



Relaxed complex scheme suggests novel inhibitors for the lyase activity of DNA polymerase beta

Khaled Barakat^a, Jack Tuszynski^{a,b,*}

^a Department of Physics, University of Alberta, Edmonton, AB, Canada

^b Department of Oncology, University of Alberta, Edmonton, AB, Canada

ARTICLE INFO

Article history:

Received 12 October 2010

Received in revised form 2 December 2010

Accepted 6 December 2010

Available online 13 December 2010

Keywords:

BER

DNA

Pol β

Virtual screening

Clustering

NCI database

DrugBank

Fragment

Ionizing radiation

Bleomycin

Monofunctional alkylating agents and

cisplatin

Pamoic acid

ABSTRACT

DNA polymerase beta (pol β), the error-prone polymerase of base excision repair, plays a significant role in chemotherapeutic agent resistance. Its over expression reduces the efficacy of anticancer drug therapies including ionizing radiation, bleomycin, monofunctional alkylating agents and cisplatin. Small-scale studies on different types of cancer showed that pol β is mutated in approximately 30% of tumors. These mutations further lower pol β fidelity in DNA synthesis exposing the genome to serious mutations. These findings suggested pol β as a promising therapeutic target for cancer treatment. More than 60 pol β -inhibitors have been identified so far, however, most of them are either not potent or specific enough to become a drug. Here, we applied the relaxed complex scheme virtual screening (RCSVS) to allow for the full receptor flexibility in filtering the NCI diversity set, DrugBank compounds and a library of ~9000 fragmental compounds for novel pol β inhibitors. In this procedure we screened the set of ~12,500 compounds against an ensemble of 11 dominant-receptor structures representing the essential backbone dynamics of the 8 kDa domain of pol β . Our results predicted new compounds that can bind with higher affinity to the lyase active site compared to pamoic acid (PA), a well-known inhibitor of DNA pol β .

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

After several decades of cancer research, the scientific community is faced with an astounding fact. That is, DNA repair pathways serve as an important determinant of cancer cells' sensitivity to anticancer agents and a major means of acquiring antitumor drug resistance [1,2]. This is mainly because, most anticancer treatments in clinical use today, directly or indirectly damage DNA by causing DNA single- or double-stranded breaks or by interfering with the functions of crucial DNA interacting proteins [3]. Following the detection of the damaged DNA, a number of multiple and overlapping DNA repair pathways attempt to restore the genome, allowing the cell to survive or, if the damage is too heavy, induce apoptosis, forcing the cell to die.

In the middle of these pathways, base excision repair (BER) is the major cellular pathway that is responsible for the recovery of single strand breaks (SSB) and removal of damaged bases such as oxidized-reduced, alkylated and de-aminated bases [4]. These DNA modifications can occur spontaneously by exposing cells to environmental mutagens [2], or synthetically as a result of anticancer treatments using alkylating agents or ionizing radiation. Although, in the former case, BER protects normal cells from lethal mutations, in the latter, it acts as a dominant way for tumor cells to reduce the efficacy of a growing list of DNA damaging agents including bleomycin [5] monofunctional alkylating agents [4], cisplatin [6] and other platinum-based compounds. Therefore, it has been broadly proposed that regulating the BER pathway via small molecule inhibitors can reduce the required dosage of such DNA damaging agents while potentiating their efficacy in eradicating cancer cells [7].

Fortunately, most of the proteins involved in the BER process have been identified, cloned and crystallized allowing the rational design of small molecule inhibitors for their activity. One of these proteins, DNA pol β , plays an important role in chemotherapeutic agent resistance, as its over-expression reduces the efficacy

* Corresponding author at: 2-04 CEB, Department of Physics, University of Alberta, Edmonton, AB, Canada T6G2G7.

E-mail addresses: kbarakat@ualberta.ca (K. Barakat), jackt@ualberta.ca (J. Tuszynski).

of anticancer drug therapies including bleomycin and cisplatin [8,9]. Furthermore, small-scale studies on different types of cancer showed that pol β is mutated in approximately 30% of tumors, which in turn reduces pol β fidelity in DNA synthesis exposing the genome to serious and often deleterious mutations [10,11]. The uncomplicated small structure of the protein (39 kDa) made it a standard model that helped in understanding the functional mechanisms of other DNA polymerases. In addition, there is a large body of evidence that pol β plays an important role throughout cell's life. For example, a “knock-out” of the gene that encodes for pol β in mice results in embryonic lethality, confirming the importance of the protein during fetal development [12]. Based on these findings, pol β , the error-prone polymerase of BER, has been seriously considered as a promising therapeutic target for cancer treatment. Many inhibitors of DNA pol β have been identified during the last two decades (reviewed in Refs. [13,14]). To name but a few, this list includes polypeptides [15], fatty acids [16], triterpenoids [17], sulfolipids [18], polar lipids [19], secondary bile acids [20], phenalenone-derivatives [21], anacardic acid [22], harbinic acid [23], flavanoid derivatives [24], and pamoic acid [25]. However, most of these inhibitors are not potent enough or lack sufficient specificity to eventually become approved drugs.

Receptor-based virtual screening (VS) is a well-established technique to uncover novel molecular inhibitors that complement a target protein in terms of shape, charge and several additional biophysical or biochemical properties [26]. Although VS applied against a crystal or relaxed receptor structure is a commonly used approach in structure-based drug design, the dynamic changes due to receptor flexibility are always neglected to the detriment of the predictive success of these methods. For instance, a ligand can induce significant conformational changes to its target, ranging from local reorganization of side-chains to hinge movement of domains [27]. Sampling these conformational changes during docking is impractical, as they involve too large a number of degrees of freedom. One successful approach that has been reported by McCammon et al. is the relaxed complex scheme (RCS) [28,29]. In this method, molecular dynamics (MD) simulations are applied to explore the conformational space of the protein receptor, while docking is subsequently used for the fast screening of drug libraries against an ensemble of receptor conformations. This methodology has been successfully applied to a number of cases [28,30,31]. An excellent example is an HIV inhibitor, raltegravir [32], which became the first FDA approved drug targeting HIV integrase. MD simulations played a significant role in discovering a novel binding site, and compounds that can exchange between the two binding sites have formed a new generation of HIV integrase inhibitors.

Here, we applied the RCS technique to account for the dominant backbone dynamics in screening for inhibitors of the lyase activity of DNA pol β . Using pamoic acid as a positive control, an inhibitor that targets pol β [25], we employed AUTODOCK version 4.0 [33], to filter a library of ~12,500 small molecules for their activity against the lyase domain of pol β . This library was comprised of the National Cancer Institute (NCI) diversity set, DrugBank small molecules and a set of ~9000 small fragments with drug-like properties. The full set of compounds has been screened against an ensemble of pol β structures. The ensemble incorporated the NMR structure of the 8 kDa domain of the protein (PDB code 1DK3) [34], in addition to the 10 principal conformations as extracted from MD simulations. In this way, AUTODOCK was used to place the compounds within the specified binding site and to search for their minimal energy conformations. Finally, the irredundant top 300 hits from AUTODOCK screening were rescored using the molecular mechanics Poisson–Boltzmann surface area (MM–PBSA) method [35]. It is hoped that our results will eventually be used in the design of more potent and particularly pol β -specific inhibitors, aimed at

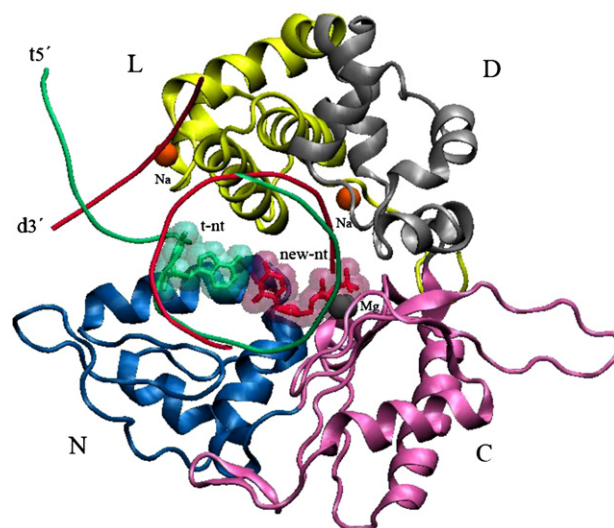


Fig. 1. Structure of DNA polymerase beta. The enzyme is a 39-kDa protein with 335 amino acids in its sequence. The full-length enzyme consists of an amino-terminal lyase domain (8 kDa) (L), shown in yellow, connected by a short protease-sensitive fragment to a carboxyl-terminal polymerase domain (31 kDa). The 31-kDa domain is further subdivided into, D- (duplex DNA binding sub-domain), shown in gray, C- (catalytic sub-domain), shown in mauve and N- (nascent base pair binding sub-domain), shown in blue. The overall structure of pol β resembles the shape of a right hand, with fingers (C-subdomain), thumb (D-subdomain) and palm (N-subdomain) arrangements. The active enzyme requires a single-stranded DNA template (green) and two divalent metal Mg^{2+} ions (gray) for its polymerase activity. It employs two major substrates, namely, a 2'-deoxynucleoside 5'-triphosphate (new-nt), shown in red and a template-primer DNA, shown in green. The 8-kDa domain interacts with the downstream duplex (red), where the 5'-phosphate on the downstream strand is located close to the dRP lyase active site. The lyase domain cooperates with the N-subdomain in order to form a doughnut-shaped structure that surrounds the DNA molecule. The enzyme also utilizes two helix-hairpin-helix (HhH) motifs that unspecifically interact with the DNA backbone. The two HhH motifs, located within the lyase domain (residues 55–79) and the D-subdomain (92–118), incorporate Na^+ ions (orange) and interact with each end of the incised DNA strand, namely, the downstream and primer strands.

improving existing cancer therapies including ionizing radiation, bleomycin, monofunctional alkylating agents and cisplatin.

