



OPEN ACCESS

EDITED BY

Ana Cristina Gonçalves,
University of Coimbra, Portugal

REVIEWED BY

Ozge Sensoy,
Istanbul Medipol University, Türkiye
Teemu Johannes Haikarainen,
Tampere University, Finland
Matthew J. Cuneo,
St. Jude Children's Research Hospital,
United States

*CORRESPONDENCE

Davide Moiani,
✉ davide.moiani@gmail.com
✉ dmoiani@mdanderson.org
John A. Tainer,
✉ jatainer@gmail.com
✉ jtainer@mdanderson.org

RECEIVED 01 June 2024

ACCEPTED 28 October 2024

PUBLISHED 28 November 2024

CITATION

Moiani D and Tainer JA (2024) A goldilocks computational protocol for inhibitor discovery targeting DNA damage responses including replication-repair functions. *Front. Mol. Biosci.* 11:1442267. doi: 10.3389/fmolb.2024.1442267

COPYRIGHT

© 2024 Moiani and Tainer. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A goldilocks computational protocol for inhibitor discovery targeting DNA damage responses including replication-repair functions

Davide Moiani^{1,2,3*} and John A. Tainer^{1,2,4*}

¹Department of Molecular and Cellular Oncology, University of Texas MD Anderson Cancer Center, Houston, TX, United States, ²Department of Cancer Biology, University of Texas MD Anderson Cancer Center, Houston, TX, United States, ³Moiani Research Inc., Missouri City, TX, United States, ⁴Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA, United States

While many researchers can design knockdown and knockout methodologies to remove a gene product, this is mainly untrue for new chemical inhibitor designs that empower multifunctional DNA Damage Response (DDR) networks. Here, we present a robust Goldilocks (GL) computational discovery protocol to efficiently innovate inhibitor tools and preclinical drug candidates for cellular and structural biologists without requiring extensive virtual screen (VS) and chemical synthesis expertise. By computationally targeting DDR replication and repair proteins, we exemplify the identification of DDR target sites and compounds to probe cancer biology. Our GL pipeline integrates experimental and predicted structures to efficiently discover leads, allowing early-structure and early-testing (ESET) experiments by many laboratories. By employing an efficient VS protocol to examine protein-protein interfaces (PPIs) and allosteric interactions, we identify ligand binding sites beyond active sites, leveraging *in silico* advances for molecular docking and modeling to screen PPIs and multiple targets. A diverse 3,174 compound ESET library combines Diamond Light Source DSI-poised, Protein Data Bank fragments, and FDA-approved drugs to span relevant chemotypes and facilitate downstream hit evaluation efficiency for academic laboratories. Two VS per library and multiple ranked ligand binding poses enable target testing for several DDR targets. This GL library and protocol can thus strategically probe multiple DDR network targets and identify readily available compounds for early structural and activity testing to overcome bottlenecks that can limit timely breakthrough drug discoveries. By testing accessible compounds to dissect multi-functional DDRs and suggesting inhibitor mechanisms from initial docking, the GL approach may enable more groups to help accelerate discovery, suggest new sites and compounds for challenging targets including emerging biothreats and advance cancer biology for future precision medicine clinical trials.

KEYWORDS

cancer, DNA repair, DNA damage response, DNA replication, precision oncology, computational pipeline, computational docking, *in silico* testing

1 Introduction

Cost and efficiency bottlenecks can limit breakthrough drug discoveries and divert efforts toward repurposing or improving existing cancer drugs (Weth et al., 2024). Yet, cancer heterogeneity and fast mutation rates make therapeutic resistance a major medical challenge, which compels oncology researchers to seek innovative solutions for new precision chemotherapeutics. As part of these efforts, the DNA Damage Response (DDR) (Pilie et al., 2019; Voutsadakis and Digkila, 2023) involving DNA repair pathways (Voutsadakis and Digkila, 2023; Wang et al., 2021) combined with their interface to DNA replication and transcription (Ye et al., 2024) have emerged as important targets for precision cancer therapy. However, these dynamic networks have numerous different potential protein targets (Brownlie et al., 2023; Choi et al., 2022; Groelly et al., 2023; da Costa et al., 2023; Helleday and Rudd, 2022; Hopkins et al., 2022; Saxena et al., 2024). So, enabling a protocol to successfully evaluate multiple targets and inhibitors may advance actionable mechanistic knowledge of cancer biology and help overcome resistance.

Furthermore, advances in large-scale production of recombinant proteins and their complexes mean that sample availability is often no longer rate-limiting (Gradia et al., 2017). These observations suggest that efficient assessment capable of examining multiple targets with accessible pre-clinical inhibitors that can be structurally tested and validated in cancer cells and models may be a superior strategy to choosing and focusing on a single candidate. Thus, we reason that providing more laboratories with robust, practical computational methods to initiate early structural and activity testing for new chemical inhibitor design may empower multifunctional DDR networks where inhibitors may show different phenotypes than knockout or knockdown strategies that completely remove a multifunction protein component.

To efficiently identify novel inhibitor tools and target sites for early structure and activity testing to help meet cancer resistance challenges, here we present the development and testing for an *in silico* Goldilocks' (GL) protocol. Compared to canonical methods that rely on extensive HT screening of large compound libraries, GL protocol leverages a compact curated virtual screening (VS) compound library and a straightforward docking protocol with Glide as exemplary accessible docking software. GL novelty includes efficient ESET library use, combining Diamond Light Source DSI-poised, Protein Data Bank fragments, and FDA-approved drugs. ESET compounds were chosen to span relevant chemotypes and be suitable for both structure determinations and cell testing. This GL approach may provide a compact library and a focused and manageable screening process, facilitating early structure and cell testing without requiring extensive synthetic chemistry resources.

GL methodology can harness advanced *in silico* techniques for molecular docking and molecular modeling to explore protein-protein interfaces (PPIs) (Lu et al., 2020; Scott et al., 2016; Tsuihiji et al., 2022) including allosteric interactions and active sites (Nussinov and Tsai, 2013). By looking beyond targeting these protein regions and transition state analogs, researchers can likely explore both larger target sites and greater opportunities to reduce cross-reactivity (Moiani et al., 2018). Thus, the GL protocol aims to efficiently and computationally examine

potential inhibitor binding sites and identify ligands suitable for experimental structural and biological testing to ultimately enable the modulation of protein functions within DDR pathways in predetermined ways.

The GL protocol presented here aims to offer three general potential advantages: 1) reduce the need for an extensive experimental screening and medicinal chemistry effort for preclinical target testing, 2) efficiently obtain VS hits suitable for focused experimental tests, and 3) going wider on interactome targets rather than deeper to harness the power of defined genome networks and the many PDB and AF3 structures. The presented GL protocol seamlessly incorporates improving technologies for accurate structure prediction of fold and biomolecular interactions (Abramson et al., 2024) and experimental results. GL results should aid efficient applications of experimental structural advances in X-ray scattering (Rosenberg et al., 2022), Small-Angle Neutron Scattering (SANS) (Adhikari et al., 2023), mass spectrometry (Filandr et al., 2024; Syed et al., 2023), crystallographic fragment screening (Wilson et al., 2021), and cryo-electron microscopy (Duan et al., 2024; Cebi et al., 2024), which can enable experimental measures of flexibility and ligand-binding site interactions.

Despite its smaller size, GL results suggest that the ESET library can identify high-quality hits with strong binding affinities and favorable binding modes. There are certainly limitations for targeting PPIs by FDA-approved compounds. Yet, we found that FDA-approved novobiocin can function as a non-competitive allosteric inhibitor for PolQ helicase, which is now in cancer clinical trials (Syed et al., 2023; Zhou et al., 2021). These and other findings support the inclusion of FDA-approved compounds in our library, as we suggest that "off-target" effects are more common than widely recognized and may uncover valuable therapeutic interactions. Our findings indicate that the ESET library achieves effective enrichment rates, underscoring its efficiency and practicality. Additionally, the ESET library yields a manageable number of false positives that our docking and scoring protocols can efficiently filter out to improve hit quality and facilitate subsequent experimental validations. By providing multiple hits and an accessible chemical library, GL enables early structural and cell testing to offer a viable alternative for researchers with limited resources, broadening access to effective drug discovery methodologies. Notably, the streamlined computational approach and compact library can accommodate multiple individual candidate proteins to infer potential impacts on functional pathways and networks.

Computationally evaluating multiple targets, allostery, and protein-protein Interfaces (PPIs) may help identify optimal targets to control pathway selection, reduce resistance, and encompass potential metabolic-epigenetic axes (Shibata et al., 2014). Specifically, we tested the GL protocol to examine emerging cancer targets within the DDR (O'Connor, 2015), including its interface with the immune response (Lee et al., 2024; Ye et al., 2024; Yuan et al., 2022; Ye et al., 2021a; Hou et al., 2020; Brosey et al., 2017), and replication (Syed and Tainer, 2018; Ye et al., 2024; Longo et al., 2023). These results suggest target sites and chemotypes for tool inhibitors and potential precision cancer drugs to control phenotypes and create synthetic lethality in cancer cells (Setton et al., 2021). In fact, this GL protocol generally empowers rapid initial

inhibitor identification making it suitable for timely coordinated responses to address emerging pathogens and biothreats as well as cancer. The GL pipeline thus aims to identify chemotypes that can bind to selected sites, laying the groundwork for efficient experimental validation and optimization with structural biophysics and cytology. Ultimately, the GL protocol, coupled with subsequent experimental validation cycles, aspires to empower the design of mechanism-based interventions by many oncology and cancer biology researchers to accelerate the identification of novel agents for lower toxicity and higher precision cancer therapy.

2 Materials and methods

This section describes the GL approach, encompassing virtual chemical libraries, target preparation, molecular docking, energy evaluation, molecule selection, binding mode analysis, and chemical analysis. Systematic GL methodology provides the computational platform to identify compounds for efficient biophysical and cytological testing to discover and develop chemical tools and potential cancer therapeutics targeting DDR and Repair pathways.

2.1 Virtual chemical libraries

Description and Preparation: A compact virtual chemical library comprising diverse small molecules was compiled to span chemotypes and structural diversity and be suitable for biophysical and cytological testing. GL protocol employs three chemical libraries and the Ligprep utility of the Schrödinger package to generate the virtual assembly for virtual screening (Sastry et al., 2013; Schrodinger, 2017; Lu et al., 2021). We term the combined library as the Early Structure-Early Testing (ESET) library, after the Egyptian goddess of healing.

Library I is the DSI-poised developed at Diamond Light Source, United Kingdom (Diamond) and Structural Genomic Consortium, United Kingdom (SGC) in the frame of one of the iNEXT Joint Research Activities constituted by 860 fragments described in the original paper (Cox et al., 2016). Library II contains 203 PDB ligands extracted and filtered from the Protein Data Bank (Revillo Imbernon et al., 2023). Library III is the LH1 (LH1 is the commercial code used by TargetMol for this library selection) FDA-approved drug from the AD part of the Tainer group GLAD library (Moiani et al., 2021). It comprises 2,111 approved FDA drugs, natural products, and chemicals in pre-clinical and clinical studies, as updated in 2020 (Zhu et al., 2021). All three libraries were independently prepared for computational studies using the Moiani et al., 2021 protocol (Moiani et al., 2021).

A novel aspect of our GL protocol comes from the Early Structure-Early Testing (ESET) library, a compact virtual chemical library curated to span chemotypes and structural diversity suitable for biophysical and cytological testing without requiring extensive high-throughput screening resources. The ESET library's design facilitates the early identification of high-potential compounds, enabling structure and cell testing at an early stage. This approach contrasts with traditional methods that require large libraries and extensive synthetic chemistry, making it accessible to researchers with limited resources.

2.2 Target selection and preparation for *in silico* studies

Target Identification: The concept of synthetic lethality comes from functional defects within the DNA Damage Response (DDR) and Repair pathways that are lethal in combination with cancer cell defects. This notion suggests targeting functional defects in pathways or networks rather than focusing on a single protein. Thus, multiple DDR proteins were selected as the focus of this study, including Poly (ADP-ribose) polymerase (PARP) (Zandarashvili et al., 2020), Poly (ADP-ribose) glycohydrolase (PARG) (Houl et al., 2019), Base Excision Repair (BER) proteins (Bacolla et al., 2021; Tsutakawa et al., 2013; Tubbs et al., 2009; Hitomi et al., 2007), and Nucleotide Excision Repair (NER) proteins (Bralic et al., 2023; Yu et al., 2023; Tsutakawa et al., 2020a; Fan et al., 2008), Double-Strand Break Repair (DSBR) proteins (e.g., MRE11-Rad50-Nbs1 complex) (Williams et al., 2008; Rotheneder et al., 2023), and DNA polymerase theta (Pol θ) (Zhou et al., 2021; Syed et al., 2023).

Target Preparation: The three-dimensional structures of selected targets were obtained from experimental databases. Targets without available experimental structures can use homology modeling by AlphaFold 2 or AlphaFold 3. These structures were prepared by removing water molecules and adding hydrogen atoms to ensure proper protonation states. In selected docked complexes, subsequent examination of bound water, atomic hydrophilicity, and protein surface topography may guide chemical optimizations as the addition or removal of a water molecule can inhibit reactions (Kuhn et al., 1992).

2.3 PDB structure selection and docking grid generation

For *in silico* analysis, we selected an exemplary series of available protein database (PDB) structures from the Protein structure database (<https://www.rcsb.org/>), mainly obtained by X-ray, with some using cryo-EM. To prepare for VS, we optimized target structures *in silico* by selecting high-resolution structures from experimental databases and ensuring they were suitable for computational analysis through the Protein Preparation Workflow in the Schrödinger suite (Schrödinger) (Moiani et al., 2021; Moiani et al., 2018).

We selected at least one relevant structure for a protein or complex involved in a DDR pathway of interest. We generated Grid for Docking runs aimed at protein-protein interaction domains, allosteric domains, active sites, and protein-nucleic acid interaction domains.

For PARP1, we chose PDB 7ONR, resolution 2.05 Å (Johannes et al., 2021); for PARG, we used the series of three PDBs presented in the previous methodology development (Moiani et al., 2021): 4B1H A, 4B1H B, resolution 2 Å, 6OAK, resolution 1.7 Å (Houl et al., 2019).

For the base-excision repair (BER) pathway, we selected UDG first for its relevance, as its activity is required to initiate the base-excision repair pathway and remove uracil from DNA (Parikh et al., 2000a; Putnam et al., 1999; Mol et al., 1995b), and second, because we have been studying its structural biology for two decades

TABLE 1 List of VS Targets and Grids. All the protein targets (left column) are listed with structure PDB ID (center column), and domain or amino acid are for VS grid (right column).

