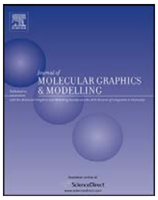
Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM



Characterization of an inhibitory dynamic pharmacophore for the ERCC1–XPA interaction using a combined molecular dynamics and virtual screening approach

Khaled H. Barakat [a], J. Torin Huzil [b], Tyler Luchko [a], Lars Jordheim [c], Charles Dumontet [c], Jack Tuszynski [a,b,*]

[a] Department of Physics, University of Alberta, Edmonton, AB, Canada
[b] Department of Oncology, University of Alberta, Edmonton, AB, Canada
[c] Faculté de Médecine, Université Claude Bernard Lyon 1, Lyon, France

ARTICLE INFO

Article history:
Received 22 November 2008
Received in revised form 17 April 2009
Accepted 22 April 2009
Available online 3 May 2009

Keywords:
Virtual screening
Clustering
NCI dtp database
DrugBank
UCN01
XPA
ERCC1
Pharmacophore

ABSTRACT

Combination chemotherapy involving Cisplatin is a standard treatment for many cancers. However, following an initial positive response, patients will often relapse, presenting with Cisplatin-resistant disease. One possible mechanism for the acquired resistance to Cisplatin is an increase in DNA repair through the up-regulation of ERCC1, an essential component of the nucleotide excision repair complex. Recruitment of ERCC1 to the site of DNA damage is coordinated through its interaction with a protein known as XPA. As there are currently no effective inhibitors of this interaction, inhibition of the ERCC1/XPA interaction may provide an effective strategy for overcoming the development of Cisplatin-resistant cancers. To discover small molecule inhibitors of this interaction, we have screened both the NCI diversity set of ligands and DrugBank-small molecules against the XPA binding site in ERCC1. These compounds were screened using two different techniques in AUTODOCK to account for receptor flexibility. First, using a set of flexible residues, as determined from MD simulations of the XPA/ERCC1 complex and second, using the relaxed complex scheme implemented by performing independent docking experiments against an ensemble of target conformations that were generated from MD simulations. Lowest energy poses from the two different methods were then used to construct a pharmacophore model, which was then validated by comparison to UCN-01, a weak inhibitor of ERCC1 mediated nucleotide excision.

1. Introduction

Cisplatin, cis-diaminedichloroplatium(II), is a component of standard therapy regimens for numerous cancers [1,2]. Cisplatin is a member of a large group of platinating agents that covalently bind DNA, most often producing 1,2-intrastrand cross-links between purine bases within the DNA double helix [3,4]. The formation of Cisplatin–DNA adducts, often as little as one adduct per 100,000–500,000 nucleotides, results in significant conformational changes to the underlying DNA structure [5] and eventually leads to the non-specific inhibition of DNA synthesis and cellular apoptosis [6–8].

While generally efficient at inducing apoptosis, acquired resistance to platinum compounds has limited their efficacy and therefore use of these agents [9,10]. Many factors are implicated in the development of resistance to DNA modifying agents; however the up-regulation of specific DNA repair mechanisms is emerging as a significant player in this process [11–13]. In particular, nucleotide excision repair (NER) is thought to be a key pathway involved in mediating resistance to platinum chemotherapy treatment [11]. This is supported by the observation that cancers, initially responsive to Cisplatin later develop resistance, exhibiting increased NER activity [13]. Cisplatin treatment is also extremely effective at killing those cells which lack functional NER [14,15].

Resistance to Cisplatin has been previously correlated with the over expression of both the excision repair cross-complementation group 1 (ERCC1) and xeroderma pigmentosum complementation group F (XPF) proteins [15–17]. These proteins form a hetero-dimeric endonuclease complex, which is recruited to DNA through a secondary interaction between ERCC1 and the xeroderma pigmentosum complementation group A (XPA) protein [18–20]. Although ERCC1 has no intrinsic enzymatic activity, it is possible to

[*] Corresponding author at: Division of Experimental Oncology, Cross Cancer Institute, University of Alberta, 11560 University Avenue, Edmonton, AB, T6G-1Z2, Canada. Tel.: +1 780 432 8906; fax: +1 780 432 8892.
E-mail address: jtus@phys.ualberta.ca (J. Tuszynski).

1093-3263/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2009.04.009

regulate its role in NER by targeting interactions within the DNA-bound complex. Here, we have chosen to focus on the interaction between ERCC1 and XPA for several reasons. No cellular function beyond NER has been observed for XPA [23] and competitive inhibition of the XPA interaction with peptide fragments is effective at disrupting NER [24]. Also, clinically, patients that have been shown to have low expression levels of either XPA or ERCC1 demonstrate higher sensitivity to Cisplatin treatment [21,22]. These observations, coupled with an available crystal structure of this interaction [24] make ERCC1 and XPA an extremely attractive target for computationally based development of small molecule inhibitors that are targeted for use in combination therapies involving Cisplatin.

Structure-based drug design of inhibitors to the ERCC1/XPA interaction requires the implementation of several computational techniques as applied to atomic-level interactions within the complex. Due to advances in software and high performance computing, this approach has undergone notable development over the past 30 years [25]. For example, the past decade has witnessed the development of receptor flexibility and a variety of scoring functions for different screening packages [26,27]. Even so, introducing efficient models for receptor flexibility into docking algorithms is still a very active area of research [28–31]. For instance, a ligand can induce significant conformational changes to its target, ranging from local reorganization of side-chains to hinge movement of domains [32]. Sampling these conformational changes during docking is unpractical as they involve numerous degrees of freedom. Additionally, screening massive databases containing large numbers of compounds that are not expected to bind to a particular target may reveal misleading results, commonly known as "decoy" compounds. Introducing receptor flexibility may therefore prevent docking methods from selecting correct ligands from a large list of molecules during virtual screening as it allows, not only true ligands, but also decoys to fit into the binding pocket of a particular target, diluting the ensemble of true ligands [33].

The relaxed complex scheme (RCS) [31] is a hybrid method that combines docking algorithms with molecular dynamics (MD) simulations. In this technique, MD simulations are applied in order to efficiently explore the conformational space of the protein receptor, while docking is subsequently used for the fast screening of drug libraries against a pre-selected ensemble of receptor conformations. This methodology has already been successfully applied to a number of cases, including the HIV inhibitor raltegravir [31,34]. Here, we have utilized the RCS technique to describe the construction of a "dynamic pharmacophore model" targeting the ERCC1–XPA interaction.

We utilized a minimized model of the XPA binding site within ERCC1 to employ flexible residue docking as implemented in AUTODOCK 4.0. This was then followed with RCS docking, where MD simulations and root-mean-square difference (RMSD) conformational clustering were used to generate a set of forty-four representative conformations of the binding site within ERCC1. AUTODOCK was then used to screen the National Cancer Institute Diversity Set (NCIDS) and DrugBank compounds against a set of seven target conformations, composed of the six most dominant cluster-representative structures along with an equilibrated folded conformation for the binding site produced by employing principal component analysis on the ERCC1 trajectory. Top hits were rescored by docking them to the whole set of cluster-representative structures and ranked by their weighted average binding energy. The non-redundant hits from these screens were then used to identify dynamic binding-site pharmacophores that target the ERCC1–XPA interaction. The pharmacophore model was then compared to docking results for a weak inhibitor of NER, UCN-01 (7-hydroxystaurosporine) [36]. These results will undoubtedly be beneficial in the future design of specific inhibitors of the ERCC1/XPA interaction, producing novel combination chemotherapy treatments aimed at overcoming acquired resistance to platinating compounds.

2. Results and discussion

Exposure to platinum-based agents, such as Cisplatin, results in the up-regulation of ERCC1 expression [15–17], resulting in the rapid removal of the DNA adducts induced by platinum-based chemotherapy [11]. The availability of the crystallographic structure of ERCC1 in a complex with a peptide from XPA, coupled with the observation that this small XPA fragment is capable of interacting with ERCC1 to form a stable complex exhibiting submicromolar-binding affinity and potent inhibition of NER activity [24] provides an excellent scaffold on which to construct inhibitors of this interaction. Results described herein discuss the construction and validation of model pharmacophores designed to aid in the design of inhibitors of ERCC1–XPA interactions and the subsequent recruitment of the ERCC1–XPF endonuclease complex to the sites of DNA damage. Our primary use of docking techniques has been to filter large compound libraries through receptor-based virtual screening, in an attempt to identify molecules that complement the XPA binding site in terms of such parameters as shape, charge, and several additional biochemical characteristics.

2.1. Molecular dynamics simulations of the ERCC1–XPA interaction

To obtain a minimized model for library screenings and a set of flexible residues for docking, the central domain of ERCC1 was subjected to MD simulations, in both its free and XPA bound states [24]. 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 XPA binding site within ERCC1. Moreover, it is generally required to start with adequately sampled, energetically minimized models in order to eliminate unfavorable atom contacts that may have been introduced as a result of crystal packing in the original structure.

2.2. Principal component analysis and completeness of sampling

As MD simulations produce numerous conformations of the ERCC1 and XPA proteins, 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 [43,44]. To perform PCA, the entire MD trajectory must first be RMSD fitted to a reference structure and a covariance matrix is calculated from 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 50 ns simulation using atoms comprising the 22 residues contained in the ERCC1 binding site with the backbone atoms RMSD fitted to the minimized crystal structure.

Covariant analysis of the trajectories from the ERCC1-free MD simulations, successively divided into thirds, was performed using the same procedure used for the PCA. Normalized overlaps calculated between each of these thirds are reported in Table 1. The high overlap between the thirds indicates that each part of the

Table 1
PCA normalized overlap for the binding site within ERCC1.

1st vs. 2nd	0.87
1st vs. 3rd	0.86
2nd vs. 3rd	0.87

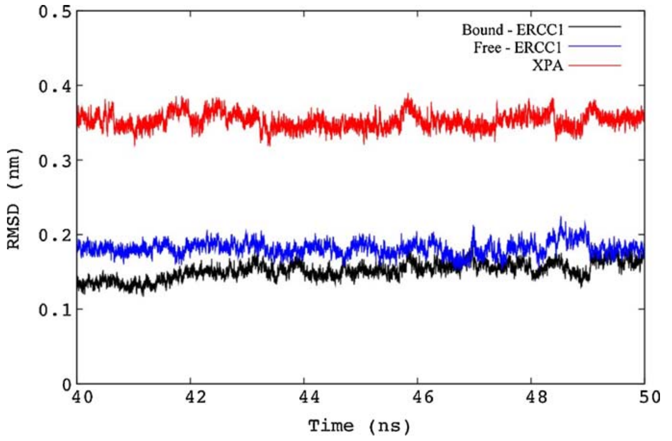


Fig. 1. Plot of the RMSD of the backbone atoms from the reference structure as a function of simulation time in XPA-peptide, ERCC1-free and XPA–ERCC1 complex.

simulation samples approximately the same conformational space, and it is unlikely that there are unexplored regions missed earlier in the run. Although there is no guarantee that complete equilibrium sampling is given, we have concluded that the observed overlap is acceptable and that adequate sampling within the MD trajectory for the binding site had been obtained.