2. Results and discussion

The faithfulness of BER is dependent on the polymerization step, where the major BER DNA polymerase, pol β , must incorporate the correct Watson–Crick base paired nucleotide into the one-nucleotide repair gap. The enzyme has been identified as a 39-kDa protein with 335 amino acids in its sequence (see Fig. 1) [36]. Its small size compared to other polymerases, made it the smallest and simplest cellular DNA polymerase found. Although pol β lacks the proof-reading 3'- or 5'-exonuclease activities, which usually are found in high fidelity enzymes, it possesses 5'-dRP lyase and AP lyase activities instead [37]. The active enzyme is a stable monomer in solution, folded into distinct domains and subdomains that exhibit a variety of functions essential for its activity. These functions include single-stranded (ss) and double-stranded (ds) DNA binding, nucleoside triphosphate (dNTP) binding, and the dRP lyase and nucleotidyl transferase catalytic activities [38]. Essentially, the full-length enzyme consists of an amino-terminal lyase domain (8 kDa), connected by a short protease-sensitive fragment to a carboxyl-terminal polymerase domain (31 kDa). The 31-kDa domain is further subdivided into C- (catalytic), D- (duplex DNA binding), and N- (nascent base pair binding) subdomains. Interestingly, similar to other DNA polymerases, the overall structure of pol β resembles the shape of a right hand, with fingers (C-subdomain), thumb (D-subdomain) and palm

(N-subdomain) arrangements (see Fig. 1) [36,39]. The active enzyme requires a single-stranded DNA template and divalent metal ions for its polymerase activity. Moreover, it employs two major substrates, namely, a 2'-deoxynucleoside 5'-triphosphate (dNTP) and a template-primer DNA. Accordingly, the C-subdomain contributes three aspartate residues, namely ASP190, ASP192 and ASP256, which coordinate two divalent metal ions (Mg^{2+}). Nevertheless, it should be noted that the presence of the nucleotide-binding metal ion as part of the protein structure has been shown to depend on the existence of DNA and the dNTP substrate within their related binding sites in the enzyme [40].

Several crystal structures of pol β at different stages of the catalytic cycle have been resolved. Considerably, these structures revealed the significant conformational changes that take place within the various subdomains of the protein [41]. These conformational dynamics processes are obvious when comparing the structure of the apo-enzyme [42] to other structures that encompass DNA and the two substrates of the enzyme [43]. The fully loaded pol β structure implied that the 8-kDa domain interacted with the downstream duplex, where the 5'-phosphate on the downstream strand is located close to the dRP lyase active site. Furthermore, the lyase domain cooperates with the N-subdomain in order to form a doughnut-shaped structure that surrounds the DNA molecule (see Fig. 1). These notable interactions and functions of the lyase domain indicate that, a small molecule that can bind to the lyase active site, especially, within the ssDNA binding pocket should be able to affect the polymerization activity of pol β as well. Besides the lyase domain dynamics, the N-subdomain seems to exhibit considerable movements once the correct dNTP substrate is bound to the enzyme [43]. Additionally, as illustrated in Fig. 1, there is a major conformational change within the structure of the DNA substrate.

Analogously to most DNA-binding proteins, pol β utilizes the well-known helix-hairpin-helix (HhH) motifs that unspecifically interact with the DNA backbone. These HhH motifs are located within the lyase domain (residues 55–79) and the D-subdomain (92–118) and interact with each end of the incised DNA strand, namely, the downstream and primer strands. In addition, like other HhH motifs, they encompass monovalent metals, which in the case of pol β have been identified to be Na^+ ions. However, as mentioned above and similarly to the nucleotide-binding metal Mg^{2+} ion, their presence within the structure of the protein is mainly dependent on the existence of the bound DNA substrate [44].

During the last two decades, many attempts were made to isolate and identify a small molecule inhibitor that can specifically bind to pol β . These efforts resulted in the discovery of more than 60 molecules that can bind to DNA polymerases in general with a few of them that can target pol β in particular (reviewed in Refs. [13,14]). Among the compounds that exhibited promising activity against pol β , with a well-defined binding mode, is pamoic acid (PA), discovered by Hu and his co-workers [25]. In their study, Hu et al. [25] used NMR analysis to identify the binding interface between PA and the 8-kDa domain of pol β and decompose its residue contributions. Their findings suggest that the binding pocket of the compound within the surface of pol β is located between the two helices, helix-2 and helix-4 of the 8-kDa domain. Interestingly, the same region has been recognized in different studies to be essential in the DNA binding and deoxyribose phosphate lyase activities of the enzyme [43]. Their results revealed that PA inhibits the deoxyribose phosphate lyase and DNA polymerase activities of purified pol β on a BER substrate. However, it should be noted that the inhibition of these two activities of pol β by PA was only observed when the compound had been pre-incubated with the enzyme before initiation of the BER reactions. These observations may indicate that the binding reaction between pol β and PA is very slow and requires an apo-enzyme to be completed. More

importantly, the compound also was able to enhance the efficacy of methyl methanesulfonate (MMS), a monofunctional methylating agent that targets the DNA whose induced damage is mainly repaired by BER. These interesting findings were further pursued by a different group from the Centre National de la Recherche Scientifique (CNRS) in France, to understand and identify the precise interactions between PA and the 8-kDa domain of pol β [45]. In this study, Hazan et al. used a combination protocol of blind docking and NMR analysis to identify the binding site of PA within the surface of the lyase domain of pol β and to suggest its binding conformation. These results confirmed the earlier findings of Hu et al. [25] and revealed that PA binds to a site formed by helix-2 and helix-4, which also corresponds to the single-stranded DNA binding site [43]. Particularly, the aromatic groups of pamoic acid formed favorable hydrophobic interactions with the residues TYR39, ALA42, GLY64 and GLY66 within the identified binding site. Furthermore, the presence of many lysine residues in the binding pocket allowed favorable electrostatic interactions for the two carboxyl groups of PA. In their proposed model, one of the carboxyl groups is oriented towards HIS34 and LYS35 making close contacts with ILE69 amide proton, while the other carboxyl group formed hydrogen bonds with the amide proton of Lys68 and with the hydroxyl group of THR67.

In the present work, we focused our search space on the binding site of PA, using it as our positive control. Our aim was to discover more potent drug candidates through filtering a library of ~12,500 structures via a RCSVS protocol. The molecules we tested included the NCI diversity set (1883 compound), the DrugBank set of small molecules (1566 compound) and more than 9000 fragment structures with drug-like properties extracted from ZINC database. The top 300 hits that showed strong affinity for pol β have been rescored using a more robust scoring function, the MM-PBSA method. Our primary use of docking techniques has been to filter the three medicinal chemistry compound libraries through receptor-based virtual screening, in an attempt to uncover molecules that complement the DNA binding site within the lyase domain of pol β in terms of such parameters as shape, charge, and several additional biochemical characteristics.

2.1. Molecular dynamics simulations of the apo-lyase domain

To generate an ensemble of equilibrated pol-models for chemical library screenings, the 8-kDa domain of pol β (PDB code 1DK3) [34] was subjected to MD simulations. The proper equilibration of these systems was essential in order to perform virtual screening on a set of rigid receptor models that represent approximately the whole conformational space of the PA binding site (which coincides with the DNA binding site) [43] within the lyase domain of pol β . It should be noted that, as we used docking as a preliminary filtering step in screening the full set of compounds for pol β -inhibitors, it was essential to generate an ensemble of pol β structures in order to partly incorporate protein flexibility during docking. In this context, we selected the top 300 compounds that can bind to their appropriate pol β conformation for post-docking analysis using MD simulations to introduce the full flexibility for both the ligand and its selected target structure.

2.2. Principal component analysis and completeness of sampling

As MD simulations produce numerous conformations of the lyase domain of pol β , we have utilized Principal Component Analysis (PCA) to transform the original space of correlated variables into a reduced set of independent variables comprising the essential dynamics of the system [46,47]. To perform PCA, the entire MD trajectory must first be RMSD (root-mean-square-deviation) fitted to a reference structure and a covariance matrix is calculated from

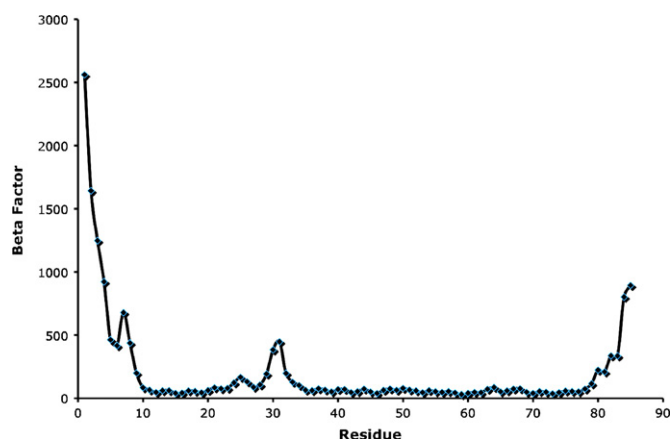


Fig. 2. Plot of the B-factors averaged over the protein backbone atoms as a function of residue number in the simulations of pol β .

its Cartesian co-ordinates. The resulting eigenvectors constitute the essential vectors of the motion, where the larger an eigenvalue, the more important its corresponding eigenvector in the collective motion. PCA was performed over the entire set of MD simulations of the 8-kDa domain using the heavy atoms comprising the 21 residues contained in the PA (DNA) binding site (residues numbered: 30, 31, 32, 33, 34, 35, 37, 38, 39, 40, 41, 42, 43, 63, 64, 65, 66, 67, 68, 69 and 70) with the backbone atoms RMSD fitted to the minimized crystal structure of the starting configurations. Resulting eigenvectors were sorted by descending eigenvalues, which represent the variance of the motion along the principal components. The 10 dominant eigenvalues are shown in Fig. S1a. The first eigenvalue has a magnitude that is significantly higher than those of the other eigenvalues. The components with the largest eigenvalues represent correlated motions of the binding site with the most significant standard deviations of the motion along the corresponding orthogonal directions.