Target	PDB	ID GRID
PARP1	7ONR	H862
PARG	4B1H A	Y795, F875, F902
PARG	4B1H B	Y795, F875, F902
PARG	6OAK	Y795, F875, F902
UDG	6VBA	Q144
XPD	6RO4	S506, T540
MRE11	8BAH	E92
GRB2	1GRI	R207
POLQ	5AGA	F422, V147, S148
POLQ	8E24	I2362, C2386

(Nguyen et al., 2021; Slupphaug et al., 1996; Mol et al., 1995a; Parikh et al., 2000b). The PDB selected for the *in silico* study is 6VBA, with a resolution of 1.8 Å (Nguyen et al., 2021).

For the nucleotide excision repair (NER) pathway, we selected XPD as a pivotal helicase and the lynchpin of the NER complex (Tsutakawa et al., 2020b) that we evaluated previously (Fan et al., 2008; Mui et al., 2011; Tsutakawa et al., 2021; Bralic et al., 2023; Yu et al., 2023). The PDB selected for the *in silico* study is 6RO4 solved at 3.5 Å resolution by cryo-EM (Kokic et al., 2019).

For the DNA double-strand break repair (DSBR) pathway, we evaluated two targets in the signal transduction cascade after our recent publication (Ye et al., 2021b), specifically the interface of Mre11-Nbs1 for the MRN complex and the GRB2 adaptor plastic dimerization site. The region selected for targeting the MRE11 dimer was at the NBS1 contact with the selected PDB code 8BAH solved at 4.13 Å resolution by cryo-EM (Rotheneder et al., 2023). The GRB2 structure selected was the dimeric structure solved at 3.1 Å resolution in PDB code 1GRI (Maignan et al., 1995).

For the cancer-relevant alternative end joining (AtI-EJ) break repair pathway, we selected POLQ (DNA Polymerase Theta), which is a cancer target undergoing clinical trials (Syed et al., 2023; Zhou et al., 2021). For this target, we organized the *in silico* study for two different protein sites: the putative contact of Novobiocin (Hambo Guo, 2023), where Nucleotide binds during polymerase activity, and an allosteric site of the protein. We used two structures: PDB ID 5AGA for the helicase domain at 2.9 Å resolution and PDB ID 8E24 for the catalytic domain at 2.34 Å resolution (Newman et al., 2015; Bubenik et al., 2022).

We generated a docking grid for each target structure with the generation function of the receptor grid in the Schrödinger suite. The grid box was sized to ensure coverage of the entire binding site, with dimensions typically set to 20 × 20 × 20 Å, centered on the binding pocket (Moiani et al., 2021). Table 1 shows the residues and position of the Grid generated for each structure.

2.4 Two VS approach and molecular docking parameters

Docking Protocols: Molecular docking studies were conducted using Glide, which is included in the Schrödinger drug discovery suite and employs well-established protocols and parameters (Yang et al., 2021). For each target, both XP (extra precision) and SP (standard precision) docking runs were performed.

Our GL approach combines two virtual screening runs per the target's Grid. We repeated our approach for each chemical library (see Results). The first VS run in the Virtual Screening workflow section of Glide, in the Docking section, enables the first three parameters: *Enhance planarity of conjugated pi groups*, *Use Epik state penalties for docking write interaction scores for residues within 12 Å of the grid center*. The scaling factor is 0.8, and the partial charge cutoff is 0.15. At the same time, we enabled Glide HTVS, Glide SP (option generates up to three compounds per state), and Glide XP; we also enabled postprocessing with Prime MM-GBSA. The second VS differed from the previous one because we disabled the Dock with Glide HTVS and Dock with Glide XP capabilities in the same Docking section. The combination of two virtual screening runs per chemical library, and target's Grid gave us informative analyses for the site evaluated for ligand binding. The first VS (VS1) docking provided a single top-ranked binding pose, while the second VS (VS2) generated multiple ligand binding poses. VS2 was used to refine the initial and only hit from VS1, providing a comprehensive view of potential binding modes. Alternative methods like MD simulations (see protocol in (Moiani et al., 2009), PDBs of VS1 complexes for exploring MD simulations are available in supplementary info) and binding pocket analysis tools can complement this approach. A cutoff of 10% was applied for the second VS results to select the most promising binding poses for further energy ranking analysis.

Energy Evaluation: The binding energies of the selected results of the second VS binding poses were evaluated using appropriate scoring functions and MM-GBSA energetic analysis to assess the strength of ligand-protein interactions, as presented in Moiani et al. (2021).

2.5 Ranking top molecules for structural biology techniques

Based on the docking results and energy evaluations, the top-performing molecules, which imply reasonable binding affinity and favorable binding modes, are proposed for future investigation by structural biology techniques and *in vitro* binding assays. Selecting favorable binding modes involved assessing the docking scores, binding affinities, and interaction profiles. Molecules with strong binding affinities and favorable interaction profiles could be prioritized for further investigation. These criteria were used to imply selected compounds with the potential for effective binding and inhibition of the target proteins. The calibration process involved re-ranking based on docking scores and MM-GBSA binding energies. The MM-GBSA calculation is an energy evaluation of complexes at 0 nanoseconds of MD simulation; such energy profile can be extended during validation with MD simulations and calculated at each time to monitor *in silico* binding affinity. The discussion paragraph explains the reasons for the top group and their chemo-diversity. The optimal result from

the VS1 was combined with a selection of results from the VS2. Future studies should involve structural biology to reveal density or putative ligand binding and *in vitro* binding assays to confirm binding affinities in the lower μM range, followed by functional assays to test the activity of top-ranked ligands.

2.6 Graphical representation combines VS1-VS2 results

The VS1 ranks ligands and gives a single result, representing the absolute minimum in the binding landscape. The addition of the ten percent cutoff of results of VS2 as density reveals the local minimum around the absolute and the possible binding landscape of the same ligand in different conformations or other ligands, effectively probing the site. As shown in the Results, the overlap of all these poses with our graphical analyses outlines the binding surface we employ to optimize with further medicinal chemistry efforts. Allosteric interaction domains and transient protein-protein interfaces during the catalytic mechanism of studied enzymes and complex enzymes can include small or large regions, so this graphical analysis can help define the size and type of chemical tools to be designed and improved.

2.7 Binding mode evaluation

Binding modes were scrutinized to assess their potential to interfere with Protein-Protein Interaction (PPI) surfaces and allosteric regions within target proteins to suggest ligands that may disrupt complexes and conformations to regulate protein function. For proteins in macromolecular complexes, the best results aim to define the interaction of the ligand binding mode with proper complex assembly and functional conformational changes during catalytic mechanisms. DNA repair complexes are typically plastic with flexible domains and modular interfaces that can show vulnerabilities during conformational changes (Chinnam et al., 2024a). Binding mode evaluation involved analyzing protein-ligand interactions, conformational changes, and functional impacts using structural and biochemical data. This analysis of the population of specific chemotypes binding specific interfaces can help elucidate mechanisms or other allosteric interactions unknown before.

To validate the protocol, we assessed the enrichment of active compounds during docking, as shown in the PARG and PARP1 examples. These results demonstrate that the docking and ESET library achieves adequate enrichment of hits with a manageable number of false positives. This observation supports using the ESET library for efficient and practical drug discovery.

Our approach aims to detect potential ligand binding modes that interfere with functional mechanisms. The same overlap can be made with nucleic substrates (DNA, RNA, or DNA/RNA hybrid) or other substrates involved in the catalytic functions of our protein targets.

2.8 Chemical analysis to suggest chemical optimizations

To suggest further optimization efforts, selected molecules' chemical properties and interactions were analyzed for hydrogen

bonding, hydrophobic interactions, and electrostatic interactions. 2D interactive maps were plotted for the single result of VS1 for each target and library.

2.9 Focus library generation guidance

Based on the chemical analysis and binding modes results, the GL approach can suggest design-focused libraries of compounds for each target for commercial or medicinal chemistry collaborators to probe and optimize ligand design and chemical optimization for subsequent studies.

3 Results

The GL computational methodology flow chart (Figure 1) outlines a practical VS protocol that may be assessable to many groups. We illustrate the application of the GL protocol by presenting results for each target in graphic representation, showing the overlap of VS1 and VS2 for each library. The complete chemical characterization of ligands and their ranking are fully described in the Supplementary information for each protein target. The VS1 result provides a ranking of ligands and includes 2D interactive maps. Employing two VS allows users to test if the VS1 single result shows a unique chemotype among targets or if we may have a screening artifact or promiscuous binder to consider as a false positive. VS2 results probing the site are shown as a surface, including the VS1 result (the top binder). Details of each molecule, ranking, and *in silico* affinity are tabulated in Supplementary information.

3.1 Targeting PARP1 and PARG PARylation and dePARylation

PARP1 has been successfully targeted with four approved drugs, and new-generation drugs are in clinical trials. Our *in silico* results re-explore chemotypes for binding to the active site area (Figure 2). Specifically, we start with the 2D interactive map of the VS1 unique binder, analyzing the type of chemo-group and interaction with specific residue around the aimed grid. On the right side, the surface generated by the multiple VS2 results, which includes the VS1 result, shows where the minimum energy binder can grow the unique binder in multiple directions. Such a view has been repeated for the results of the three ESET component libraries (see the supplementary information for details on VS2 targeting PARP1).

In the human genome, the Glycohydrolase PARG is the only homolog of its class for dePARylation, and its singularity makes it attractive for specificity. PARG is an emerging target with a few compounds in early-stage clinical trials, and compounds form our group in preclinical testing (Brosey et al., 2021; Houli et al., 2019; Moiani et al., 2021). Here, we demonstrate our methodology for improving chemical scaffolds for toxicity or other aspects, although toxicity testing was not performed in this study. In Figure 3 (panels A, B, and C), we present respectively the results for PARG structure 4B1H_A (panel A), 4B1H_B (panel B), and 6OAK (panel C). In

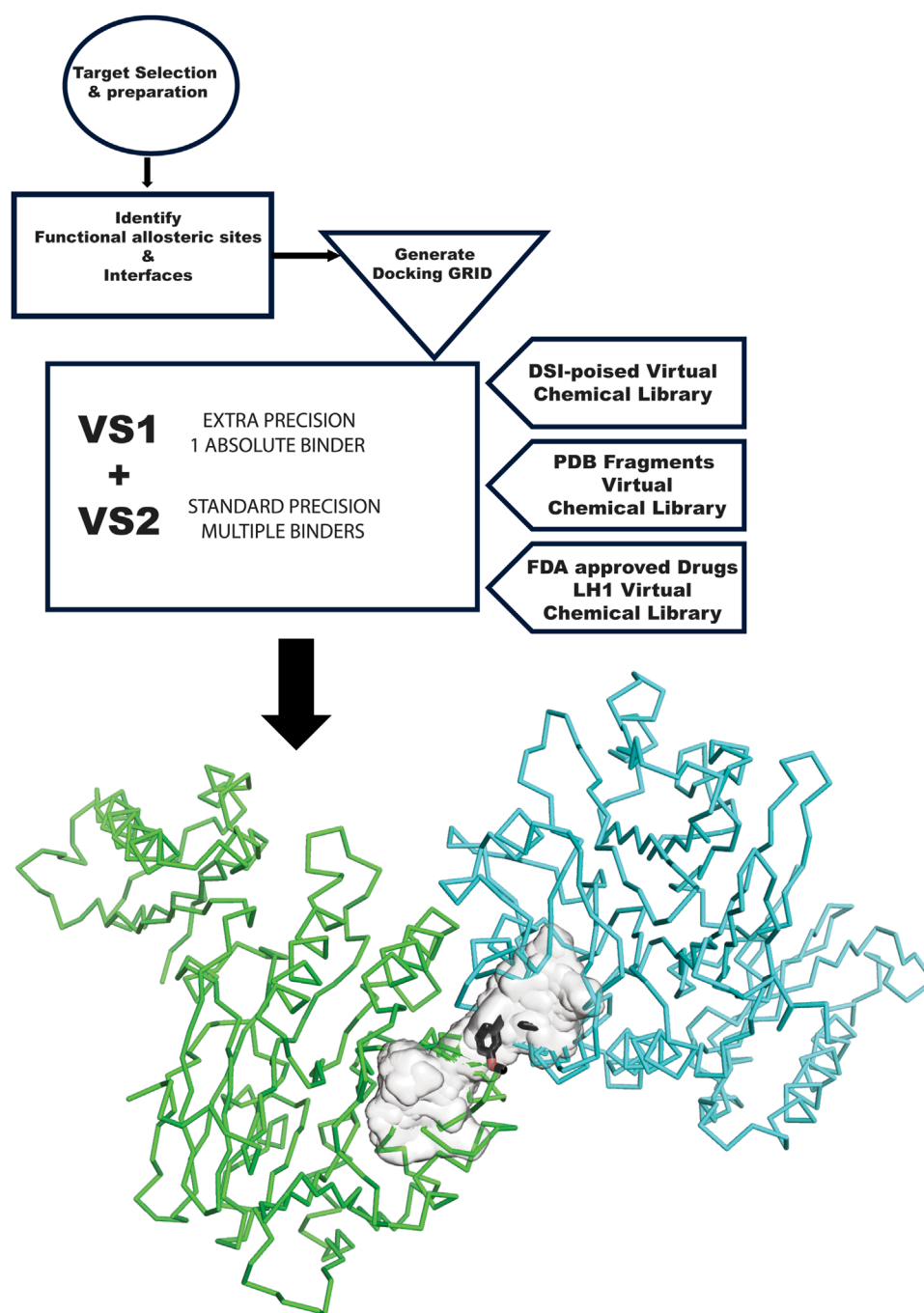


FIGURE 1

GL VS flow chart. The flow chart starts at the top left, with the target selection and appropriate structure. The structure needs to be prepared for *in silico* experiments. The next step is analyzing the target and the surface of contact with other proteins involved in its complexes and another allosteric site relevant to the mechanism. These sites become the center of the GRID generation. VS1 and VS2 are run with the three libraries used in this study: DSI Poised, PDB Fragments, and FDA-approved drugs LH1. Results are then graphically plotted and evaluated.

this case, we used three structures for the same target, focusing on a conformational change of an aromatic residue in the catalytic domain and the effect on the *in silico* results (the reasons were explained in (Moiani et al., 2021)). The three panels show results for the three different libraries (A) DSI-poised (B) PDB fragments, (C) LH1 FDA-approved drugs. Each panel has three rows: 1)

representing the result for PARG structure 4B1H_A, 2) for PARG structure 4B1H_B, and 3) for PARG structure 6OAK. Collectively, these unveil the complexity of examining allosteric and transient conformational change and how to feed the *in silico* screening. Considering the same protein target, the same region of GRID, and just changing the orientation of one amino acid, which is crucial for

