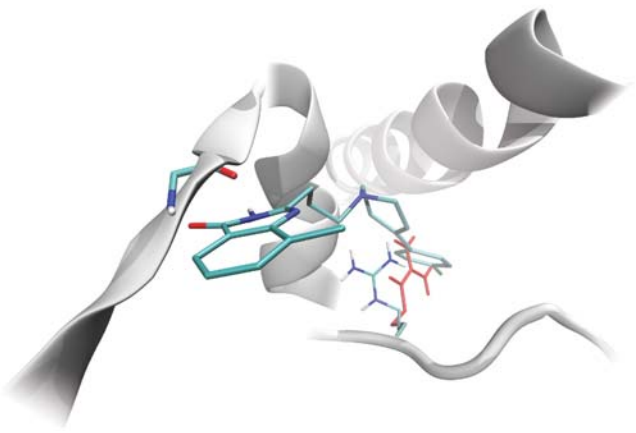

Fig. 2 Structural fingerprints of inhibitor binding by PARP: the correlated hydrogen bonds with conserved G769 residue and the placement of ligand's tail in the gorge of the binding site. Positions of R878 side chain characteristic for 1uk1 (colored by atom name) and 1efy (red) structures of PARP1 are outlined

always correctly in the 1efy model, while incorrect positioning of ligand's tail was the prevailing reason of docking failures with 1efy. Incorrectly docked ligands oriented their tails toward the solvent medium instead of penetrating the gorge of the active site, indicating there were some steric hindrances. A superposition of 1uk1 and 1efy structures revealed one notable difference regarding the organization of the active site gorge in those two models, namely the position of the residue R878. In 1efy model, that residue was slightly shifted and that reduced the effective volume of the gorge and its capacity to accommodate ligands with longer tails (Fig. 2).

We decided to test if we could improve the 1efy model by realigning the mentioned residue. For that purpose, we performed covalent docking of the R878 side chain to find its alternative energy-optimal positions. It appeared that the new top-scoring position of R878 was much closer to that of the 1uk1 model. Encouraged by this result, we performed docking of the full ligand set on the updated 1efy model. This time the results were much better (Table 1). Only 2 ligands, 2-24 and 5–43, were docked incorrectly in all 20 attempts, while 121 ligands (85% of all ligands) were docked correctly with the probability of 0.5 or greater. A closer examination of the incorrectly docked ligands revealed that ligand 2–24 was too large (it was probably the largest of all studied ligands) and it could not accommodate its rigid pyrrolocarbazole pharmacophore in the active site. Ligand 5–43 was incorrectly docked due to a scoring function error of the Lead Finder program, which overestimated the hydrogen bonding energy with the ligand's nitro group and erroneously placed it in contact with G769.

Unfortunately, we could not identify a reasonably simple way of improving the quality of the 1uk1 model. It was evident that some extra space in the pharmacophore binding sub-site could improve docking. Unlike in the 1efy case, we could not find any particular residues that we could move and therefore acquire the needed space. Probably a more complex procedure like molecular dynamics optimizations could relax the 1uk1 structure in a desired way, but we decided to stop further experiments since we had already identified a high quality PARP1 structure model.

Binding energy estimations

Free energies of ligand binding provided by the dG-scoring function of Lead Finder were compared with the experimental data extracted from the literature. In order to improve reliability of such comparisons, the value of predicted binding energy was calculated as an average dG-score of correctly docked poses in 20 independent docking attempts for each ligand. For ligands 2–24 and 5–43 that were incorrectly docked in all 20 attempts, the lowest dG-score value was taken. Our objectives in conducting such analysis were: a) to compare binding energies of the correctly docked ligand poses, and b) to perform the comparison in a statistically meaningful way. The results of comparison (Table 1) indicate that Lead Finder demonstrated impressive accuracy in the binding energy prediction with *rmsd* of 1.67 kcal mol^{-1} on the set of 142 inhibitors.

Interestingly, a recent study where 330 diverse protein-ligand complexes were used to benchmark the binding energy prediction by Lead Finder[15] returned an *rmsd* value of 1.50 kcal mol^{-1} that is fairly close to the results of the current study. This suggests that Lead Finder can consistently predict dG values regardless of the nature of a particular protein or a ligand.