Fig. S1b represents the spatial distributions of occupancies for the conformational states over the planes spanned by the three dominant principal components of the binding site for the simulated system. The PA (DNA) binding site within pol β adopted limited conformations throughout the MD simulations, indicating the rigidity of the protein. The grouping of MD trajectories into a limited number of clusters suggests the presence of favored folded conformations with significant basins of attraction. This observation was also confirmed by results presented in Fig. 2, where the main-chain B-factors (averaged over heavy atoms) for almost all residues constructing the 8-kDa domain being rigid except for a number of residues that are located on the protein termini or which have no direct influence on the binding site. These are residues 1–9, 28–31, and 80–87.

Covariant analysis of the trajectories from the MD simulations, successively divided into thirds, was performed using the same procedure used for PCA. Normalized overlaps calculated between each of these thirds range from 0.79 to 0.81. The high overlap between the thirds indicates that each part of the simulation samples approximately the same conformational space, and it is unlikely that there are unexplored regions missed earlier in the runs. Although there is no guarantee that complete equilibrium sampling is given, we have concluded that the observed overlap is acceptable and adequate sampling within the MD trajectories for the binding site had been obtained.

2.3. Ensemble-based virtual screening

While integrating receptor flexibility into docking reduces the risk of unfavorable ligand–target interactions, accommodating full

receptor flexibility during a docking experiment is impractical. One solution for this problem is to allow only those parts of the receptor that affect the protein–ligand interactions to be flexible during docking. Due to the increased sampling requirements and limited computational resources, flexible parts are generally restricted to the principal side chains that are believed to be involved in binding interactions. This procedure allows for localized protein movement, resulting in an improved fit of the ligand. However, an obvious drawback when considering only the flexibility of restricted protein fragments is that the collective motion of the complete receptor backbone is neglected. To overcome this deficiency we have used an ensemble of protein conformations for docking as an alternative approach to introduce a feature of global protein flexibility. This ensemble could describe the entire conformational space of the binding site, yet it must still be represented by a set of limited conformations in order to save computational screening time.

To generate a reduced set of representative models of the PA(DNA) binding site, we applied the RMSD conformational clustering to the generated trajectory. Fig. 3 shows the evaluation of the Davies–Bouldin index (DBI) and the percentage of variance explained by the data (SSR/SST) for different cluster counts (see Section 3). DBI for the apo-system exhibited local minima at cluster counts of 45, 60, 80 and 105. However, as the percentage of variance explained by the data started to plateau after 35 clusters, we concluded that 45 clusters is a reasonable cut-off for pol β structures.

In this study, we constructed an ensemble of 11 distinct conformations to perform ensemble-based virtual screening on pol β against the full set of ligand compounds. This ensemble incorporated the 10 most dominant structures that comprised ~85% of trajectory in addition to the relaxed NMR pol β conformation. The ultimate goal was to reduce the number of representative structures included in the ensemble-based screening and concurrently comprise most of the conformational space of the binding site. Fig. 4 represents the 11 structures used in this work. The 8-kDa structure adopted different conformational changes demonstrating the significance of introducing receptor flexibility during the docking procedure.

2.4. Pose clustering

As mentioned above, 11 independent virtual screening experiments were performed against the full set of database compounds. Screening of the full set of compounds contained in the NCIDS, DrugBank and the fragmental databases (more than 12,500 molecules), against the 11 target structures, produced a total of ~13 million distinct poses that required classification. While AUTODOCK is capable of clustering these poses into subgroups depending on RMSD, the total number of clusters and the population of each cluster is mostly dependent on the RMSD cut-off that is initially chosen. As such, there is no adequate means to anticipate an optimum cut-off for the RMSD to produce the best quality result. Since we are dealing with a diverse set of input ligands, this clustering method does not provide an accurate means of comparing resulting populations and binding energies between ligands, making it difficult to score compounds accurately.

The optimal number of clusters required for grouping similar conformations of a ligand in a typical docking run would depend on factors such as the binding mode, shape, and flexibility of the ligand. To be truly successful, a dynamic clustering technique must take full advantage of observable differences between the diverse conformations adopted by the ligand within the binding site. Moreover, it should adapt the cluster count to extract and make sense of the information inherited with these conformations. In this study, we automated the clustering approach used for clustering the MD

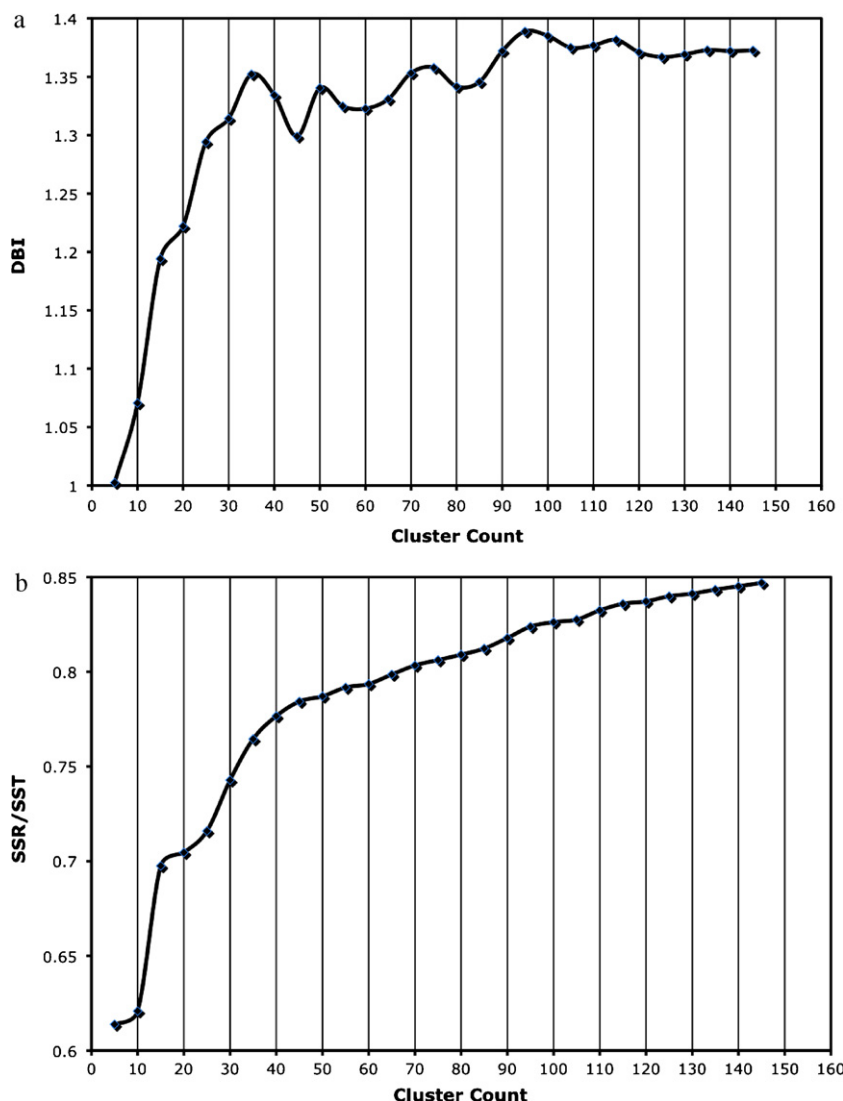


Fig. 3. Clustering analysis for the pol β trajectory. A high-quality clustering is obtained when a local minimum in DBI correlates with saturation in the SSR/SST ratio (cluster count of 45).

trajectory (see Section 3) to extract the optimal number of clusters from docking results. This required the selection of a parameter to measure the quality of the clusters produced and to represent a convergence criterion for the clustering program. Of the various clustering metrics described in the literature, we used the elbow criterion as a measure for clustering convergence (see Methodology) due to its simple implementation. Also, visualization of the clustering is rapid and obvious. Here, we have applied this methodology by calculating the percentage of variance found within the data after each attempt to extract a new cluster from the system. As the number of clusters exceeds the optimal number, the percentage of variance should plateau indicating a complete extraction of the significant information included in the system.

2.5. Ranking of pol β -ensemble-based screening using AUTODOCK scoring function

For each virtual screening experiment, we have ranked significant poses for each of the molecules contained in the database by using the results from the elbow criterion and the lowest energy that corresponds to the most populated cluster. Once all poses from each ligand entry were clustered, we then filtered all of the clusters so that only those containing at least 25% of the total population

were considered as top hits. Top hits were collected from the 11 screening simulations by first extracting the largest cluster from each individual screening followed by ranking the clusters according to their binding energies. This produced a set of non-redundant hits ranked by their binding energies of the most populated cluster. The top 20 hits of AUTODOCK screening are shown in Table S1.

The apparent K_D value for PA binding to pol β is 9 μM [25]. The binding constant (K_D) can be used to evaluate the observed free energy change of binding, β , through the relation:

$$\Delta G = RT \ln K_D \quad (1)$$

where R is the gas constant, $R = 1.987 \text{ cal K}^{-1} \text{ mol}^{-1}$ and T is the absolute temperature. Using the AUTODOCK scoring function, we obtained a value of -6.2 kcal/mol for the binding energy of PA to the 8-kDa domain. Although this value is in excellent agreement with the experimental value (-6.9 kcal/mol) as calculated using the above equation, it has been widely reported in the literature that empirical scoring functions including the AUTODOCK scoring function, are not efficient in discriminating false positives in VS experiments and are biased toward their training set of compounds [48,49]. Consequently, in this work, the top 300 hits were rescored using the MM-PBSA method and re-ranked for their binding to the protein (see below).