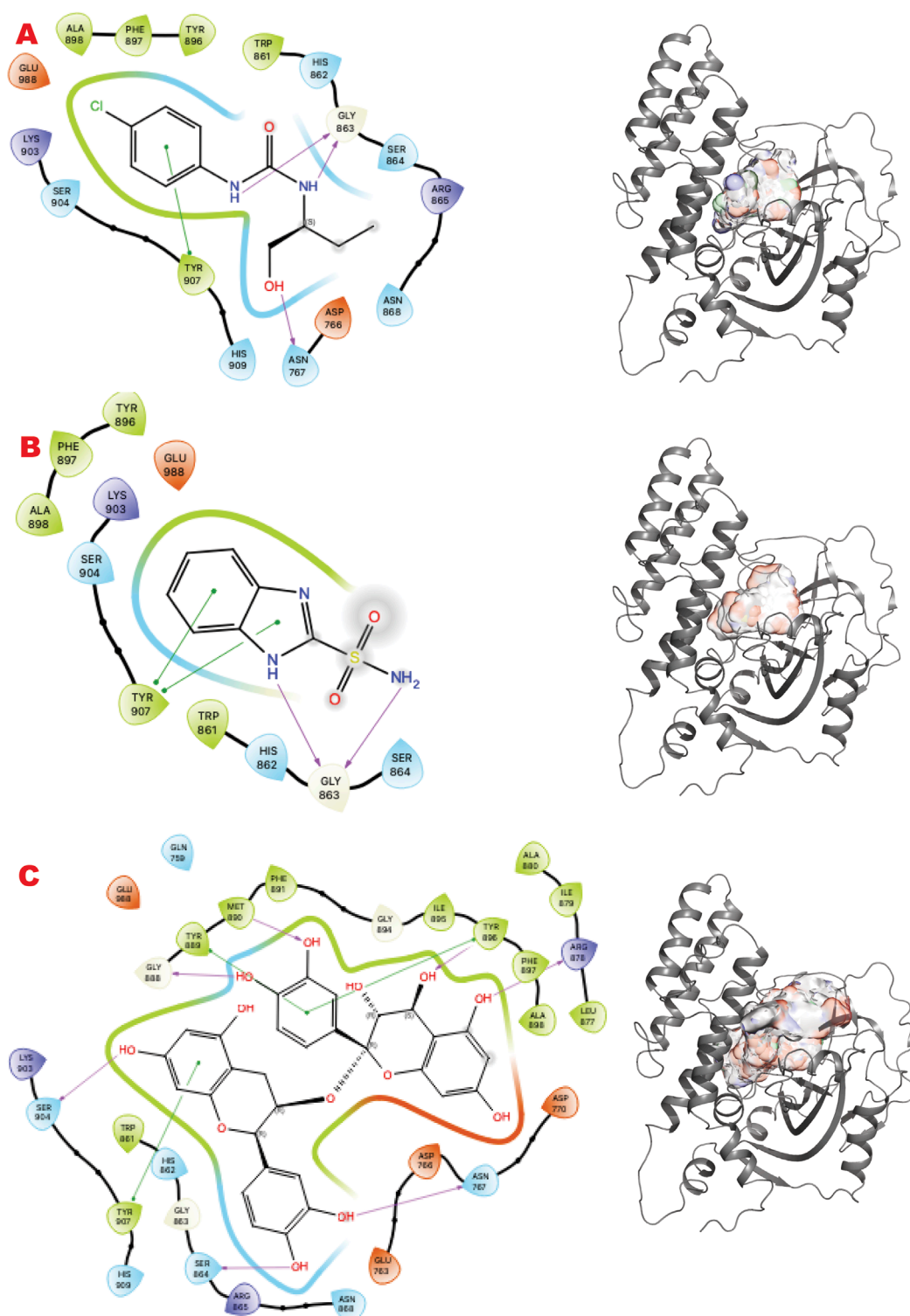


FIGURE 2

PARP1 VS targeting results. (A) VS1 result, from DSI-poised library, from left 2D ligand interactive map of VS1 result, compound 1-(4-chlorophenyl)-3-(1-hydroxybutan-2-yl)urea, right VS2+VS1 results electrostatic maps from DSI-poised library targeting PARP1 catalytic domain; (B) VS1 result, from PDB Fragment library targeting PARP1 catalytic domain; (C) VS1 result from FDA approved drugs library, left 2D ligand interactive map of VS1 result, compound Proanthocyanidins, right VS2+VS1 results electrostatic maps from FDA approved drugs library targeting PARP1 catalytic domain.

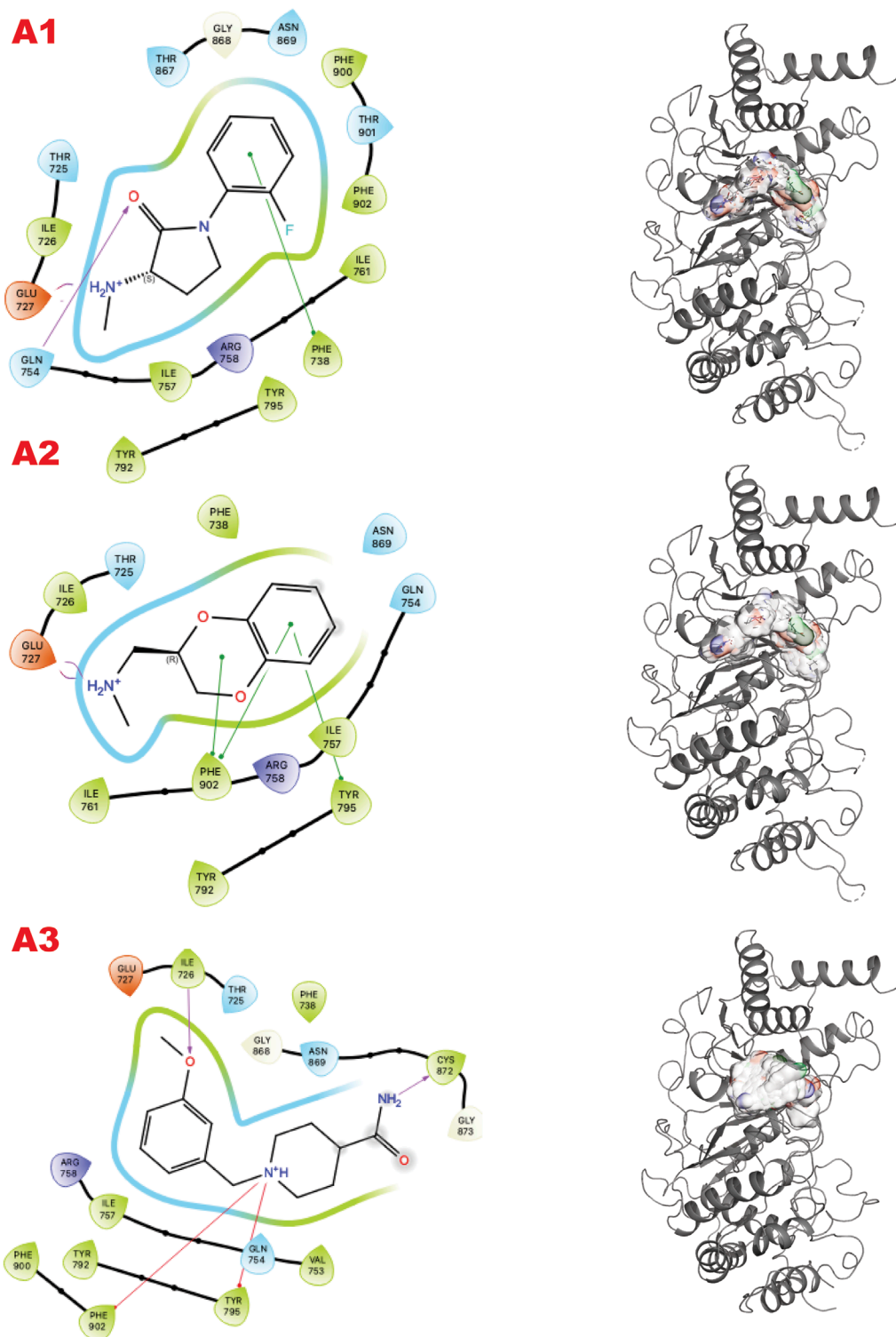


FIGURE 3
(Continued).

the mechanism, the results of the three panels are different for the same libraries and protocols.

We analyzed the enrichment of active compounds for PARP1 and PARG, demonstrating the protocol's

effectiveness in identifying high-potential hits. Specifically, our findings highlight the manageable false positive rates and the quality of hits suitable for further experimental validation.

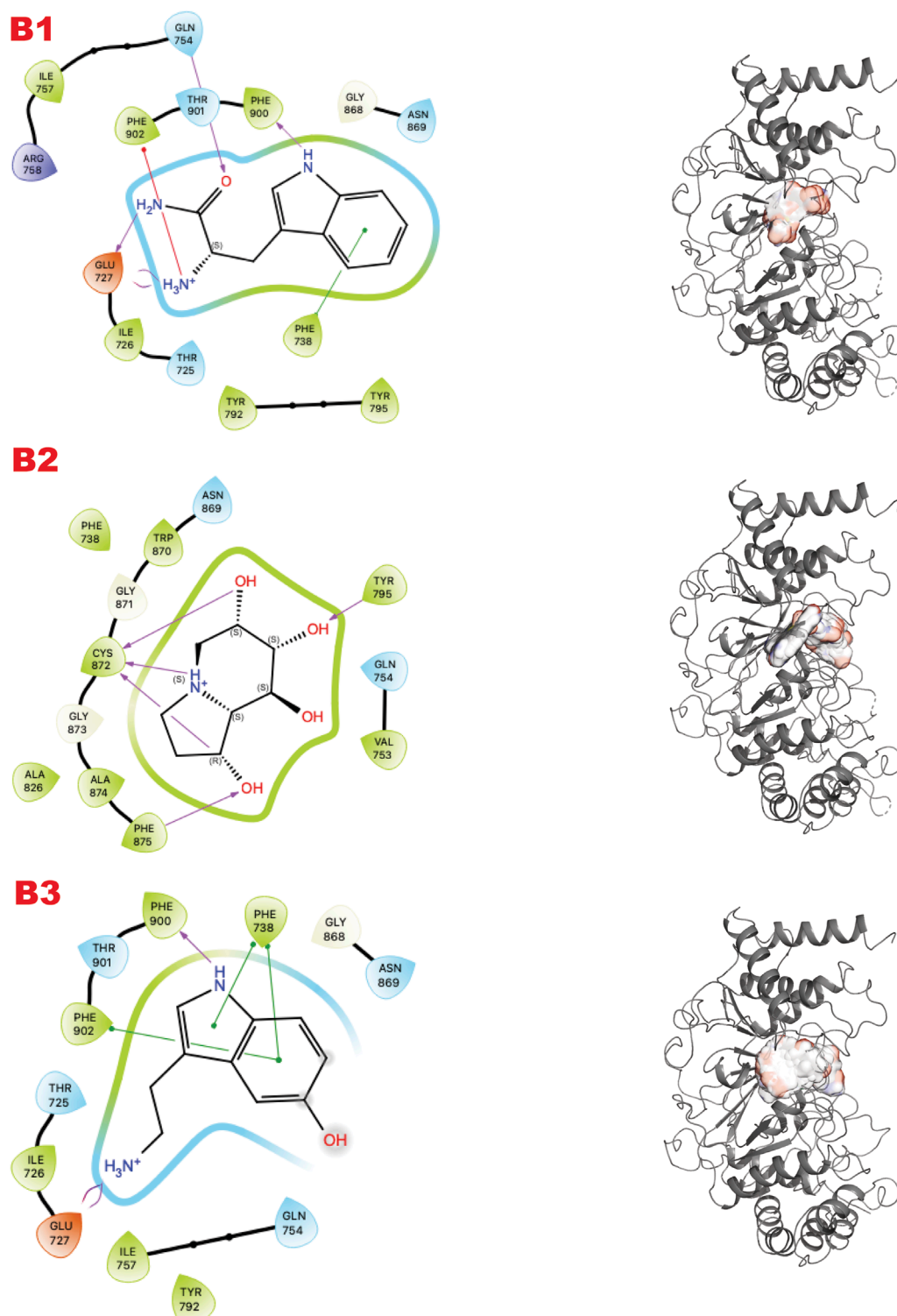


FIGURE 3
(Continued).

3.1.1 Analysis of enrichment of active compounds for PARP1 and PARG

We analyzed the PDB fragment results for all three PARG target structures and PARP1 to validate the enrichment. We observed

that results of PARG_4B1H_B, PARG_6OAK, and PARP1 showed at least thirty percent or more of the results are redundant with the compound octahydroindolizine-1,6,7,8-tetrol (except for PARG_4B1H_A where this compound has not been found due