Adjustment of the scoring function for binding energy calculations

Without regard to the fact that the accuracy of binding energy prediction with the standard dG-scoring function was already fairly convincing, we were able to demonstrate that the prediction accuracy could be further improved for a particular set of ligands. In other words, Lead Finder's dG scoring function can be easily customized by end users to obtain more accurate estimations of the binding energy for a particular set of ligands. The Lead Finder dG scoring function represents a linear combination of different energy terms:

$$\Delta G = \Sigma k_i E_i \quad (3)$$

where E_i stand for different types of interactions (e.g., van der Waals, Coulomb, hydrogen bonding, etc.) and k_i-are the corresponding scaling coefficients. Therefore, one can alter the scaling coefficients to modify the target function in a desired way. From a practical point of view, the customization of a scoring function is easy. The energy components (E_i) provided by Lead Finder for a set of ligands of interest are exported into a table where one can adjust the scaling coefficients to fine-tune the target function. The procedure can be automated or performed manually in spreadsheet processing software.

In the current work we tried to customize the dG scoring function to improve the binding energy prediction for the PARP enzyme. We did that manually in MS Excel, avoiding automation for two reasons: we wanted to have full control over the parameters during the optimization stage and we wanted to demonstrate that such customization can be easily done by an end user.

Once the scaling coefficients were adjusted by using the entire set of 142 ligands as a training set, the *rmsd* of the predicted binding energy dropped from 1.67 kcal mol^{-1} to

Table 1 Experimental and calculated binding energies (in kcal mol^{-1}) of PARP1 inhibitors. Ligands, which were docked correctly with probability <0.5 are underlined; ligands misdocked in all 20 trials are typed in bold

Ligand	ΔG_{exp}	ΔG_{calc}	Ligand	ΔG_{exp}	ΔG_{calc}	Ligand	ΔG_{exp}	ΔG_{calc}
1–14	−10.2	−10.8	3–11a	−8.4	−9.1	4–23	−11.1	−9.2
1-1a	−11.0	−8.1	*3–12a*	−7.4	−6.7	4–24	−11.4	−11.7
1-1b	−11.0	−10.2	3–13a	−9.1	−9.9	4–25	−11.6	−8.0
1-1c	−12.2	−10.3	3–14a-1	−9.4	−10.2	4–26	−8.7	−9.5
1-1d	−11.3	−11.1	3–14a-2	−9.4	−8.0	4–27	−9.8	−9.0
1-1e	−11.2	−8.3	3–15a	−9.9	−6.2	4–28	−10.9	−8.4
1-1f	−10.3	−9.7	3–16d-1	−11.6	−8.6	4–29	−11.8	−8.2
1-1 g	−11.4	−9.6	3–16e-1	−10.7	−9.5	*4–3*	−10.3	−11.5
1–2a	−10.0	−9.6	3–16e-2	−9.5	−9.8	4–30	−10.7	−7.5
1–2b	−10.4	−10.6	3–16f	−10.8	−9.9	*4–31*	−11.4	−12.0
1–3a	−9.9	−9.2	3–16g	−10.6	−10.3	4-4	−11.2	−8.6
1–3b	−10.8	−11.3	3–17c	−10.9	−12.1	4–5	−11.2	−8.8
1–3c	−10.1	−9.5	3–17e	−10.9	−10.0	4–6	−10.9	−8.4
1–3d	−10.0	−10.0	3–17g	−11.4	−12.4	*4–7*	−9.5	−7.0
1–3e	−10.3	−9.7	3–4a	−11.2	−10.9	4–8	−10.9	−8.1
1–3f	−10.2	−10.4	3–4b	−11.5	−11.8	4–9	−10.7	−9.2
1–9a	−10.0	−10.5	3–4c	−11.0	−11.0	5–17	−9.4	−7.8
1–9b	−10.5	−11.0	3–4d	−10.9	−9.6	*5–18*	−10.5	−11.1
2-1	−10.5	−8.8	3–4e	−11.3	−8.7	5–19	−8.1	−7.2
2–10	−10.4	−10.1	3–4f	−11.6	−9.6	5–20	−9.8	−7.5
2–11	−10.5	−10.5	3–4g	−11.2	−9.9	5–21	−7.9	−9.8
2–12	−10.2	−10.2	3–5a	−10.6	−8.6	5–22	−8.2	−9.7
2–13	−10.2	−11.3	3–5d	−10.1	−9.1	5–23	−9.8	−8.9
2–14	−10.3	−9.9	3–5e	−10.5	−10.0	5–24	−9.0	−6.8
2–15	−10.3	−10.6	3–6a-1	−11.2	−7.7	5–25	−8.0	−9.1
2–16	−10.8	−10.8	3–6a-2	−11.2	−8.0	5–26	−10.1	−11.1
2–17	−10.9	−11.4	3–6b	−11.1	−9.1	5–27	−9.0	−8.3
2–18	−10.8	−11.0	3–6d-1	−11.2	−8.9	5–29	−9.7	−9.9
2–19	−10.7	−10.6	3–6d-2	−9.5	−9.9	5–30	−8.7	−8.6
2-2	−10.9	−9.4	3–6e	−10.7	−8.9	5–31	−9.9	−9.9
2–20	−10.3	−9.9	3–6f	−10.9	−9.6	5–32	−8.5	−5.9
2–21	−10.2	−9.7	3–9a-1	−9.8	−9.8	5–33	−9.4	−6.3
2–22	−10.9	−10.5	3–9a-2	−9.8	−10.4	5–34	−7.9	−7.4
2–23	−10.2	−9.2	4-1	−10.4	−11.6	5–35	−8.9	−7.3
2–24	−10.3	−7.7	4–10	−11.3	−11.6	5–36	−9.8	−7.5
2–25	−10.8	−9.8	4–11	−10.3	−9.2	5–37	−9.6	−7.6
2–26	−10.4	−10.9	4–12	−10.9	−9.9	5–38	−9.8	−7.6
2–27	−10.8	−12.0	4–13	−10.6	−9.7	5–39	−9.3	−7.5
2–28	−10.6	−11.2	4–14	−10.3	−9.6	5–40	−9.1	−7.6
2–29	−10.2	−11.4	4–16	−11.4	−9.0	5–41	−10.3	−7.7
2–3	−10.9	−9.5	*4–17*	−10.3	−6.7	5–42	−10.0	−7.4
2–30	−10.1	−8.0	4–18	−8.6	−8.0	**5–43**	−9.9	−5.9
2–4	−11.1	−11.7	4–19	−11.5	−11.7	5–44	−7.8	−8.1
2–5	−10.4	-11.7	4-2	−8.6	−8.5	5–45	−8.0	−7.3
2–6	−10.0	−9.3	4–20	−11.0	−10.4	5–46	−8.0	−7.6
2–7	−10.4	−7.2	4–21	−10.7	−8.5	6–17	−9.3	−7.0
2–8	−10.8	−9.1	4–22	−11.0	−8.9	6–21	−8.8	−7.9
2–9	−10.2	−10.6						

