Edinburgh Research Explorer

Comparison of ATP-binding pockets and discovery of homologous recombination inhibitors

Comparison of ATP-binding pockets and discovery of homologous recombination inhibitors

Vincent Blay [a,b,c,*], Saule Gailiunaite [a], Chih-Ying Lee [d], Hao-Yen Chang [d], Ted Hupp [e], Douglas R. Houston [a,*], Peter Chi [d,f]

[a] Institute of Quantitative Biology, Biochemistry and Biotechnology, University of Edinburgh, Edinburgh, Scotland EH9 3BF, UK
[b] Department of Microbiology and Environmental Toxicology, University of California at Santa Cruz, Santa Cruz, CA 95064, USA
[c] Institute for Integrative Systems Biology (I2Sysbio), Universitat de València and Spanish Research Council (CSIC), 46980 Valencia, Spain
[d] Institute of Biochemical Sciences, National Taiwan University, Taipei 10617, Taiwan
[e] MRC Institute of Genetics & Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, UK
[f] Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan

ABSTRACT

The ATP binding sites of many enzymes are structurally related, which complicates their development as therapeutic targets. In this work, we explore a diverse set of ATPases and compare their ATP binding pockets using different strategies, including direct and indirect structural methods, in search of pockets attractive for drug discovery. We pursue different direct and indirect structural strategies, as well as ligandability assessments to help guide target selection. The analyses indicate human RAD51, an enzyme crucial in homologous recombination, as a promising, tractable target. Inhibition of RAD51 has shown promise in the treatment of certain cancers but more potent inhibitors are needed. Thus, we design compounds computationally against the ATP binding pocket of RAD51 with consideration of multiple criteria, including predicted specificity, drug-likeness, and toxicity. The molecules designed are evaluated experimentally using molecular and cell-based assays. Our results provide two novel hit compounds against RAD51 and illustrate a computational pipeline to design new inhibitors against ATPases.

1. Introduction

The nucleic acid adenosine triphosphate (ATP) is the energy currency of the cell, and many enzymes require it to carry out their function. All ATP catalysing enzymes belong to the P-loop nucleotide triphosphate hydrolases (NTPases) class and can be further grouped in two divisions: kinase-GTPases and Additional Strand, Catalytic Glutamate (ASCE) enzymes.[1] Many kinases are successful drug targets,[2] and type I kinase inhibitors compete with ATP for binding to the protein. However, many of these inhibitors lack specificity and bind to a range of off-target ATPases.[2,3] ATP binding sites tend to have conserved regions, including the Walker A motif (a loop with sequence [A/G]xxxxGK[S/T], where × stands for any amino acid) and the Walker B motif (hhhhD[D/E], where h stands for any hydrophobic residue).[4] The work with kinases points out that, in order to achieve specificity towards an ATP binding site, it is key to leverage less conserved residues.[5,6]

Proteins in the ASCE superfamily have two defining characteristics.

First, they have an additional β-sheet between the Walker A and Walker B motifs. Second, they contain a highly conserved glutamate residue that interacts with a nucleophilic water molecule and attacks the γ-phosphate of ATP. These proteins also display additional conserved structural features, such as sensor 1, sensor 2, and an arginine finger.[7,8] ASCE ATPases can be further grouped into families, based on other structural features, such as the presence of a C-terminal hairpin or helical bundles.[9] For example, AAA + ATPases (ATPases Associated with various cellular Activities) is an ASCE protein family with a C-terminal helical bundle. Generally, AAA + proteins form helical hexamers and process the ligand through a central channel, using ATP hydrolysis.[10] Further divisions of families into clades are also used.

Compared to kinases, ASCE ATPases show a broader range of functions and are less understood as drug targets. Many compounds against them target pockets other than the ATP binding site, such as DNA binding sites or protein–protein interaction surfaces.[11] Given their involvement in a wide range of conditions, including various cancers, it

---

would be highly desirable to improve target selection and design of inhibitors against ASCE ATPases. In this work, we explore the ATP binding sites of ASCE ATPases in search of structural differences that could be exploited by novel therapeutics. We propose structural approaches, direct and indirect, to compare binding pockets and facilitate target selection. We then select a subset of ATPases and conduct a computer-assisted drug design (CADD) campaign around RAD51, including the identification and experimental confirmation of two novel inhibitors of homologous recombination.

2. Methods

2.1. Selection of structures

A diverse set of 26P-loop ATPase structures were included in the study, listed in Table S1. These structures were selected to span a variety of subfamilies,[12,13] focusing on human proteins, and all containing the conserved Walker-A motif [A/G]XXXXGK[S/T].[4] The proteins belong to a variety of clades within the ASCE superfamily. A few kinesin structures were also chosen to enrich the diversity of the study (1bg2, 1ii6, 5wdh). The latter fall in the Kinase-GTPase division (TRAFAC class).[1,14,15]

2.2. Distance-based pocket signatures

A new method for structural pocket comparison was proposed, which leverages the presence of a conserved Walker-A motif across all ATPases studied. The approach avoids the difficulties of aligning pockets with different shapes or compositions. First, the Walker-A motif was identified using the regular expression [AG]XXXXGK[ST]. The S or T residue, and sometimes a Walker-B motif, are known to help stabilize the binding of the Mg^{2+} ion present in MgATP/MgADP and are therefore adjacent to the ATP binding pocket. Thus, we decided to compute a characteristic radial distance profile from the α-carbon of this S or T residue to the α-carbon of other residues, which we use as the reference point. We decided to consider only hydrophilic residues (RHKDESTNQCWY) in this case because i) such residues are more likely to be found on the protein's surface,[16] and ii) they can engage in H-bonds, polar or electrostatic interactions and could have the ability to 'steer' the binding of the ligand.[17] This structural profile or "signature" was computed up to a distance of 12 Å from the reference point. It is expected that a drug-sized ligand binding within this volume would likely interfere with the binding of the native ATP substrate. Characteristic radial distance profiles can be readily compared across enzymes and provide an indication of their structural similarity around the ATP binding site.

Additional profiles were generated by looking for patches of hydrophilic residues within the 12-Å sphere defined above. The idea was to try to highlight the residues located towards the hinge region of the ATPase. We reasoned that such a region might contain a higher density of hydrophilic residues than other regions close to the Walker-A motif. Thus, we identified the patch within a 5 Å-radius sphere contained within the larger sphere with the largest possible number of hydrophilic residues. A characteristic radial distance profile was thus generated for that patch of residues for every enzyme studied. Obviously, all the parameters above (cutoffs, types of residues) could be modified based on the information of interest. The method was implemented as custom-made scripts in MATLAB 2020a. Scripts are available from the authors upon request.

2.3. Pairwise pocket similarities

We applied the software APoc 1.0[18] to compute pairwise similarities of ATP binding pockets. Similar to the distance-based signatures, the ATP binding pockets were delimited by drawing a 12-Å sphere around the end of the Walker-A motif. The results were visualized using the multi-dimensional scaling in scikit-learn for Python.

2.4. Virtual libraries

Two different virtual libraries were used in this study. To probe for structural differences between ATP-binding sites of different proteins (indirect comparison method, see below), we used molecules from ChemBridge's KINASet library. We reasoned that excessively flexible or functionalized compounds need not be better at discriminating subtle differences between different ATP binding sites. Therefore, the library was pre-filtered using a lead-like filter[19] using Filter-it (Silicos-it), resulting in 2,458 molecules for docking onto each of the 26 targets (Supporting Information file mmc2.xlsx). A second, larger library was built for virtual screening by combining ChemBridge's KINASet and MedChemExpress' Kinase Hinge Binders. The library was filtered for drug-likeness,[20] resulting in 26,567 molecules for docking onto selected targets (file mmc2.xlsx).

2.5. Docking

Molecular docking was conducted using the recent software PSOVina-2.0. This software combines Vina's forcefield and scoring functions with an improved global optimization algorithm.[21] Every compound was evaluated to an equivalent "exhaustiveness" value of 64, which increases the chance of identifying the lowest energy pose. The search box was defined by using the ligand in the crystal structure, typically MgADP or MgATP, and padding it in every x,y,z direction. For the few structures that did not contain a ligand, MgADP was manually docked onto the active site using AutoDock and used similarly. For binding site comparison, a 4 Å padding was used, for identification of inhibitors on selected targets, a padding of 6 Å was used. The docking results were ranked based on the scoring function values, which can be considered an estimate of ΔG, the dissociation free energy of the compound-target complex. To estimate the specificity of a compound for a target *t* over a series of targets *T*, the following formula was used[22]:

$$\text{Specificity}_t = \frac{e^{-\Delta G_t}}{\sum_{j \in T} e^{-\Delta G_j}} \tag{1}$$

For each compound, the lowest ΔG for all PSOVina docking attempts on the same target was used in the expression. The estimates for the Gibbs free energies of binding are reported as dissociation constants at 298 K.