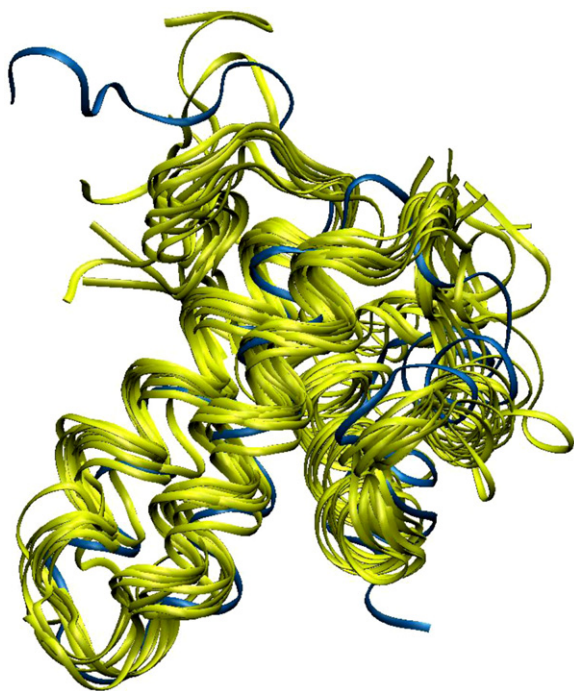


Fig. 4. Eleven dominate conformations for pol β . This ensemble comprised the NMR structure (blue) and 10 structures extracted from the MD trajectory (yellow). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

2.6. Ranking of pol β top hits using the MM–PBSA scoring function

As scoring of the poses is crucial in prioritizing and ranking of the compounds, it is important to use sensitive and accurate scoring functions that can replicate and predict experimental data. This is normally achieved using an objective scoring function that directs the conformational search algorithm in predicting the native conformation and ultimately estimates the binding affinity of the protein–ligand complex. Nevertheless, as mentioned above, it has been broadly demonstrated that docking scoring functions are less successful at predicting the actual binding affinities and at discriminating true binders from inactive (decoy) compounds [50]. These puzzling results are direct outcomes of many factors that have been misused while analyzing the binding interactions of the resulting poses as a compromise to speed up the docking process. These factors mostly involve the lack of proper solvation, the neglect of protein flexibility and the bias toward the training set of structures that have been used in optimizing the scoring process [51]. In fact, developing new scoring functions and innovative ranking schemes is still a largely unexplored area of research in the field of docking. Although a more precise scoring method can be practically implemented within docking, a large computational cost that is associated with such a function will become a prohibitive barrier preventing researchers from using it.

To overcome these difficulties, one can use MD simulations as a post-docking tool in order to refine the final docked complexes and eliminate their impractical interactions. Although this procedure requires extensive computational resources, it tends to improve the protein–ligand interactions and enhance the complementarity between them. Moreover, the stability of the complexes over the simulation time is a direct measure of the consistency of binding, since improperly docked structures are expected to produce unstable trajectories. Finally, MD simulations provide a direct means of assessing the role of water molecules in mediating the contacts

between the protein target and ligand molecules which helps quantify their effect on protein–ligand binding.

While this procedure leads to the prediction of the correct conformation of the protein/ligand complexes by incorporating their induced fit interactions, another essential constraint for a successful VS experiment is to accurately predict their binding energies. This requires a consistent scoring method that can efficiently separate active from non-active compounds and classify the hits according to their relative binding strength. Here, we used the MM–PBSA method, introduced by Kollman et al. [52] to measure the binding energies of the top 300 hits relative to that of the positive control, i.e. PA, and compare their MM–PBSA-ranking to that of AUTODOCK calculations. This technique has been previously adopted as an alternative approach to rescore docking results and shown to produce accurate free energies at a reasonable computational cost [30,53]. Its main advantages are the lack of adjustable parameters and the option of using a single MD simulation for the complete system to determine all energy values. As has been reported earlier by other groups, the most computationally demanding step is the calculation of the solute entropy using the normal mode (NMODE) method. Although this component can be neglected if only relative binding (relative ranking) of compounds is required [52], we calculated the entropy contributions for all the top 300 hits using 200 snapshots extracted from their 2 ns MD trajectories (see Section 3). In our calculations, this part ranged from 10 to 15 kcal/mol, indicating its significance in predicating the overall binding energies. The top 34 hits according to MM–PBSA calculations are shown in Table 1. As expected, the MM–PBSA analysis predicted the binding energy of PA more accurately than the AUTODOCK scoring function. Furthermore, AUTODOCK ranking of the top hits has been partially altered when compared to MM–PBSA calculations (see Table 1). These results also illustrate the limitations of the AUTODOCK scoring function in eliminating from the set of active compounds false positive ligands, i.e. compounds that are predicted to bind the target but fail to do so in validation assays. We noticed this in our calculations where a number of compounds highly ranked by AUTODOCK exhibited very low, and in some cases positive, binding energies using the MM–PBSA analysis, indicating their weak binding to the protein (see Table 1). The identified binding mode of PA is shown in Fig. 5a. The atomic distances between PA and the two residues ALA42 and ILE69 are in excellent agreement with what was shown by Hu et al. [25] and Hazan et al. [45], confirming the successful docking of PA within the DNA binding site of pol β . Furthermore, Fig. 5 demonstrates the binding modes of the top three hits of the MM–PBSA ranking. Similar to a substantial number of our suggested top hits, the shown compounds are small in size, however, they are occupying a considerable portion of the DNA-binding pocket. These lead compounds can be employed as the basis for a further fragment-based drug design step, in order to construct potent and more specific pol β inhibitors.

3. Methods and materials

3.1. Molecular dynamics simulations

The 8-kDa domain of pol β (residues 1–87) was taken from the PDB entry 1DK3 [34]. MD simulations were carried out using the NAMD program [54], at a mean temperature of 310 K and physiological pH (pH 7) using the all-hydrogen AMBER99SB force field [55]. Protonation states of all ionizable residues were calculated using the program PDB2PQR [56]. Following parameterization, the protein was immersed in the center of a TIP3P water cube after adding hydrogen atoms to the initial protein structure. The cube dimensions were chosen to provide at least a 20 Å-wide buffer of 17,605 water molecules around the systems. To neutralize and

Table 1
Top 30 hits according to MM–PBSA ranking. Compounds are ranked by their binding energies as were calculated using the MM–PBSA method and compared to their ranking using the AUTODOCK scoring function.

MM–PBSA scoring		Ensemble-based scoring		ID	Chemical structure
Rank	BE \pm 1.5 (kcal/mol)	Rank	BE \pm 2.2 (kcal/mol)		
1	–12.5	180	–7.2	ZINC19229065	
2	–12.4	263	–7.0	ZINC00020243	
3	–11.9	33	–8.2	ZINC04102187	
4	–11.3	24	–8.4	NSC#372280	
5	–10.6	5	–9.6	NSC#210627	
6	–10.4	19	–8.6	NSC#327705	
7	–9.9	14	–8.8	NSC#116654	
8	–9.8	40	–8.1	NSC#12363	

Table 1 (Continued)

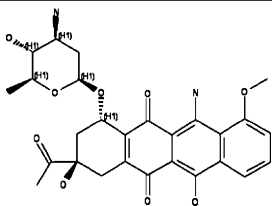
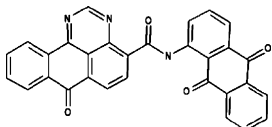
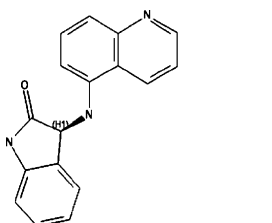
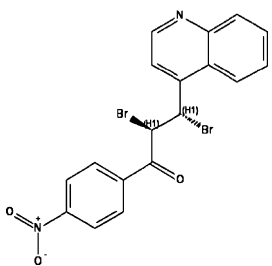
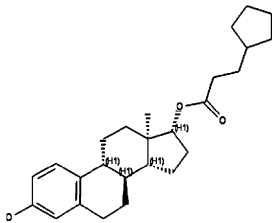
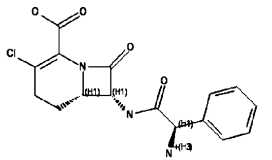
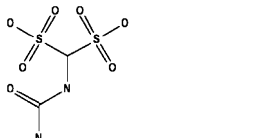
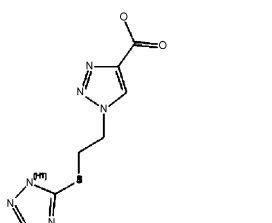
MM-PBSA scoring		Ensemble-based scoring		ID	Chemical structure
Rank	BE \pm 1.5 (kcal/mol)	Rank	BE \pm 2.2 (kcal/mol)		
9	−9.6	23	−8.4	NSC#254681	
10	−9.3	22	−8.5	NSC#299137	
11	−9.2	202	−7.2	NSC#117198	
12	−9.2	88	−7.6	NSC#150289	
13	−9.0	77	−7.7	NSC#3354	
14	−8.9	264	−7.0	ZINC01530992	
15	−8.8	13	−8.8	ZINC05368838	
16	−8.6	245	−7.0	ZINC20596577	

Table 1 (Continued)