0.88 kcal mol^{-1}. As can be seen from Table 2, only 2 types of energy contributions were notably changed-the hydrogen bond energy and the so called Born energy associated with the electrostatic component of ligand desolvation. As for the hydrogen bond energy, all corresponding contributions, namely the constant term equalizing ligand's hydrogen bonds in the solution and in the protein-bound states; the penalty term on polar atoms that do not form hydrogen bonds; the energies of the hydrogen bonds formed in polar and non-polar environments; the extra energy of the correlated hydrogen bonds-were increased. Such transformation of the scaling coefficients have a clear physical sense-the hydrogen bond interactions are crucial for specific binding with PARP, and the optimization of the dG scoring function confirmed that assumption quantitatively. As for the Born energy, many PARP1 inhibitors were positively charged and therefore an additional adjustment of this term made clear sense. However, it should be stressed that a significant change of a particular coefficient may not have a significant influence on the resulting binding energy, since energy contributions can differ by orders of magnitude. For example, a several-fold strengthening of the hydrogen bonding contribution was offset by a minor change in the Van der Waals coefficient.

Speaking of the methodology of scoring function customization, current results bring evidence that fine-tuning of the energy scaling coefficients for the particular protein or ligand set is closely related to the physical characteristics of the modeled system. This observation suggests that an automatic customization must somehow account for the linkage between scaling coefficients and their physical sense and take into account the natural limits for the variation of scaling coefficients. These issues will be a likely subject of our future studies devoted to the customization of Lead Finder's scoring functions.

Modeling selectivity of PARP1 and PARP2 inhibition

Finally, we investigated if the currently available accuracy of binding energy calculation was adequate to model the selectivity of inhibitor binding between the two PARP isoforms-PARP1 and PARP2 [24]. Despite the fact that PARP1 is a predominant form of the enzyme, it has been shown recently that targeting PARP2 might have a distinct therapeutic potential [3, 16]. The major issue we faced regarding the selectivity modeling was the absence of an experimentally defined structure of the human PARP2. However, the inhibitor binding interface of PARP2 had only 2 amino acid substitutions V762I and L769G (residue numbers correspond to the PARP1 sequence) compared to PARP1. We decided to probe if a side-chain optimization of those residues could produce a good model of PARP2 and if we could attain selectivity by using our docking and scoring approach. To evaluate performance of selectivity modeling, experimental data from a recent study devoted to the synthesis of PARP2 inhibitors were used[16]. The