2.6. Ligandability

Ligandability of the different targets was assessed using two approaches (Fig. 1). First, two independent docking runs were conducted on each target using PSOVina with *exhaustiveness* = 32. The agreement between both scores across the library was evaluated with Pearson R^2 and Spearman R^2 values. We hypothesized that the consistency of docking may be informative of ligandability. A less ligandable protein surface may exhibit flatter and more numerous energy minima that are more challenging for the global optimizer to navigate. This could limit the consistency of the results and decrease the ability to design potent inhibitors. Secondly, we applied established methods to evaluate target ligandability, which provide further validation of the previous approach. DoGSiteScorer uses volume, enclosure, and hydrophobicity estimates to calculate a ligandability score.[23] A higher score indicates a more ligandable target.

2.7. Indirect comparison of pockets

To compare binding sites based on docking results, the top virtual ligands for each target were extracted and compared across targets. Two different comparison approaches were developed. The first approach is based on the similarity of the top ligands' 2D chemical structure. The pairwise similarity of every top ligand for a protein to every other top
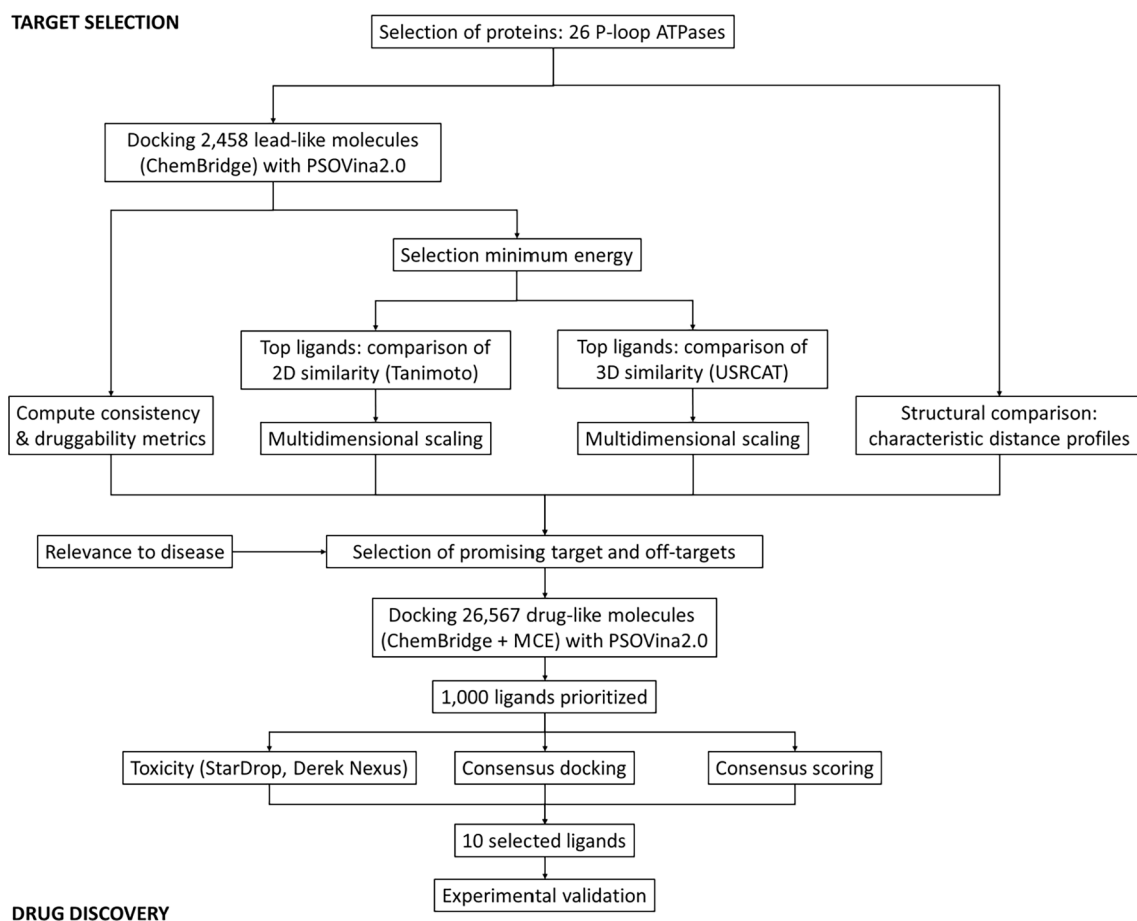
TARGET SELECTION

Fig. 1. Pipeline developed in this work for ATPase target selection and inhibitor discovery.

ligand for every other protein can be computed using Tanimoto similarity metric between their fingerprints. For this, E-state fingerprints were computed using the R package *Rcpi*.[24] The second approach is based on the similarity of the top ligands' 3D conformations. For this, we computed USRCAT descriptors for each molecule using the Python package *rdkit*. The pairwise similarity of every top ligand on a protein to every other top ligand on every other protein was computed using the function GetUSRScore, which computes a Manhattan distance metric between the corresponding descriptor vectors.[25] When necessary, similarities were converted into distances as 1-similarity.

After computing distances between individual virtual ligands, we sought to summarize them as distances between lists of ligands that would inform of the similarity between binding pockets. For this, we set to compute the pairwise distance between the centroid for each list of top ligands, typically $N = 12$ was used. The difficulty is that there is no "centroid molecule" for each list; only distances between pairs of ligands can be measured. Thus, we resorted to some mathematical results to allow us to access the distances between the theoretical centroids in an alternative manner. Let the ligands be denoted as $x_1, x_2, \ldots, x_N$. Let i be the indexes for one list of molecules I, and j the indexes for another list of molecules J. The centroids would be given as:

$$c_I = \frac{1}{|I|}\sum_{i \in I} x_i, \quad c_J = \frac{1}{|I|}\sum_{i \in I} x_i \qquad (2)$$

and we wish to find the distance $||c_I\text{-}c_J||$ in terms of the distances $d_{ij} = ||x_i\text{-}x_j||$. The following algebraic identity is used in ANOVA calculations[26]:

$$||c_I - c_J||^2 = \frac{1}{|I||J|}\left(SS(I \cup J) - (|I| + |J|)\left(\frac{1}{|I|}SS(I) + \frac{1}{|J|}SS(J)\right)\right) \qquad (3)$$

where SS refers to the sum of squares of distances between each point in a set and their centroid. The polarization identity expresses this in terms of squared distances between all points[26]:

$$SS(K) = \frac{1}{2}\sum_{l,m \in K}||x_l - x_m||^2 = \frac{1}{2}\sum_{l \in K}\sum_{m \in K}d_{lm}^2 \qquad (4)$$

Expressions (3) and (4) can be combined and rearranged to obtain the desired result. In a case when the lists of molecules have the same length, $|I| = |J| = N$, we can write:

$$||c_I - c_J|| = \frac{1}{N}\left(\sum_{i \in I}\sum_{j \in J}d_{ij}^2 - SS(I) - SS(J)\right)^{1/2} \qquad (5)$$

This way the sum of squares of distances for each list of ligands only needs to be precomputed once.

This expression thus allows computing the distance between a theoretical "centroid ligand" of a protein and that of another protein, providing a measure of "distance" or dissimilarity between the pockets themselves. Pairwise distances between pockets were represented using multidimensional scaling (MDS) with the R package *MASS*. Notice that no information is lost in this visualization, and the plotted distances correspond to the Tanimoto or USRCAT distances between centroids.

2.8. Consensus docking and consensus scoring

The consideration of multiple docking software and multiple rescoring functions has been shown to improve the accuracy of computational predictions over the use of a single software package.[27–29] Thus, the top 1,000 virtual ligands of the target of interest identified by PSOVina2.0 were re-docked independently using Vina 1.1.2 and

Autodock 4.2.6. The RMSD between the top redocked pose from each algorithm was computed. Low RMSD values were regarded as a sign of higher confidence in the predictions. The resulting top pose for each compound from Vina and Autodock was re-scored using the scoring functions X-Score 1.3,[30] NNScore 1.0,[31] NNScore 2.01,[32] DSX_CSD 0.89,[33] and RFScore-VS[34] (see file mmc2.xlsx). Compounds to which multiple independent scoring algorithms assigned a tight predicted binding affinity were favored. By considering all results, a subset of 20 compounds was manually prioritized. These compounds were then additionally re-scored using K_{DEEP}[35] on the PSOVina poses (file mmc2. xlsx), were assessed for ADMET red flags (see below), and were queried for IP claims in the SureChEMBL database. Considering all results, a subset of 10 compounds was manually selected for experimental testing.