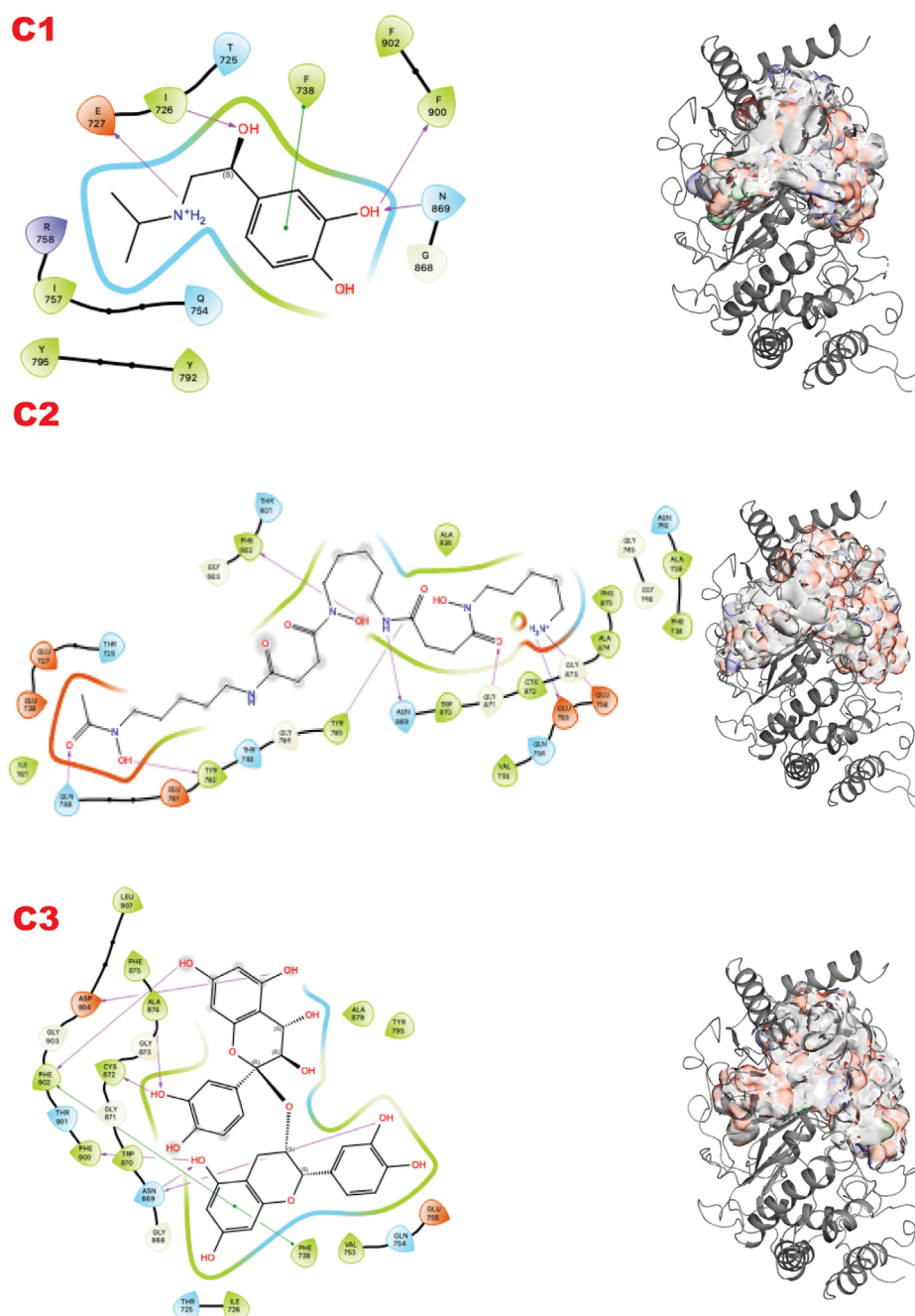


FIGURE 3

(Continued). PARG VS targeting results for two structures: 4B1H_A, 4B1H_B, 6OAK. Panel A: DSI-poised library results, Panel B: PDB: PDB Fragment library results, Panel C: LH1 FDA-approved drugs library results. (A1) VS1 result, from DSI-poised library, from left 2D ligand interactive map of VS1 results, compound: 1-(2-fluorophenyl)-3-(methylamino)pyrrolidin-2-one hydrochloride; right VS2+VS1 results in electrostatic maps targeting PARG catalytic 4B1H_A domain; (A2) VS1 result, from DSI-poised library, from left 2D ligand interactive map of VS1 results, compound: [(2,3-dihydro-1,4-benzodioxin-2-yl)methyl](methyl)amine; right VS2+VS1 results in electrostatic maps targeting PARG catalytic 4B1H_B domain; (A3) VS1 result, from DSI-poised library, from left 2D ligand interactive map of VS1 results, compound: 1-[(3-methoxyphenyl)methyl]piperidine-4-carboxamide; right VS2+VS1 results in electrostatic maps targeting PARG catalytic 6OAK domain; (B1) VS1 result, from PDB-fragments library, from left 2D ligand interactive map of VS1 results, compound: L-tryptophanamide-2; right VS2+VS1 results in electrostatic maps targeting PARG catalytic 4B1H_A domain; (B2) VS1 result, from PDB-fragments library, from left 2D ligand interactive map of VS1 results, compound: indolizidine-1,6,7,8-tetrol; right VS2+VS1 results in electrostatic maps targeting PARG catalytic 4B1H_B domain; (B3) VS1 result, from PDB-fragments library, from left 2D ligand interactive map of VS1 results, compound: Serotonin; right VS2+VS1 results in electrostatic maps targeting PARG catalytic 6OAK domain; (C1) VS1 result, from LH1 FDA-approved drugs library, from left 2D ligand interactive map of VS1 results, compound: Isoprenaline hydrochloride; right VS2 results in electrostatic maps targeting PARG catalytic 4B1H_A domain; (C2) VS1 result, from LH1 FDA-approved drugs library, left 2D ligand interactive map of VS1 results, compound: Deferoxamine Mesylate; right VS2+VS1 results in electrostatic maps targeting PARG catalytic 4B1H_B domain; (C3) VS1 result, from LH1 FDA-approved drugs library, from left 2D ligand interactive map of VS1 results, compound: Proanthocyanidins; right VS2+VS1 results in electrostatic maps targeting PARG catalytic 6OAK domain.

to a conformational block in the binding site). We identify such compound as false a positive for a molecular hit, though we consider his chemotype as a hint in the probing mode. In fact, this similarity to the sugar moiety helps analyze binding modes for putative rational substrates with sugar moieties that are part of the polynucleotide. Analyzing the top specific hit for PARG and PARP1, we found two clear families of rational binders: tryptophan-related compounds (Melatonin and Serotonin) and tyrosine-related compounds (Tyrosinamide and Resveratrol). Those compounds are related to specific neurotransmitters, Serotonin, and Dopamine, which are involved in the Circadian clock (Korshunov et al., 2017; Daut and Fonken, 2019; Morin, 1999), as well as the PARP1 (Yang et al., 2023). The rational enrichment of specific metabolites in the results from 200 down to 5 helps to probe the site, elucidate possible binders, and exclude others (Grocock et al., 2014).

3.2 UDG BER initiator

Targeting UDG is relevant to controlling the BER pathway in coordination with the APE1 endonuclease and is crucial in immunology and cancer immunotherapy (Mol et al., 2000; Parikh et al., 2000a; Slupphaug et al., 1996). Our group defined inhibition mediated by promiscuous ligand, ATA (Nguyen et al., 2021), and its implications. Here, we revisit the UDG site where the uracil binds and releases apurinic bases containing DNA to understand and find better and more specific chemical binders.

Figure 4 shows the results of the GL *in silico* methods applied in three rows. In the same order of representation for VS1 2D, and VS2 surfaces that include VS1 absolute binder, a clear binding mode in the uracil recognizing pocket emerges for all three libraries: a polar fragment with no specific orientation. This explains the previously characterized ATA, a polar ligand with three equal units that bind without specificity. In this case, directionality could be given by mimicking the ssDNA or dsDNA binding mode of the substrate of the UDG. Such observations may help to further design binders for more specificity and better functionality in targeting the enzyme's role in initiating the BER pathway.

3.3 Targeting XPD as a pivotal NER enzyme

XPD and its helicase activity are essential in the initial damage verification step of the TFIIH complex (Bralic et al., 2023). Targeting the narrowest point where DNA binds first may suppress TFIIH activity and the NER pathway. Figure 5 shows the results of the *in silico* methods applied in three rows. Specifically, the first row represents the results of the DSI Poised library targeting XPD helicase: (left) the result of VS1 shown using 2D ligand interaction map, (right) VS2 results that include VS1 absolute binder, and the sum of the electrostatic surface of all ligands evaluated represents such results. In the same order, the second row shows the results of the PDB fragments library targeting the XPD helicase domain. In the same order, the third row shows the LH1 FDA-approved drug library results targeting the XPD helicase domain. The site targeted is involved in DNA binding.

Interestingly, the VS1 binder in the first row mimics a tyrosine and a phosphotyrosine and may also reveal a phosphotyrosyl DNA-adduct binding site. Learning from top binders' chemotype can suggest future ligand customization for better preclinical studies.

3.4 Targeting MRE11 dimer and NBS1 interfaces to control DSB repair

MRE11, part of the MRE11-RAD50-NBS1 (MRN) complex, participates in multiple DNA Repair pathways (Syed and Tainer, 2018). Here, we sought to target the MRE11-NBS1 interface that regulates the symmetric/asymmetric dimeric conformation of MRE11-RAD50, which is crucial for mechanism and pathway selection (Shibata et al., 2014; Moiani et al., 2018).

Figure 6 shows the results of the *in silico* methods applied in three rows. Specifically, from the VS1 results of the three libraries, the chemistry that should outcompete the NBS1 binding has some characteristics already explored by previous ligands from the PFM analogs. The DSI-poised top binder is a chemotype that combines an aromatic cycle and a non-aromatic heterocycle connected by carbon, which gives flexibility to the two units. The complex FDA drug that is indeed a cyclopeptide has a termination as ILE, as has been explored in the MRE11 endonuclease inhibitors. All observed dockings are at the interface of the MRE11 dimer and may impact symmetric or asymmetric dimerization.

3.5 Stabilizing inactive GRB2 adaptor to control DNA break repair

GRB2 is a plastic adapter of signal transduction that is active as a monomer and inactive as a dimer (Ye et al., 2021b; Ye et al., 2024; Ahmed et al., 2015). Although GRB2 has been studied for decades, attempts to find small molecules to control its activity have yet to reach clinical evaluation. Its participation in signaling in the DDR pathway connected to the MRN complex recently suggested new ways to modulate its functions and activities. Here, our results identify sites and chemotypes for potential modulators.

Figure 7 shows the results of the GL *in silico* methods applied in three rows. Specifically, the cavity created internally by such dimerization only allows a specific chemical and sterical configuration, as the VS1 ligands show. A combination of multi-fragment chemicals, such as three aromatic connected in the results of the FDA-approved drug or an aromatic bridged to a nonaromatic ring by a urea moiety, can usually be nucleotides or polynucleotides or vitamin-based cofactors.

3.6 Controlling cancer dependency on microhomology end-joining by targeting POLQ helicase-polymerase

The POLQ enzyme comprises three principal domains, as shown in Figure 8 panel A: the Helicase domain, a central linker domain, and the polymerase domain. PolQ acts in Alt-EJ and replication fork restart (Syed et al., 2023). Our strategy aims at two targets using the GL *in silico* methodology. A) The helicase

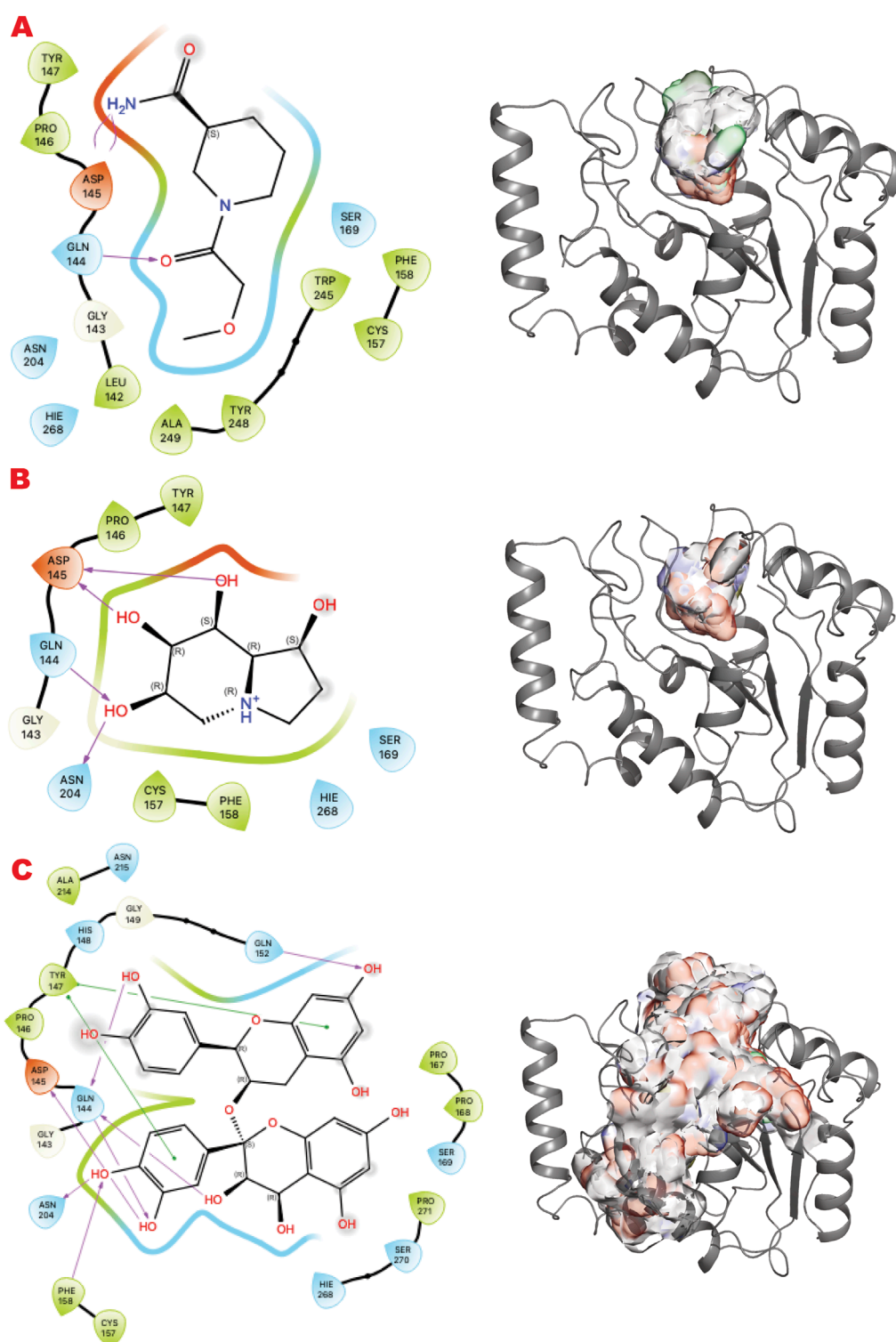


FIGURE 4

UDG VS targeting results. **(A)** VS1 result, from DSI-poised library, from left 2D ligand interactive map of VS1 result, compound 1-(2-methoxyacetyl)piperidine-3-carboxamide, right VS2+VS1 results electrostatic maps from DSI-poised library targeting UDG catalytic domain; **(B)** VS1 result, from PDB-fragments library, from left 2D ligand interactive map of VS1 result, compound octahydroindolizine-1,6,7,8-tetrol, right VS2+VS1 results electrostatic maps from PDB-fragments library targeting UDG catalytic domain; **(C)** VS1 result, from LH1 FDA-approved drugs library, from left 2D ligand interactive map of VS1 result, compound Proanthocyanidins, right VS2+VS1 results electrostatic maps from LH1 FDA-approved drugs library targeting UDG catalytic domain.