Plots of the RMSDs for the backbone atoms from the initial coordinates of the XPA peptide and ERCC1-free and -bound to XPA for the last 10 ns of the simulation illustrate the inherent stability of the complex (see Fig. 1). For the XPA-free simulations, the protein backbone RMSD fluctuated about a mean of 1.8 Å. When XPA was bound to ERCC1, the protein backbone RMSDs fluctuated about a mean of 1.6 Å, indicating a stabilizing effect induced by XPA interactions within ERCC1. Relative to values calculated for the ERCC1 backbone, the XPA backbone RMSD fluctuated around a higher mean of 3.5 Å, illustrating the greater mobility of the XPA peptide as compared to the ERCC1 protein. This observation was also confirmed by results presented in Fig. 2a, where unbound ERCC1 main-chain B-factors (averaged over heavy atoms) are generally higher than the corresponding bound values. This, again, suggests the relative flexibility of the model in this region, especially residues 105–119, 140–160 and 168–177 which constitute the XPA binding site. Within the XPA–ERCC1 model, residues 178–183 were shown to have more flexibility than those in the free model suggesting that they are not involved in the protein–protein interaction. Overall, the 22 residues defining the binding site seem to be relatively rigid during the MD simulation in the XPA–ERCC1 models. In particular, residues 72–75 of XPA (see Fig. 2b) are more rigid than other XPA residues, suggesting their critical participation in binding to ERCC1.

2.3. Energy decomposition of the XPA–ERCC1 interaction

Current docking methodologies provide a mechanism for the inclusion of flexible receptor side chains within the docking grid [27,49]. However, without a clear understanding as to which residues should be introduced as flexible, this can quickly become an intractable problem. As there was no specific information with regards to which residues contribute to the ERCC1/XPA binding interaction, we have calculated the free energy of binding for each residue in ERCC1 that has been shown to interact with XPA. Equilibration of both the holo and apo forms of the ERCC1 binding site allowed us to obtain free energy profiles for each of the amino acids involved in the ERCC1–XPA interaction (see Table 2 and Fig. 3c).

The binding between the ERCC1 binding site and the XPA peptide is due mainly to the favorable solute–solute electrostatic and hydrophobic interactions ($\Delta E^{ele}_{gas} \approx -42$ kcal/mol, and $\Delta E_{gas} \approx -39$ kcal/mol), which outweighed the unfavorable solute–solvent electrostatic interaction ($\Delta G^{ele}_{solv} \approx 45$ kcal/mol) by ≈−36 kcal/mole. Although the non-polar contribution to the solvation energy ($\Delta G^{nonele}_{solv} \approx -6$ kcal/mol) is favorable for binding, it does not contribute significantly to the binding affinity. The most significant binding contributions between ERCC1 and XPA were determined to be mediated primarily by five residues from the XPA peptide, namely, G72, G73, G74, F75 and I76, contributing

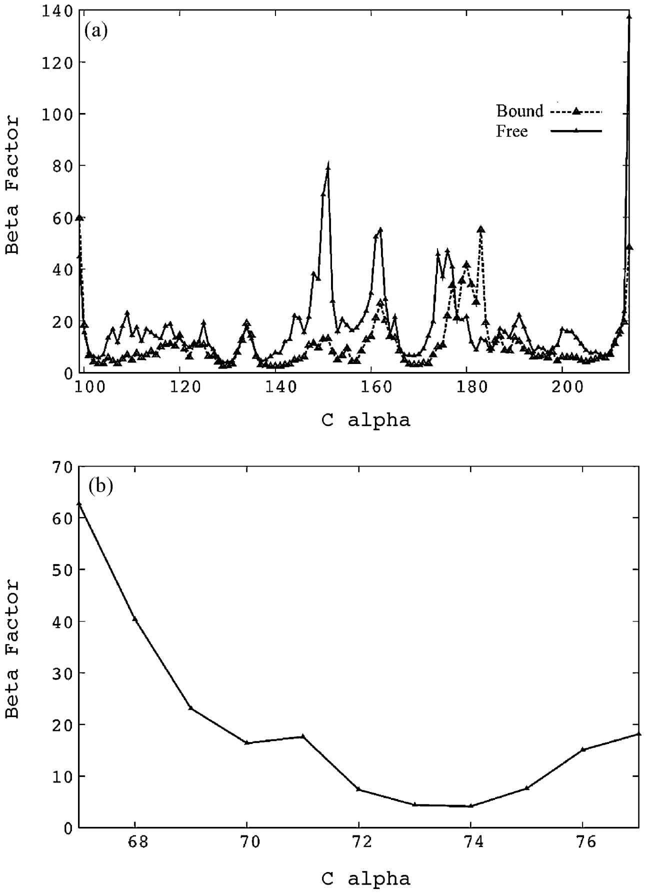


Fig. 2. Plot of the B-factors averaged over the protein backbone atoms as a function of residue number in the simulations of (a) ERCC1-free and ERCC1-bound and (b) XPA peptide. The solid and dotted lines correspond to ERCC1-free and ERCC1-bound, respectively.

Table 2
Binding energy decomposition into key residues mediating the XPA–ERCC1 interaction.

	Residue i	ΔG_i	$\Delta E^{ele}_{i.gas}$	$\Delta E_{i.gas}$	$\Delta G^{ele}_{i.sol}$	$\Delta G^{nonele}_{i.sol}$
ERCC1	ARG106	−1.71	−7.16	−2.16	8.01	−0.41
	GLN107	−2.51	−0.98	−1.76	0.35	−0.12
	GLY109	−1.27	−0.63	−2.11	1.93	−0.46
	ASN110	−1.96	−0.72	−2.50	1.13	−0.13
	PRO111	−1.36	−0.78	−1.23	0.94	−0.29
	PHE140	−1.66	−0.72	−1.95	1.17	−0.16
	LUE141	−1.90	−1.56	−1.17	0.85	−0.02
	SER142	−1.98	−2.59	−0.82	1.53	−0.10
	TYR145	−5.58	−5.19	−2.92	3.12	−0.57
	TYR152	−1.19	−3.61	−1.14	3.68	−0.12
XPA	GLY72	−3.03	−10.72	−1.15	9.28	−0.44
	GLY73	−3.18	−2.52	−3.33	3.06	−0.39
	GLY74	−4.56	−5.02	−4.62	5.78	−0.70
	PHE75	−6.15	0.54	−8.17	2.24	−0.76
	ILE76	−3.82	−0.79	−4.10	1.86	−0.79
Total energy (kcal/mol)		∼−42	−42.45	−39.13	44.93	−5.46

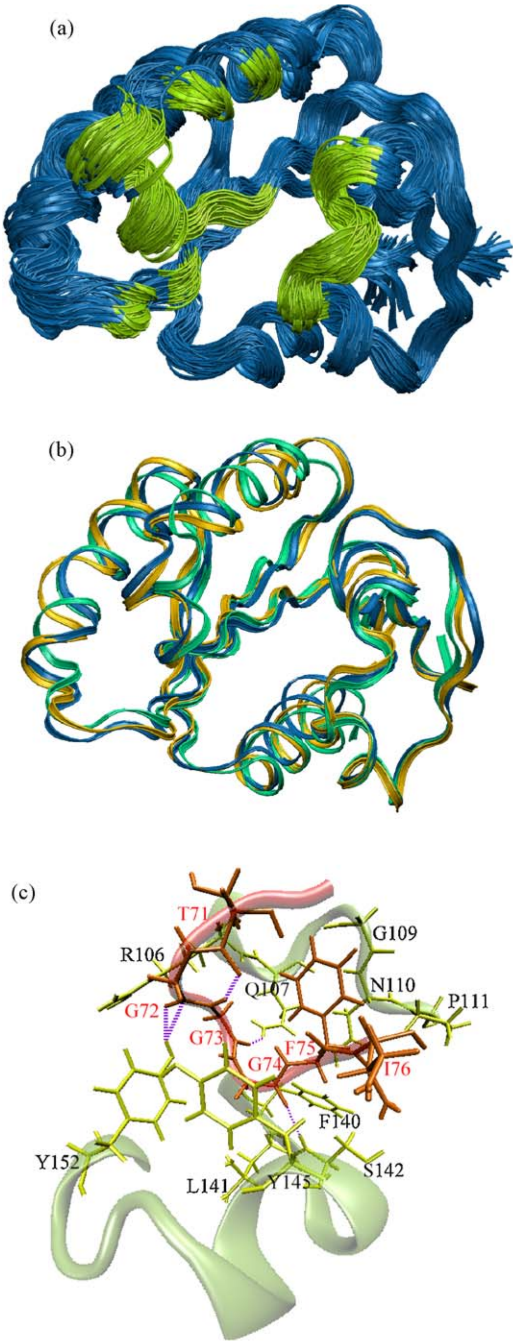


Fig. 3. The interaction between XPA-peptide (red) and ERCC1 binding site. (a) Forty-four representative structures for the ERCC1 binding site. The binding site is colored in green and (b) three representative apo structures: the minimized crystal structure (green), the most dominant cluster representative (blue) and the PCA equilibrated model (yellow), (c) the binding between ERCC1 (teal) and XPA (red) is primarily mediated by 5 residues from XPA peptide, namely; G72, G73, G74, F75 and I76. On the other hand, the contribution from the ERCC1 binding site is distributed among 10 residues; R106, Q107, G109, N110, P111, F140, L141, S142, Y145 and Y152.

approximately 50% of the total binding energy. Within the ERCC1 binding site, Y145 contributed approximately −5 kcal/mol to the binding energy with the remainder of the −42 kcal/mol being distributed among other 8 residues (R106, Q107, G109, N110, P111, L141, S142 and Y152). Overall, the main contributors to the binding energy were Y145 from ERCC1 and F75 from XPA, which stacks against N110 from ERCC1, and contributed ∼−2 kcal/mol (see Fig. 3c). In our model, the hydroxyl group of Y152 in ERCC1 also forms two hydrogen bonds with the backbone carbonyl of G72 in XPA. Two hydrogen bonds were also observed between S142 of ERCC1 and G72 of XPA, and Q107 of ERCC1 and G73 of XPA. An intramolecular hydrogen bond was observed between T71 and G73 within XPA and is in agreement with experimental findings detailing critical residues mediating the XPA–ERCC1 interaction [24]. All of these residues explicitly define the binding site within ERCC1 and therefore the coformation of the potential inhibitor that should mimic the XPA peptide (see Fig. 3c).

2.4. Flexible docking virtual screening

While integrating receptor flexibility into docking reduces the risk of unfavorable interactions between the ligand and its target, accommodating full receptor flexibility during a docking experiment is unpractical [27]. One solution of 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. Although this procedure ignores important changes in the peptide backbone, it allows for localized protein movement, resulting in an improved fit of the ligand. Decomposition of the total binding energy from our models into individual residue contributions allowed us to identify key residues that mediate the ERCC1–XPA interaction. As Y145 contributed about 26% of the total ERCC1 binding energy (see Table 2), for our initial docking runs, we have used the minimized crystal structure of the ERCC1–XPA complex, with Y145 being the only flexible residue during the virtual screening procedure.