Table 2 Relative change (expressed in the folds of magnitude) of the energy scaling coefficients obtained upon the dG-scoring function customization

Coefficient	Description[a]	Relative change
k_{VdW}	Van der Waals interactions	0.75
k_{sol}	Nonpolar solvation energy (volume-based term)	0.85
$k_{polar\text{-}P}$	Surface of polar ligand's atoms contact with protein	1
$k_{polar\text{-}S}$	Surface of polar ligand's atoms contact with the solvent	1
$k_{nonpolar\text{-}P}$	Surface of nonpolar ligand's atoms contact with protein	1
$k_{nonpolar\text{-}S}$	Surface of nonpolar ligand's atoms contact with the solvent	1
$k_{elec,0}$	Electrostatic energy of interactions of buried charges	1
$k_{elec,1}$	Electrostatic energy of interactions of intermediately buried charges	1
$k_{elec,2}$	Electrostatic energy of interactions of surface charges	1
k_{lig_born}	Born energy of ligand's charges transfer from the solvent to protein	4
$k_{HB,\ lig\text{-}pen}$	Penalty on ligand's polar atoms that do not form H-bonds with protein	2
$k_{HB,\ prot\text{-}pen}$	Penalty on protein's polar atoms that do not form H-bonds with ligand	2.5
$k_{HB\text{-}polar}$	Energy of H-bonds in polar environment	3
$k_{HB\text{-}nonpolar}$	Energy of H-bonds in nonpolar environment	3
$k_{HB,\ corr}$	Energy of correlated H-bonds	5.5
k_{metal}	Energy of ligand interaction with metal ions	1
k_{nb}	Difference in the ligand's non-bonded energy upon binding	1
k_{1-4}	Difference in the ligand's energy of 1–4 interactions upon binding	1
$k_{dihedral}$	Energy of dihedral angles	1
k_{tors}	Energy of entropic losses associated with ligand's rotatable bonds	1

[a] definitions of the energy terms can be found in Ref. [15]

experimental data included inhibition constants of 20 compounds (series 1 on Fig. 1). Approximately half of them were selective for PARP1 and another half-for PARP2 (Table 3). The cited work is the only public source of the PARP2 inhibition constants at present and its data are especially valuable for the selectivity benchmarks since all measurements were performed in similar conditions.

First, we attempted to prepare a PARP2 model by simply deleting the side chain of L769 and attaching a methyl group to V762, reproducing mutations L769G and V762I correspondingly. No energy optimization of the mutated or neighboring residues was performed. Then, inhibitors were docked to the resulting model and we observed no correlation between the experimental and calculated selectivity, despite the fact that the overall agreement with the PARP2 binding energies was satisfactory (*rmsd* 0.90 kcal mol^{-1}). It became obvious that selectivity modeling was more sensitive to the structural features of the protein. We decided to further refine the PARP2 model by optimizing positions of the mutated residues and their immediate neighborhood. For practical reasons, we used Lead Finder's covalent docking functionality to optimize side-chains, thereby avoiding more complex procedures such as molecular dynamics computations. First, no side chain optimization was necessary for the L769G and V762I mutations (in the latter case the residue was tightly packed within the protein interior, so it had no additional degrees of freedom). However, the introduction of isoleucine instead of valine led to a steric clash with residues D766 and Y889, which were hydrogen-bonded to each other in the PARP1 structure. After optimization, D766 residue formed a salt bridge with R878, while Y889 formed a hydrogen bond with the main chain carbonyl of G892 (Fig. 3). Obviously, the introduced steric constraints disrupted the hydrogen bonding interaction between D766 and Y889, however those residues were able to compensate for that loss. Interestingly, D766 could not form a salt bridge with R878 in PARP1, because L769 sterically shielded that interaction; however, when leucine was replaced by glycine