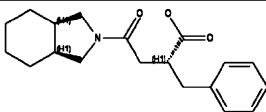
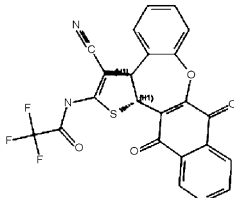
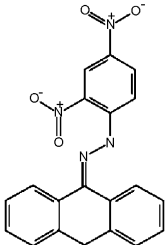
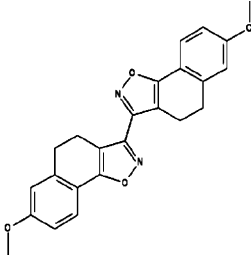
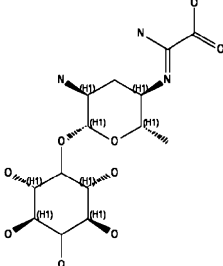
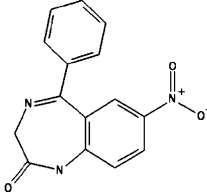
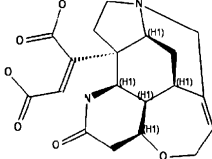
MM-PBSA scoring		Ensemble-based scoring		ID	Chemical structure
Rank	BE \pm 1.5 (kcal/mol)	Rank	BE \pm 2.2 (kcal/mol)		
17	−8.4	87	−7.6	ZINC11616579	
18	−8.2	226	−7.1	NSC#686365	
19	−7.9	6	−9.5	NSC#201873	
20	−7.8	79	−7.7	NSC#371688	
21	−7.7	160	−7.3	NSC#100858	
22	−7.6	258	−7.0	ZINC03812992	
23	−7.5	9	−9.0	NSC#123420	

Table 1 (Continued)

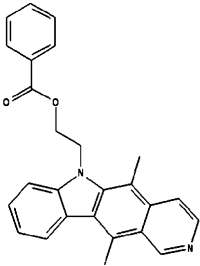
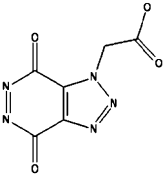
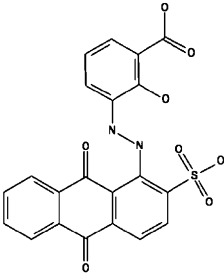
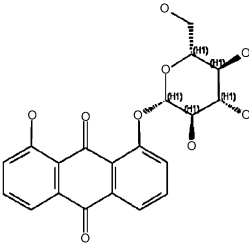
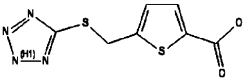
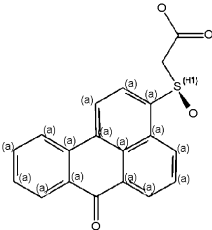
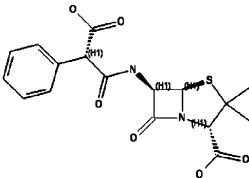
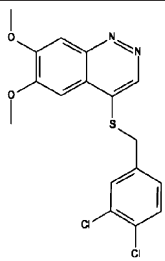
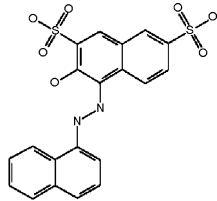
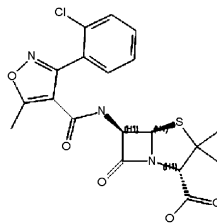
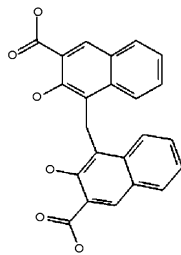
MM–PBSA scoring		Ensemble-based scoring		ID	Chemical structure
Rank	BE \pm 1.5 (kcal/mol)	Rank	BE \pm 2.2 (kcal/mol)		
24	–7.5	214	–7.1	NSC#125908	
25	–7.5	123	–7.4	ZINC16958839	
26	–7.4	3	–9.7	NSC#125908	
27	–7.1	18	–8.7	NSC#255980	
28	–7.0	142	–7.3	ZINC20395500	
29	–7.0	155	–7.3	NSC#16211	
30	–6.9	143	–7.3	ZINC03978033	

Table 1 (Continued)

MM-PBSA scoring		Ensemble-based scoring		ID	Chemical structure
Rank	BE \pm 1.5 (kcal/mol)	Rank	BE \pm 2.2 (kcal/mol)		
31	−6.9	139	−7.3	NSC#64814	
32	−6.8	4	−9.6	NSC#45583	
33	−6.7	28	−8.3	ZINC03875417	
34	−6.7	278	−6.2	PA	

prepare the simulated system under a physiological ionic concentration, 41 chloride and 32 sodium ions were, respectively, added by replacing water molecules having the highest electrostatic energies on their oxygen atoms. The number of counter ions for each case was calculated by first estimating the amount of ions that is needed to set up the solvated system under normal physiological conditions (pH 7), followed by adding the number of chloride ions required to bring its net charge to zero. The fully solvated protein was then minimized and subsequently heated to the simulation temperature with heavy restraints placed on all backbone atoms. Following heating, the system was equilibrated using periodic boundary conditions for 100 ps and energy restraints reduced to zero in successive steps of the MD simulation. The simulations were then continued for 92 ns during which atomic coordinates were saved to the trajectory every 2 ps. The total simulation time was determined by visualizing the quality of sampling as predicted by PCA (see below). The RMSD (data not shown) and B-factors (see Fig. 2) for the protein backbone were then computed over the last 10 ns of the MD simulation using the PTRAJ utility within AMBER10 [57]. Hydrogen bond analyses were performed by computing the average distance between donor and acceptor atoms. A hydrogen bond was defined by a heavy donor–heavy acceptor distance ≤ 3.4 Å, a light donor–heavy acceptor distance ≤ 2.5 Å, and a deviation of less than $\pm 60^\circ$ from linearity.

Following the same MD protocol mentioned above we prepared 300 MD simulations for each top hit that resulted from the ensemble-based screening (see Section 2). That is, each ligand in complex with the protein was solvated, heated and then equilibrated at the 310 K. Parameters for ligands were assigned using the generalized AMBER force field [58] and partial charges were calculated with the AM1-BCC method [59] using Antechamber in the AMBER10 package. Following parameterization, the protein/ligand complexes were subjected to MD simulations for a production phase of 2 ns. Snapshots were extracted every 2 ps and used for the MM–PBSA binding energy analysis.

3.2. RMSD clustering to extract representative MD structures

MD simulations on the 8-kDa domain of pol β produced numerous structures that explored their conformational space. Although performing VS against each snapshot of these trajectories is the most accurate way to account for full receptor flexibility, implementing this technique is unfeasible since it requires massive computational resources. A suitable way to accommodate receptor flexibility and concurrently, reduce the required computational time is to dock the ligand library against a set of representative structures that describes the conformational space of the target [60,61]. Although this approach significantly reduces the number

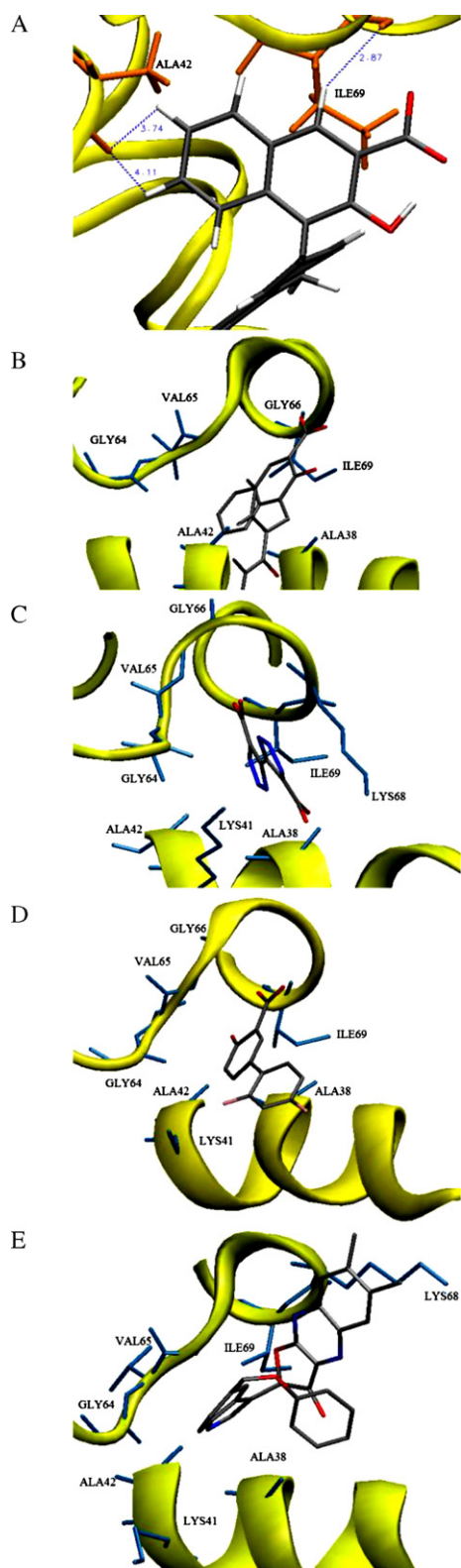


Fig. 5. Binding modes of selected hits. The binding mode of PA (A, B) and the top three hits from the MM-PBSA ranking (C–E). Pol β is shown in yellow, important protein residues are shown in orange (A) and blue (B–D), and the different atoms of the bound compounds are shown by their representative colors (carbon in gray, oxygen in red, nitrogen in blue and hydrogen in white). The atomic distances (in angstroms) of PA are in excellent agreement with what was described in (Refs. [25,49]). Compounds shown in (C and D) are small in size, similar to a number of our top hits, however, they occupy a considerable portion of the binding pocket.

of VS experiments, care should be taken in extracting these representative structures. A common approach to generate such an ensemble of representative models is to perform RMSD conformational clustering on the whole trajectory [60]. Unfortunately, there is no universally accepted clustering algorithm that can be used to extract all of the information contained within the MD simulation. However, recent studies suggest that a number of clustering algorithms, such as average-linkage, means and self-organizing maps (SOM) can be used in clustering MD data [62]. The clustering quality can be anticipated by calculating a number of clustering metrics that can deduce the optimal number of clusters to be extracted and their population size. These are the Davies–Bouldin index (DBI) [63] and the “elbow criterion” [62]. A high-quality clustering scheme is expected when high DBI values are calculated. On the other hand, using the elbow criterion, the percentage of variance explained by the data is expected to plateau for cluster counts exceeding the optimal number [62]. Using these metrics, by varying the number of clusters, one should expect for adequate clustering, a local minimum for DBI and a horizontal line for the percentage of variance explained by the data.