2.5. Ensemble-based virtual screening

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 as a target for docking as an alternative approach to introduce a feature of global protein flexibility. Such an ensemble at the extreme, is capable of describing the entire conformational space of the binding site, yet 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 ERCC1 binding site, we applied RMSD conformational clustering to the apo-binding site trajectory obtained from the MD simulation. Using the average-linkage algorithm, we obtained a total of 44 clusters which represent the complete MD trajectory (see Fig. 3a). Of these 44, the six most dominant clusters represented approximately 48%, 8%, 6%, 5.5%, 4% and 3.8% of the entire ensemble, respectively (see Fig. 4). We concluded that these six dominant clusters were sufficient to describe the collective conformational changes in the apo-ERCC1 trajectory for subsequent screening experiments.

The accumulation of approximately half of the MD trajectory conformations into the first dominant cluster motivated us to use PCA in order to extract the lowest energy conformation of the binding site. This conformation was then appended to the six dominant structures to perform an ensemble-based virtual screening against the full set of ligand compounds. Fig. 5 represents the spatial distributions of occupancies for the conformational states over the planes spanned by the dominant principal components of the binding site. The grouping of conformations into a single cluster suggests the presence of a global minimum and a significant basin of attraction, indicative of a low energy conformation for the binding site. A representative structure for the folded conformation was then calculated by

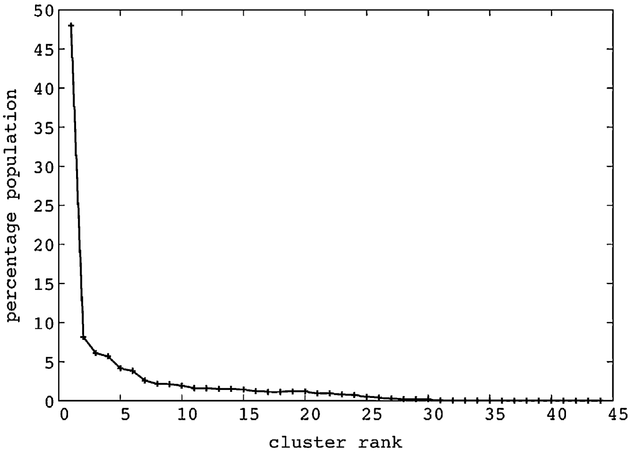


Fig. 4. Percentage population of the 44 clusters generated by RMSD conformational clustering of the apo MD trajectory. The most dominant six clusters constitute about 75% of the whole ensemble and are more populated than the rest of the 44 conformations.

collecting all conformations contained within the three minima. The backbone atoms of the binding site of these conformations were then RMS fitted to the reference structure and the centroid of the RMS fit was used as an additional representative conformation to the six dominant structures in virtual screening experiments.

It is noteworthy that, while the two equilibrated structures were produced through two different methods, the RMSD between the two models was only 1.12 Å, which is quite low when compared to values calculated previously (see Fig. 1). The high degree of similarity between the two conformations (see Fig. 3b) reveals the significance of the PCA methodology as an alternative way to explore the energy landscape and to construct more accurate equilibrium structures.

2.6. Pose clustering

In this study, we performed eight screening experiments against the full set of database compounds. The first was against the minimized crystal structure with a flexible Y145, while the other seven constituted the ensamble based screen. Screening of all 3450 compounds contained in the NCDIS and Drugbank databases, against the eight target structures, produced a total of 2.76 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 population for each cluster is mostly dependent on the RMSD cutoff that is initially chosen. As such, there is no adequate means to anticipate an optimum cutoff for the RMSD to produce the best quality result. As 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 of the ligand 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. This requires 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 [42]. This is because it is simple to implement and 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 [42]. This is illustrated in Fig. 6, where the elbow criterion for the top three hits from ensemble screening suggested different cluster counts for the different poses. This clustering methodology proposed three clusters with three different representative conformations for the planar molecule characterizing the top hit (see Table 3), while 14 clusters were suggested for the third hit which includes more torsional degrees of freedom. We propose this clustering method as an alternative dynamic technique to be used for virtual

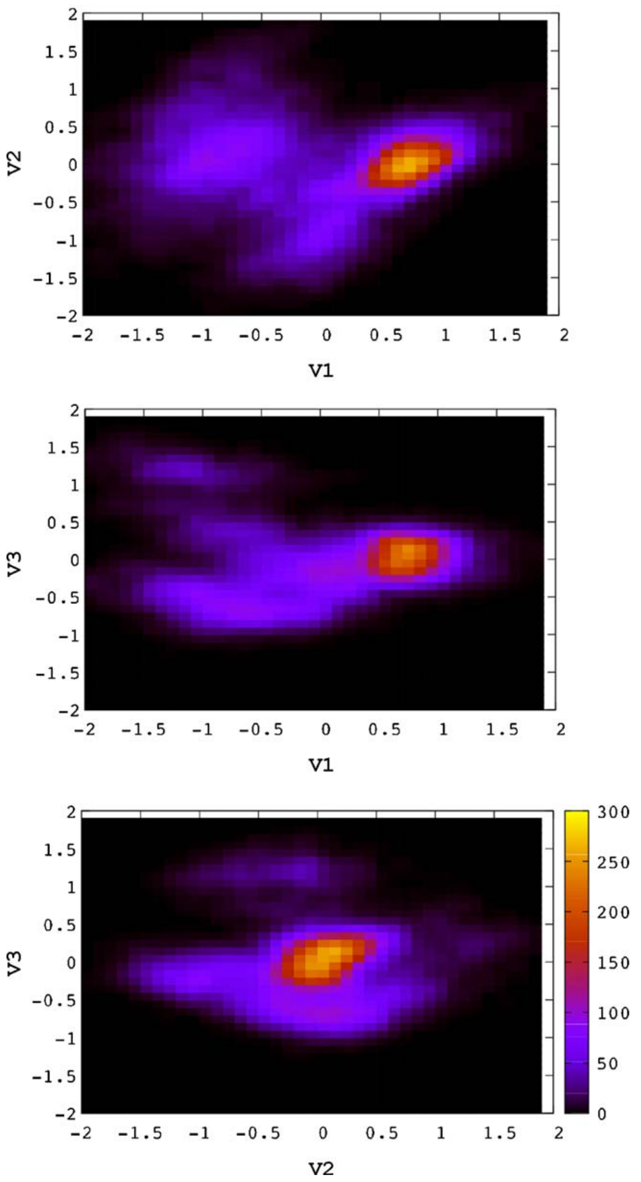


Fig. 5. PCA of the ERCC1 binding site. Projections of the ensemble of conformations onto the planes of the three most important principal components. The first and second, the first and third and the second and third principal components are plotted on the *x* and *y* axes, respectively. The histograms represent the occupancies of the corresponding conformational states, with lighter colors indicating more frequently visited areas. The three histograms reveal a global minimum indicating the convergence of sampling and folding of the ERCC1 binding site.

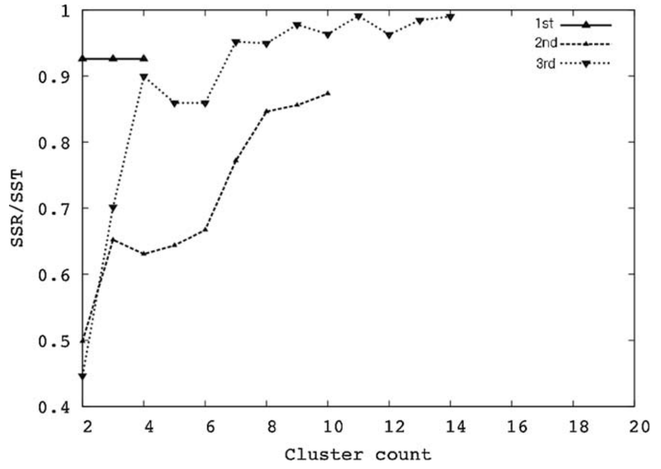


Fig. 6. Percentage of variance for the top 3 hits from RCS VS experiment. The SSR/SST is expected to plateau for cluster counts exceeding the optimal number of clusters.

screening as opposed to clustering all the poses with a single RMSD value for all the docked compounds.

2.7. Pose ranking

For each virtual screening experiment, we have ranked significant poses for each of the 3450 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 are considered as top hits. For the flexible screening experiment, top hits were ranked by their binding energies of the largest cluster. Top hits from the ensemble-based screening experiments were collected from the seven experiments by first extracting the largest cluster from each individual screening flowed by ranking the clusters by their binding energies. This produced a set of non-redundant hits ranked by their binding energies of the most populated cluster.

Table 3 shows the top ten hits from the ensemble-based screening ranked by their binding energies and compared to their docking results from the flexible run. It is clear that the two methods produce dissimilar ranking for most of the compounds, with several hits being excluded from the flexible screening due to the 25% cutoff on the largest cluster population. Although the flexible docking showed low binding energies, even lower than the ensemble based calculations, the poor clustering of these hits suggested that they did not fit properly into the binding pocket. This observation indicates the importance of backbone dynamics and side-chain movement as compared to allowing only one residue to be flexible during docking. In order to refine the ensemble-based screening results and consider all possible target conformations, we docked the top 50 hits obtained from the ensemble docking results to the complete set of receptor

Table 3
Top ten hits from the ensemble screening. The ranking and binding energies are compared to the flexible screening.