2.9. Prediction of ADMET properties

StarDrop 6.6.4 was used to estimate key ADMET-related properties, including partition coefficient, polar surface area, and the pIC_{50} for hERG receptor binding. Selected compounds were also analyzed for potential toxicity endpoints using Derek Nexus v.6.1.0. More than three warnings or two plausible toxicity flags meant that the molecule was not included for further study.

2.10. Test compounds

9 shortlisted compounds (Table 2) were sourced directly from the manufacturer (Enamine, Ukraine), with purity > 90%, generally > 95%. 5 mg of each compound were acquired in powder form and dissolved in DMSO as stocks for the different assays.

2.11. RAD51-p53 ELISA assay

96-well plates were coated with p53 protein overnight (about 200 ng per well) in carbonate buffer. The protein was then blocked with BSA and washed with PBS. Compound mixes were prepared by diluting 2 μl of test compound (10 mM in DMSO) along with 100 ng of RAD51 (Abcam, ab63808) to 100 μl with PBS. Positive controls were prepared by diluting 2 μl ATP or AMP-PNP (10 mM in DMSO) along with 100 ng of RAD51 to 100 μl with PBS. A negative control was prepared by diluting 100 ng of RAD51 to 100 μl with PBS. A DMSO control was similarly prepared by diluting 2 μl DMSO along with 100 ng of RAD51 to 100 μl with PBS. Each mix was then added to a separate p53-coated well, incubated for two hours, and washed with PBS. Primary antibodies against RAD51 were introduced (Cell Signaling Technology, Rabbit mAb #8875), incubated for 45 min, and washed 3 times with PBS. Then, fluorescent secondary antibodies were introduced, incubated for 45 min, and washed 3 times with PBS. Fluorescence was then measured and normalized by the signal from the normalization well. In the absence of ATP or other inhibitors, RAD51 binds to p53, and the antibodies detect it. The fluorescent signal is approximately proportional to the amount of bound RAD51. If ATP or another inhibitor is present, then RAD51 cannot bind, it is washed away, and the fluorescence signal is reduced.

2.12. Cell culture

U2OS cells were cultured in Dulbecco's modified Eagle medium (DMEM, Gibco) supplemented with 10% fetal bovine serum (FBS, Biological Industries) at 37 °C and 5% CO_2.

2.13. Homologous recombination assay

U2OS cell line carrying the direct repeats green fluorescent protein (DR-GFP) reporter was used to analyze HR activity as described previously.[36] Briefly, cells were seeded in 6-well plates and treated with the appropriate concentration of test compound for 24 h, followed by transfection with pCBASceI using Lipofectamine 3000 transfection reagent (Invitrogen) according to the manufacturer's instructions. 24 h after transfection, the medium was changed for fresh medium with the test compound. 48 h after transfection, cells were then trypsinized, and the faction of GFP-positive cells was evaluated using a FACSCalibur flow cytometer (BD Biosciences).

2.14. Immunoblotting analysis

Cells were incubated in lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% NP-40, 10% glycerol, 1 mM EDTA, 10 μg/ml aprotinin, 10 μg/ml chymostatin, 10 μg/ml leupeptin, 10 μg/ml pepstatin A, and 1.5 mM PMSF) on ice for 15 min and sonicated afterwards. The lysate was clarified by centrifugation at 13,700 rcf for 15 min. Then protein concentration was determined by the BCA protein assay kit (Pierce). After running in sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) consisting of 10% acrylamide, proteins were transferred onto polyvinylidene fluoride (PVDF) membrane. The membrane was blocked with 5% bovine serum albumin (BSA) in PBS containing 0.01% Tween-20 (PBST) at room temperature for 1 h and it was then incubated with primary antibodies at 4 °C overnight. The next day, the membrane was incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies at room temperature for 1 h after washing three times with PBST. Then, the membrane was washed three times before being incubated with enhanced chemiluminescent horseradish peroxidase substrate (ThermoFisher) for 3 min. Images were acquired using the BioSpectrum imaging systems (UVP).

The primary antibodies used were anti-HA (Santa Cruz, sc-7392, for HA-I-SceI, 1:1000 dilution) and anti-α tubulin (GeneTex, GTX112141, 1:10,000 dilution). The HRP-conjugated secondary antibodies were anti-rabbit (GeneTex, GTX213110–01, 1:5000 dilution) and anti-mouse (GeneTex, GTX213111–01, 1:5000 dilution).

2.15. Cell cycle analysis

Cells were washed with PBS and fixed in 70% ethanol at 4 °C overnight. This was followed by washing and staining in PBS containing propidium iodide (PI, 20 μg/ml), ribonuclease A (200 μg/ml), and Triton X-100 (0.1% v/v) at 37 °C for 30 min. Cell cycle distribution was calculated and analyzed by detecting the signal of DNA-bound PI fluorescence using a FACSCalibur flow cytometer (BD Biosciences). In principle, The G2/M cells were expected to have a 2-fold fluorescence intensity than the G1 cells. Thus, the fluorescence intensity was used to analyze the DNA content, delineating 2 N (G1 phase), 2 N-4 N (S phase), and ≥ 4 N (G2/M phase) cells.

3. Results and discussion

3.1. Structural comparisons of ATP binding pockets and target selection

Since a drug may interact with a range of ATP-binding proteins in the body, to achieve specificity it can be advantageous to select a target with unique features in or around the binding pocket. Thus, we considered a diverse set of ATPases (Table S1) and explored several strategies to compare their ATP binding sites. We devised two structural comparison approaches aimed at identifying ATP binding pockets that display distinct structural features. We argue that such distinctiveness would be a desirable trait when selecting a target, as it could allow the design of more specific drugs against it.

As a first structural comparison approach, we generated an orientation-independent distance profile or "signature" of residues around the ATP-binding site for each protein in the study (see Methods and Figure S1). The results allow a visual comparison of the degree of similarity of ATP-binding sites containing a Walker-A motif. For example, results in Figure S1 suggest that the ATP-binding site pockets in structures 5tf7 and 2www are very similar, whereas those in 2xsz and 2fh5 are significantly different. Similarly, the ATP-binding pocket in

2fh5 looks distinct from the rest, as do those in 3d8b, 2zan, or 1bg2. This kind of information can be useful in target selection.

A second approach was based on the computation of pairwise distances between pockets and the subsequent visualization of the resulting matrix of distances. The software APoc implements powerful alignment-based ideas to compute a normalized distance score between any pair of protein pockets.[18] We defined the pockets based on a distance cutoff to the end of the Walker-A motif and used Apoc to compare all possible pairs between the 26 pockets. The resulting pairwise distance matrix was then projected using a metric multidimensional scaling (MDS) algorithm and visualized. The results are shown in Figure S2 and suggest that pockets like those in 2fh5, 5nwl, 4rh7, and 4f93 are quite distinct from the rest, in reasonable agreement with the method above.

An alternative indirect comparison was carried out by comparing results from docking a small library of compounds against the targets. For this, we used a diverse set of 2,458 lead-like compounds. We then assessed the similarity of the top virtual hits for each protein. The centroids defined by the top hits was used to compute dissimilarities (distances) between every pair of proteins in two different ways: i) using a Tanimoto-based distance, which considers the 2D chemical structure of the ligands, and ii) using USRCAT-based distance, which takes into account the 3D poses of the ligands.[25,37] The resulting distances were projected for visualization, as shown in Figure 2. The plot facilitates the comparison of ATP-binding pockets. For example, it seems that the

Fig. 2. Indirect comparison of ATP binding pockets of selected ASCE ATPases. 2D multi-dimensional scaling (MDS) projection depicting **A,** Tanimoto-based distances between the centroids of the top 12 ligands onto each target, **B,** USRCAT-based distances between the centroids of the top 10 virtual ligands for each target. The targets selected for subsequent studies are highlighted in blue.

pockets in 4rh7 and 5tf7 behave similarly in docking, whereas those of 3d8b and 1bg2 are much more different. These plots can be further summarized in 1D metrics, as shown in Table 1.

Thanks to the use of centroids, this method is more robust than comparing individual ligands. We did not observe a significant clustering of the structures depending on the presence of different ligands in the native crystal structure or the resolution, suggesting that other differences in the binding pockets are being captured by the method as more relevant in this case. However, if these factors were a concern when applying the method to other scenarios, one could relax the protein structures using molecular dynamics, or include multiple conformations of each protein to help identify one that can targeted specifically.[38,39]

Although there are differences depending on the specifics of the comparison method (e.g., compare Fig. 2a and 2b), distinct ATP-binding pockets are identifiable in both indirect approaches, such as, those in 1bg2, 2qz4, 5nwl, or 3d8b. Note that 1bg2 and 3d8b had also been identified as interesting from the direct structure-based methods above.