To generate a reduced set of representative pol β models, we performed RMSD conformational clustering with the average-linkage algorithm as implemented in the PTRAJ utility of AMBER10 using cluster counts ranging from 5 to 150 clusters. Structures were extracted at 2 ps intervals over the entire simulation times. All C_{α} -atoms were RMSD fitted to the minimized initial structure in order to remove overall rotation and translation. RMSD-clustering was performed on the 21 residues contained in the PA (DNA) binding site (residues numbered: 30, 31, 32, 33, 34, 35, 37, 38, 39, 40, 41, 42, 43, 63, 64, 65, 66, 67, 68, 69 and 70). These residues were clustered into groups of similar conformations using the atom-positional RMSD of the entire amino acid, including side chains and hydrogen atoms, as the similarity criterion. The optimal numbers of clusters were chosen after evaluation of the two clustering metrics, described above, for different cluster counts (see Section 2). A total of 45 clusters were extracted from the trajectory. The centroid of each cluster, the structure having the smallest RMSD to all members of the cluster, was chosen as the cluster representative structure and the most dominant structures were used as rigid templates for the ensemble-based docking experiments (see Section 2).

3.3. Principal component analysis

PCA can transform the original space of correlated variables from a large MD simulation into a reduced space of independent variables comprising the essential dynamics of the system [46,47]. For a typical protein, the system's dimensionality is thereby reduced from tens of thousands to fewer than 50 degrees of freedom. To perform PCA for a subset of N atoms, the entire MD trajectory is RMSD fitted to a reference structure, in order to remove all rotations and translations. The covariance matrix can then be calculated from their Cartesian atomic co-ordinates as:

$$\sigma_{ij} = \langle (r_i - \langle r_i \rangle) (r_j - \langle r_j \rangle) \rangle \quad (2)$$

where r_i represents the three Cartesian co-ordinates (x_i, y_i or z_i) and the eigenvectors of the covariance matrix constitute the essential vectors of the motion. It is generally accepted that the larger an eigenvalue, the more important its corresponding eigenvector in the collective motion. PCA can also be employed to predict the completeness of sampling during the MD simulation. A method proposed by Hess [64] divides an MD trajectory into separate parts, and their normalized overlap is calculated using the covariant

matrices for each pair of parts:

$$\text{normalized overlap}(C_1, C_2) = 1 - \frac{\sqrt{\text{tr}((\sqrt{C_1} - \sqrt{C_2})^2)}}{\sqrt{\text{tr}(C_1) + \text{tr}(C_2)}} \quad (3)$$

where C_1 and C_2 are the covariant matrices, and the symbol tr denotes the trace operation. If the overlap is 0, then the two sets are considered to be orthogonal, whereas an overlap of 1 indicates that the matrices are identical. To ensure completeness of sampling for MD simulations of the 8-kDa domain of pol β , PCA of the binding-site residues was performed using the positions of all heavy atoms. The MD trajectory was divided into three parts and the normalized overlap between each pair was calculated to determine the completeness of sampling.

4. Selection of ligand database

The National Cancer Institute Diversity Set (NCDIS) [65], DrugBank small molecules [66], and a set of 9135 fragment structures were used as our test libraries of compounds. The NCIDS is a collection of approximately 2000 compounds that are structurally representative of a wide range of molecules, representing almost 140,000 compounds that are available for testing at the NCI. Unfortunately, a number of ligands containing rare earth elements could not be properly parameterized and were excluded, leaving a total of 1883 compounds for analysis. Here, we used a version of the NCIDS formatted for use in AUTODOCK which was prepared by the AUTODOCK Scripps team. The DrugBank small molecule library is a set of 1488 FDA-approved small molecule drugs downloaded from the ZINC database. Some of these molecules were present in more than one protonation state adding another 78 structures to the docked ligands. In addition, we included a set of 9135 clean-fragments compounds, downloaded from the ZINC database. These fragments have a Tanimoto coefficient of 70%, molecular weight lower than 250 Da, xlogP lower than 2.5, number of rotatable bonds less than 5 and only a single stereoisomer and protonation state for each compound.

4.1. Ligand screening

Virtual screening on the PA binding site (which coincides with the DNA binding site) [43] within pol β was performed using AUTODOCK, version 4.0 [33]. Hydrogen atoms were added to the protein and ligands and partial atomic charges were then assigned using the Gasteiger–Marsili method [67]. Atomic solvation parameters were assigned to the atoms of the protein using the AUTODOCK 4.0 utility ADDSOL. Docking grid maps with $118 \times 98 \times 104$ points and grid point spacing of 0.26 Å was then centered on the PA binding site within pol β using AUTOGUID4.0 program [33]. Rotatable bonds of each ligand were then automatically assigned using the AUTOTORS utility of AUTODOCK4.0. Docking was performed using the Lamarckian Genetic Algorithm (LGA) method with an initial population of 400 random individuals; a maximum number of 10×10^6 energy evaluations; 100 trials; 50,000 maximum generations; a mutation rate of 0.02; a crossover rate of 0.80 and the requirement that only one individual can survive into the next generation. A total of 11 independent virtual screening runs were performed against the full set of docked ligands with all residues of the receptors set rigid during docking experiments. This set of pol β models comprises one structure that represents the minimized NMR conformation of pol β and 10 conformations that represent $\sim 85\%$ of the MD trajectory (see Section 2).

4.2. Clustering of docked poses

The previously described virtual screening experiments involved millions of conformations of each ligand bound to pol

β . AUTODOCK can cluster these output poses into subgroups depending on their RMSD values referred to a reference structure. Although this approach shows the possible binding modes of a ligand to the binding site, the number of clusters and the population size for each cluster depends heavily on the RMSD cut-off used. It is not possible to anticipate an optimum cut-off for the RMSD in order to produce a clustering pattern with the highest confidence, motivating us to use alternative approaches in performing the clustering analysis. In this study, we used an automated approach to couple one of the commonly used clustering metrics, the elbow criterion [62], with the clustering module of PTRAJ utility of AMBER10. This method exploits the fact that the percentage of variance exhibited by the data (λ), is expected to plateau for cluster counts exceeding the optimal number.

The percentage of variance is defined by:

$$\lambda = \frac{\text{SSR}}{\text{SST}} \quad (4)$$

where (SSR) is the sum-of-squares regression from each cluster summed over all clusters and (SST) is the total sum of squares. Here, we used the SOM algorithm as implemented in the PTRAJ utility of the AMBER10 program to cluster the docking results. This modified clustering program increases the number of clusters required until the percentage of variance exhibited by the data (λ) plateaus. This can be determined by calculating the first and second derivatives of the percentage of variance with respect to the clusters number ($d\lambda/dN$ and $d^2\lambda/dN^2$) after each attempt to increase the cluster counts. The clustering process then stops at an acceptable value for these derivatives that is close to 0. Consequently, the clustering procedure depends only on the system itself and adjusts itself to arrive at the optimal clustering pattern for that specific system.

4.3. Rescoring of top hits using MM–PBSA

Binding free energies were calculated using the molecular mechanics Poisson–Boltzmann surface area (MM–PBSA) method [35] as implemented in AMBER10. The total free energy is the sum of average molecular mechanical gas-phase energies (E_{MM}), solvation free energies (G_{solv}), and entropy contributions ($-TS_{solute}$) of the binding reaction:

$$G = E_{MM} + G_{solv} - TS_{solute} \quad (5)$$

The total molecular mechanical energies can be further decomposed into contributions from electrostatic (E_{ele}), van der Waals (E_{vdw}) and internal energies (E_{int}):

$$E_{MM} = E_{ele} + E_{vdw} + E_{int} \quad (6)$$

In this work, the molecular mechanical (E_{MM}) energy of each snapshot was calculated using the SANDER module of AMBER10 with all pair-wise interactions included using a dielectric constant (ϵ) of 1. The solvation free energy (G_{solv}) was estimated as the sum of electrostatic solvation free energy, calculated by the finite-difference solution of the Poisson–Boltzmann equation in the Adaptive Poisson–Boltzmann Solver (APBS) program as implemented in AMBER10 and non-polar solvation free energy, calculated from the solvent-accessible surface area (SASA) algorithm. The solute entropy was approximated using the normal mode analysis. Applying the thermodynamic cycle for each ligand–pol β complex, the binding free energy between an arbitrary ligand molecule and the pol β protein can be approximated by:

$$\Delta G^\circ = \Delta G_{gas}^{pol\beta\text{-ligand}} + \Delta G_{solv}^{pol\beta\text{-ligand}} - \{\Delta G_{solv}^{ligand} + \Delta G_{solv}^{pol\beta}\} \quad (7)$$

Here, ($\Delta G_{gas}^{pol\beta\text{-ligand}}$) represents the free energy per mole for the non-covalent association of the ligand–pol β complex in vacuum (gas phase) at 310 K, while ($-\Delta G_{solv}$) stands for the work required

to transfer a molecule from its solution conformation to the same conformation in vacuum at 310 K (assuming that the binding conformation of the ligand–pol β complex is the same in solution and in vacuum).