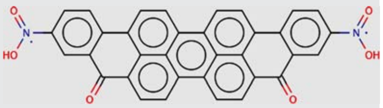
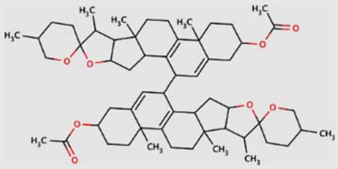
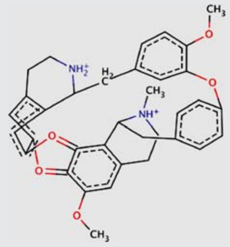
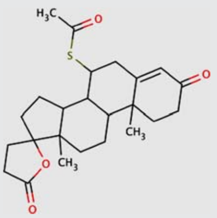
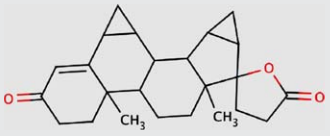
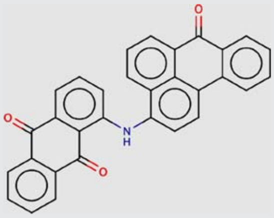
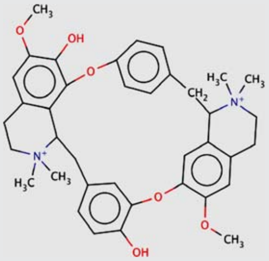
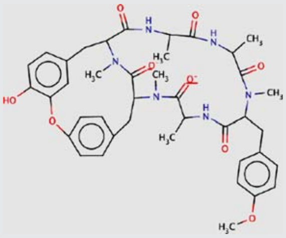
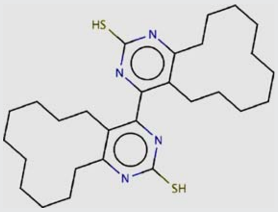
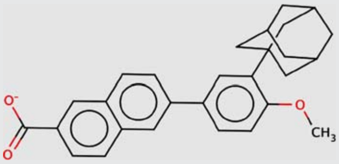
ENS Rank	FLEX rank	ENS mean (kcal/mol)	FLEX BE (kcal/mol)	ID	Structure
1	1	−8.79	−11.67	NSC #51535	
2	EXCLUDED	−8.21	−9.28	NSC #93352	 H3C H3C H3C H3C H3C O O O O H3C O O O O CH3 CH3 CH3 CH3
3	17	−7.55	−9.83	NSC #181486	 CH3 O NH2+ H2 C CH3 NH+ O O O CH3
4	7	−7.51	−10.24	ZINC03861599	
5	57	−7.49	−9.11	ZINC03927200	

Table 3 (*Continued*)

ENS Rank	FLEX rank	ENS mean (kcal/mol)	FLEX BE (kcal/mol)	ID	Structure
6	15	−7.49	−9.88	NSC #13987	
7	EXCLUDED	−7.46	−9.34	NSC #36387	 CH3 O OH O H3C CH3 CH2 N+ N+ H3C CH3 O O CH3 OH
8	76	−7.41	−8.98	NSC #259969	 O H N O H N CH3 H3C N H3C O CH3 N O O N—CH3 HO H N O H3C O H3C—O
9	13	−7.39	−9.95	NSC #372060	
10	8	−7.38	−10.19	ZINC03784182	

representative structures and applied the relaxed complex scheme to re-score the poses.

2.8. Rescoring using the RCS

The non-redundant 50 hits obtained from the ensemble-based screening experiments were re-docked into all of the 44 clusters representing the apo-ERCC1 MD ensemble. For each compound (see Fig. 7) the RCS weighted average and minimum binding energies were compared to the ensemble screening average and minimum binding energies. To compare how these ligands were docked to the crystal structure, we also included the binding energies of the most populated cluster from the flexible screening. For most of the compounds, the RCS weighted-mean and the ensemble-mean differ by less than 0.7 kcal/mol, indicating that our selection of only the first six representative structures was sufficient to describe the conformational space of the binding site (see Fig. 7). Furthermore, those compounds which ranked very low when docked to a number of target structures including the most dominant conformation (−3.52 kcal/mol), could be excluded in the RCS scoring.

While each conformation in the ensemble screening contributed one-seventh of the total binding energy, the relaxed complex scheme evaluates the total energy by scaling individual energies by the percentage population of the docked structures. In this context, for a compound to be ranked high in the RCS score, it must be docked properly against most of the target ensemble, particularly, against the most dominant structures. This scoring technique enables the RCS method to identify and eliminate decoy compounds from the list of ligands. Some hits showed lower RCS minimum-binding energies than their representative values calculated using the ensemble approach, suggesting their binding to a more accepting representative structure. This is the main advantage of using RCS to re-score the top hits; as some compounds may bind to a rarely visited receptor conformation.

Fig. 8 shows the average clustering of the top hits for the three different methods. Once more, the RCS and the ensemble methods showed the same results for most of the ligands. The average

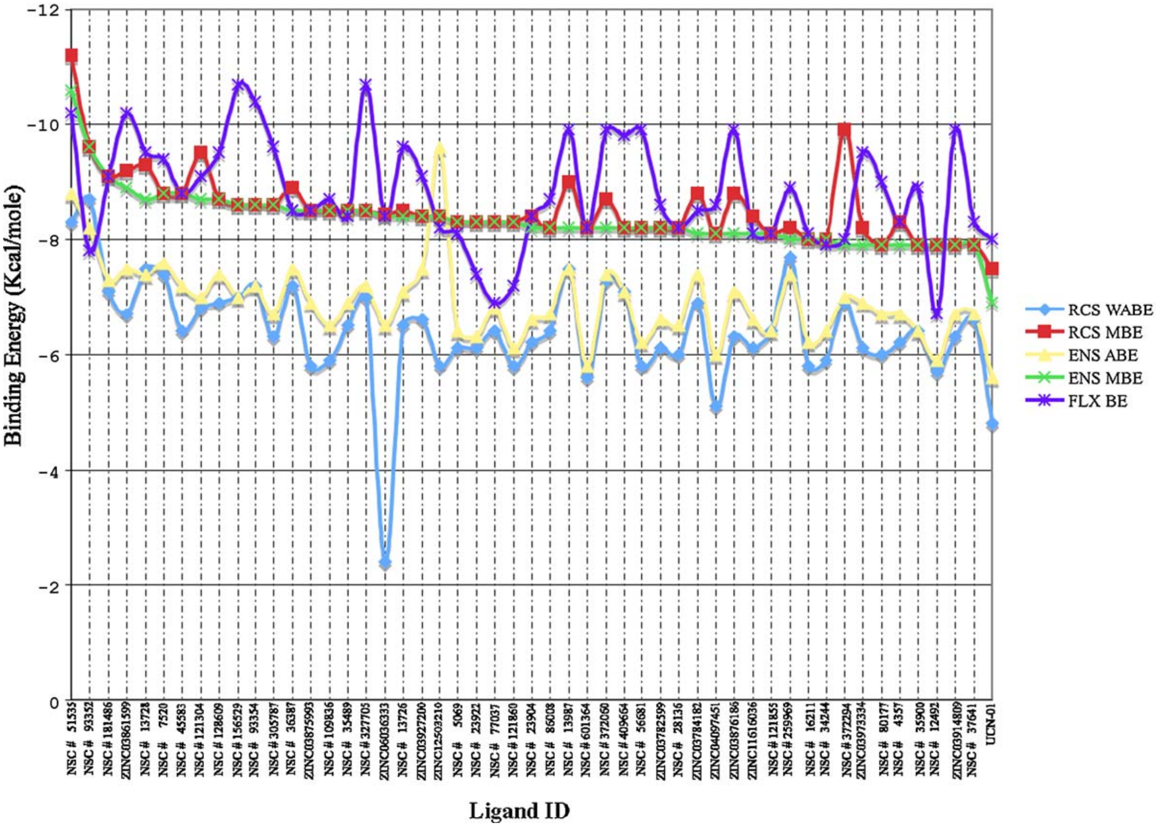


Fig. 7. Binding energy statistics for the irredundant top 50 hits suggested by the ensemble-based screening. The RCS weighted-average (RCS WABE) and the ensemble average binding energies (ENS ABE) showed similar behavior for most of the compounds. The RCS minimum energy (RCS MBE) and the ensemble minimum binding energy (ENS MBE) plateau with small number of hits showing lower BE for the RCS screening indicating their binding to infrequent conformations. The Flexible screening binding energies fluctuated around −9 kcal/mol showing lower binding energies than the other methods. The RCS excluded the ZINC06036333 compound from the top hits suggesting that it is not a true binder.

cluster populations for the two methods are more than 30 poses for about 56% of the top hits, indicating their binding for most of the representative conformations. It is noteworthy that more than 50% of the ensemble screening's hits were excluded from the flexible screening due to the 25% cutoff criterion on the largest cluster population.

2.9. Electrostatic surface calculations

The binding mode for three selected top hits within their most favored binding site conformations is shown in Fig. 9. Electrostatic surface maps are included to provide an additional perspective of the charge distribution in the ERCC1 cavity. The binding cleft is mainly positively charged with small negatively charged spots on boundaries of the binding site. This electrostatic potential distribution indicates that the binding site may exhibit a weak positive electrostatic potential. Although, the charge distribution changed slightly between the two representative binding sites indicating the perseverance of its overall shape, the positive potential is apparent in the closed conformation. Moreover, the charge complementarity between the binding site and the top hits is apparent from Fig. 9 and is indicative of a proper binding mode.

2.10. Pharmacophore and binding characterization

Having obtained a comprehensive description of the ERCC1–XPA binding interaction and a diverse set of ligand interactions, we turned our attention to the creation of a model describing key chemical features of both the binding site and ligands. These models are commonly known as pharmacophores and represent chemical functions, valid not only for a currently bound molecule, but also for the putative binding characteristics of unknown molecules. Due to its overall simplicity, this method can be extremely computationally efficient and is exceptionally well suited for interpreting the virtual screening results of large compound libraries. In general, a single ligand bound to a protein's active site is able to provide sufficient information to start the construction of a pharmacophore model. This approach is generally used in the analysis of a known X-ray or NMR structure of a ligand–receptor complex. It is also possible to develop a pharmacophore from a set of ligands which bind to the same region within the target. In the first, structure-based approach, critical chemical features are recognized from the known complex. The second, ligand-based approach extracts a common set of chemical features by exploring properties of the bound ligands.

The top hits from NCI diversity set included in the scored compounds were filtered for drug likeness and recorded in Table 4 along with the DrugBank top poses. The binding energies for the hits ranged from −7.66 to −5.11 kcal/mol. Note, at this energy range, the UCN-01 compound is not selected among the top hits since its RCS score was −4.81 kcal/mol (see Fig. 7). The top hits showed an overall similar structure including planer hydrophobic rings mostly located on the two edges of the ligands with hydrogen bond donors and acceptors in the middle of the structures. These hits generally mimic the XPA peptide (see Fig. 3c) in its interaction with the ERCC1 binding site. Most of the filtered compounds showed almost the same ranking in the ensemble based calculations. This is a confirmation that the six representative structures were sufficient to substitute the full set of the 44 representative conformations. To further reduce the complexity of