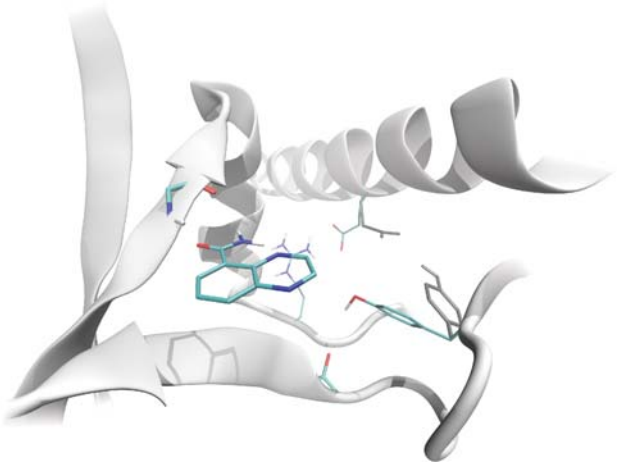

Fig. 3 The modeled binding site of PARP2. Residues D766 and Y889 (gray-before optimization, colored by name-after optimization) are outlined

Table 3 Experimental and calculated data on the selectivity of PARP1 and PARP2 inhibition by the compounds of Series 1

Ligand	Experiment			Calculation		
	$K_{i(PARP1)}$, nM	$K_{i(PARP2)}$, nM	Selectivity[a]	$K_{i(PARP1)}$, nM	$K_{i(PARP2)}$, nM	Selectivity
1–14	41	60.5	−0.17	5.1	95.3	−1.27
1-1a	10.5	437.8	−1.62	38.3	344.6	−0.95
1-1b	11.5	439.2	−1.58	7.6	710.7	−1.97
1-1c	1.5	62.6	−1.62	23.8	14.5	0.21
1-1d	6.5	360	−1.74	14.7	41.7	−0.45
1-1e	8	120.2	−1.18	45.8	150.5	−0.52
1-1f	34	453.6	−1.13	4.1	37.8	−0.96
1-1 g	5.5	216	−1.59	102.9	n/a[b]	n/a
1–2a	60	50.4	0.08	59.1	347.2	−0.77
1–2b	30	59.8	−0.30	9.0	10.2	−0.05
1–3a	65.5	10.1	0.81	313.0	28.7	1.04
1–3b	16.5	5.0	0.52	42.4	8.5	0.70
1–3c	50.5	5.8	0.94	59.9	12.0	0.70
1–3d	59	7.9	0.87	86.3	15.2	0.75
1–3e	35.5	5.8	0.79	80.9	22.4	0.56
1–3f	43.5	6.5	0.83	28.5	28.7	0.00
1–9a	60	64.8	−0.03	25.2	3.9	0.81
1–9b	24.5	60.5	−0.39	3.6	8.6	−0.38

[a] selectivity is expressed as the $\log_{10}$ of the ratio of binding constants for PARP1 and PARP2

[b] ligand was misdocked and excluded from the selectivity evaluation

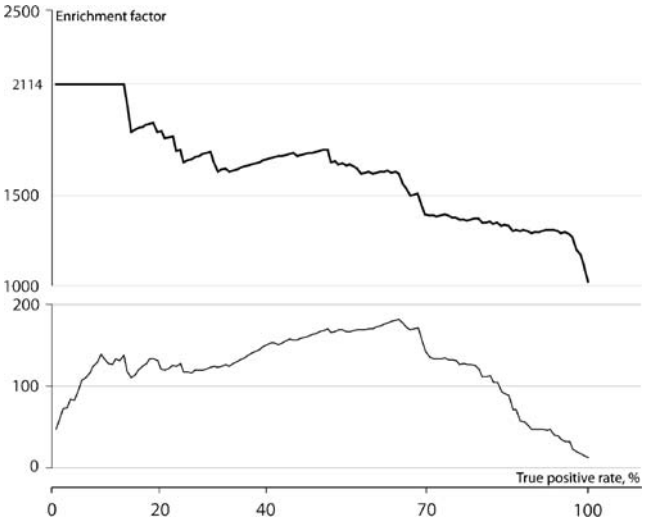

Fig. 4 Enrichment factor plot of the virtual screening for PARP1. Light gray curves correspond to the default virtual screening performance; dark gray curves describe the performance after additional structural filtration of the docked ligand poses

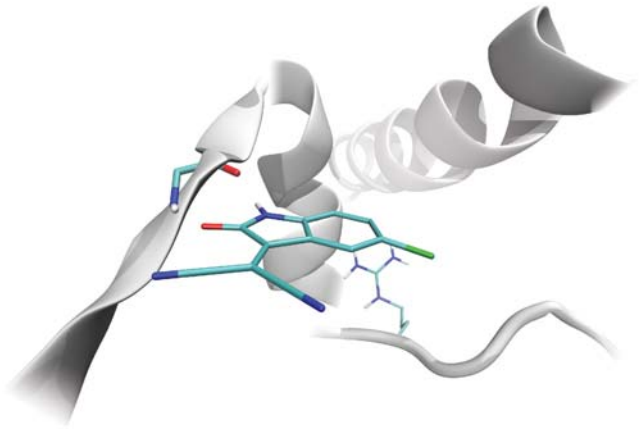

Fig. 6 An example of novel potential PARP1 inhibitor recovered by virtual screening. The structure is disclosed with a permission from the Vitas-M Laboratory