As an additional criterion to select a protein target for drug discovery, we considered the ligandability of each protein pocket, that is, the facility to identify potent small-molecule ligands binding to it. For this, we considered the software DoGSiteScorer,[23] with some key results summarized in Table 1. We also investigated the consistency of docking results as a potential indicator of ligandability. Molecular docking solves a complex energy optimization problem between the protein and a small molecule by varying the conformation of the latter. We hypothesized an energy function with a deeper global minimum (i.e., more ligandable) may be easier to find tight binders for, and therefore docking results would be more consistent, than pockets that are flatter and have shallower energy local minima. Indeed, this is what we observe in Fig. 3, with structures for which docking results are more consistent leading to the prediction of tighter binders in a given library, suggesting that the targets are more ligandable. In fact, we do observe some correlation between the consistency of docking and the DoGSiteScore ligandability metric (Figure S3). We also find that trends observed in small libraries (Figure S4) are also informative of trends in larger libraries (Figure S5), in line with previous observations on ligandability.[40]

Based on the results from the direct and indirect structural comparisons, preliminary docking, and ligandability assessments, a subset of target protein structures was selected for further study: 5nwl, 2xsz, 3d8b, and 4ay2. In particular, 5nwl (RAD51) displayed a distinct structural signature (Figure S1) and also bound tightly to compounds that looked relatively different from its closest neighbors (Fig. 2), both when considered in 2D or 3D, and it also seemed to have good ligandability (Table 1 and Fig. 3). Structures 2xsz, 3d8b, and 4ay2 were selected as potential off-targets based on their relative proximity to 5wnl (Fig. 2), reasonably diverse structural signatures (Figure S1), and promising predicted ligandability (Table 1 and Fig. 3).

Structure 2xsz corresponds to a dodecameric complex of pontin (RuvBL1) and reptin (RuvBL2), which has a wide range of functions.[41] 3d8b is a structural model of fidgetin-like protein 1 (FIGNL1), which is involved in DNA double-strand break (DSB) repair, among others.[42] FIGNL1 interacts with RAD51 through a conserved binding domain and can modulate filament formation.[43] 4ay2 corresponds to retinoic acid-inducible gene I (RIG-I), involved in identifying viral RNA and initiating acute innate immune response.[44,45] 5nwl is a structural model of human RAD51. When a double-strand DNA break occurs, the cell can utilize the sister chromatid as a template to fill in the missing information and close the gap, a process called homologous recombination (HR).[46] For this, 3′ ssDNA overhangs around the break are produced by a nuclease, and these are rapidly coated by replication protein A (RPA) to prevent degradation or formation of secondary structures. Subsequently, RAD51 displaces RPA and assembles around the ssDNA to form a presynaptic filament (PSF), similar to RecA in *E. coli*. This process is controlled by mediator proteins such as BRCA2.[47,48] RAD51 fragments elongate via multiple nucleation events followed by growth, which stops
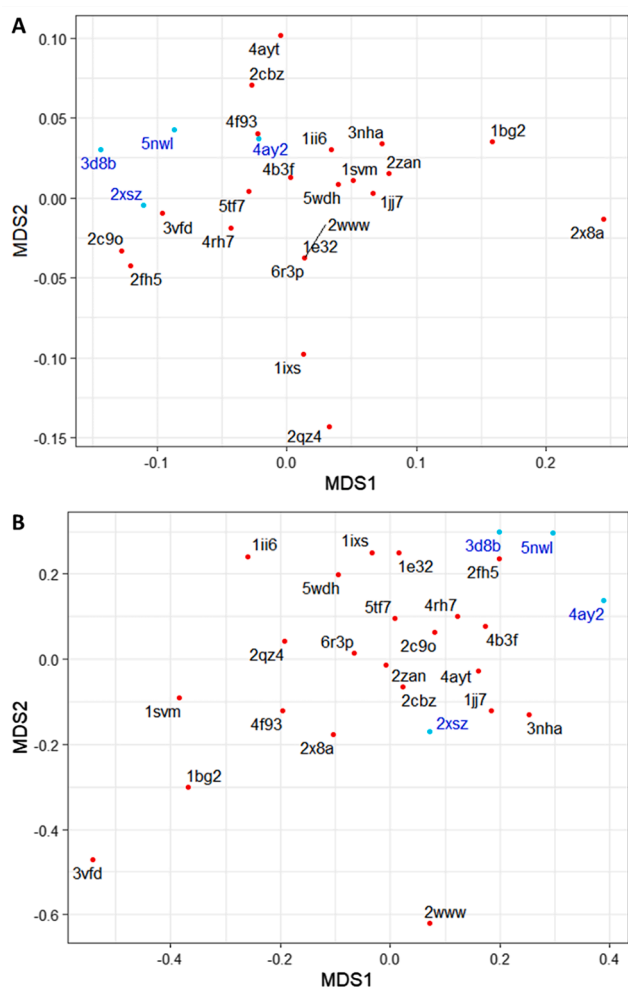
Table 1

Summary of druggability (ligandability) and similarity metrics across the ATP-binding pockets of the different ATPases. Targets used for virtual screening are highlighted in bold.

Structure	DoGSite pocket volume (Å³)	DoGSite *Druggability Score*	Docking consistency (Pearson R^2)	Docking consistency (Spearman R^2)	Tanimoto distance to closest centroid	USRCAT distance to closest centroid
1bg2	561	0.75	0.552	0.533	0.129	0.248
1e32	704	0.8	0.694	0.673	0.000	0.236
1ii6	369	0.7	0.736	0.715	0.032	0.237
1ixs	1101	0.81	0.775	0.759	0.095	0.241
1jj7	117	0.13	0.626	0.603	0.113	0.204
1svm	751	0.81	0.824	0.801	0.032	0.255
2c9o	1163	0.81	0.880	0.865	0.060	0.227
2cbz	305	0.48	0.610	0.579	0.046	0.229
2fh5	507	0.77	0.752	0.735	0.060	0.220
2qz4	671	0.8	0.667	0.639	0.113	0.251
2www	191	0.41	0.548	0.529	0.000	0.283
2x8a	312	0.55	0.607	0.578	0.187	0.235
2xsz	–	–	0.862	0.856	0.079	0.239
2zan	735	0.77	0.619	0.592	0.060	0.241
3d8b	2856	0.81	0.840	0.830	0.065	0.220
3nha	190	0.29	0.663	0.642	0.077	0.204
3vfd	778	0.83	0.797	0.782	0.077	0.254
4ay2	150	0.24	0.711	0.694	0.046	0.227
4ayt	463	0.86	0.637	0.613	0.095	0.213
4b3f	670	0.84	0.713	0.689	0.082	0.242
4f93	386	0.61	0.738	0.720	0.119	0.235
4rh7	204	0.42	0.682	0.664	0.080	0.227
5nwl	1132	0.78	0.936	0.932	0.077	0.250
5tf7	312	0.72	0.585	0.575	0.098	0.233
5wdh	709	0.85	0.684	0.671	0.058	0.237
6r3p	546	0.79	0.657	0.626	0.000	0.242

when the fragment reaches about 2 μm long.[49] The functional filament structure requires ATP binding[50] and has a secondary site that binds the dsDNA template, allowing the 3′ ssDNA overhang to be searched along for homology.[51] When the region of homology is found, RAD51 catalyzes strand exchange. The missing sequence region is completed by DNA polymerase, followed by migration of the branches and resolution of the Holliday junctions to complete HR.

3.2. Computational drug design against RAD51

Novel potent inhibitors of HR would be highly desirable for cancer treatment, and the ATP binding pocket of RAD51 offers a promising target to this end.[52–54] Tumor cells often undergo high mutation rates, which are supported by overexpression of enzymes like cytidine deaminase and increased DNA damage.[55] Inhibiting HR can lead to the death of cells experiencing hypermutation.[52] RAD51 is overexpressed in many tumors, including familial BRCA1-deficient breast tumors, where it may compensate for the lack of functional BRCA1 or other DNA repair proteins[56–58]. Furthermore, inhibiting RAD51 may also increase the susceptibility of tumor cells to therapies that act by inducing interstrand crosslinks or double-strand breaks, such as cisplatin or radiotherapy[59]. However, RAD51 is proving to be a challenging target. For example, a recent study screening the whole NIH Small Molecule Repository (202,556 compounds) against RAD51 encountered a relatively low hit rate (0.09% compounds) and the most promising compound had an IC_{50} around 27 μM, which improved only marginally after SAR efforts.[60] Some small-molecule inhibitors have been reported in studies over the years (Table S2). Unfortunately, most display limited potency and make poor lead compounds for further optimization due to toxicity or non-lead-like physicochemical profiles. Thus, the application of novel computational approaches considering multiple design dimensions could be helpful to guide the discovery of novel inhibitors.