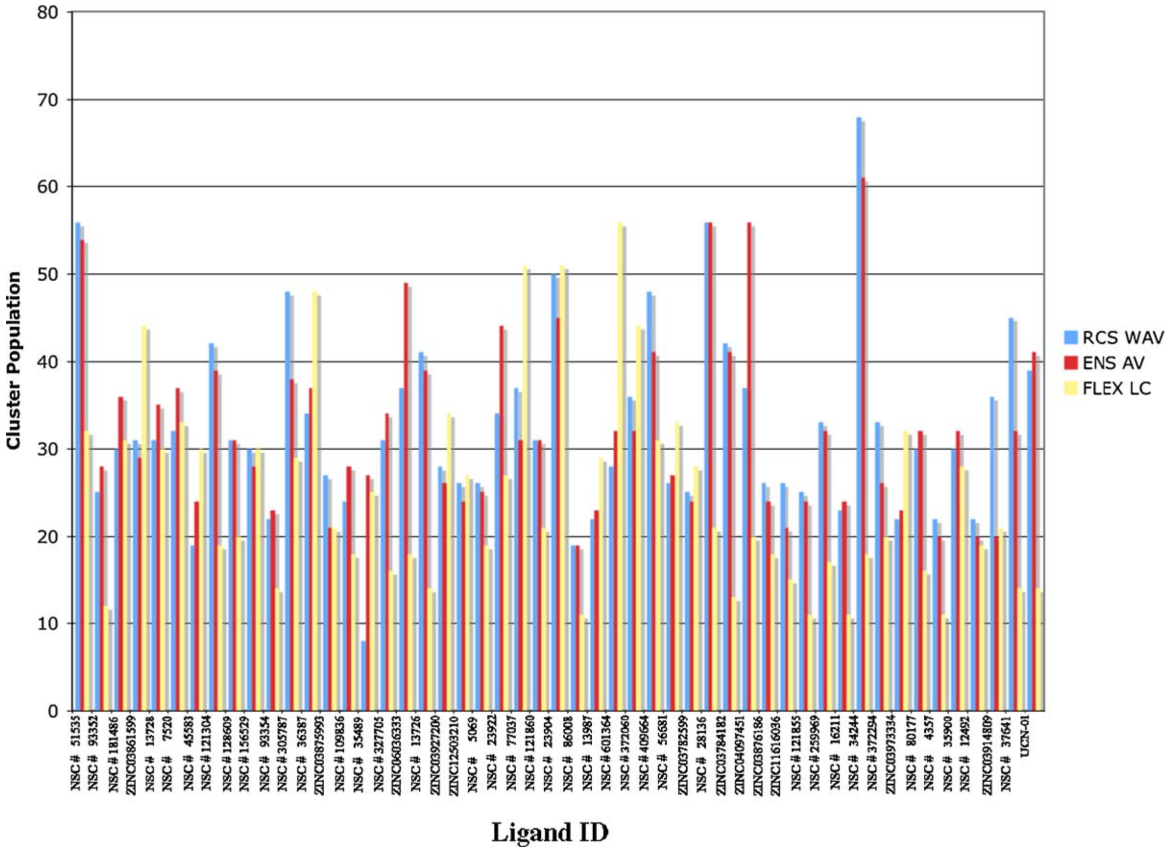


Fig. 8. Clustering of the irredundant top 50 hits suggested by the ensemble-based screening. The RCS weighted-average (RCS WAV) and the ensemble average (ENS AV) population showed the same clustering for most of the top hits. For more than 50% of the hits, the flexible largest cluster population (FLEX LC) is lower than the 25% cutoff, indicating that they have been excluded from the flexible-screening ranking.

the pharmacophore generation, those ligand atoms that did not fall within a cutoff of 25% occupancy were removed and remaining atoms were used to construct the excluded shell describing the ligands as bound to ERCC1 (see Fig. 10a).

Using the pharmacophore feature in Discovery Studio 2.1, we created individual binding site pharmacophore interaction models for the top 30 hits from each of the six ensemble screening experiments. The ERCC1 binding site pharmacophore model (see Fig. 10b) consists of two spatially separated areas of hydrophobic interaction encompassing residues R108, F140, L141, Y152 and I153. In addition to the hydrophobic interactions, there are also two regions of possible aromatic stacking with residues Y145 and Y152. We note that residue Y145 was identified as a critical interaction in the ERCC1–XPA binding energy calculations and was set as flexible in the virtual screening experiments. A critical observation was that the Y154 side chain occupies a position on the floor of the binding pocket, while the Y145 side chain sits above the binding pocket, resulting in the formation of a shallow cavity. This configuration of tyrosine side chains presents the likelihood for forming an aromatic sandwich, a feature that is observed in the active site of monoamine oxidases [57]. In addition to hydrophobic and aromatic features, the binding pocket is also defined by three hydrogen bond acceptor regions (residues R108, R106, N110 and S142) and, two smaller hydrogen bond donor regions (residues P105, R106 and a small region of F140).

One of the most interesting findings of this exercise was that the majority of ligands from the virtual screening experiments were extremely symmetrical, a feature that is reflected in the pharmacophore model (see Fig. 10c). The most significant chemical feature within the pharmacophore is the naphthalene group, which forms an aromatic sandwich between Y145 and Y152 within the ERCC1 binding site. The positioning of hydroxyl groups within

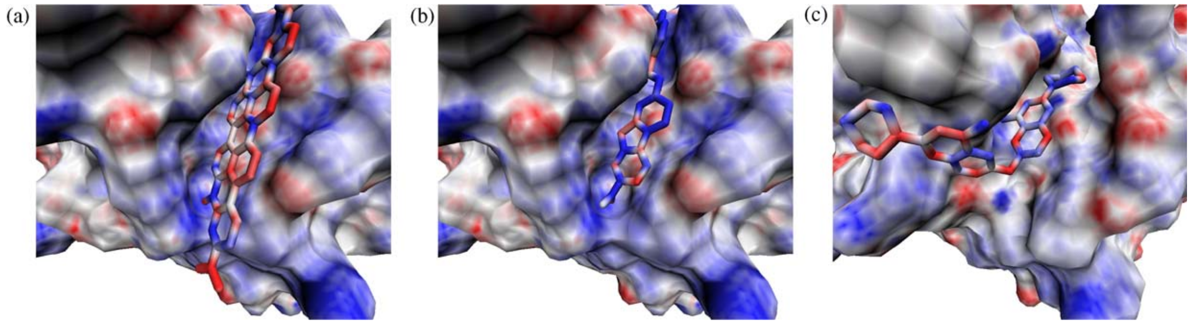


Fig. 9. Three selected hits within their preferred binding site conformations. The binding cleft within ERCC1 and the top hits are colored by residue electrostatic potential with coloring scale of −10 kT/e (red) to +10 kT/e (blue). The closed structure (a and b) is more positive than the open structure (c).

Table 4
Drug-like compounds from the NCI set and top hits from the DrugBank ranked by their RCS score and compared to the ensemble screening rank.

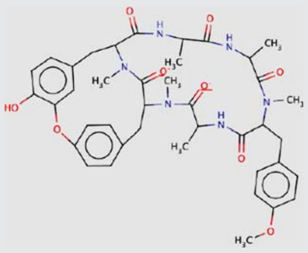
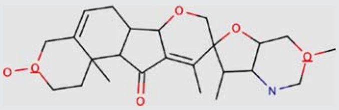
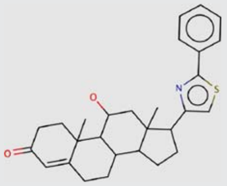
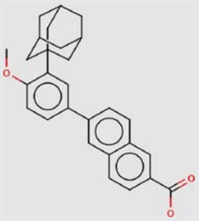
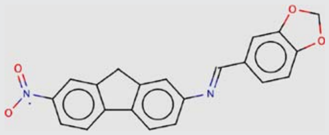
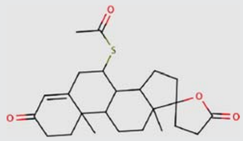
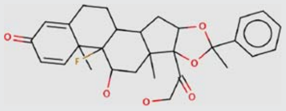
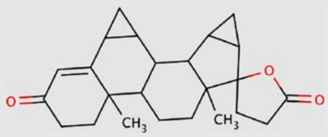
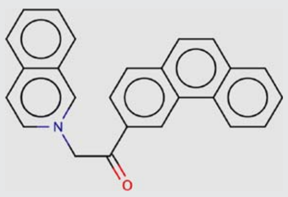
RCS Rank	ENS Rank	RCS WABE (kcal/mol)	NSC #	Structure
3	8	−7.66	259969	
6	3	−7.37	7520	
8	14	−7.21	93354	
15	10	−6.93	ZINC03784182	
17	22	−6.76	121304	
18	4	−6.66	ZINC03861599	
19	27	−6.62	37641	
20	5	−6.57	ZINC03927200	
21	23	−6.51	35489	

Table 4 (*Continued*)

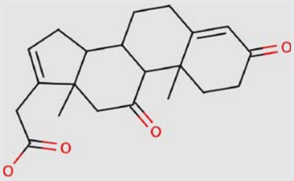
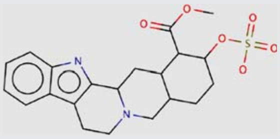
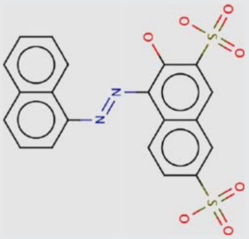
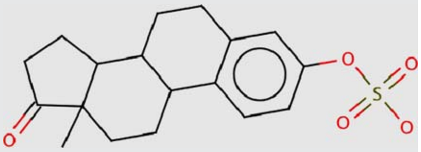
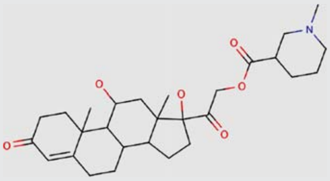
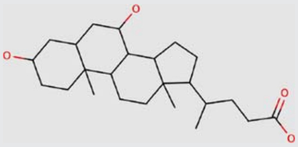
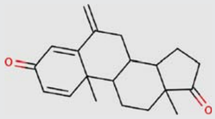
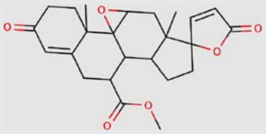
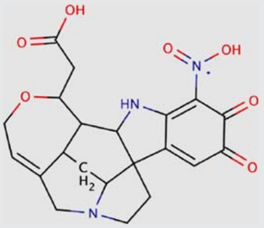
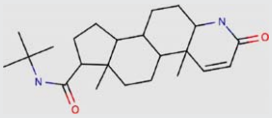
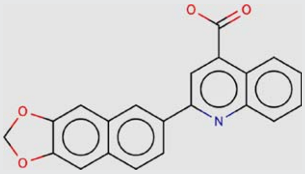
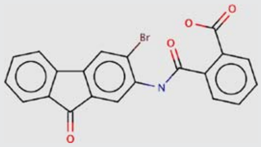
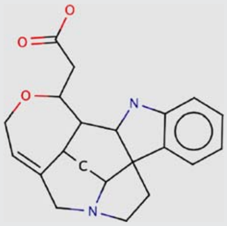
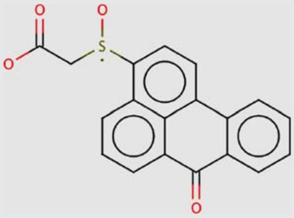
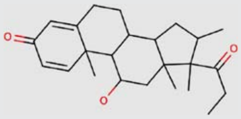
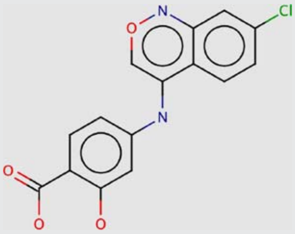
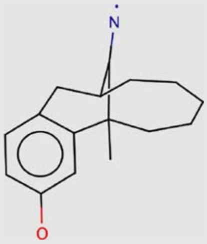
RCS Rank	ENS Rank	RCS WABE (kcal/mol)	NSC #	Structure
23	28	−6.40	86008	
25	41	−6.37	121855	
27	16	−6.35	45583	
28	19	−6.31	ZINC03876186	
30	33	−6.24	23904	
31	30	−6.24	ZINC03914809	
33	25	−6.24	ZINC03973334	
34	32	−6.11	ZINC11616036	
36	40	−6.07	5069	

Table 4 (*Continued*)

RCS Rank	ENS Rank	RCS WABE (kcal/mol)	NSC #	Structure
37	35	−6.07	ZINC03782599	
40	42	−5.99	134244	
42	45	−5.82	56681	
43	47	−5.82	121860	
45	46	−5.79	16211	
46	34	−5.76	ZINC12503210	
47	49	−5.70	12492	
49	48	−5.11	ZINC04097451	

the napthalene group also results in the formation of favorable hydrogen bonding interactions with R106 and S142 in ERCC1. The hydrocarbon linker region between the two hydrophobic ring functionalities spans the 6 Å seperating hydrophobic patches in the ERCC1 binding site and also provides additional rotational flexibility required for proper ring orientation.