in PARP2, such contact became feasible. Using the refined model of PARP2 we obtained a reasonable correlation of the selectivity (Table 3) that could be explained at the molecular level. The movement of D766 in PARP2 changed the shape of the active site gorge, which allowed more comfortable accommodation of ligands with a large rigid pharmacophore. At the same time, ligands with a long flexible tail fitted into PARP1 structure better, probably because D766 did not shield R878 that capped the end of the gorge.

Virtual screening performance

To further characterize Lead Finder's capacity to model the PARP1 inhibition we set up a virtual screening experiment where the already mentioned set of 142 inhibitors was mixed with a library of 300,000 compounds, and the resulting set of ligands was screened against PARP1 and rank-ordered by their predicted VS-scores. The resulting enrichment curves depicting the enrichment factor as a function of the percentage of recovered inhibitors, and the true positive rate as a function of the percentage of screened decoys are presented in Figs. 4 and 5. The enrichment curves start with a rapid rise attributed to the correct placement of true PARP1 inhibitors in the top of the rank-ordered library. A stable high enrichment is observed for 90% of the inhibitors recovered; after that, the enrichment gradually declines since the residual fraction of the ligands was not docked correctly and therefore their VS-score values were no better than those of the decoys. Despite the fact that a minor fraction of the ligands was not correctly docked, the commonly cited indicators of virtual screening were impressive: EF20=120, EF40=150, EF70=140, TP0.5=70, TP2=90, and TP5=98. We would like to point out that a significant part of the set of 142 PARP1 inhibitors were micromolar binders (Table 1). It is well known that many benchmarking studies rely on the use of nanomolar

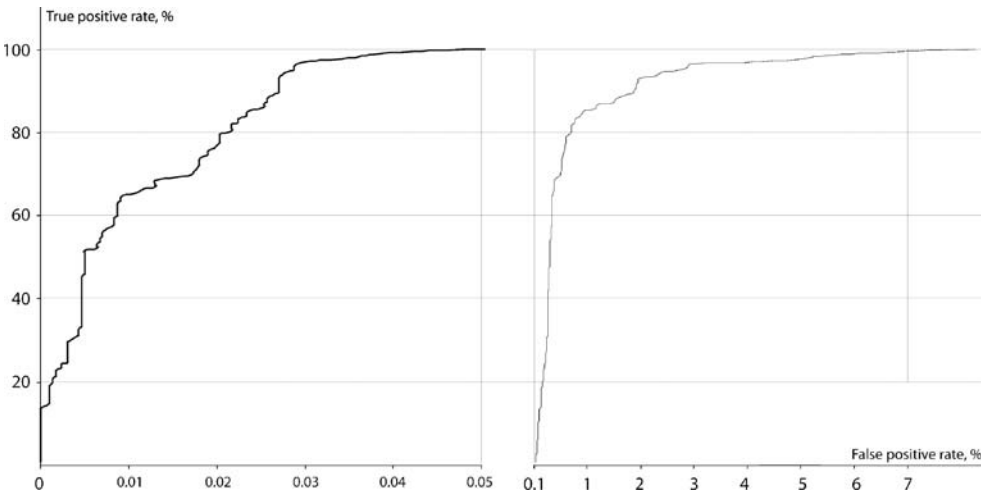

Fig. 5 True positive rate plot of the virtual screening for PARP1. Light gray curves correspond to the default virtual screening performance; dark gray curves describe the performance after additional structural filtration of the docked ligand poses. Only the informative part of the plot where the true positive rate is less than 100% is depicted

binders that present a less significant challenge for docking software. Therefore, the used set of inhibitors was closer to the real-life ligand sets, and Lead Finder successfully withstood that challenge.