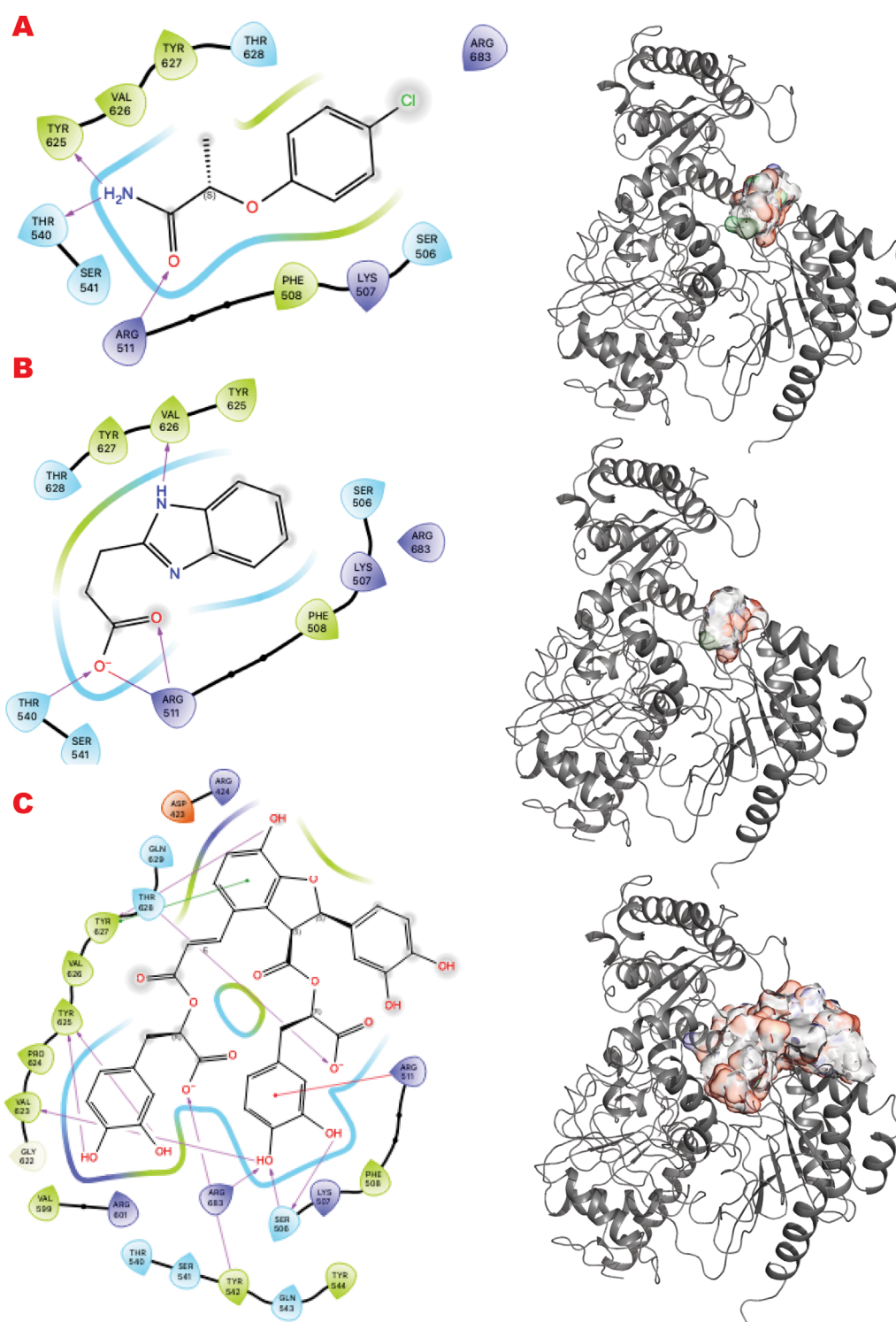


FIGURE 5
XPD VS targeting results. **(A)** VS1 result, from DSI-poised library, from left 2D ligand interactive map of VS1 result, compound 2-(4-chlorophenoxy)propanamide, right VS2+VS1 results electrostatic maps from DSI-poised library targeting XPD helicase domain; **(B)** VS1 result, from PDB-fragments library, from left 2D ligand interactive map of VS1 result, compound Procodazole, right VS2+VS1 results electrostatic maps from PDB-fragments library targeting XPD helicase domain; **(C)** VS1 result, from LH1 FDA-approved drugs library, from left 2D ligand interactive map of VS1 result, compound Salvanolic Acid B, right VS2+VS1 results electrostatic maps from LH1 FDA-approved drugs library targeting XPD helicase domain.

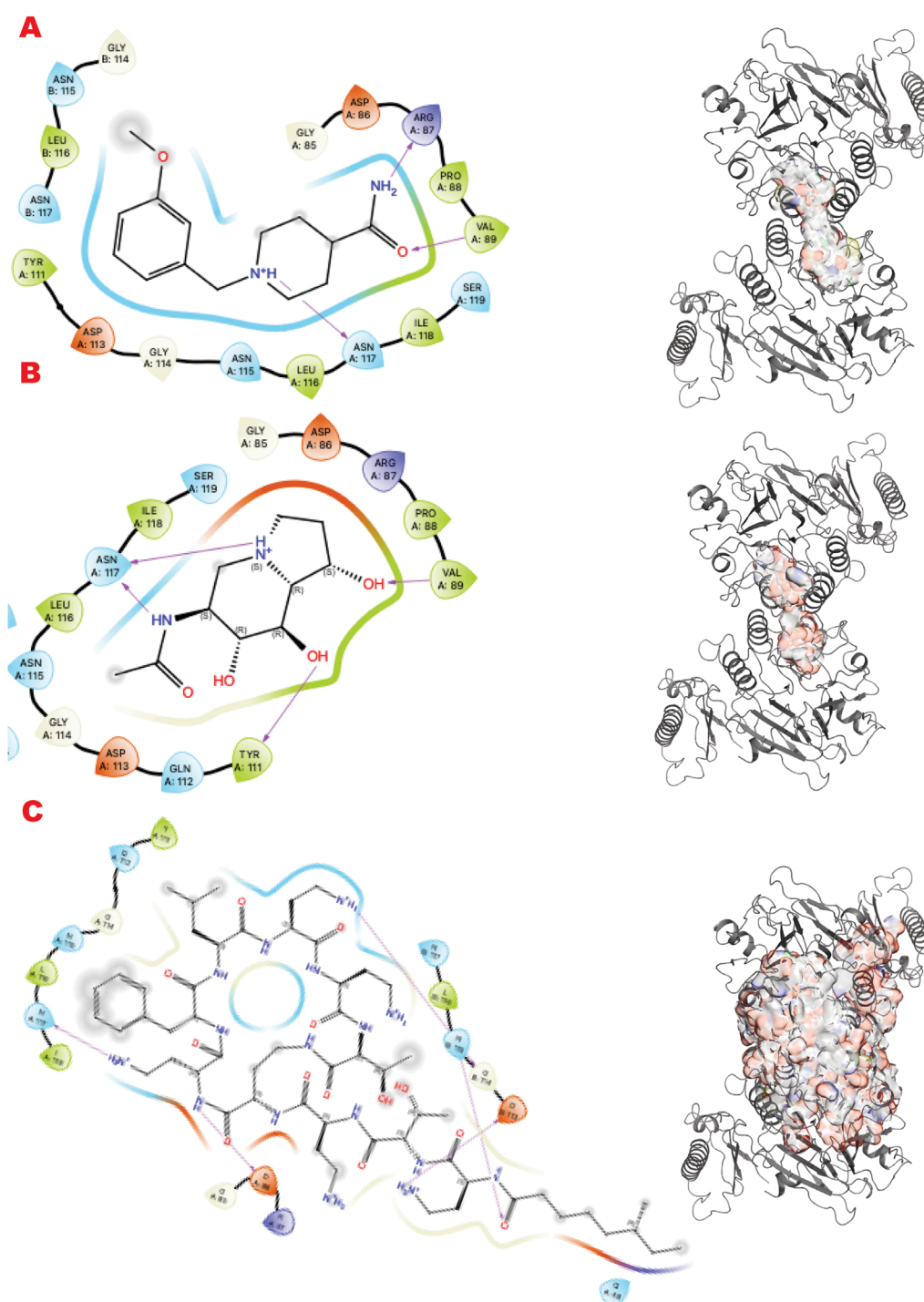
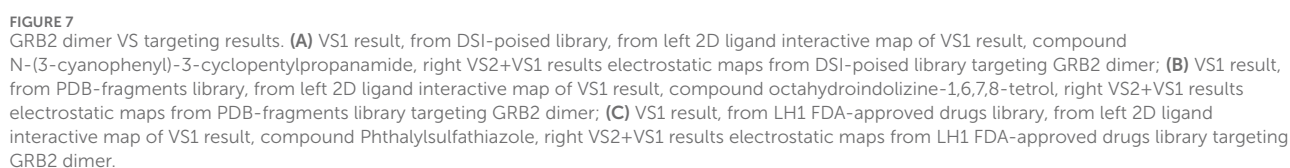


FIGURE 6
MRE11 dimer VS targeting results. **(A)** VS1 result, from DSI-poised library, from left 2D ligand interactive map of VS1 result, compound 1-[(3-methoxyphenyl)methyl]piperidine-4-carboxamide, right VS2+VS1 results electrostatic maps from DSI-poised library targeting MRE11 dimer; **(B)** VS1 result, from PDB-fragments library, from left 2D ligand interactive map of VS1 result, compound 6-acetamido-6-Deoxy-Castanospermine, right VS2+VS1 results electrostatic maps from PDB-fragments library targeting MRE11 dimer; **(C)** VS1 result, from LH1 FDA-approved drugs library, from left 2D ligand interactive map of VS1 result, compound Polymyxin B sulfate, right VS2+VS1 results electrostatic maps from LH1 FDA-approved drugs library targeting MRE11 dimer.

domain where Novobiocin (Newman et al., 2015; Syed et al., 2023; Zhou et al., 2021) is implicated to bind as an allosteric inhibitor of ATPase activation. This site still lacks a high-resolution structure in literature but evidently blocks DNA binding and

ATPase stimulation. B) An allosteric site in the polymerase domain (Bubenik et al., 2022).

Figure 8 (panels A and B) shows the results of the GL silico methods applied to the two grids of POLQ as seen in the scheme



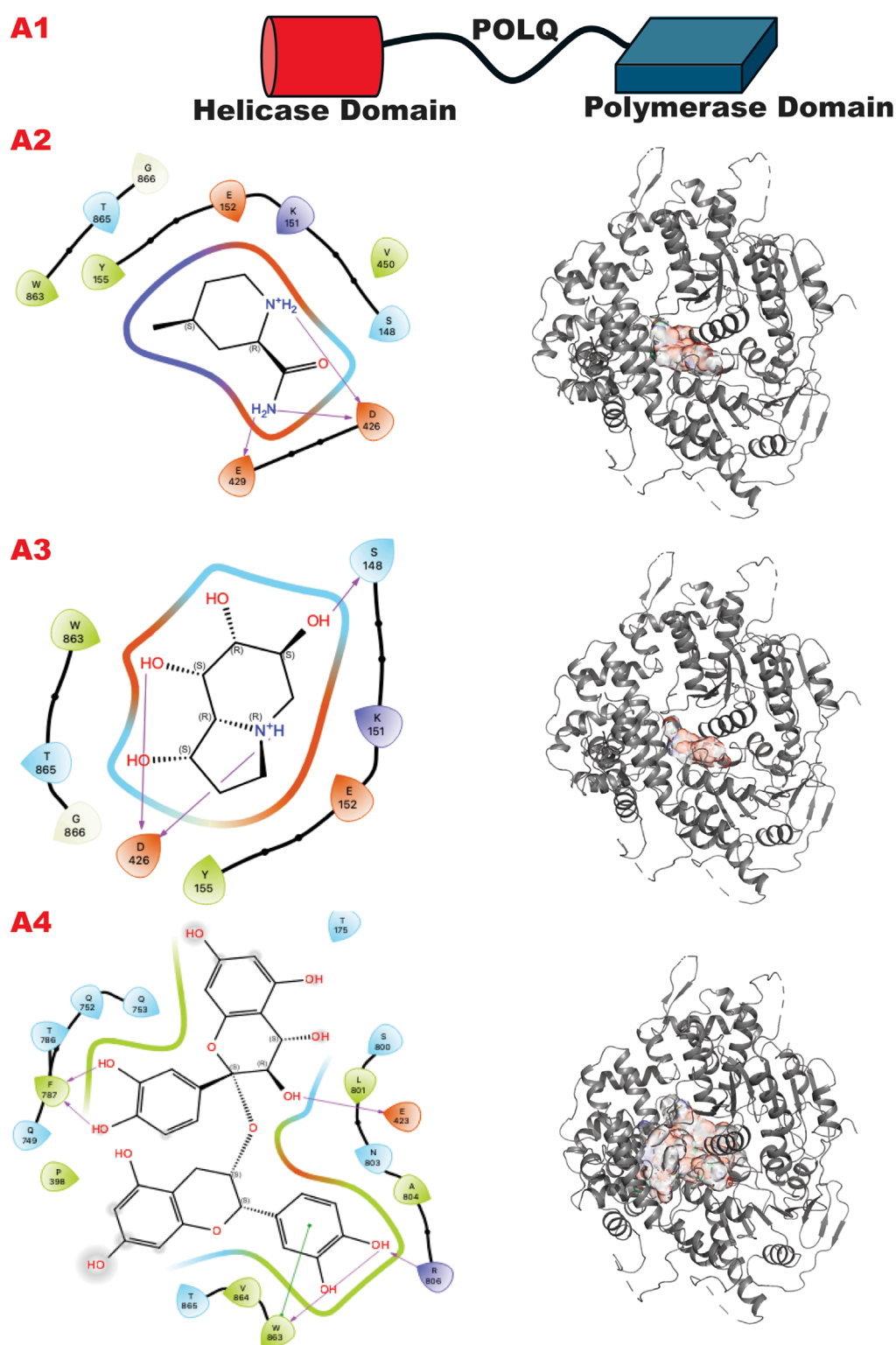


FIGURE 8
(Continued).

in A1: specifically, panel A shows results for the POLQ helicase domain, and panel B for the POLQ allosteric domain. Figure 8, panels A and B show a clear difference in chemotypes targeting the

two domains. Specifically, panel A shows a chemotype that we have similarly explored in another target while aiming at a region where polynucleotides bind, which could help us control the function of