While little is known about the precise binding properties for the weak inhibitor of NER, UCN-01 (see Fig. 10d), we can propose a

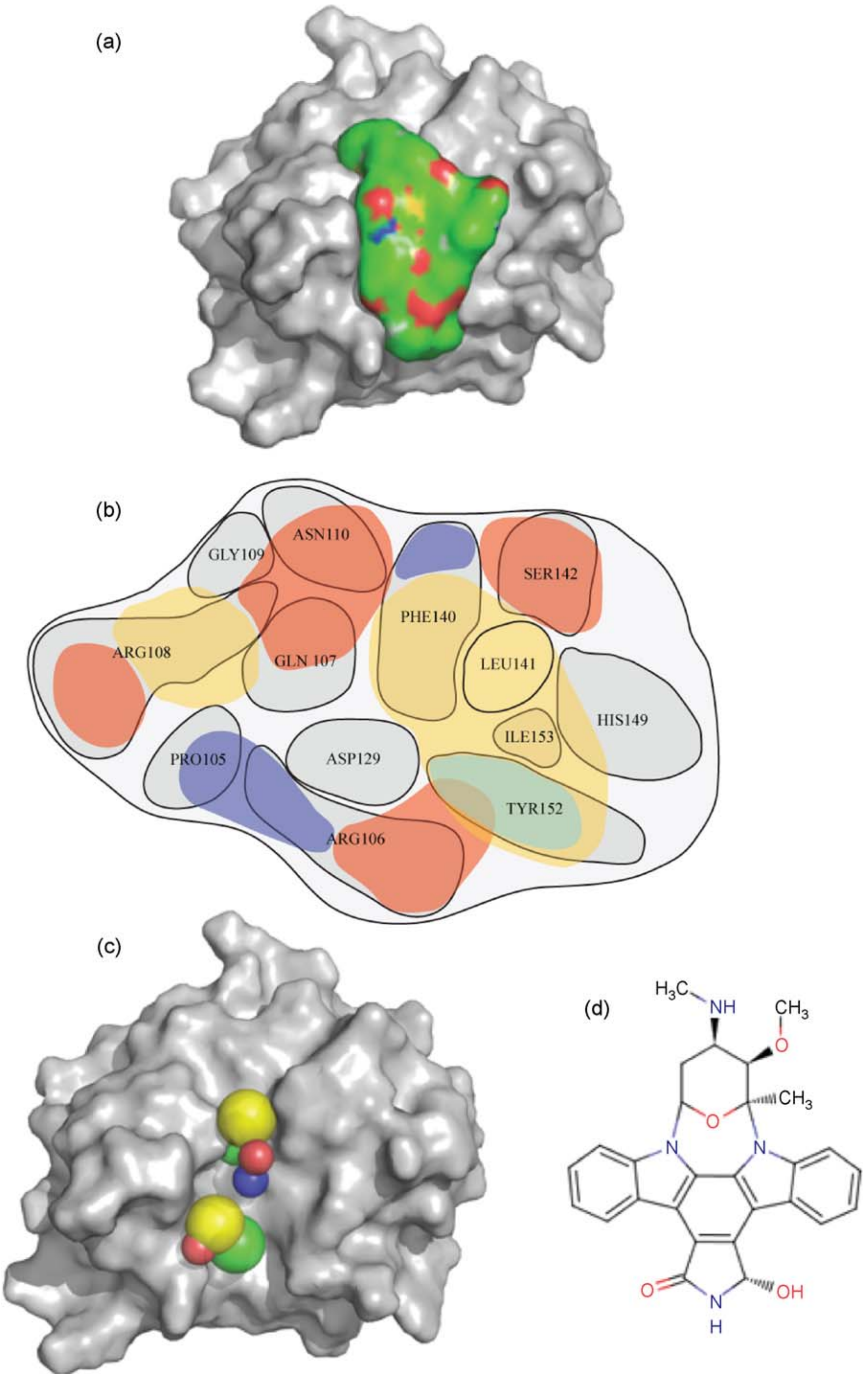


Fig. 10. Pharmacophore determination. (a) The equilibrated ERCC1 (grey surface) showing the excluded volume occupied by atoms from ligands obtained in the virtual screening experiments (green surface). Atoms included in this image were obtained by clustering the top ligands, from virtual screening experiments, and omitting those that were outside of a 90% RMSD cutoff. (b) Pharmacophores from each of the top 30 ligands were created with their interactions in the ERCC1 binding site. The type of pharmacophore interactions, with each residue was scored and is represented schematically. Yellow patches indicate hydrophobic interactions with the pocket, red and blue patches represent hydrogen bond acceptor and donors respectively, while green patches indicate aromatic interactions. Orientation of the binding site is the same as in panel a. Tyrosine 145 and Histidine 149 (indicated by an asterisk *) do not lie on the bottom of the pocket but are observed within a lip that overhangs the pocket (see panel a). (c) The averaged pharmacophore model obtained from the docked poses from virtual screening. Each sphere represents a specific chemical entity with the size being representative of the overall contribution at each position. Coloring is identical to that described for panel a. (d) The chemical structure of UCN-01.

binding mode that is similar to the pharmacophore (see Fig. 10c). Like the molecules used to construct the pharmacophore, UCN-01 is quite symmetrical and also presents the possibility for stacking of either of the phenyl rings between Y145 and Y152 in the ERCC1 binding site. There are also several hydrogen bond donors that could effectively stabilize interactions with the pyrrolidine ring. Interestingly, the overall size of UCN-01 is not able to occupy the entire ERCC1 binding site pharmacophore model and its rigidity is quite high due to the central ring structure and absence of a flexible linker. These factors may influence the relatively weak effect that UCN-01 has on NER inhibition [36]. However, without detailed data describing binding to ERCC1, these comments are only speculative.

3. Conclusion

The elucidation of the mechanisms by which cells develop resistance to Cisplatin is an area of intense research activity since it presents one of the major barriers to the continued clinical success of this drug. There is a great need for novel combination chemotherapy design that would specifically alleviate known deficiencies in current Cisplatin-based treatment by augmenting them with ERCC1-targetting compounds. To date, only one such compound, UCN-01, has been characterized and tested pre-clinically [36]. UCN-01 was originally described as a cell cycle checkpoint abrogator targeting phosphokinases that was also shown to non-specifically inhibit nucleotide excision repair. As a consequence, it is expected to potentiate Cisplatin toxicity based on its suspected effect on ERCC1. UCN-01 is still in Phase I clinical trials as a general chemotherapeutic agent but not linked to Cisplatin resistance effects and its future success is far from guaranteed. Therefore, the search for other compounds with similar functionality is warranted. Accordingly, the purpose of this study was to undertake a computational search for potential inhibitors of this pathway by inhibiting the XPA–ERCC1 protein–protein interaction that is involved in the final stages of this pathway.

Using MD simulations and binding energy analysis we identified the key residues constituting the binding pocket within ERCC1 for its interaction with XPA. Moreover, by employing principal component analysis on the binding site within ERCC1, we have confirmed the completeness of sampling of the apo MD trajectory and generated an equilibrated structure for the ERCC1 pocket. Subsequently, we have used conformational RMSD clustering to extract 44 different representative structures that describe the whole conformational space of the ERCC1 pocket.

The eight screening experiments employed the NCIDS [35] library and DrugBank small-molecules as the set of putative ligands and AUTODOCK4 as the docking engine. The docked compounds comprised a set of 3450 different chemical structures with predetermined pharmacological suitability. The docked poses were clustered using a proposed dynamic technique that adapts the number of clusters to the optimal clustering pattern. Ranked by the binding energy of the most populated cluster, the non-redundant top 50 compounds resulted from the ensemble-based screening were rescored using the RCS by redocking them to the 44 structures that describe the whole MD trajectory. The RCS and the ensemble-based screens identified almost the same results for the top hits, indicating that the cutoff to six representative structures was adequate to describe the whole MD trajectory. Only one compound was rejected by the RCS indicating its ability to exclude false binders from the top hits. On the other hand, only two hits from the flexible screening have been identified among the non-redundant set with more than 50% of the RCS top hits having been excluded due to their low binding energies. Accordingly, it is apparent that, although flexible docking allowed important side chains to move during docking, the lack of backbone dynamics prevented many promising compounds from being selected among the top hits. Moreover, flexible docking was not able to pull back any of the top hits suggested by the ensemble-based screen.

The pharmacophore model we have described here can be used in the subsequent identification of novel ERCC1–XPA inhibitors from the rapid virtual screening of multi-conformational three-dimensional structure databases. Moreover, it can be employed as the basis for the rational design of specific inhibitors for the XPA–ERCC1 interaction that would ultimately result in the development of a Cisplatin-based combination therapy for a broad range of cancers.

4. Methods and materials

4.1. Molecular dynamics simulations

The central domain of ERCC1 (residues 99–214), both free and bound to an 11-residue fragment of XPA (residues 67–77) was taken from PDB entry 2JNW [24]. Molecular dynamics (MD) simulations were carried out using the NAMD program [37] at a mean temperature of 300 K and physiological pH (pH 7) using the all-hydrogen AMBER99SB force field [38]. Protonation states of all ionizable residues were calculated using the program PDB2PQR [39]. Following parameterization, the ERCC1 protein alone or in complex with the XPA peptide were immersed in the center of a TIP3P water [40] cube after adding hydrogen atoms to the initial protein structure. The cube dimensions were chosen to provide at least a 20 Å buffer of 16596 (15323) water molecules around the systems. To neutralize and prepare the XPA-bound or (free systems) under a physiological ionic concentration, 32 (29) chloride and 30 (27) sodium ions were respectively added by replacing water molecules having the highest electrostatic energies on their oxygen atoms. 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 50 ns during which atomic coordinates were saved to the trajectory every 2 ps. The RMSD and B-factors for the protein backbone were then computed over the last 10 ns of the MD simulation using the PTRAJ utility within AMBER10 [41]. 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 greater than 3.4 Å, a light donor–heavy acceptor distance greater than 2.5 Å, and a deviation of less than 60° from linearity.