Structural filtration improves the virtual screening performance

We also hypothesized that the virtual screening efficiency could be further improved if we exploited the knowledge about the specific interactions formed by PARP1 with its binders. As our docking studies revealed, all of the 142 inhibitors we considered were capable of forming 2 correlated hydrogen bonds with residue G769 (except for the two ligands that could not be correctly docked). Obviously, we assumed that such hydrogen bonding to G769 was mandatory for strong binding with PARP, and we searched through docked poses of the entire library of ligands to filter out compounds that did not conform to such structural criterion. The structural filtration revealed that only 1526 out of 300,000 ligands satisfied the criterion. From a practical viewpoint, this means that the enrichment factor could be increased by a factor of 197 through the reduction of library size by structural filtration of the docked ligand poses. An additional enrichment was attained by ranking the selected ligands according to their VS-score. The resulting enrichment plot (Figs. 4 and 5) illustrates the impact of structural filtration on the virtual screening performance. The updated indicators of the enrichment after structural filtration were: EF20=2000, EF40=1800, EF70=1400, TP0.5=100, TP2=100, and TP5=100.

We argue that the structural filtration of docked ligand poses has a clear advantage over the ligand-based or 3D-QSAR approaches of improving the virtual screening performance. Obviously, ligand-based searches are limited to the predefined chemical scaffolds, raising concerns about the discovery of novel compounds. Moreover, structural similarity of a ligand pharmacophore does not guarantee its proper accommodation in the protein binding site since its shape and flexibility do matter. In order to demonstrate how docking-based methods can retrieve novel chemical scaffolds, we are providing an illustration of a sample ligand we found in a virtual screening experiment-a ligand that nicely fills the active site cavity and forms crucial hydrogen bonds with the G769 (Fig. 6). The 3D-QSAR approaches may also start from examining the docked ligand poses to infer particular structural features favoring binding, however the obtained correlations may have little or no physical sense. In our view, the refinement of virtual screening results (e.g., structural filtration) must originate from physically meaningful structural features of the protein binding site. In the case of PARP, a displacement of the solvent (from G769) by the ligand was obviously favorable: the ligand could form 2 hydrogen bonds without any entropic losses, while only tight packing of the water molecules could render the same interactions, and probably the enthalpic losses were also inevitable in that case. We tentatively hypothesize that such observations might be viewed as a general route for creating structural filters and improving virtual screening performance for arbitrary proteins.

It is worth mentioning that we used different approaches to improve the binding energy prediction and the virtual screening performance. Since binding energy evaluation was meaningful only for true PARP1 binders, a slight refinement of Lead Finder's dG scoring function (i.e., the adjustment of energy scaling coefficients) was enough. At the same time, improvement of the virtual screening performance required a more radical approach that was not directly connected with the VS scoring function. Clearly, these results point to the imperfections of the scoring functions, especially in evaluating the affinity of non-binders. However, we have demonstrated that the scoring function customization and the structural filtration of docked ligand poses can dramatically improve docking and virtual screening results and be easily implemented by Lead Finder users.

Conclusions

In summary, we have demonstrated that the currently available docking and scoring approaches implemented in the Lead Finder software are capable of modeling PARP inhibition with the degree of accuracy approaching the experiment. Moreover, we have shown how the basic performance of binding energy estimations and virtual screening can be improved by the scoring function customization and the structural filtration of docked ligand poses, which can be easily implemented by Lead Finder users. We also hope that the obtained results will aid the experimental discovery of novel classes of PARP1 ligands, as well as development of PARP1 or PARP2 isoform-selective inhibitors.

Acknowledgments Our special thanks to Prof. Viktor Gergel and Alexander Grishagin, the Center of Supercomputer Technologies of the N.I. Lobacevsky State University of Nizhni Novgorod, for our access to the high-performance computing cluster (Nizhni Novgorod segment of the SKIF-grid program).

References

1. D'Amours D, Desnoyers S, D'Silva I, Poirier GG (1999) Poly (ADP-ribosyl) ation reactions in the regulation of nuclear functions. Biochem J 342:249–268
2. Hassa PO, Haenni SS, Elser M, Hottiger MO (2006) Nuclear ADP-ribosylation reactions in mammalian cells: where are we today and where are we going? Microbiol Mol Biol Rev. doi:10.1128/MMBR.00040-05