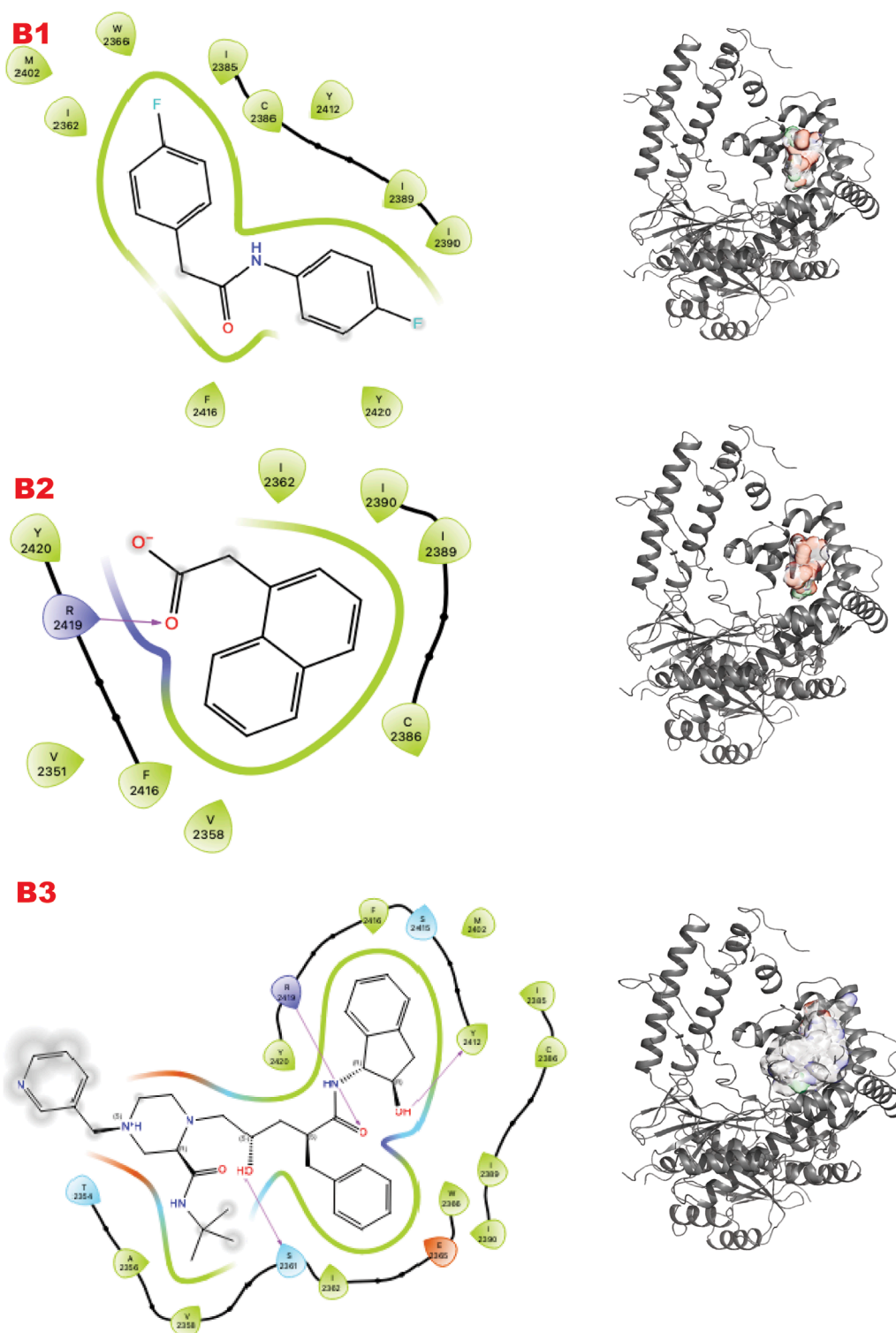


FIGURE 8

(Continued). POLQ VS targeting results for its two different functional domains. Panel A. **(A1)** Schematic representation of POLQ and its functional domains; **(A2)** VS1 result, from DSI-poised library, from left 2D ligand interactive map of VS1 result, compound 4-methylpiperidine-2-carboxamide, right VS2+VS1 results electrostatic maps from DSI-poised library targeting POLQ helicase domain; **(A3)** VS1 result, from PDB-fragments library, from left 2D ligand interactive map of VS1 result, compound octahydroindolizine-1,6,7,8-tetrol, right VS2+VS1 results electrostatic maps from PDB-fragments library targeting POLQ helicase domain; **(A4)** VS1 result, from LH1 FDA-approved drugs library, from left 2D ligand interactive map of VS1 result, compound Proanthocyanidins, right VS2+VS1 results electrostatic maps from LH1 FDA-approved drugs library targeting POLQ helicase domain. Panel B. **(B1)** VS1 result, from DSI-poised library, from left 2D ligand interactive map of VS1 result, compound N,2-bis(4-fluorophenyl)acetamide, right VS2+VS1 results electrostatic maps from DSI-poised library targeting POLQ allosteric domain; **(B2)** VS1 result, from PDB-fragments library, from left 2D ligand interactive map of VS1 result, compound 2-(1-naphthyl)acetic acid, right VS2+VS1 results electrostatic maps from PDB-fragments library targeting POLQ allosteric domain; **(B3)** VS1 result, from LH1 FDA-approved drugs library, from left 2D ligand interactive map of VS1 result, compound Indinavir sulfate, right VS2+VS1 results electrostatic maps from LH1 FDA-approved drugs library targeting POLQ allosteric domain.

the helicase domain of POLQ. The other chemotypes targeting the allosteric site behind the ATPase region show clear chemotypes with aromatic rings or two condensed rings that can properly bind the flexible site's hydrophobic site. An analysis of the surfaces of the VS2 can help to understand the area where this ligand can be expanded, making them more specific for the target and targeted region.

4 Discussion

There is a significant recent shift towards embracing computational technologies in academia and pharma. Furthermore, the computational timeline (2–12 months) is favorable compared to longer standard times for gene-to-lead discovery (4–6 years) (Sadybekov and Katritch, 2023). We therefore developed and evaluated an efficient computational approach for a cancer hallmark and potential Achilles' heel for therapy: genomic instability that depends upon DNA damage responses in the context of stresses from non-B-DNA, replication, transcription, and mitochondrial dysfunction including changes to the epigenetic-metabolic regulatory axis plus the development of apoptotic resistance and metastasis (Bacolla et al., 2019; Bacolla et al., 2016; Das et al., 2024; Singh et al., 2020; Provasek et al., 2024; Fu et al., 2024; Tsutakawa et al., 2020a; Tsutakawa et al., 2017; Tsutakawa et al., 2011). These DDR targets tend to be hard-to-drug, dynamic proteins, and multi-functional complexes that would benefit from approaches that address both these challenges and the opportunities from advances in both protein preparation (Gradia et al., 2017) and structural biology (Fraser and Murcko, 2024; Dyer et al., 2014; Kuhlbrandt, 2014). What may be critical bottlenecks in structure-based drug design for DDR?

For academic research or biotech teams interested in obtaining inhibitor tools for actionable knowledge of biology and mechanisms for preclinical drug research, three main problem areas may be chemical synthesis, structure-based optimizations, and biological testing. For example, although a vast library size and thousands of virtual hits may be the most pragmatic area to optimize, doing so may not improve success if it comes at the cost of making the other areas more difficult. While larger libraries can yield more hits, our compact library approach focuses on quality (its small size aids accurate curation) and efficiency, aiming to identify high-potential compounds suitable for early structural and activity testing and allowing for further development.

The docking score improves with library size, reflecting that libraries are still far smaller than parameter space. As a result, molecular libraries for virtual screening have increased from 3.5 million to over 29 billion compounds. Yet, for huge libraries, hits chosen from among the top-ranking molecules can become dominated by artifacts and incorrect tautomerization (Lyu et al., 2023). Even ignoring this problem, superior initial hits from vast libraries come at the cost of doing more chemistry, challenging curation, structural biology, and biological assays to distinguish biologically valuable hits. The resources of most research teams generally cannot be arbitrarily expanded to curate huge libraries, synthesize vast numbers of hits, process the many synthesized hit structures, or do thousands of biological assays in multiple animals or cell types. We find that even relatively compact academic libraries can be sufficient for identifying binders with

the potential to control allostery and provide ligands for X-ray structures for further optimizations, pending experimental verification (Brosey et al., 2024).

Furthermore, we know that flexibility impacts binding even for protein-antibody recognition (Tainer et al., 1984), that stable inactive complexes can divert functional pathways (Shibata et al., 2014; Tubbs et al., 2009), and that expanding flexible inhibitor pockets can improve specificity and affinity for otherwise undruggable targets (Garcin et al., 2008). What is the most valuable *in silico* approach for preclinical drug discovery involving DDR? We argue that it may be a protocol facilitating the validation of multiple target sites with knowledge of mechanism and biological activity supported by structures and tool compounds to probe target sites and to dissect multi-functionality.

Preclinical biological assays are best done initially in cell assays and later in animal models—these are usually restricted by practicality to selected compounds, so thousands of hits may not prove valuable for cancer biology. What is potentially more valuable is a strategy and method to identify a limited set of hits suitable for biological testing and optimization. In this way, initial hits may serve as tools to test the target site and the biology of cells. We suggest that our compact ESET library and practical GL may facilitate initial hit identification for subsequent validation in cell-based assays, providing a foundation for future animal studies.

We reason that VS could better enable researchers to efficiently harness advanced and integrative structural techniques, including NMR, X-ray Crystallography, X-ray scattering, mass spectrometry, and Cryo-EM for pharmacologically relevant targets. To do this, starting from compact libraries that may allow early structures without new chemical synthesis for compounds suitable for in-cell and potential future animal testing may be optimal. Furthermore, structural and biological advances encourage looking beyond single targets toward pathways and networks where efficient VS protocols may identify tool compounds to test multiple or pivotal targets for cancer biology and medicine more efficiently than giant library approaches. The *in silico* GL gambit, therefore, involves employing a compact chemical library exemplified by the ESET combined library employed here and two VS levels, providing a ranking of hits (VS1) and a probing of binding sites (VS2). Coupling such compact chemical libraries that are accessible without synthesis to an efficient GL *in silico* pipeline providing an objective ranking of hits and probing of sites may facilitate structure-early processes to interrogate multiple targets and biological networks to test biological outcomes, not enzymes and find pivotal targets. The ESET library comprises fragments and FDA-approved compounds for diverse chemotypes. Fragments provide starting points for optimization, while FDA-approved compounds may offer immediate potential for cell-based testing.

Actionable cancer biology is intrinsically hard to achieve due to the complexity of cells and tissues and the amount of heterogeneity and noise in biological systems. The GL protocol targets may allow researchers to probe multiple proteins to test potential impacts on broader pathways and networks, potentially aiding in understanding complex biological systems.

The GL methodology, therefore, proposes an exemplary test of a new pragmatic computational approach and pipeline. The GL approach uses the compact ESET library to probe novel interfaces and allosteric interactions that inform biology and mechanisms for

DDR targets in tumor-relevant cellular pathways, such as seen for PolQ, whose importance is greater in cancer and radiation therapy resistance than in normal cells (Dutta et al., 2017). Thus, The GL methodology aims to facilitate preclinical drug discovery for target and tool development to probe cancer biology and the mechanism for undruggable proteins and multifunctional complexes (out of our protein targets, only PARP1 is a clinically validated target).

4.1 Targeting PARylation and dePARylation

Dynamically formed and removed PARylation is a critical post-translational modification controlling the DDR signaling and repair initiation (Han and Tainer, 2002; Houl et al., 2019). Analyzing the PARP1 results, we find that only niraparib of the approved drugs (olaparib, rucaparib, talazoparib, niraparib) was ranked in our top 200 results representing a top 10 percent cutoff. To consider these *in silico* results compared to experimental binding for the specific VS2 of the LH1 FDA-approved drugs for targeting PARP1 catalytic domain, we analyzed 100 percent of results involving ~17000 poses with unique chemotypes or repeated with a difference conformation. Performing this comprehensive analysis of all binding results revealed all four approved drugs in multiple conformational poses, specifically olaparib (4 poses), Rucaparib (12 poses), niraparib (5 poses), and talazoparib (21 poses). The ranking was niraparib best, rucaparib second, olaparib third, and talazoparib fourth.

To better understand this behavior versus our VS1 result (which ranks ligands) and other better-ranked ligands, we examined all these poses superimposed with the crystal structure of each ligand bound to the PARP1 catalytic domain (PDB ID: 5DS3 (Olaparib) (Dawicki-McKenna et al., 2015), 6VKK (Rucaparib) (Zandarashvili et al., 2020), 7KK3 (Talazoparib) (Ryan et al., 2021), 7KK5 (Niraparib) (Ryan et al., 2021) (Figure 9). For the four drugs, the representation clearly matches the *in silico* calculation for the multi-aromatic unit in the center of the active site. Yet, for all drugs, the other chemical moiety shows multiple conformations around the crystal structure coordinate, yielding low energy binding poses (see Supplementary information for the complete tabulation of VS2 results). Notably, GL docking fits the approved drugs in multiple poses, with only some close to crystal structure as the energetic minimum. We expect the multiple pose representation partly reflects limitations in adequately assessing entropic impacts in our *in silico* energy evaluations. While crystal structures provide valuable details, they may not always reflect the most biologically relevant solution conformations (Chinnam et al., 2024b), which can impact docking evaluations. Also, different assembly states can be seen in the presence of protein partners as seen for proliferation proteins such as CKS and CKD2 and repair protein complexes (Parge et al., 1993; Bourne et al., 1996; Morris et al., 2002; Hammel et al., 2010). Furthermore, computational predictions are optimized to predict crystal structural conformations, as shown by comprehensive X-ray scattering analyses (Hura et al., 2019). We reason that the correct GL docking of larger drug-sized compounds can have more error than smaller chemical ligands that may better satisfy our predominantly enthalpic energetic evaluation. This notion argues in favor of fitting smaller compounds and chemotypes to reduce multiple pose interaction volumes.

We thus consider our VS2 top results that probe sites as relevant, along with the approved drugs that may rank below them, as seen here. Interestingly, new fragments binding in the catalytic site for our FDA-approved Anna Karenina principle states that one must be successful in all three arenas to avoid failure. This means that although a huge library size and thousands of virtual hits may be the most pragmatic arena to optimize, doing so may come at the cost of making the other areas more difficult.

Our top binder is proanthocyanidin, a natural polyphenol compound with cardiovascular properties. The promiscuity of compounds like tannin and flavonoids provides helpful results for *in silico* screening that may also reflect the biological factor that metabolites, RNA, and other small molecules provide regulation and balance in cells and tissues (Das et al., 2024). To optimize *in silico* preclinical hits, added structure-based and biology-based testing is necessary: this is the premise of the early-structure-early-testing ESET combined library and GL protocol. Furthermore, to propose new strategic ligands for DDR *in vitro* analysis, GL methodology may help find optimal natural compounds that can inform target and biological mechanisms. Such compounds are valuable for having optimal solubility, cell permeability, and low toxicity. Most PARP1-approved drugs after treatment involve resistance mechanisms (Noordermeer and van Attikum, 2019), prompting efforts to find other targets in the BRCA-PARP network, such as MRE11, GRB2, ASCC1, and PARG (Brosey et al., 2021; Houl et al., 2019; Ye et al., 2021b; Ye et al., 2024; Chinnam et al., 2024a).

When we explored potential inhibitors for the dePARylation enzyme PARG, we observed that gallotanin was an initially potent inhibitor. This compound is also part of the same class as proanthocyanidin. In GL *in silico* methodology, such compounds populate the top rankings due to the size and type of chemo group involved in the interactions (multiple hydroxyls). The advantage of these compounds is the lower toxicity. They often have limited specificity and potency but can provide useful entry points for structure-based optimizations. Notably, the PARG macrodomain is an important enzymatic domain for many pathogenic RNA viruses, so these analyses are also advancing insights into inhibitors against emerging RNA viral pathogens and biothreats.