To discover new HR inhibitors, two focused libraries commercially available were combined and filtered for drug-likeness, resulting in a combined library of 26,567 chemical entities (see Methods). The compounds were then docked against the four structures selected (5nwl, 2xsz, 3d8b and 4ay2) using PSOVina2.0 with high exhaustiveness

settings. For every target, monotonic relationships were found between predicted binding energy and specificity (Fig. 4), which agrees with previous studies[61,62] and highlights the importance of considering ligandability in target selection.
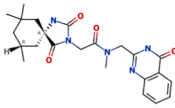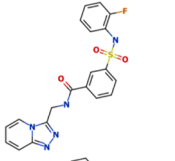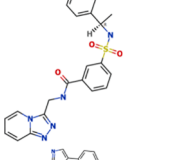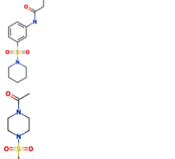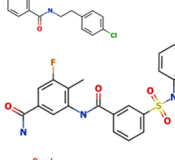
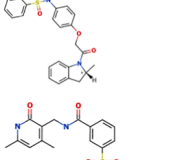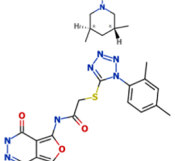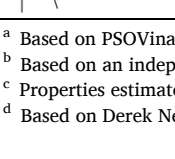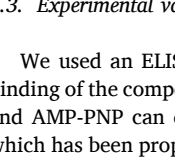The docked poses for compounds EN7883 and EN9632 are illustrated in Figure S6. Both compounds are predicted to establish contacts with two RAD51 units, mimicking the structural role of ATP. The imidazole ring in EN7883 seems capable of establishing multiple hydrogen bonds with residues of the Walker A motif (K133, T131, G132). The compound would also engage in hydrogen bonds with S317 and R310. On the other hand, compound EN9632 is predicted to fill more extensively the binding pocket and engage in many Van der Waals interactions. Two oxo- groups, from its carbonyl and sulfone groups, are predicted to participate in hydrogen binding with K133 and E322, respectively.

Previous work on consensus docking has shown that when multiple docking programs agree on a predicted binding pose, the pose is more likely to be correct.[27,63,64] Therefore, we also conducted independent docking of the top 1,000 virtual ligands on RAD51 using the software Vina and AutoDock. Similarly, consensus scoring schemes have demonstrated superior predictive accuracy and enrichment for true binders in virtual screens compared to individual scoring functions.[28] Therefore, the compounds were re-evaluated using a variety of scoring functions (see Methods and file mmc2.xlsx).

The potential toxicity of the molecules was considered using the software Derek Nexus v.6.1.0, a knowledge-based system that identifies potential toxicophores using pattern matching against curated libraries for a range of human toxicity endpoints. Most top compounds raised little toxicity concerns (Table S3), and those that raised more serious concerns were discarded (see Table S4 for examples). Predicted hERG binding was also considered using the QSAR models in StarDrop. The compounds proposed were predicted to have hERG IC_{50} values above 10 μM, which are considered acceptable for plasma-available drugs.[65] A subset of the top virtual ligands was manually selected based on the different predictions, shown in Table 2. None of these compounds had any patent conflicts regarding HR inhibition, according to SureChEMBL.

Table 2

Predicted properties for selected compounds on the ATP binding site of 5nwl after docking the extended library (26,567 drug-like compounds) on a subset of related targets. Hydrogens are not displayed in the structures.

Compound structure	Compound id.	Consensus docking	Consensus scoring[a]	Specificity for 5nwl[a]	ADMET properties
	EN7426	PSOVina2.0 K_i (nM) = 1.1 Vina K_i (nM) = 1.3 AutoDock K_i (nM) = 1.8 RMSD[b] < 2.0 Å	NNScore K_i > 25 µM K_{DEEP} K_i (nM) = 1.1 CS = 7	77%	MW[c] = 439.5 Da logP[c] = 1.97 TPSA[c] = 115.5 hERG pIC$_{50}$[c] = 4.37 Toxicity[d] = low
	EN9928	PSOVina2.0 K_i (nM) = 1.3 Vina K_i (nM) = 1.1 AutoDock K_i (nM) = 3.1 RMSD[b] < 2.0 Å	NNScore K_i < 25 µM K_{DEEP} K_i (nM) = 17 CS = 15	82%	MW[c] = 425.4 Da logP[c] = 2.07 TPSA[c] = 105.5 hERG pIC$_{50}$[c] = 5.63 Toxicity[d] = low
	EN16731	PSOVina2.0 K_i (nM) = 2.0 Vina K_i (nM) = 1.1 AutoDock K_i (nM) = 1.8 RMSD[b] > 2.0 Å	NNScore K_i < 25 µM K_{DEEP} K_i (nM) = 6.7 CS = 44	80%	MW[c] = 435.5 Da logP[c] = 2.56 TPSA[c] = 105.5 hERG pIC$_{50}$[c] = 5.38 Toxicity[d] = low
	EN21991	PSOVina2.0 K_i (nM) = 6.1 Vina K_i (nM) = 3.1 AutoDock K_i (nM) < 1 RMSD[b] < 2.0 Å	NNScore K_i > 25 µM K_{DEEP} K_i (nM) = 24 CS = 9	74%	MW[c] = 439.5 Da logP[c] = 3.36 TPSA[c] = 92.51 hERG pIC$_{50}$[c] = 5.43 Toxicity[d] = low
	EN4999	PSOVina2.0 K_i (nM) = 7.3 Vina K_i (nM) = 8.6 AutoDock K_i (nM) < 1 RMSD[b] < 2.0 Å	NNScore K_i < 25 µM K_{DEEP} K_i (nM) = 9.7 CS = 4	76%	MW[c] = 450 Da logP[c] = 2.03 TPSA[c] = 86.8 hERG pIC$_{50}$[c] = 5.05 Toxicity[d] = low
	EN19056	PSOVina2.0 K_i (nM) < 1 Vina K_i (nM) < 1 AutoDock K_i (nM) < 1 RMSD[b] > 2.0 Å	NNScore K_i < 25 µM K_{DEEP} K_i (nM) = 21 CS = 33	77%	MW[c] = 445.4 Da logP[c] = 2.52 TPSA[c] = 118.4 hERG pIC$_{50}$[c] = 5.01 Toxicity[d] = low
	EN9632	PSOVina2.0 K_i (nM) = 5.5 Vina K_i (nM) = 6.1 AutoDock K_i (nM) < 1 RMSD[b] < 2.0 Å	NNScore K_i > 25 µM K_{DEEP} K_i (nM) = 47 CS = 11	66%	MW[c] = 436.5 Da logP[c] = 3.22 TPSA[c] = 66.9 hERG pIC$_{50}$[c] = 5.78 Toxicity[d] = low
	EN3339	PSOVina2.0 K_i (nM) = 4.3 Vina K_i (nM) = 3.7 AutoDock K_i (nM) < 1 RMSD[b] < 2.0 Å	NNScore K_i < 25 µM K_{DEEP} K_i (nM) = 8.5 CS = 2	74%	MW[c] = 431.5 Da logP[c] = 2.71 TPSA[c] = 99.3 hERG pIC$_{50}$[c] = 4.47 Toxicity[d] = low
	EN7883	PSOVina2.0 K_i (nM) = 2.2 Vina K_i (nM) = 2.6 AutoDock K_i (nM) = 6.4 RMSD[b] > 2.0 Å	NNScore K_i > 25 µM K_{DEEP} K_i (nM) = 18 CS = 369	81%	MW[c] = 425.5 Da logP[c] = 2.66 TPSA[c] = 131.6 hERG pIC$_{50}$[c] = 4.34 Toxicity[d] = low

[a] Based on PSOVina2.0 docking.
[b] Based on an independent run with the original Vina and AutoDock algorithms.
[c] Properties estimated using StarDrop 6.6.4.
[d] Based on Derek Nexus predictions (see Table S3 for details).

3.3. Experimental validation of HR inhibitors

We used an ELISA-based assay as a first approach to evaluate the binding of the compounds to RAD51. Previous studies reported that ATP and AMP-PNP can disrupt the interaction between hRAD51 and p53, which has been proposed to have potential interest in cancer therapy.[66] We reasoned that compounds that bind to the same binding pocket as ATP may be similarly effective at disrupting the interaction. A preliminary screen at a fixed concentration (Figure 5), confirms that ATP and AMP-PNP disrupt the interaction between hRAD51 and p53. Notably, we observe that compounds EN19056, EN9632, EN3339, and EN7883 also lead to a disruption of the interaction. These compounds were thus prioritized for further cellular study.