3. Hassa PO, Hottiger MO (2008) The diverse biological roles of mammalian PARPs, a small but powerful family of poly-ADP-ribose polymerases. Front Biosci. doi:10.2741/2909
4. Jagtap P, Szabo C (2005) Poly(ADP-ribose) polymerase and the therapeutic effects of ITS inhibitors. Nat Rev Drug Discov. doi:10.1038/nrd1718
5. Pacher P, Szabo C (2007) Role of Poly(ADP-ribose) polymerase 1 (PARP-1) in cardiovascular diseases: the therapeutic potential of PARP inhibitors. Cardiovasc Drug Rev. doi:10.1111/j.1527-3466.2007.00018.x
6. Niedergang C, Okazaki H, Mandel P (1979) Properties of purified calf thymus poly(adenosine diphosphate ribose) polymerase. Comparison of the DNA-independent and the DNA-dependent enzyme. Eur J Biochem 102:43–57
7. Dillon KJ, Smith GC Martin NM (2003) A flashplate assay for the identification of PARP-1 inhibitors. J Biomol Screen. doi:10.1177/1087057103008003013
8. Decker P, Miranda EA, de Murcia G, Muller S (1999) An improved nonisotopic test to screen a large series of new inhibitor molecules of poly(ADP-ribose) polymerase activity for therapeutic applications. Clin Cancer Res 5:1169–1172
9. Brown JA, Marala RB (2002) Development of a high-throughput screening-amenable assay for human poly(ADP-ribose) polymerase inhibitors. J Pharmacol Toxicol Method. doi:10.1016/S1056-8719(02)00223-X
10. Lee S, Koo HN, Lee BH (2005) Development of a miniaturized assay for the high-throughput screening program for poly(ADP-ribose) polymerase-1. Methods Find Exp Clin Pharmacol. doi:10.1358/mf.2005.27.9.939334
11. Putt KS, Hergenrother PJ (2004) An enzymatic assay for poly (ADP-ribose) polymerase-1 (PARP-1) *via* the chemical quantitation of NAD^+: application to the high-throughput screening of small molecules as potential inhibitors. Anal Biochem. doi:10.1016/j.ab.2003.11.015
12. Perkins E, Sun D, Nguyen A et al. (2001) Novel inhibitors of poly (ADP-ribose) polymerase/PARP1 and PARP2 identified using a cell-based screen in yeast. Cancer Res 61:4175–4183
13. Costantino G, Macchiarulo A, Camaioni E, Pellicciari R (2001) Modeling of poly(ADP-ribose)POLYMERASE (PARP) inhibitors. Docking of ligands and quantitative structure-activity relationship analysis. J Med Chem. doi:10.1021/jm010116l
14. Bellocchi D, Macchiarulo A, Costantino G, Pellicciari R (2005) Docking studies on PARP-1 inhibitors: insights into the role of a binding pocket water molecule. Bioorg Med Chem. doi:10.1016/j.bmc.2004.11.024
15. Stroganov OV, Novikov FN, Stroylov VS et al. (2008) Lead finder: an approach to improve accuracy of protein-ligand docking, binding energy estimation, and virtual screening. J Chem Inf Model. doi:10.1021/ci800166p
16. Ishida J, Yamamoto H, Kido Y et al. (2006) Discovery of potent and selective PARP-1 and PARP-2 inhibitors:SBDD analysis *via* a combination of X-ray structural study and homology modeling. Bioorg Med Chem. doi:0.1016/j.bmc.2005.09.061
17. Dunn D, Husten J, Ator MA and Chatterjee S (2007) Novel poly (ADP-ribose) polymerase-1 inhibitors. Bioorg Med Chem Lett. doi:10.1016/j.bmcl.2006.10.010
18. Hattori K, Kido Y, Yamamoto H et al. (2007) Rational design of conformationally restricted quinazolinone inhibitors of poly(ADP-ribose)polymerase. Bioorg Med Chem Lett. doi:10.1016/j.bmcl.2007.07.091
19. Hattori K, Kido Y, Yamamoto H et al. (2004) Rational approaches to discovery of orally active and brain-penetrable quinazolinone inhibitors of poly(ADP-ribose)polymerase. J Med Chem. doi:10.1021/jm0499256
20. PARP ligands (2009) BioMolTech Corp, Toronto. http://www.biomoltech.com/downloads/papers/supp_inf_parp_ligands_2008_12.zip. Accessed 25 Jan 2008
21. Simonin F, Pochj O, Delarue M, de Murcia G (1993) Identification of potential active-site residues in the human poly(ADP-ribose) polymerase. J Biol Chem 268:8529–8535
22. STK library (2007) Vitas-M Laboratory, Moscow. http://www.vitasmlab.com/compound-libraries-2.htm. Accessed 25 Jan 2008
23. Jain AN, Nicholls A (2008) Recommendations for evaluation of computational methods. J Comput Aided Mol Des. doi:10.1007/s10822-008-9196-5
24. Ame JC, Rolli V, Schreiber V et al. (1999) PARP-2, a novel mammalian DNA damage-dependent Poly(ADP-ribose) polymerase. J Biol Chem 274:17860–17868