5. Conclusions

Family-X of DNA polymerases in general and pol β in particular are the foremost elements of the BER pathway [37b]. Fortunately, a lot of structural data and biological information about pol β are currently available, making it the first DNA polymerase enzyme whose structural description is complete [41,68]. These observations stimulated research aimed at finding regulators of BER through the discovery of inhibitors of the polymerization step of the pathway. The fundamental principle behind this objective is to maximize the ionizing radiation or chemotherapeutic-induced damage within the genome of cancer cells in order to potentiate the efficacy of these DNA-damaging agents and hence, induce apoptosis [13]. Surprisingly, in spite of more than 20 years of extensive research focused on inhibitors for pol β and the high number of various structures that have been isolated and extracted, there is no single molecule that can be identified as a pol β -specific inhibitor. This is mainly because most of the investigated compounds target other polymerases or enzymes and a considerable number of them cannot enter the cell due to solubility problems. More importantly, even with the large number of pol β crystal structures deposited into the PDB database, no *de novo* drug design or virtual screening study for molecular structures that can target specific domains of the enzyme has been reported in the literature to date. This motivated us to take on a complete computational perspective in this work, in order to identify specific drug candidates that can target the lyase activity of pol β .

In this context, we applied the RCS technique to account for the full receptor flexibility in screening for inhibitors of the lyase activity of DNA pol β . Our library of screening compounds comprised of the National Cancer Institute (NCI) diversity set, DrugBank small molecules and a set of ~9000 small fragments with drug-like properties. The full set of compounds (~12,500) has been screened against 10 pol β structures. AUTODOCK was used to place the compounds within the specified binding site and to search for their minimal energy conformations. Then, the irredundant top 300 hits from AUTODOCK screening were rescored using the MM-PBSA method. We used pamoic acid (PA), a well-known pol β -inhibitor, as our positive control. This is because it was the subject of recent extensive studies and the information on its binding to pol β has been determined with a high degree of accuracy. Although, more than 12,500 compounds have been screened in this study, we suggest a future use of larger libraries (on the order of 100,000–1,000,000 compounds) may be even more successful in discovering higher affinity hits.

Our results confirmed the experimental findings concerning the binding of PA to the DNA binding cleft within the 8-kDa domain of pol β with an affinity that is close to the reported experimental data. Furthermore, we suggested a set of compounds that can target the DNA-binding site within the 8-kDa with higher affinity than PA. It is hoped that our results will eventually be used in the design of more potent and particularly pol β -specific inhibitors, aimed at improving existing cancer therapies including ionizing radiation, bleomycin, monofunctional alkylating agents and cisplatin.

Acknowledgments

All of the molecular dynamics simulations and virtual screening experiments were produced using the SHARCNET, AICT (University of Alberta cluster) and WESTGRID computational facilities. Funding

for this work was obtained from NSERC, Alberta Cancer Foundation, Alberta Cancer Research Institute, Canadian Breast Cancer Foundation, Alberta Advanced Education and Technology and the Allard Foundation.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2010.12.003.

References

- [1] G. Lehne, E. Elonen, M. Baekelandt, T. Skovsgaard, C. Peterson, Challenging drug resistance in cancer therapy – review of the First Nordic Conference on Chemoresistance in Cancer Treatment, October 9th and 10th, 1997, *Acta Oncol.* 37 (5) (1998) 431–439.
- [2] J.W. Harper, S.J. Elledge, The DNA damage response: ten years after, *Mol. Cell* 28 (5) (2007) 739–745.
- [3] B. Seruga, P.C. Hertz, L.W. Le, I.F. Tannock, Global drug development in cancer: a cross-sectional study of clinical trial registries, *Ann. Oncol.* (2009).
- [4] G. Xu, M. Herzig, V. Rotrek, C.A. Walter, Base excision repair, aging and health span, *Mech. Ageing Dev.* 129 (7–8) (2008) 366–382.
- [5] G. Damia, M. D'Incalci, Targeting DNA repair as a promising approach in cancer therapy, *Eur. J. Cancer* 43 (12) (2007) 1791–1801.
- [6] G.Y. Liu, Q.X. Qu, R.R. Mi, J. Qi, Influence of mifepristone on DNA repair genes and cisplatin sensitivity in human ovarian cancer drug-resistance cells, *Zhonghua Fu Chan Ke Za Zhi* 43 (2) (2008) 132–135.
- [7] L. Liu, Y. Nakatsuru, S.L. Gerson, Base excision repair as a therapeutic target in colon cancer, *Clin. Cancer Res.* 8 (9) (2002) 2985–2991.
- [8] J.S. Hoffmann, M.J. Pillaire, D. Garcia-Estefania, S. Lapalu, G. Villani, In vitro bypass replication of the cisplatin-d(GpG) lesion by calf thymus DNA polymerase beta and human immunodeficiency virus type I reverse transcriptase is highly mutagenic, *J. Biol. Chem.* 271 (26) (1996) 15386–15392.
- [9] V. Bergoglio, Y. Canitrot, L. Hogarth, L. Minto, S.B. Howell, C. Cazaux, J.S. Hoffmann, Enhanced expression and activity of DNA polymerase beta in human ovarian tumor cells: impact on sensitivity towards antitumor agents, *Oncogene* 20 (43) (2001) 6181–6187.
- [10] D. Starcevic, S. Dalal, J.B. Sweasy, Is there a link between DNA polymerase beta and cancer? *Cell Cycle* 3 (8) (2004) 998–1001.
- [11] K. Chan, S. Houlbrook, Q.M. Zhang, M. Harrison, I.D. Hickson, G.L. Dianov, Overexpression of DNA polymerase beta results in an increased rate of frameshift mutations during base excision repair, *Mutagenesis* 22 (3) (2007) 183–188.
- [12] H. Gu, J.D. Marth, P.C. Orban, H. Mossman, K. Rajewsky, Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting, *Science* 265 (5168) (1994) 103–106.
- [13] S.M. Hecht, Inhibitors of the lyase activity of DNA polymerase b, *Pharmaceut. Biol.* 41 (2003) 10.
- [14] K. Barakat, M. Gajewski, J. Tuszynski, DNA polymerase beta (pol β) inhibitors: a comprehensive overview (in press).
- [15] I. Husain, B.S. Morton, W.A. Beard, R.K. Singhal, R. Prasad, S.H. Wilson, J.M. Besterman, Specific inhibition of DNA polymerase beta by its 14 kDa domain: role of single- and double-stranded DNA binding and 5'-phosphate recognition, *Nucleic Acids Res.* 23 (9) (1995) 1597–1603.
- [16] Y. Mizushima, N. Tanaka, H. Yagi, T. Kurosawa, M. Onoue, H. Seto, T. Horie, N. Aoyagi, M. Yamaoka, A. Matsukage, S. Yoshida, K. Sakaguchi, Fatty acids selectively inhibit eukaryotic DNA polymerase activities in vitro, *Biochim. Biophys. Acta* 1308 (3) (1996) 256–262.
- [17] N. Tanaka, A. Kitamura, Y. Mizushima, F. Sugawara, K. Sakaguchi, Fomitellid acids, triterpenoid inhibitors of eukaryotic DNA polymerases from a basidiomycete, *fomitella fraxinea*, *J. Nat. Prod.* 61 (9) (1998) 1180.
- [18] Y. Mizushima, I. Watanabe, K. Ohta, M. Takemura, H. Sahara, N. Takahashi, S. Gasa, F. Sugawara, A. Matsukage, S. Yoshida, K. Sakaguchi, Studies on inhibitors of mammalian DNA polymerase alpha and beta: sulfolipids from a peridophyte, *Athyrium niponicum*, *Biochem. Pharmacol.* 55 (4) (1998) 537–541.
- [19] A. Ogawa, T. Murate, S. Izuta, M. Takemura, K. Furuta, J. Kobayashi, T. Kamikawa, Y. Nimura, S. Yoshida, Sulfated glycolipid from archaeobacterium inhibits eukaryotic DNA polymerase alpha, beta and retroviral reverse transcriptase and affects methyl methanesulfonate cytotoxicity, *Int. J. Cancer* 76 (4) (1998) 512–518.
- [20] A. Ogawa, T. Murate, M. Suzuki, Y. Nimura, S. Yoshida, Lithocholic acid, a putative tumor promoter, inhibits mammalian DNA polymerase beta, *Jpn. J. Cancer Res.* 89 (11) (1998) 1154–1159.
- [21] M. Perpelescu, J. Kobayashi, M. Furuta, Y. Ito, S. Izuta, M. Takemura, M. Suzuki, S. Yoshida, Novel phenalenone derivatives from a marine-derived fungus exhibit distinct inhibition spectra against eukaryotic DNA polymerases, *Biochemistry* 41 (24) (2002) 7610–7616.
- [22] J.Z. Chen, Y. Wang, L. Sucheck, S. Snow, A.S. Hecht, Inhibitors of DNA polymerase b from *Schoepfia Californica*, *J. C. S. Chem. Commun.* (1998) 2769–2770.
- [23] J.Z. Deng, S.R. Starck, S.M. Hecht, C.F. James, M.E. Hemling, Harbinatic acid, a novel and potent DNA polymerase beta inhibitor from *Hardwickia binata*, *J. Nat. Prod.* 62 (7) (1999) 1000–1002.