It has been well documented that an ATP nucleotide cofactor is a prerequisite for RAD51 to assemble a functional nucleoprotein filament and conduct homology-directed double-strand break (DSB) repair.[50] Our newly discovered compounds could compete with ATP to RAD51. Thus, we sought to examine whether these compounds could inhibit RAD51 enzymatic activity at the cellular level. A RAD51-mediated HR reporter in the human cell line was utilized to evaluate whether compounds EN19056, EN9632, EN3339, and EN7883 could suppress HR activity. In this reporter assay, human bone osteosarcoma epithelial cells (U2OS) were engineered with a cassette containing an I-SceI endonuclease cleavage site. The design of the cassette leads to a fluorescent protein produced if the cleavage is repaired through RAD51-mediated HR (Fig. 6A). The results evidenced that the fraction of fluorescent cells in the assay, and therefore the extent of HR, is greatly diminished when the compound EN7883 or EN9632 is present in the media in the
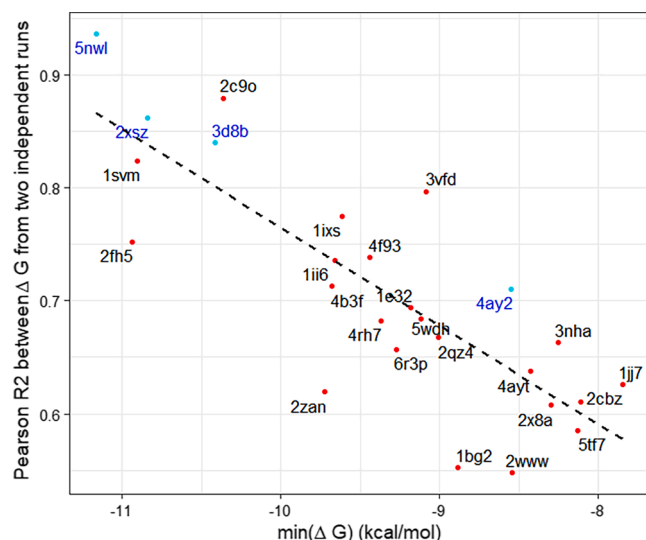
Fig. 3. Comparison of the consistency of docking results, measured based on Pearson R^2 between binding energies from two independent docking runs (each with exhaustiveness = 32) of 2,458 compounds, and the strongest binding affinity identified, min(ΔG).
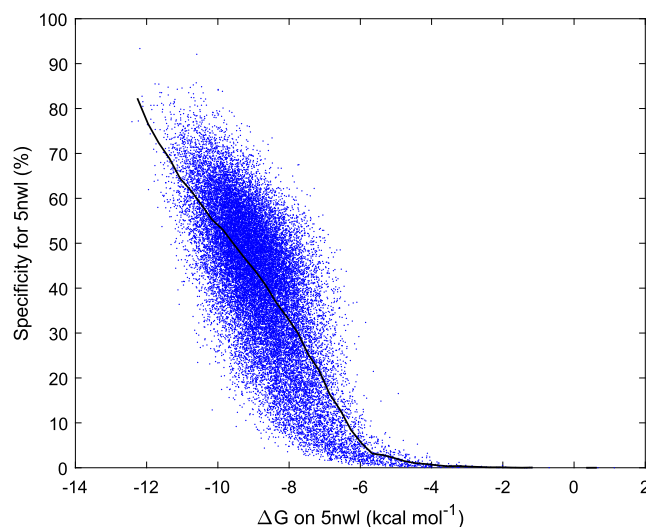

Fig. 4. Relationship between specificity for 5nwl over other selected targets and binding affinity across 26,567 drug-like compounds. The solid line represents the moving average with a 0.3 kcal mol$^-$ [1] window.

Fig. 5. Fluorescent ELISA assay results when incubating RAD51 and p53 in the presence of different test compounds. RAD51 and p53 interact with each other, and this interaction is inhibited by ATP and AMP-PNP (positive controls). Each well contained 200 ng of p53 and was dispensed 100 ng of RAD51 and 200 μM of the corresponding test compound. Fluorescence is expressed relative to the control with only RAD51 and p53 (PBS, negative control). Lower fluorescence values indicate a more effective disruption of the p53-RAD51 interaction. Compounds corresponding to the white bars were selected for further study.

4. Conclusion

ATP binding pockets are present in many enzymes beyond kinases. ATP hydrolysis is often used to drive conformational changes in the enzymes that are necessary for a high turnover frequency. Some of these ATPases are related to important medical conditions, and small-molecule inhibitors against the corresponding ATP binding pocket could offer novel therapeutic possibilities. In this case, the selection of the proper ATPase target is key to minimizing binding to off-targets, which could cause undesired effects.

We demonstrate a rational target selection approach adapted to ATPases, in which the structures of ATP binding pockets are considered along with their ligandability and therapeutic potential. We demonstrate different computational tools to compare structures of ATP binding sites, including a distance-based pocket signature. We also show how molecular docking allows comparing binding pockets and assessing their relative ligandability. In particular, the consistency of individual docking runs can inform about the predicted affinity of virtual hits identifiable against the target pursued. Based on the results, we decided to pursue human RAD51 in a computer-driven campaign.

An updated virtual screening pipeline was applied to human RAD51, illustrating how potential off-targets can be considered, and how multiple docking and scoring results can be integrated in a consensus. Our approach also considers ADME and toxicity predictions to filter out problematic compounds *in silico*. Despite starting with a pre-filtered, focused chemical library, some of the virtual hits had likely toxicity liabilities, which are one of the major causes for attrition in drug discovery. This suggests that, besides the conventional filters based on physicochemical properties, it would also be advisable to filter virtual libraries based on the presence of probable toxicophores. Ten compounds were selected for experimental target-based assays, and four were prioritized for cell-based assays. Two of the compounds studied were able to inhibit homologous recombination in human cells. These two hit compounds have not been identified before and may warrant follow-up studies.

More generally, the pipeline demonstrated serves as a blueprint for broader drug discovery campaigns targeting the ATP binding pockets of ATPases. In future efforts, it could be interesting to use larger set of ATPases, including novel structures recently made available through predictive modeling. The pipeline might also be applied to target ATPases in pathogens, such as viruses or bacteria. To improve the

low micromolar concentration range (Fig. 6B). We ruled out the possibility that attenuation of HR could have stemmed from the poor expression of endonuclease using I-SceI immunoblot analysis. It is worth noting that higher concentration (80 μM) of test compound did affect the I-SceI protein level. Since HR is conducted mainly in S/G2 phase cells, we examined whether those chemicals affected cell cycle progression. The results of the cell cycle analysis (Fig. 6C) indicate that compounds EN7883 and EN9632 do not have any significant effect on cell cycle progression.

In conclusion, the results indicate that compounds EN7883 and EN9632 are novel promising hits against RAD51-mediated homologous recombination, warranting medicinal chemistry efforts to further optimize them. The results also evidence the ability of the pipeline proposed to identify promising ATPase targets and to discover interesting chemical matter to modulate them.
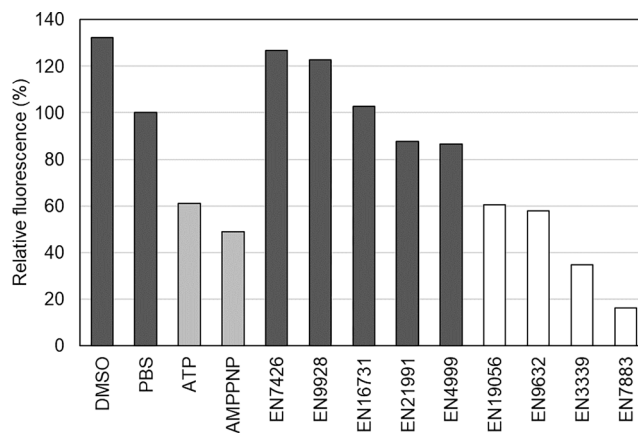
accuracy of the results, multiple conformations of ATP binding pockets could be considered. Machine-learning-based scoring functions also have potential to increase the accuracy and enrichment achieved in computer-assisted drug discovery.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bmc.2022.116923.