4.2 Targeting DNA base excision and transferase activities

BER is the critical DNA repair pathway that deals with most DNA base damage, and uracil DNA glycosylase (UDG) acts in DDR and its interface of replication by removing uracil incorporated into DNA from cytidine deamination or misincorporation in duplex DNA and at replication forks (Parikh et al., 2000a). Although a DNA-mimicking protein inhibitor exists for humans and *E. coli* UDG, chemical inhibitors would be highly valuable (Putnam et al., 1999; Mol et al., 1995a). Let's analyze the top VS1 results targeting the UDG DNA binding active site (Figure 4). These suggest a mimicry of the uracil base from analyzing the top compound of the DSI poised library (Figure 4A). If we observe the top binder from the PDB fragment, the compound octahydroindolizine-1,6,7,8-tetrol-61, which may be a false positive for high recurrence, represents the binding of a sugar moiety like pentose or hexose. One could choose a pentose connected to the uracil ring for this site. It is a lower

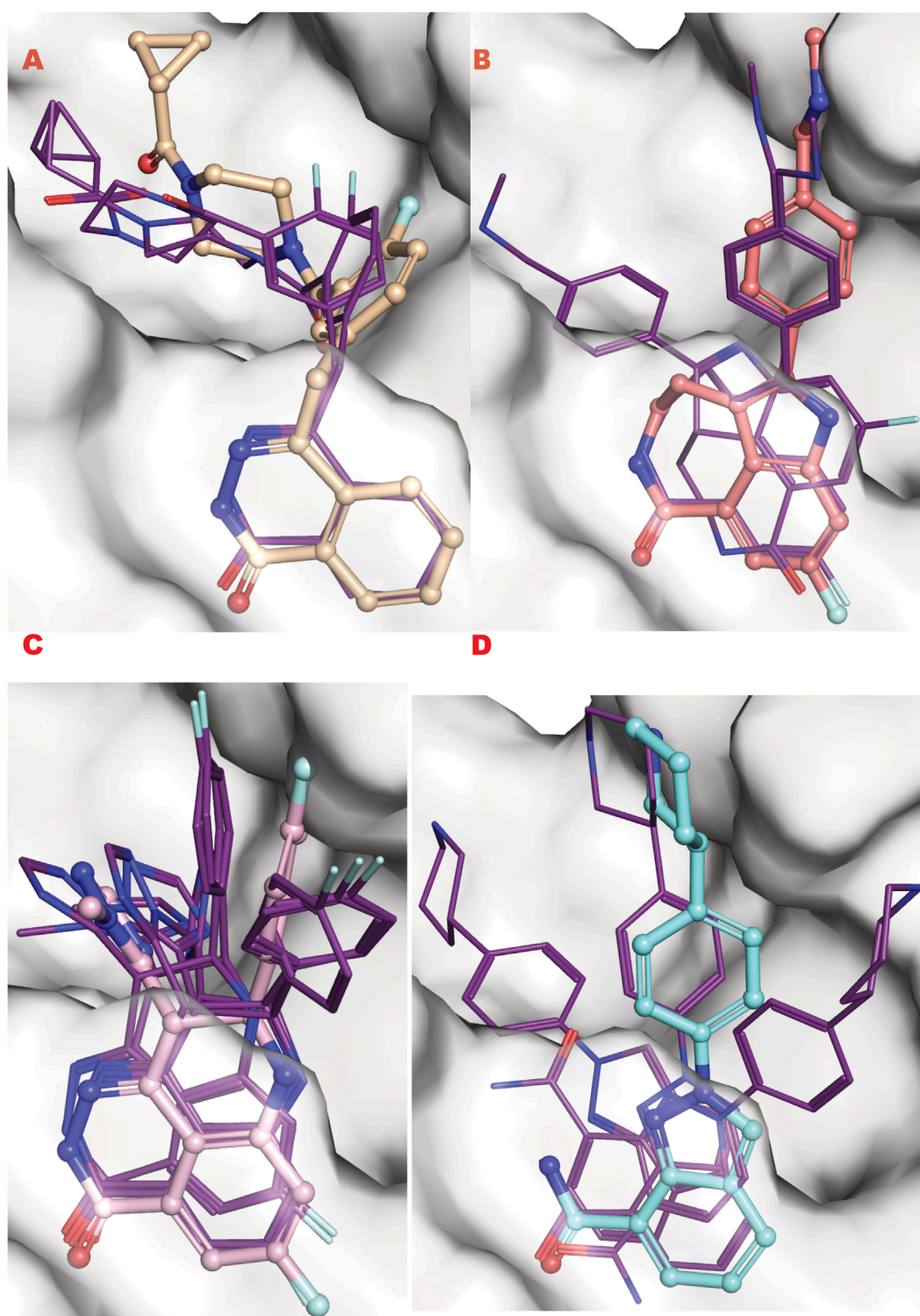


FIGURE 9

PARP1 approved FDA drugs VS results superposed onto crystallographic results. **(A)** Olaparib from crystal structure PDB ID 5DS3 (ball and sticks) superposed to PARP1 structure targeted by VS2 (PDB ID 7ONR) with best Olaparib *in silico* results bold lines; **(B)** Rucaparib from crystal structure PDB ID 6VKK (ball and sticks) superposed to PARP1 structure targeted by VS2 (PDB ID 7ONR) with best Olaparib *in silico* results bold lines; **(C)** Talazoparib from crystal structure PDB ID 7KK3 (ball and sticks) superposed to PARP1 structure targeted by VS2 (PDB ID 7ONR) with best Olaparib *in silico* results bold lines; **(D)** Niraparib from crystal structure PDB ID 7KK5 (ball and sticks) superposed to PARP1 structure targeted by VS2 (PDB ID 7ONR) with best Olaparib *in silico* results bold lines.

binder than the uracil product moiety that can remain bound after the enzyme releases the abasic-containing DNA product. The top result for the FDA approval again shows another possible false positive, the Proanthocyanidin. Still, this compound reach mimics the broader binding of a polynucleotide (double or single strands) as the previous target mimicked the binding of a poly-ADP-ribose. A deep analysis of the listed local minimums in specific regions of the targeting Grid from the VS2 run identifies several potential ways to improve the VS1 scaffold and make a compound more specific for the selected target.

We observe analogous chemical pattern results for the XPD DNA binding channel target site. The helicase XPD is an ATP-driven ssDNA transferase that binds and processes ssDNA in the TFIIH complex for the NER pathway; the site we target is the narrowest point where DNA binds first (Yu et al., 2023; Bralic et al., 2023; Yan et al., 2019). Similar pinch points in replication-repair nucleases EXO5, WRN, FEN1, and XPG suggest possible analogous targeting (Hambarde et al., 2021; Tsutakawa et al., 2020a; Tsutakawa et al., 2017). Furthermore, the plasticity and flexibility of these regions resemble the BER and PARG enzymes analyzed in terms of partial pre-organization for binding and recognition of polynucleotides or polyADPribose. As we can see in Figure 5A, the DSI top result is a tyrosine or phosphotyrosine mimic compound, which could lead to the classification of the site as potentially phospho-tyrosine-DNA adduct or just tyrosine phosphorylation or dephosphorylation site connect to DNA Repair mechanism proceeding. The PDB fragment 5B hit again resembles a nucleotide or aromatic chemotype in agreement with the other result. The FDA-approved top result (5C) is again a high molecular weight compound that is very hydrophilic, Desmopressin, a modification of the hormone vasopressin (antidiuretic). The compound has two aromatic amino acid sites, PHE-TYR, adjacent to the top binder of the other fragments for this target.

4.3 Targeting the critical adapter GRB2

GRB2 acts in the RAS-Map Tyr kinase signaling pathway for proliferation. It has recently been discovered and characterized for interactions with MRE11 nuclease (MRN complex) as well as the RAD51 recombinase (Ye et al., 2024; Ye et al., 2021b). GRB2 exemplifies a class of adaptors that link cell proliferation to DDR. As such, targeting GRB2 opens a new way to control DDR at the signal transduction level (Ye et al., 2021b). MRE11 and GRB2 have a dimeric form; MRE11 dimerizes at dsDNA binding in active form, while GRB2 dimer is inactive. The *in silico* methodology runs against the dimer structure of both with two different aims: 1) targeting the MRE11-NBS1 interface should reveal the chemistry behind the organization of MRE11 as a symmetrical or asymmetrical dimer induced by NBS1 with repercussions on its activity, 2) targeting GRB2 dimer aims to inhibit GRB2 function and MRE11 interface. Identifying tool compounds for these two ways to interfere with DSBs should help to test their ability to target tumor cells (Ye et al., 2021b; Ye et al., 2024).

Furthermore, the GRB2 analyses suggest a new way to probe substrate binding. The dimeric structure of GRB2 has been challenged as a natural solution dimer structure (Maignan et al., 1995). The plasticity of this adapter in solution is high due to its adaptivity to signal transduction substrate. It is intrinsic in its molecular organization as three domains connected by flexible linkers. We reason that

the X-ray crystal structure contains relevant information on the GRB2 inactive form that merits interpretation of what is entrapped in its homodimeric interface. Therefore, we provide a chemical interpretation for combining top results from the PDB fragments library. This illustration shows how a more significant representation can be developed to understand a potential natural substrate that may govern such an inactive form. We present the superposition of top results from the PDB fragment library and how to merge the chemotype to envision a potential substrate binder to keep such inactive dimer form in the cell (Figure 10). In the center, we have the octahydroindolizine-1,6,7,8-tetrol-61 (a false positive) representing sugar (pentose or hexose) and phosphate mimic binding. On the left side, we have a purine base mimic (Adenosyl or Guanosyl). On the ending right side, we have a 6-ring binding with polarity. In this initial evaluation of a possible binding chemotype, the combination hints at a putative substrate like the GDP-mannose related to the GRB2 cellular environment. The next step might be to search a human metabolite database, e.g., Human Metabolome (<https://hmdb.ca/>).

4.4 Targeting pol Q polymerase and helicase sites

The PolQ helicase-polymerase is central to the minor DSB pathway ATL-EJ, which becomes important for many cancer cells and therapeutic resistance (Eckelmann et al., 2020; Dutta et al., 2017; Syed et al., 2023; Zhou et al., 2021). Targeting the different helicase and polymerase sites of PolQ's distinct catalytic processes to control a specific cellular and molecular pathway is attractive but challenging. For POLQ we target the helicase site at the DNA binding site where putative Novobiocin binds (Hambo Guo, 2023) and an allosteric site near the Polymerase catalytic site.

Dissecting multiple functions of the same protein with two different chemicals has been proven possible (Shibata et al., 2014), so it can also be applied to polymerase. Here, we observe that the GL results show different chemotypes from the two domain-targeted results (Figure 8, panels A, B). Specifically, the helicase domain site is predisposed to interact with groups, as shown by the top fragment VS1, which ranks ligands. This is typically mimicking polynucleotide binding. The allosteric site allows more hydrophobic binding driven by a double aromatic ring-connected motif. Also, the FDA-approved results confirm the hydrophilicity versus hydrophobicity of the two sites. One challenge will be in the experimental evaluation of the combination of best affinity molecules and results in normal versus tumor cells where PolQ is more highly expressed.

In the VS2 targeting that probes the helicase domain site, novobiocin sodium, which is active against POLQ (Zhou et al., 2021), is ranked within the first 35 compounds (without energy re-ranking); we use energy re-ranking to calibrate novobiocin on top of the list, and we plotted the binding site with 2D map in the supplementary information (Supplementary Figure S1). The experimental validation of this site comes from mass spectrometry (Syed et al., 2023), which is consistent with a BioRxiv preprint (Hambo Guo, 2023). With the energy re-ranking ratio 0.7 (docking score) plus 0.3 (0.1 MMBSA delta G bind), Novobiocin ranked in the top 25 FDA-approved drug compounds.

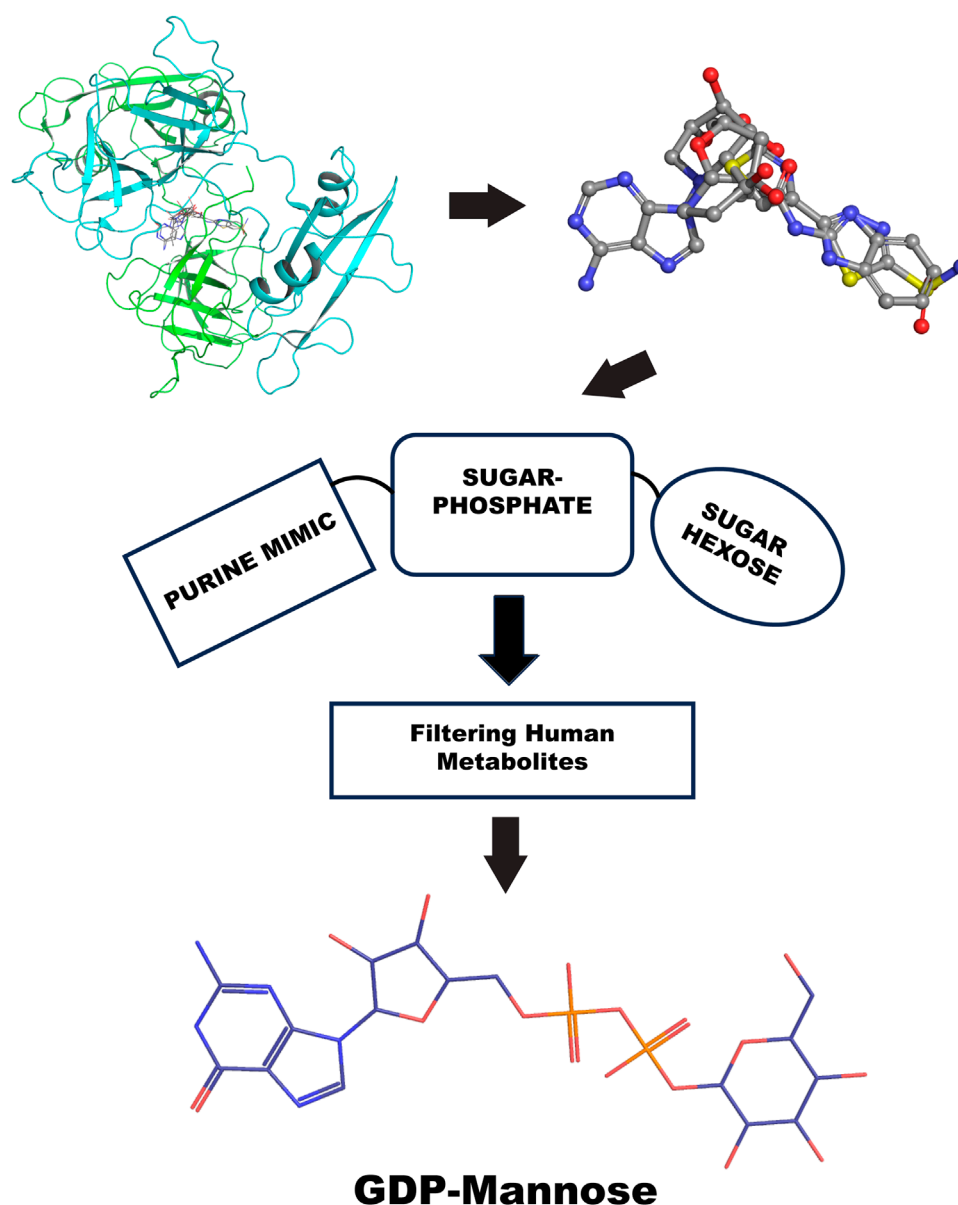


FIGURE 10
GRB2 potential substrate binding flow chart. The top left shows the top VS2 results of the PDB fragment library targeting the GRB2 dimer interface; the top right superposition of fragments results. The center represents the chemotype scheme of the potential substrate, followed by a screening process through the human metabolome; at the bottom is the potential result, represented by a 2D chemical sketch of GDP-mannose.