- [24] D.J. Maloney, J.Z. Deng, S.R. Starck, Z. Gao, S.M. Hecht, (+)-Myristinin A, a naturally occurring DNA polymerase beta inhibitor and potent DNA-damaging agent, *J. Am. Chem. Soc.* 127 (12) (2005) 4140–4141.
- [25] H.Y. Hu, J.K. Horton, M.R. Gryk, R. Prasad, J.M. Naron, D.A. Sun, S.M. Hecht, S.H. Wilson, G.P. Mullen, Identification of small molecule synthetic inhibitors of DNA polymerase beta by NMR chemical shift mapping, *J. Biol. Chem.* 279 (38) (2004) 39736–39744.
- [26] R.T. Kroemer, Structure-based drug design: docking and scoring, *Curr. Protein Pept. Sci.* 8 (4) (2007) 312–328.
- [27] C.W. Murray, C.A. Baxter, A.D. Frenkel, The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase, *J. Comput. Aided Mol. Des.* 13 (6) (1999) 547–562.
- [28] J.H. Lin, A.L. Perryman, J.R. Schames, J.A. McCammon, Computational drug design accommodating receptor flexibility: the relaxed complex scheme, *J. Am. Chem. Soc.* 124 (20) (2002) 5632–5633.
- [29] J.H. Lin, A.L. Perryman, J.R. Schames, J.A. McCammon, The relaxed complex method: accommodating receptor flexibility for drug design with an improved scoring scheme, *Biopolymers* 68 (1) (2003) 47–62.
- [30] K. Barakat, J. Mane, D. Friesen, J. Tuszynski, Ensemble-based virtual screening reveals dual-inhibitors for the p53-MDM2/MDMX interactions, *J. Mol. Graph. Model.* 28 (6) (2010) 555–568.
- [31] K.H. Barakat, J. Torin Huzil, T. Luchko, L. Jordheim, C. Dumontet, J. Tuszynski, Characterization of an inhibitory dynamic pharmacophore for the ERCC1-XPA interaction using a combined molecular dynamics and virtual screening approach, *J. Mol. Graph. Model.* 28 (2) (2009) 113–130.
- [32] M. Markowitz, B.Y. Nguyen, F. Gotuzzo, F. Mendo, W. Ratanasuwana, C. Kovacs, J. Zhao, L. Gilde, R.H.T. Isaacs, Potent antiviral effect of MK-0518, novel HIV-1 integrase inhibitor, as part of combination ART in treatment-naïve HIV-1 infected patients, in: XVI International AIDS Conference, 16th International AIDS Conference, Toronto, Canada, 2006.
- [33] M.M. Garrett, S.G. David, S.H. Robert, H. Ruth, E.H. William, K.B. Richard, J.S. Arthur, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *J. Comput. Chem.* 19 (1999) 1639.
- [34] M.W. Maciejewski, D. Liu, R. Prasad, S.H. Wilson, G.P. Mullen, Backbone dynamics and refined solution structure of the N-terminal domain of DNA polymerase beta. Correlation with DNA binding and dRP lyase activity, *J. Mol. Biol.* 296 (1) (2000) 229–253.
- [35] P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D.A. Case, T.E. Cheatham, Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum model, *Acc. Chem. Res.* 33 (2000) 889.
- [36] W.A. Beard, S.H. Wilson, Structure and mechanism of DNA polymerase beta, *Chem. Rev.* 106 (2) (2006) 361–382.
- [37] (a) S.H. Wilson, Mammalian base excision repair and DNA polymerase beta, *Mutat. Res.* 407 (3) (1998) 203–215; (b) W.A. Beard, S.H. Wilson, Structural design of a eukaryotic DNA repair polymerase: DNA polymerase beta, *Mutat. Res.* 460 (3–4) (2000) 231–244.
- [38] (a) W.A. Beard, S.H. Wilson, Purification and domain-mapping of mammalian DNA polymerase beta, *Meth. Enzymol.* 262 (1995) 98–107; (b) A. Kumar, S.G. Widen, K.R. Williams, P. Kedar, R.L. Karpel, S.H. Wilson, Studies of the domain structure of mammalian DNA polymerase beta. Identification of a discrete template binding domain, *J. Biol. Chem.* 265 (4) (1990) 2124–2131.
- [39] P.M. Burgers, E.V. Koonin, E. Bruford, L. Blanco, K.C. Burtis, M.F. Christman, W.C. Copeland, E.C. Friedberg, F. Hanaoka, D.C. Hinkle, C.W. Lawrence, M. Nakanishi, H. Ohmori, L. Prakash, S. Prakash, C.A. Reynaud, A. Sugino, T. Todo, Z. Wang, J.C. Weill, R. Woodgate, Eukaryotic DNA polymerases: proposal for a revised nomenclature, *J. Biol. Chem.* 276 (47) (2001) 43487–43490.
- [40] H. Pelletier, M.R. Sawaya, W. Wolffe, S.H. Wilson, J. Kraut, A structural basis for metal ion mutagenicity and nucleotide selectivity in human DNA polymerase beta, *Biochemistry* 35 (39) (1996) 12762–12777.
- [41] Y. Uchiyama, R. Takeuchi, H. Kodera, K. Sakaguchi, Distribution and roles of X-family DNA polymerases in eukaryotes, *Biochimie* 91 (2) (2009) 165–170.
- [42] M.R. Sawaya, H. Pelletier, A. Kumar, S.H. Wilson, J. Kraut, Crystal structure of rat DNA polymerase beta: evidence for a common polymerase mechanism, *Science* 264 (5167) (1994) 1930–1935.
- [43] H. Pelletier, M.R. Sawaya, A. Kumar, S.H. Wilson, J. Kraut, Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and dCTP, *Science* 264 (5167) (1994) 1891–1903.
- [44] H. Pelletier, M.R. Sawaya, Characterization of the metal ion binding helix-hairpin-helix motifs in human DNA polymerase beta by X-ray structural analysis, *Biochemistry* 35 (39) (1996) 12778–12787.
- [45] C. Hazan, F. Boudsocq, V. Gervais, O. Saurel, M. Ciais, C. Cazaux, J. Czaplicki, A. Milon, Structural insights on the pamoic acid and the 8 kDa domain of DNA polymerase beta complex: towards the design of higher-affinity inhibitors, *BMC Struct. Biol.* 8 (2008) 22.
- [46] A.E. Garcia, Large-amplitude nonlinear motions in proteins, *Phys. Rev. Lett.* 68 (17) (1992) 2696–2699.
- [47] A. Amadei, A.B. Linssen, H.J. Berendsen, Essential dynamics of proteins, *Proteins* 17 (4) (1993) 412–425.
- [48] D. Tondi, U. Slomczynska, M.P. Costi, D.M. Watterson, S. Ghelli, B.K. Shoichet, Structure-based discovery and in-parallel optimization of novel competitive inhibitors of thymidylate synthase, *Chem. Biol.* 6 (5) (1999) 319–331.
- [49] I. Halperin, B. Ma, H. Wolfson, R. Nussinov, Principles of docking: an overview of search algorithms and a guide to scoring functions, *Proteins* 47 (4) (2002) 409–443.
- [50] (a) B.K. Shoichet, R.M. Stroud, D.V. Santi, I.D. Kuntz, K.M. Perry, Structure-based discovery of inhibitors of thymidylate synthase, *Science (New York, NY)* 259 (5100) (1993) 1445–1450; (b) G. Schneider, H.J. Bohm, Virtual screening and fast automated docking methods, *Drug Discov. Today* 7 (1) (2002) 64–70; (c) R. Abagyan, M. Totrov, High-throughput docking for lead generation, *Curr. Opin. Chem. Biol.* 5 (4) (2001) 375–382.
- [51] A.S. Reddy, S.P. Pati, P.P. Kumar, H.N. Pradeep, G.N. Sastry, Virtual screening in drug discovery – a computational perspective, *Curr. Protein Peptide Sci.* 8 (4) (2007) 329–351.
- [52] P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D.A. Case, T.E. Cheatham 3rd, Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models, *Acc. Chem. Res.* 33 (12) (2000) 889–897.
- [53] B. Kuhn, P. Gerber, T. Schulz-Gasch, M. Stahl, Validation and use of the MM-PBSA approach for drug discovery, *J. Med. Chem.* 48 (12) (2005) 4040–4048.
- [54] K. Laxmikant, S. Robert, B. Milind, B. Robert, G. Attila, K. Neal, P. James, S. Aritomo, V. Krishnan, NAMD2: greater scalability for parallel molecular dynamics, *J. Comput. Phys.* 151 (1999) 283–312.
- [55] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple Amber force fields and development of improved protein backbone parameters, *Proteins* 65 (3) (2006) 712–725.
- [56] T.J. Dolinsky, P. Czodrowski, H. Li, J.E. Nielsen, J.H. Jensen, G. Klebe, N.A. Baker, PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations, *Nucleic Acids Res.* 35 (2007) (Web Server issue), W522–5.
- [57] D.A. Case, T.E. Cheatham 3rd, T. Darden, H. Gohlke, R. Luo, K.M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, The Amber biomolecular simulation programs, *J. Comput. Chem.* 26 (16) (2005) 1668–1688.
- [58] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, *J. Comput. Chem.* 25 (9) (2004) 1157–1174.
- [59] A. Jakalian, D.B. Jack, C.I. Bayly, Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation, *J. Comput. Chem.* 23 (16) (2002) 1623–1641.
- [60] L.S. Cheng, R.E. Amaro, D. Xu, W.W. Li, P.W. Arzberger, J.A. McCammon, Ensemble-based virtual screening reveals potential novel antiviral compounds for avian influenza neuraminidase, *J. Med. Chem.* 51 (13) (2008) 3878–3894.
- [61] R.M. Knegtel, I.D. Kuntz, C.M. Oshiro, Molecular docking to ensembles of protein structures, *J. Mol. Biol.* 266 (2) (1997) 424–440.
- [62] J. Shao, S.W. Tanner, N. Thompson, T.E. Cheatham, Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms, *J. Chem. Theory Comput.* 3 (2007) 2312.
- [63] D.L. Davies, Bouldin DW, A cluster separation measure, *IEEE Trans. Pattern Anal. Mach. Intell.* 1 (1979) 224.
- [64] B. Hess, Convergence of sampling in protein simulations, *Phys. Rev. E: Stat. Nonlinear Soft Matter Phys.* 65 (3 Pt 1) (2002) 031910.
- [65] <http://dtp.nci.nih.gov/branches/dscb/diversity.explanation.html> (last checked May, 20, 2010).
- [66] <http://zinc.docking.org/vendor0/dbsm/index.html> (last checked May 20, 2010).
- [67] J. Gasteiger, M. Marsili, Iterative partial equalization of orbital electronegativity: a rapid access to atomic charges, *Tetrahedron* 36 (1980) 3219.
- [68] K. Bebenek, T.A. Kunkel, Functions of DNA polymerases, *Adv. Protein Chem.* 69 (2004) 137–165.