References

[1] Shalaeva DN, Cherepanov DA, Galperin MY, Mulkidjanian AY. Comparative analysis of active sites in P-loop nucleoside triphosphatases suggests an ancestral activation mechanism. Bioarxiv 2018;(Journal Article):https://doi.org/10.1101/439992.
[2] Hong L, Sklar LA. Targeting GTPases in Parkinson's disease: comparison to the historic path of kinase drug discovery and perspectives. *Front Mol Neurosci.* 2014;7(52):1–10.
[3] Zhang J, Yang PL, Gray NS. Targeting cancer with small molecule kinase inhibitors. Nat Rev Cancer 2009;9(Journal Article):28–39.
[4] del Toro D, Ortiz D, Ordyan M, et al. Walker-A motif acts to coordinate ATP hydrolysis with motor output in viral DNA packaging. *J Mol Biol.* 2016;428(13):2709–2729.
[5] Young PR, McLaughlin MM, Kumar S. et al. Pyridinyl imidazole inhibitors of p38 mitogen-activated protein kinase bind in the ATP site. The Journal of Biological Chemistry 1997;**272**(Journal Article):12116–21.
[6] Huang D, Zhou T, Lafleur K, Nevado C, Caflisch A. Kinase selectivity potential for inhibitors targeting the ATP binding site: A network analysis. *Bioinformatics.* 2010;26(2):198–204.
[7] Snider J, Thibault G, Houry WA. The AAA+ superfamily of functionally diverse proteins. *Genome Biol.* 2008;9(4):1–8.
[8] Miller JM, Enermark EJ. Fundamental Characteristics of AAA+ Protein Family Structure and Function. Archaea 2016;**2016**(Journal Article):9294307.
[9] Iyer LM, Makarova KS, Koonin EV, Aravind L. Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: Implications for the origins of chromosome segregation, cell division and viral capsid packaging. *Nucleic Acids Res.* 2004;32(17):5260–5279.
[10] Pi F, Vieweger M, Zhao Z, Wang S, Guo P. Discovery of a new method for potent drug development using power function of stoichiometry of homomeric biocomplexes or biological nanomotors. *Expert Opinion in Drug Delivery.* 2016;13(1):23–36.
[11] Hengel SR, Spies MA, Spies M. Small molecule inhibitors targeting DNA repair and DNA repair deficiency in research and cancer therapy. *Cell Chem Biol.* 2017;24(9):1101–1119. https://doi.org/10.1016/j.chembiol.2017.08.027.
[12] Chène P. ATPases as Drug Targets: Learning From Their Structure. *Nat Rev Drug Discov.* 2002;1(9):665–673.
[13] Vetter IR, Wittinghofer A. Nucleoside Triphosphate-Binding Proteins: Different Scaffolds to Achieve Phosphoryl Transfer. *Q Rev Biophys.* 1999;32(1):1–56.
[14] Wendler P, Ciniawsky S, Kock M, Kube S. Structure and function of the AAA+ nucleotide binding pocket. *Biochimica et Biophysica Acta (BBA) - Molecular. Cell Res.* 2012;1823(1):2–14.
[15] Tafoya S, Bustamante C. Molecular switch-like regulation in motor proteins. *Phil Trans R Soc B.* 2018;373(1749):20170181.
[16] Schwartz R, King J. Frequencies of hydrophobic and hydrophilic runs and alternations in proteins of known structure. *Protein Sci.* 2006;15(1):102–112.
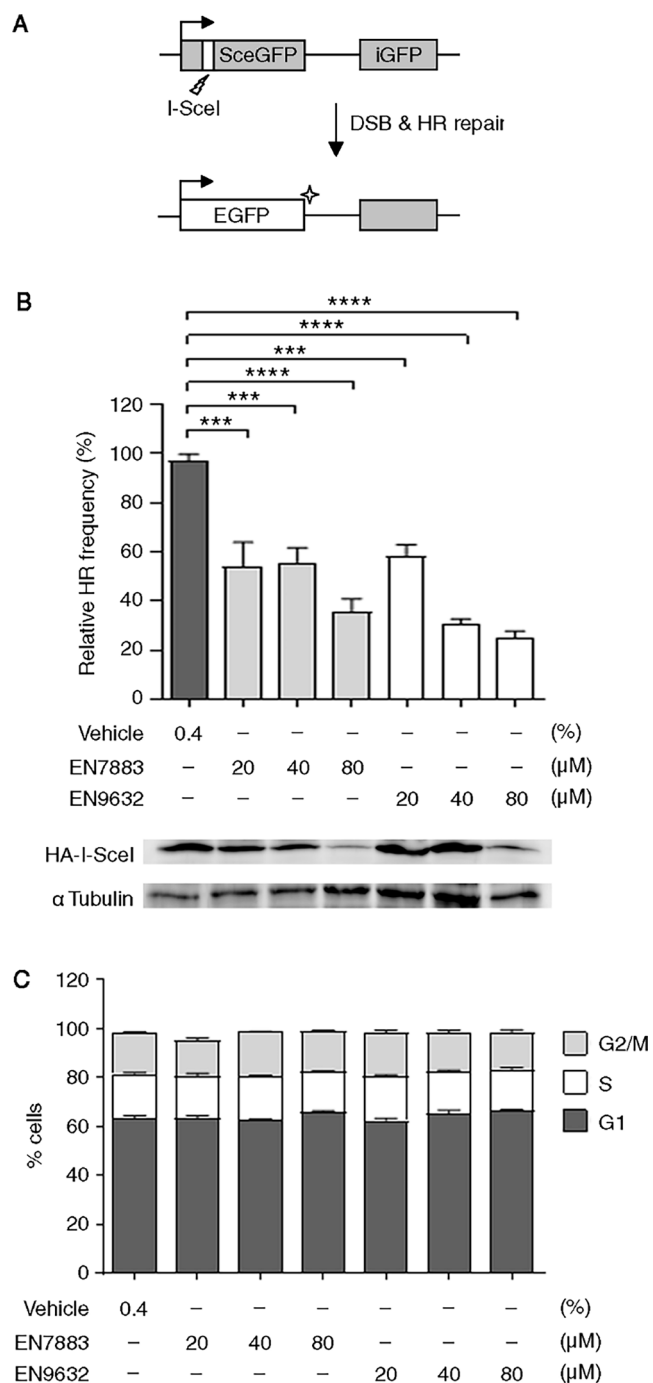[17] Pantsar T, Poso A. Binding Affinity via Docking: Fact and Fiction. *Molecules.* 2018;23(8):1899.

Fig. 6. Cell-based assays of compounds EN7883 and EN9632 (see Table 2 for structure). **A,** Schematic of the DR-GFP reporter assay to quantify HR activity. Mechanistically, the reporter cassette DR-GFP harbors two copies of the inactive *EGFP* gene, SceGFP and iGFP. Then, the DNA double-strand breaks (DSB) repair is triggered by the expression of the I-SceI endonuclease, which generates a site-specific DSB within the SceGFP sequence. A functional *EGFP* sequence can be expressed after the HR pathway is used. **B, C,** U2OS cells were treated with the indicated concentration of compounds for 24 h before I-SceI transfection. After 48 h of transfection, GFP $^+$ cells and cell cycle profiles were quantified by flow cytometry. The percentage of GFP$^+$ cells was normalized by that of cells treated with vehicle (0.4% DMSO). In addition, protein expression of HA-tagged I-SceI and tubulin are shown. ***P $<$ 0.001; ****P $<$ 0.0001. Bars indicate the mean $\pm$ s.e.m. from three independent experiments (n = 3). Statistical analyses consisted of one-way ANOVA with Tukey's post hoc test.

[18] Skolnick J, Gao M. Interplay of physics and evolution in the likely origin of protein biochemical function. *PNAS*. 2013;110(23):9344–9349. https://doi.org/10.1073/pnas.1300011110.

[19] Hann MM, Oprea TI. Pursuing the Leadlikeness Concept in Pharmaceutical Research. *Curr Opin Chem Biol*. 2004;8(3):255–263.

[20] Oprea TI, Gottfries J, Sherbukhin V, Svensson P, Kühler TC. Chemical information management in drug discovery: optimizing the computational and combinatorial chemistry interfaces. *J Mol Graph Model*. 2000;18(4–5):512–524.

[21] Tai HK, Jusoh SA, Siu SWI. Chaos-embedded particle swarm optimization approach for protein-ligand docking and virtual screening. Journal of Cheminformatics 2018;10(Journal Article):62.

[22] Blay V, Li MC, Ho SP, Stoller ML, Hsieh HP, Houston DR. Design of drug-like hepsin inhibitors against prostate cancer and kidney stones. Acta Pharmaceutica Sinica B 2019;(Journal Article). Doi: 10.1016/j.apsb.2019.09.008.

[23] Volkamer A, Kuhn D, Rippmann F, Rarey M. DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics*. 2012;28(15):2074–2075.

[24] Cao DS, Xiao N, Xu QS, Chen AF. Rcpi: R/Bioconductor package to generate various descriptors of proteins, compounds and their interactions. *Bioinformatics*. 2014;31(2):279–281.

[25] Kumar A, Zhang KYJ. Advances in the Development of Shape Similarity Methods and Their Application in Drug Discovery. Front Chem 2018;6(Journal Article):315.

[26] Paolella MS. *Linear Models and Time-Series Analysis: Regression, ANOVA, ARMA and GARCH*. Wiley; 2018.

[27] Houston DR, Walkinshaw MD. Consensus docking: improving the reliability of docking in a virtual screening context. *J Chem Inf Model*. 2013;53(2):384–390.