4.2. RMSD clustering to extract representative MD structures

To generate a reduced set of representative models of the ERCC1 binding site, we performed RMSD conformational clustering with the average-linkage algorithm as implemented in the PTRAJ utility of AMBER10 using a critical distance of 1.3 Å [42]. For the apo ERCC1 simulation, structures were extracted at 2 ps intervals over the entire 50 ns simulation. 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 22 residues that line the XPA binding site, namely those numbered: 106–112, 129, 140–146, 148, 149, 152, 153, 156, 172, and 174. 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 cutoff was chosen after evaluation of the dependence of cluster populations against the total number of clusters using a range of 0.9–1.4 Å. Forty-four clusters were obtained and the six most dominant clusters represented approximately 48%, 8%, 6%, 5.5%, 4% and 3.8% of the whole ensemble, respectively. 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 was used as rigid template for docking experiment.

4.3. Principal component analysis

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 [43,44]. For a typical protein, the system's dimensionality is thereby reduced from tens of thousands to fewer than fifty 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 coordinates as

$$\sigma_{ij} = \langle (r_i - \langle r_i \rangle)(r_j - \langle r_j \rangle) \rangle \tag{1}$$

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 [45] divides an MD trajectory into separate parts, and their normalized overlap is calculated using the covariant matrices for each pair of parts:

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

where $C1$ and $C2$ are the covariant matrices, and the symbol tr is used to denote 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 ERCC1, 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.

4.4. Equilibrated ERCC1 model

A detailed representation of the conformational dynamics can be obtained by projecting the trajectory onto the planes spanned by the most dominant eigenvectors of the covariance matrix. The higher the occupancy of a conformational state in this projection, the lower the free energy of that state [46,47]. Therefore, by observing the regions at which many conformations cluster, one can predict the minimal energy conformations visited by an MD trajectory and estimate a representative conformation for these structures. The entire MD trajectory was projected onto the planes spanned by the first and second, the first and third and the second and third principal components. The conformations residing within the global minimum region were used to predict an equilibrated binding site template. The equilibrated model was compared to the most dominant cluster representative structure and has been appended to the set of conformations used in VS experiments.

4.5. Energy evaluation of the ERCC1/XPA interaction

The trajectory of the ERCC1/XPA MD simulation was analyzed using the MMPBSA utility of AMBER10 to calculate the individual binding energies between residues within the XPA peptide and the ERCC binding site. The binding energy was further divided into individual residue contributions to recognize the key residues in the interaction between the two proteins. Following the identification of the significant residues, we carried out alanine scanning on these residues by performing MD simulation on substituted models and calculating the resultant binding energies. Consequently, residues essential for the interaction between ERCC1 and XPA were determined for subsequent flexible docking.

4.6. Binding free energy

Binding free energies were calculated using the molecular mechanics Poisson–Boltzmann surface area (MM–PBSA) method as implemented in AMBER10 [48]. 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:

$$G = E_{MM} + G_{solv} - TS_{solute} \tag{3}$$

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

$$E_{MM} = E_{ele} + E + E_{\text{int}} \tag{4}$$

Using Eq. (1), the average binding free energy can be calculated by averaging over snapshots captured from the MD trajectory, that is

$$\Delta\hat{G} = \hat{G}_{ERCC1\cdot XPA} - \{\hat{G}_{ERCC1} + \hat{G}_{XPA}\} \tag{5}$$

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.

Applying the thermodynamic cycle for the ERCC1–XPA complex, the binding free energy between the XPA67-77 and the ERCC199-214 binding site can be approximated by

$$\Delta G^\circ = \Delta G_{gas}^{ERCC1\cdot XPA} + \Delta G_{solv}^{ERCC1\cdot XPA} - \{\Delta G_{solv}^{ERCC1} + \Delta G_{solv}^{XPA}\} \tag{6}$$

Here, ($\Delta G_{gas}^{ERCC1\cdot XPA}$) represents the free energy per mole for the non-covalent association of the ERCC1–XPA complex in vacuum (gas phase) at 300 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 300 K (assuming that the binding conformation of ERCC1–XPA is the same in solution and in vacuum).

The gas phase free energy can be further decomposed into electrostatic and non-electrostatic components:

$$\Delta G_{gas}^{ERCC1\cdot XPA} \approx \Delta E_{gas} + \Delta E_{gas}^{ele} \tag{7}$$

Furthermore, the solvation free energy can be expressed as a sum of non-electrostatic and electrostatic contributions:

$$\Delta G_{solv} \approx \Delta G_{solv}^{nonele} + \Delta G_{solv}^{ele} \tag{8}$$

The non-electrostatic part was approximated by a linear function of the (SASA). That is

$$\Delta G_{solv}^{nonele} = \gamma \times \text{SASA} \tag{9}$$

where $\gamma = 7.2$ cal/mol/Å^2.

4.7. Selection of ligand database

The National Cancer Institute Diversity Set (NCDIS) [35] and DrugBank-small-molecules were used as our test libraries of compounds to obtain a lead pharmacophore for inhibiting the ERCC/XPA interaction. 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. In this study, we used a version of the NCIDS formatted for use in AUTODOCK which was prepared by the AUTODOCK Scripps team. The DrugBank 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 compounds to the screening set. In addition, UCN-01, a compound that has been previously demonstrated to weakly inhibit the ERCC1–XPA interaction, was used as a comparison during the screening experiments and for subsequent pharmacophore validation [36].

4.8. Ligand screening

Virtual screening on the ERCC1 binding site was performed using AUTODOCK, version 4.0 [49]. Hydrogen atoms were added to ERCC1 and ligands and partial atomic charges were then assigned using the Gasteiger–Marsili [50] method. Atomic solvation parameters were assigned to the protein atoms using the AUTODOCK 4.0 utility ADDSOL. A docking grid map with $70 \times 70 \times 70$ points and grid point spacing of 0.375 Å was then centered on the XPA binding site within the ERCC1 receptor using AUTOGRID4.0 program [49]. Rotatable bonds of each ligand were then automatically assigned using AUTOTORS utility of AUTODOCK4.0. Docking was performed using the Lamarckian genetic algorithm (LGA) method with an initial population of 150 random individuals; a maximum number of 20×10^6 energy evaluations; 100 trials; 27,000 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 eight independent virtual screening runs were performed against the full set of docked ligands. The first used the minimized holo crystal structure of the ERCC1–XPA binding site with key residues, determined from previous MD experiments, set as flexible during the docking experiment. The other seven experiments utilized the six representative conformations of the dominant clusters produced from the clustering analysis along with the equilibrated model of the ERCC1 binding site as determined using principal component projections.

4.9. Clustering of docked poses

The previously described virtual screening experiments produced numerous conformations of each ligand bound to ERCC1. The AUTODOCK program 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 cutoff used. It is not possible to anticipate an optimum cutoff 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. Unfortunately, there is no universally accepted clustering algorithm that can be used to extract all of the information contained within the docked conformations. 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 trajectories [42]. The clustering quality can be anticipated by calculating a clustering metric that can deduce the optimal number of clusters to be used. One of the commonly used methods is the elbow criterion [51], in which the percentage of variance, explained by the data (λ), is expected to plateau for cluster counts exceeding the optimal number. The percentage of variance is defined by

$$\lambda = \frac{\mathrm{SSR}}{\mathrm{SST}} \tag{10}$$

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 explained 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.

4.10. Relaxed complex scheme (RCS) and receptor flexibility

While ligand flexibility is well accounted for in virtual screening experiments, the inclusion of receptor flexibility remains an important challenge. The RCS is a hybrid technique that combines the rewards of docking algorithms with dynamic structural information provided by MD simulations, therefore explicitly accounting for the flexibility of both the receptor and the docked ligands. To employ the RCS we performed 7 independent screening runs against rigid templates of the binding site within ERCC1. This set of conformations comprises the central member cluster structures of the 6 dominant clusters that constitute about 75% of the whole MD trajectory. In addition to these structures we have added the equilibrated model produced from PCA to the set of docked structures. Docking results have been sorted by the lowest binding energy of the most populated cluster using the proposed ligand clustering technique (see above). We only consider a compound among the top hits if the most populated cluster includes at least 25% of all docked conformations. The top 50 hits from each system were combined to produce an irredundant set of promising compounds. To validate and refine the virtual screening results, redocking experiments were performed on the combined hits into the rest of the 44 clustering representative structures, to account for 100% of the ensemble of the apo MD trajectory. Following the same docking procedure and parameter set described in the previous sections, docking poses were ranked using their weighted average binding energies and were used for further analysis.

For the weight means calculations, the following formula was used

$$\text{Weighted Average Binding Energy (WABE)} = \sum_{i}^{M} \text{percent distribution}(i) \times \text{binding energy}(i)$$

where i is the index number of each ensemble cluster, whose percent distribution sums up to 100% and M is number of different structures included in the ensemble.

4.11. Electrostatic surface calculations

Electrostatic potentials for ERCC1 were calculated using the APBS program [52] and mapped onto a reduced molecular surface with the VMD visualization program [53]. ERCC1 was treated as a low dielectric medium ($\varepsilon_{in} = 1$) surrounded by a high dielectric solvent ($\varepsilon_{out} = 80$ for water). The ionic strength was set to 0.1 M. The low-dielectric region of the protein was defined as the region inaccessible to contact by a 1.4 Å sphere rolling over the molecular surface, defined by atomic co-ordinates of the MD structure and radii taken from the all-hydrogen AMBER99SB force field. The electrostatic potential calculations employed a 200 × 200 × 200 grid with a spacing of 0.5 Å.

4.12. ERCC1/XPA binding site pharmacophore

Pharmacophore construction is based on the detection of relevant chemical interaction points between a ligand and a protein. Several algorithms are able to successively recognize these interactions and classify them into a collection of chemical features or, as it is more commonly described, a pharmacophore model. To produce a final pharmacophore for the XPA ERCC1 binding interaction, we used Accelrys Discovery Studio 2.1 (Accelrys Inc., 2008) to construct models for the top 30 poses collected from each of the top six representative ERCC1 conformations.

Pharmacophore generation involved using the catalyst/HipHop program to generate feature-based 3D pharmacophore alignments [54,55]. This was accomplished by examining each separate pose for the presence of certain chemical features, followed by the determination of a three-dimensional configuration of the chemical features. Catalyst provides a predefined dictionary of chemical features found to be important in drug–enzyme/receptor interactions. These are hydrogen bond donors, hydrogen bond acceptors, hydrophobic groups, ring aromatic and positive/negative ionizable groups. For the pharmacophore modeling runs, common features selected for the run were ring aromatic, hydrogen bond donor and acceptor, hydrophobic group and ionizable groups. Since we do not have access to any activity data for any of the compounds being screened, we adopted a strategy, where HipHop assumes that differences in activities will be related to the differences in other relevant factors like conformational energies, but not due to the absence of any important features required for binding. Merging and overlaying of each of the resulting pharmacophores was then accomplished using the clique detection algorithm combined with the Kabsch alignment approach [56]. However, due to the vast structural dissimilarity of each of the ligands obtained from the docking stage, chemical features associated with each residue found within the ERCC1 binding site were tabulated and a pattern of interaction was then determined manually.