[28] Perez-Castillo Y, Sotomayor-Burneo S, Jimenes-Vargas K, et al. CompScore: Boosting Structure-Based Virtual Screening Performance by Incorporating Docking Scoring Function Components into Consensus Scoring. *J Chem Inf Model*. 2019;59(9):3655–3666.

[29] Li H, Sze KH, Lu G, Ballester PJ. Machine-learning scoring functions for structure-based virtual screening. Wires Comput Mol Sci 2020;(Journal Article):e1478.

[30] Wang R, Lai L, Wang S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput Aided Mol Des*. 2002;16(1):11–26.

[31] Durrant JD, McCammon JA. NNScore: A Neural-Network-Based Scoring Function for the Characterization of Protein−Ligand Complexes. *J Chem Inf Model*. 2010;50(10):1865–1871. https://doi.org/10.1021/ci100244v.

[32] Durrant JD, McCammon JA. NNScore 2.0: A Neural-Network Receptor-Ligand Scoring Function. *J Chem Inf Model*. 2011;51(11):2897–2903. https://doi.org/10.1021/ci2003889.

[33] Neudert G, Klebe G. DSX: A Knowledge-Based Scoring Function for the Assessment of Protein-Ligand Complexes. *J Chem Inf Model*. 2011;51(10):2731–2745. https://doi.org/10.1021/ci200274q.

[34] Wójcikowski M, Ballester PJ, Siedlecki P. Performance of machine-learning scoring functions in structure-based virtual screening. *Sci Rep*. 2017;7(1):46710. https://doi.org/10.1038/srep46710.

[35] Jiménez J, Skalic M, Martínez-Rosell G, De Fabritiis G. KDEEP: Protein–Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks. J Chem Inf Model 20018;58(2):287–96.

[36] Pierce AJ, Johnson RD, Thompson LH, Jasin M. XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes Dev*. 1999;13(20):2633–2638. https://doi.org/10.1101/gad.13.20.2633.

[37] Shin WH, Zhu X, Bures MG, Kihara D. Three-Dimensional Compound Comparison Methods and Their Application in Drug Discovery. *Molecules*. 2015;20(7):12841–12862.

[38] De Paris R, Quevedo CV, Ruiz DD, Norberto de Souza O, Barros RC. Clustering Molecular Dynamics Trajectories for Optimizing Docking Experiments. Computational Intelligence and Neuroscience 2015;2015:e916240. Doi: 10.1155/2015/916240.

[39] Basciu A, Malloci G, Pietrucci F, Bonvin AMJJ, Vargiu AV. Holo-like and Druggable Protein Conformations from Enhanced Sampling of Binding Pocket Volume and Shape. *J Chem Inf Model*. 2019;59(4):1515–1528. https://doi.org/10.1021/acs.jcim.8b00730.

[40] Edfeldt FNB, Folmer RHA, Breeze AL. Fragment screening to predict druggability (ligandability) and lead discovery success. *Drug Discov Today*. 2011;16(7–8):284–287. https://doi.org/10.1016/j.drudis.2011.02.002.

[41] Matias PM, Baek SH, Bandeiras TM, et al. The AAA+ proteins Pontin and Reptin enter adult age: From understanding their basic biology to the identification of selective inhibitors. Front Mol Biosci 2015;2(Journal Article):17.

[42] Ma J, Li J, Yao X, et al. FIGNL1 is overexpressed in small cell lung cancer patients and enhances NCI-H446 cell resistance to cisplatin and etoposide. *Oncol Rep*. 2017;37(4):1935–1942.

[43] Matsuzaki K, Kondo S, Ishikawa T, Shinoara A. Human RAD51 paralogue SWSAP1 fosters RAD51 filament by regulating the anti-recombinase FIGNL1 AAA+ ATPase. Nat Commun 20019;10(Journal Article):1407.

[44] Kell AM, Gale M. RIG-I in RNA virus recognition. *Virology*. 2017;37(4):1935–1942.

[45] Rawling D, Fitzgerald ME, Pyle AM. Establishing the role of ATP for the function of the RIG-I innate immune sensor. ELife 2015;4(Journal Article):1–21.

[46] Vilenchik MM, Knudson AG. Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer. *Proc Natl Acad Sci USA*. 2003;100(22):12871–12876.

[47] Sullivan MR, Bernstein KA. RAD-ical New Insights into RAD51 Regulation. *Genes*. 2019;9(12):629.

[48] Afshar N, Argunhan B, Palihati M, Taniguchi G, Tsubouchi H, Iwasaki H. A novel motif of Rad51 serves as an interaction hub for recombination auxiliary factors. *ELife*. 2021;10, e64131. https://doi.org/10.7554/eLife.64131.

[49] Candelli A, Holthausen JT, Depken M, et al. Visualization and quantification of nascent RAD51 filament formation at single-monomer resolution. *Proc Natl Acad Sci USA*. 2014;111(42):15090–15095.

[50] Andriuskevicius T, Kotenko O, Makovets S. Putting together and taking apart: assembly and disassembly of the Rad51 nucleoprotein filament in DNA repair and genome stability. *Cell Stress*. 2018;2(5):96–112. https://doi.org/10.15698/cst2018.05.134.

[51] Ito K, Murayama Y, Kurokawa Y, et al. Real-time tracking reveals catalytic roles for the two DNA binding sites of Rad51. *Nat Commun*. 2020;11(1):2950. https://doi.org/10.1038/s41467-020-16750-3.

[52] Budke B, Lv W, Kozikowski AP, Connell PP. Recent Developments Using Small Molecules to Target RAD51: How to Best Modulate RAD51 for Anticancer Therapy? *ChemMedChem*. 2016;11(22):2468–2473.

[53] Chen Q, Cai D, Li M, Wu X. The homologous recombination protein RAD51 is a promising therapeutic target for cervical carcinoma. *Oncol Rep*. 2017;38(2):767–774.

[54] Ferguson PJ, Vincent MD, Koropatnick J. Synergistic Antiproliferative Activity of the RAD51 Inhibitor IBR2 with Inhibitors of Receptor Tyrosine Kinases and Microtubule Protein Peter J. Ferguson, Mark D. Vincent and James Koropatnick. Journal of Pharmacology and Experimental Therapeutics 2018;364(1):46–54.

[55] Mills K, Cyr A, Maclay T, Day M, Hasham MG, Khalil A. A Small Molecule RAD51 Inhibitor Preferentially Affects Cells Expressing High Cytidine Deaminase Activity. Blood 2017;130(Journal Article):4627.

[56] Xia SJ, Shammas MA, Shmookler Reis RJ. Elevated recombination in immortal human cells is mediated by HsRAD51 recombinase. Mol Cell Bio 1997;17(Journal Article):7151–8.

[57] Maacke H, Opitz S, Jost K, et al. Over-expression of wild-type Rad51 correlates with histological grading of invasive ductal breast cancer. Int J Cancer 2000;88(Journal Article):907–13.

[58] Raderschall E, Stout K, Freier S, Suckow V, Schweiger S, Haaf T. Elevated levels of Rad51 recombination protein in tumor cells. Cancer Res 2002;62(Journal Article):219–25.

[59] Helleday T. Homologous recombination in cancer development, treatment and development of drug resistance. Carcinogenesis 2010;31(Journal Article):955–60.

[60] Huang F, Motlekar NA, Burgwin CM, Napper AD, Diamond SL, Mazin AV. Identification of specific inhibitors of human RAD51 recombinase using high-throughput screening. *ACS Chem Biol*. 2011;6(6):628–635.

[61] Eaton BE, Gold L, Zichi DA. Let's Get Specific: The Relationship Between Specificity and Affinity. *Chem Biol*. 1995;2(10):633–638.

[62] Yan Z, Zheng X, Wang E, Wang J. Thermodynamic and kinetic specificities of ligand binding. Chem Sci 2013;4(Journal Article):2387–95.

[63] Tuccinardi T, Poli G, Romboli V, Giordano A, Martinelli A. Extensive consensus docking evaluation for ligand pose prediction and virtual screening studies. *J Chem Inf Model*. 2014;54(10):2980–2986.

[64] Poli G, Martinelli A, Tuccinardi T. Reliability analysis and optimization of the consensus docking approach for the development of virtual screening studies. *J Enzyme Inhib Med Chem*. 2016;31(167–173).

[65] Pollard CE, Skinner M, Lazic SE, et al. An Analysis of the Relationship Between Preclinical and Clinical QT Interval-Related Data. *Toxicol Sci*. 2017;159(1):94–101.

[66] Buchhop S, Gibson MK, Wang XW, Wagner P, Stürzbecher HW, Harris CC. Interaction of p53 with the human Rad51 protein. *Nucleic Acids Res*. 1997;25(19):3868–3874. https://doi.org/10.1093/nar/25.19.3868.