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 the Alberta Cancer Foundation, the Allard Foundation and NSERC.

References

[1] T. Boulikas, M. Vougiouka, Cisplatin and platinum drugs at the molecular level, Oncol. Rep. 10 (2003) 1663–1682.
[2] J. Lokich, N. Anderson, Carboplatin versus cisplatin in solid tumors: an analysis of the literature, Ann. Oncol. 9 (1998) 13–21.
[3] E. Raymond, S. Faivre, J.M. Woynarowski, S.G. Chaney, Oxaliplatin: mechanism of action and antineoplastic activity, Semin. Oncol. 5 (1998) 4–12.
[4] L.A. Zwelling, K.W. Kohn, Mechanisms of action of cis-dichlorodiamineplatinum (II), Cancer Treat Rep. 63 (1979) 1439–1444.
[5] M.F. Pera, C.J. Rawlings, J.J. Roberts, The role of DNA repair in the recovery of human cells from cisplatin toxicity, Chem. Biol. Interact. 37 (1981) 245–261.
[6] J.A. Rice, D.M. Crothers, A.L. Pinto, The major adduct of the antitumor drug cis-dichlorodiamineplatinum with DNA bends the duplex by approximate equal to 40 degrees toward the major groove, Proc. Natl. Acad. Sci. U.S.A. 85 (1988) 4158–4161.
[7] H.N. Fraval, J.J. Roberts, Excision repair of cis-dichlorodiamineplatinum-induced damage to DNA of Chinese hamster cells, Cancer Res. 39 (1979) 1793–1797.
[8] E. Reed, R.F. Ozols, R. Tarone, S.H. Yuspa, M.C. Poirier, Platinum–DNA adducts in leukocyte DNA correlate with disease response in ovarian cancer patients receiving platinum-based chemotherapy, Proc. Natl. Acad. Sci. U.S.A. 84 (1987) 5024–5028.
[9] W. McGuire, R.F. Ozols, Chemotherapy of advanced ovarian cancer, Semin. Oncol. 25 (1998) 340–348.
[10] R.F. Ozols, P.J. O'Dwyer, T.C. Hamilton, Clinical reversal of drug resistance in ovarian cancer, Gynecol. Oncol. 51 (1993) 90–96.
[11] A. Cara, M. Rabik, E. Dolan, Molecular mechanisms of resistance and toxicity associated with platinating agents, Cancer Treat Rev. 33 (2007) 9–23.
[12] R. Altaha, X. Liang, J.J. Yu, E. Reed, Excision repair cross complementing group 1; gene expression and platinum resistance, Int. J. Mol. 14 (2004) 559–570.
[13] M. Dabholkar, J. Vionnet, F. Bostick-Bruton, J.J. Yu, E. Reed, Messenger RNA levels of XPAC and ERCC1 in ovarian cancer tissue correlate with response to platinum-based chemotherapy, J. Clin. Invest. 94 (1994) 703–708.
[14] X. Wu, W. Fan, S. Xu, Y. Zhou, Sensitization to the cytotoxicity of cisplatin by transfection with nucleotide excision repair gene xeroderma pigmentosum group A antisense RNA in human lung adenocarcinoma cells, Clin. Cancer Res. 9 (2003) 5874–5879.
[15] G. Giaccone, Clinical perspectives on platinum resistance, Drugs 5 (2000) 9–17.
[16] R. Metzger, C.G. Leichman, K.D. Danenberg, P.V. Danenberg, H.J. Lenz, K. Hayashi, S. Groshen, D. Salonga, H. Cohen, L. Laine, P. Crookes, H. Silberman, J. Baranda, B. Konda, L. Leichman, ERCC1 mRNA levels complement thymidylate synthase mRNA levels in predicting response and survival for gastric cancer patients receiving combination cisplatin and fluorouracil chemotherapy, J. Clin. Oncol. 16 (1998) 309–316.
[17] J.T. Reardon, A. Vaisman, S.G. Chaney, A. Sancar, Efficient nucleotide excision repair of cisplatin, oxaliplatin, and bis-aceto-ammine-dichloro-cyclohexylamine-platinum(IV) (JM216) platinum intrastrand DNA diadducts, Cancer Res. 59 (1999) 3968–3971.
[18] L. Li, S.J. Elledge, C.A. Peterson, E.S. Bales, R.J. Legerski, Specific association between the human DNA repair proteins XPA and ERCC1, Proc. Natl. Acad. Sci. U.S.A. 91 (1994) 5012–5016.
[19] G.W. Buchko, N.G. Isern, L.D. Spicer, M.A. Kennedy, Human nucleotide excision repair protein XPA: NMR spectroscopic studies of an XPA fragment containing the ERCC1-binding region and the minimal DNA-binding domain (M59-F219), Mutat. Res. 486 (2001) 1–10.
[20] M. Saijo, I. Kuraoka, C. Masutani, F. Hanaoka, K. Tanaka, Sequential binding of DNA repair proteins RPA and ERCC1 to XPA in vitro, Nucleic Acids Res. 24 (1996) 4719–4724.
[21] B. Koberle, J.R. Masters, J.A. Hartley, R.D. Wood, Defective repair of cisplatin-induced DNA damage caused by reduced XPA protein in testicular germ cell tumours, Curr. Biol. 9 (1999) 273–276.
[22] C. Welsh, R. Day, C. McGurk, J.R. Masters, R.D. Wood, B. Köberle, Reduced levels of XPA, ERCC1 and XPF DNA repair proteins in testis tumor cell lines, Int. J. Cancer 110 (2004) 352–361.
[23] E. Rosenberg, M.M. Taher, N.B. Kuemmerle, J. Farnsworth, K. Valerie, A truncated human xeroderma pigmentosum complementation group A protein expressed from an adenovirus sensitizes human tumor cells to ultraviolet light and cisplatin, Cancer Res. 61 (2001) 764–770.
[24] V.T. Oleg, I. Dmitri, O. Barbara, S. Lidija, S. Ilana, O. Robert, D.S. Orlando, W. Gerhard, E. Tom, Structural basis for the recruitment of ERCC1–XPF to nucleotide excision repair complexes by XPA, EMBO J. 26 (2007) 4768–4776.
[25] R.E. Hubbard, Structure-based Drug Discovery: An Overview, first ed., RSC Publishing, Cambridge, 2007, p. 261.
[26] R. Huey, G.M. Morris, A.J. Olson, D.S. Goodsell, A semiempirical free energy force field with charge-based desolvation, J. Comput. Chem. 28 (2007) 1145–1152.
[27] S.F. Sousa, P.A. Fernandes, M.J. Ramos, Protein-ligand docking: current status and future challenges, Proteins 65 (2006) 15–26.
[28] H.A. Carlson, K.M. Masukawa, K. Rubins, F.D. Bushman, W.L. Jorgensen, R.D. Lins, J.M. Briggs, J.A. McCammon, Developing a dynamic pharmacophore model for HIV-1 integrase, J. Med. Chem. 43 (2000) 2100–2114.
[29] C.N. Cavasotto, R.A. Abagyan, Protein flexibility in ligand docking and virtual screening to protein kinases, J. Mol. Biol. 337 (2004) 209–225.
[30] C.N. Cavasotto, J.A. Kovacs, R.A. Abagyan, Representing receptor flexibility in ligand docking through relevant normal modes, J. Am. Chem. Soc. 127 (2005) 9632–9640.
[31] 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 (2002) 5632–5633.
[32] 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 (1999) 547–562.
[33] R. Abagyan, M. Totrov, High-throughput docking for lead generation, Curr. Opin. Chem. Biol. 5 (2001) 375–382.
[34] M. Markowitz, B.Y. Nguyen, F. Gotuzzo, F. Mendo, W. Ratanasuwan, C. Kovacs, J. Zhao, L. Gilde, R. Isaacs, H. Teppler, Potent antiviral effect of MK-0518, novel HIV-1 integrase inhibitor, as part of combination ART in treatment-naive HIV-1 infected patients, in: Proceedings of the XVI International AIDS Conference; 16th International AIDS Conference, Toronto, Canada, August 13–18, 2006, no. ThLB0214.
[35] http://dtp.nci.nih.gov/branches/dscb/diversity_explanation.html (Last checked October 19, 2008).
[36] J. Hong, Y. Li-Ying, Cell cycle checkpoint Abrogator UCN-01 inhibits DNA repair, Cancer Res. 59 (1999) 4529–4534.
[37] L. Kalé, NAMD2: greater scalability for parallel molecular dynamics, J. Comput. Phys. 151 (1999) 283–312.
[38] 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 (2006), 65,712–65,725.

[39] 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) W522–W525.

[40] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, J. Chem. Phys. 79 (1983), 926-923.

[41] D.A. Case, T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, The Amber biomolecular simulation programs, J. Comput. Chem. 26 (2005) 1668–1688.

[42] 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) 2312–2334.

[43] A.E. Garcia, Large-amplitude nonlinear motions in proteins, Phys. Rev. Lett. 68 (1992) 2696–2699.

[44] A. Amadei, A.B. Linssen, H.J. Berendsen, Essential dynamics of proteins, Proteins 17 (1993) 412–425.

[45] B. Hess, Convergence of sampling in protein simulations, Phys. Rev. E. Stat. Nonlin. Soft Matter Phys. 65 (2002) 31910–31920.

[46] H. Grubmuller, Predicting slow structural transitions in macromolecular systems: conformational flooding, Phys. Rev. E. Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top. 52 (1995) 2893–2906.

[47] I. Kosztin, B. Barz, L. Janosi, Calculating potentials of mean force and diffusion coefficients from nonequilibrium processes without Jarzynski's equality, J. Chem. Phys. 124 (2006) 64106–64111.

[48] 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) 889–897.

[49] 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) 1639–1662.

[50] J. Gasteiger, M. Marsili, Iterative partial equalization of orbital electronegativity: a rapid access to atomic charges, Tetrahedron 36 (1980) 3219–3228.

[51] T. Mitchel, Machine Learning, McGraw-Hill, Boston, 1997.

[52] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, J.A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome, Proc. Natl. Acad. Sci. U.S.A. 98 (2001) 10037–10041.

[53] H. William, D. Andrew, S. Klaus, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.

[54] P.W. Sprague, Automated chemical hypothesis generation and database searching with catalyst, in: K. Müller (Ed.), Perspectives in Drug Discovery and Design, vol.3, ESCOM Science Publishers B.V., Leiden, The Netherlands, 1995 , pp. 1–20.

[55] J. Greene, S. Kahn, H. Savoj, P. Sprague, S. Teig, Chemical function queries for 3D database search, J. Chem. Inf. Comput. Sci. 34 (1994) 1297–1308.

[56] W. Kabsch, A solution for the best rotation to relate to sets of vectors, Acta Crystallogr. A32 (1976) 922–923.

[57] R.M. Geha, K. Chen, J. Wouters, F. Ooms, J.C. Shih, Analysis of conserved active site residues in monoamine oxidase A and B and their three-dimensional molecular modeling, J. Biol. Chem. 277 (2002) 17209–17216.