4.5 Limitations of this study

This is a manuscript on the GL computational approach and ESET library. The GL approach thus provides both target probing (to select promising points in pathway networks) and tool inhibitors suitable for optimization due to the chemistry selected for the ESET library. The computational results identify testable tool inhibitors, but these are supposed to be considered hypotheses until tested experimentally. However, experimental tests are noted for earlier versions of the GL protocol in several cited papers. This cited published work with multiple related efforts suggests that the reported GL-type approach can generate inhibitor tools for several systems as validated by both

affinity measures and crystal structures. Notably, validation will require new experimental support for newly discovered inhibitors. Furthermore, crystal structures may not always reflect the most biologically relevant solution conformations (Chinnam et al., 2024a). However, they are still a gold standard in drug design and the best starting point for docking and evaluating the VS hits proposed here.

5 Future perspectives

Significant advances in the structural biochemistry of DNA repair and replication networks promote an ongoing renaissance

in biophysics and molecular biology stimulated by their interface with cancer biology (Zhang et al., 2021). In the context of these experimental and computational advances, we suggest that there remains a lag in novel drug development. Furthermore, we reason that added novelty could come from better enabling structural and biological researchers to employ their knowledge to identify and test tool inhibitors probing fundamental mechanisms and new targets. These observations motivated us to develop an enabling practical GL protocol with two VS protocols (VS1 for ligand ranking and VS2 for site probing) combined with an ESET compact chemical library that may harness more biological expertise to stimulate increasing future breakthroughs. To our knowledge, this is the only VS study targeting multiple DDR pathways and targets that supports its efficiency and ability to consider networks rather than single targets.

Currently, many groups can design knockdown and knockout methodologies to remove a gene product, but relatively few can design the new chemical inhibitor tools needed to dissect multifunctional DDR networks. Large chemical libraries will continue to identify leads for structure-based and biology-based optimizations. However, an alternative GL strategy is to provide a compact library of cell-accessible compounds for efficient VS for early structures and early testing (ESET) on multiple targets. Based upon the exemplary GL VS results presented here, a practical path to obtain promising hits involves the following steps as also supported by independent observations (Zhao, 2024): 1) test a diverse set of fragments that have less flexibility and torsional variation to identify favorable subsite interactions, 2) focus on well-ranked poses for relatively rigid and higher molecular weight compounds, and 3) select compounds available without chemical synthesis and suitable to complement VS rankings with subsequent experimental affinity measurements, activity assays, and structure determinations.

Complex DDR systems challenge canonical approaches that typically focus on a single target (deep VS efforts) due to the efforts required for large libraries and subsequent steps. Here, the notion presented is that the simple, robust GL protocol lets researchers test multiple proteins (go wider for pre-clinical searches) compared to larger library screens that yield more hits that need more time for each target. We, therefore, envision that the GL approach will allow more researchers to test networks and activities by targeting allostery, interfaces, and active sites. The GL protocol thereby provides an efficient compass to guide biology-driven inhibitor tool development by many groups. Currently, *in silico* researchers often seek to discover functional ligands by running calculations against well-established systems already of interest to Pharma and Biotech teams. The presented GL protocol aims to leverage a basic science approach enabling the evaluation of new experimental and modeled structures and interactions by developing inhibitors that can probe mechanisms and biological differences separating pathways and interactions networks in normal and cancer cells. For example, researchers can apply the ESET library and GL protocol to find inhibitor tools to probe cancer impacts for point mutations and for functionality-implicated variants of unknown significance in pockets and channels revealed by cancer genome sequencing and experimental structures or AlphaFold modeling (Longo et al., 2023; Tsutakawa et al., 2021; Fan et al., 2008; Tsutakawa et al., 2020a). We anticipate that the collective methods and results presented here, which unveil the GL protocol, compact ESET chemical library, and exemplary DDR target sites and compounds,

may allow more researchers to probe mechanisms and advance chemical tools for actionable knowledge to develop novel future therapeutic inventions. Viral biothreats are another representative emerging area where the ability to efficiently test multiple targets and dissect network interactions is critically important. Thus, GL-enabled inhibitor results may help address the missing novelty in drug development that is key to precision oncology and likely also to biopreparedness and efficient response capabilities.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

DM: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing—original draft, Writing—review and editing. JT: Funding acquisition, Resources, Conceptualization, Supervision, Validation, Writing—review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by National Cancer Institute grants P01 CA092584 and R35 CA220430, a Robert A. Welch Chemistry Chair (G-0010), the Cancer Prevention and Research Institute of Texas RP180813, and Department of Energy DE-AC02-05CH11231 Taskforce 5.

Acknowledgments

We thank Albino Bacolla, Zamal Ahmed, Runze Shen, Aleem Syed, Min Liu, and Chi-Lin Tsai for the discussions.

Conflict of interest

Author DM was employed by Moiani Research Inc.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of

their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630, 493–500. doi:10.1038/s41586-024-07487-w
- Adhikari, L., Mishra, H., Semalty, M., and Semalty, A. (2023). Small angle Neutron scattering in drug discovery research: a novel tool for advanced study of structures of biological macromolecules. *Curr. Drug Discov. Technol.* 20, e150523216942. doi:10.2174/1570163820666230515162614
- Ahmed, Z., Timsah, Z., Suen, K. M., Cook, N. P., Lee, G. R. T., Lin, C. C., et al. (2015). Grb2 monomer-dimer equilibrium determines normal versus oncogenic function. *Nat. Commun.* 6, 7354. doi:10.1038/ncomms8354
- Bacolla, A., Sengupta, S., Ye, Z., Yang, C., Mitra, J., De-Paula, R. B., et al. (2021). Heritable pattern of oxidized DNA base repair coincides with pre-targeting of repair complexes to open chromatin. *Nucleic Acids Res.* 49, 221–243. doi:10.1093/nar/gkaa1120
- Bacolla, A., Tainer, J. A., Vasquez, K. M., and Cooper, D. N. (2016). Translocation and deletion breakpoints in cancer genomes are associated with potential non-B DNA-forming sequences. *Nucleic Acids Res.* 44, 5673–5688. doi:10.1093/nar/gkw261
- Bacolla, A., Ye, Z., Ahmed, Z., and Tainer, J. A. (2019). Cancer mutational burden is shaped by G4 DNA, replication stress and mitochondrial dysfunction. *Prog. Biophys. Mol. Biol.* 147, 47–61. doi:10.1016/j.pbiomolbio.2019.03.004
- Bourne, Y., Watson, M. H., Hickey, M. J., Holmes, W., Rocque, W., Reed, S. I., et al. (1996). Crystal structure and mutational analysis of the human CDK2 kinase complex with cell cycle-regulatory protein CksHs1. *Cell* 84, 863–874. doi:10.1016/s0092-8674(00)81065-x
- Bralic, A., Tehseen, M., Sobhy, M. A., Tsai, C. L., Alhudhali, L., Yi, G., et al. (2023). A scanning-to-incision switch in TFIIH-XPG induced by DNA damage licenses nucleotide excision repair. *Nucleic Acids Res.* 51, 1019–1033. doi:10.1093/nar/gkac1095
- Brose, C. A., Ahmed, Z., Lees-Miller, S. P., and Tainer, J. A. (2017). What combined measurements from structures and imaging tell us about DNA damage responses. *Methods Enzymol.* 592, 417–455. doi:10.1016/bs.mie.2017.04.005
- Brose, C. A., Houli, J. H., Katsonis, P., Palapiti-Modarage, L. P. F., Bommagani, S., Arvai, A., et al. (2021). Targeting SARS-CoV-2 Nsp3 macrodomain structure with insights from human poly(ADP-ribose) glycohydrolase (PARG) structures with inhibitors. *Prog. Biophys. Mol. Biol.* 163, 171–186. doi:10.1016/j.pbiomolbio.2021.02.002
- Brose, C. A., Link, T. M., Shen, R., Moiani, D., Burnett, K., Hura, G. L., et al. (2024). Chemical screening by time-resolved X-ray scattering to discover allosteric probes. *Nat. Chem. Biol.* 20, 1199–1209. doi:10.1038/s41589-024-01609-1
- Brownlie, J., Kulkarni, S., Algethami, M., Jeyapalan, J. N., Mongan, N. P., Rakha, E. A., et al. (2023). Targeting DNA damage repair precision medicine strategies in cancer. *Curr. Opin. Pharmacol.* 70, 102381. doi:10.1016/j.coph.2023.102381
- Bubenik, M., Mader, P., Mochirian, P., Vallee, F., Clark, J., Truchon, J. F., et al. (2022). Identification of **RP-6685**, an orally bioavailable compound that inhibits the DNA polymerase activity of polθ. *J. Med. Chem.* 65, 13198–13215. doi:10.1021/acs.jmedchem.2c00998
- Cebi, E., Lee, J., Subramani, V. K., Bak, N., Oh, C., and Kim, K. K. (2024). Cryo-electron microscopy-based drug design. *Front. Mol. Biosci.* 11, 1342179. doi:10.3389/fmolb.2024.1342179
- Chinnam, N. B., Thapar, R., Arvai, A. S., Sarker, A. H., Soll, J. M., Paul, T., et al. (2024a). ASCC1 structures and bioinformatics reveal a novel Helix-Clasp-Helix RNA-binding motif linked to a two-histidine phosphodiesterase. *J. Biol. Chem.* 107368. doi:10.1016/j.jbc.2024.107368
- Chinnam, N. B., Thapar, R., Arvai, A. S., Sarker, A. H., Soll, J. M., Paul, T., et al. (2024b). ASCC1 structures and bioinformatics reveal a novel helix-clasp-helix RNA-binding motif linked to a two-histidine phosphodiesterase. *J. Biol. Chem.* 300, 107368. doi:10.1016/j.jbc.2024.107368
- Choi, W., Lee, E. S., Kim, J. W., Kim, H. K., Kang, H. C., Kim, S. W., et al. (2022). Modifying effect of the interleukin-18 level on the association between *BDNF* methylation and long-term cardiovascular outcomes in patients with acute coronary syndrome. *Int. J. Mol. Sci.* 23, 15270. doi:10.3390/ijms232315270
- Cox, O. B., Krojer, T., Collins, P., Monteiro, O., Talon, R., Bradley, A., et al. (2016). A poised fragment library enables rapid synthetic expansion yielding the first reported inhibitors of PHIP(2), an atypical bromodomain. *Chem. Sci.* 7, 2322–2330. doi:10.1039/c5sc03115j
- Da Costa, A., Chowdhury, D., Shapiro, G. I., D'Andrea, A. D., and Konstantinopoulos, P. A. (2023). Targeting replication stress in cancer therapy. *Nat. Rev. Drug Discov.* 22, 38–58. doi:10.1038/s41573-022-00558-5
- Das, C., Bhattacharya, A., Adhikari, S., Mondal, A., Mondal, P., Adhikary, S., et al. (2024). A prismatic view of the epigenetic-metabolic regulatory axis in breast cancer therapy resistance. *Oncogene* 43, 1727–1741. doi:10.1038/s41388-024-03054-9
- Daut, R. A., and Fonken, L. K. (2019). Circadian regulation of depression: a role for serotonin. *Front. Neuroendocrinol.* 54, 100746. doi:10.1016/j.yfrne.2019.04.003
- Dawicki-Mckenna, J. M., Langelier, M. F., Denizio, J. E., Riccio, A. A., Cao, C. D., Karch, K. R., et al. (2015). PARP-1 activation requires local unfolding of an autoinhibitory domain. *Mol. Cell* 60, 755–768. doi:10.1016/j.molcel.2015.10.013
- Duan, J., He, X. H., Li, S. J., and Xu, H. E. (2024). Cryo-electron microscopy for GPCR research and drug discovery in endocrinology and metabolism. *Nat. Rev. Endocrinol.* 20, 349–365. doi:10.1038/s41574-024-00957-1
- Dutta, A., Eckelmann, B., Adhikari, S., Ahmed, K. M., Sengupta, S., Pandey, A., et al. (2017). Microhomology-mediated end joining is activated in irradiated human cells due to phosphorylation-dependent formation of the XRCC1 repair complex. *Nucleic Acids Res.* 45, 2585–2599. doi:10.1093/nar/gkw1262
- Dyer, K. N., Hammel, M., Rambo, R. P., Tsutakawa, S. E., Rodic, I., Classen, S., et al. (2014). High-throughput SAXS for the characterization of biomolecules in solution: a practical approach. *Methods Mol. Biol.* 1091, 245–258. doi:10.1007/978-1-62703-691-7_18
- Eckelmann, B. J., Bacolla, A., Wang, H., Ye, Z., Guerrero, E. N., Jiang, W., et al. (2020). XRCC1 promotes replication restart, nascent fork degradation and mutagenic DNA repair in BRCA2-deficient cells. *Nat. Cancer* 2, zcaa013. doi:10.1093/narcan/zcaa013
- Fan, L., Fuss, J. O., Cheng, Q. J., Arvai, A. S., Hammel, M., Roberts, V. A., et al. (2008). XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. *Cell* 133, 789–800. doi:10.1016/j.cell.2008.04.030
- Filandr, F., Sarpe, V., Raval, S., Crowder, D. A., Khan, M. F., Douglas, P., et al. (2024). Automating data analysis for hydrogen/deuterium exchange mass spectrometry using data-independent acquisition methodology. *Nat. Commun.* 15, 2200. doi:10.1038/s41467-024-46610-3
- Fraser, J. S., and Murcko, M. A. (2024). Structure is beauty, but not always truth. *Cell* 187, 517–520. doi:10.1016/j.cell.2024.01.003
- Fu, J., Ling, J., Li, C. F., Tsai, C. L., Yin, W., Hou, J., et al. (2024). Nardilysin-regulated scission mechanism activates polo-like kinase 3 to suppress the development of pancreatic cancer. *Nat. Commun.* 15, 3149. doi:10.1038/s41467-024-47242-3
- Garcin, E. D., Arvai, A. S., Rosenfeld, R. J., Kroeger, M. D., Crane, B. R., Andersson, G., et al. (2008). Anchored plasticity opens doors for selective inhibitor design in nitric oxide synthase. *Nat. Chem. Biol.* 4, 700–707. doi:10.1038/nchembio.115
- Gradia, S. D., Ishida, J. P., Tsai, M. S., Jeans, C., Tainer, J. A., and Fuss, J. O. (2017). MacroBac: new technologies for robust and efficient large-scale production of recombinant multiprotein complexes. *Methods Enzymol.* 592, 1–26. doi:10.1016/bs.mie.2017.03.008
- Groelly, F. J., Fawkes, M., Dagg, R. A., Blackford, A. N., and Tarsounas, M. (2023). Targeting DNA damage response pathways in cancer. *Nat. Rev. Cancer* 23, 78–94. doi:10.1038/s41568-022-00535-5
- Grocock, L. M., Nie, M., Prudden, J., Moiani, D., Wang, T., Cheltsov, A., et al. (2014). RNF4 interacts with both SUMO and nucleosomes to promote the DNA damage response. *EMBO Rep.* 15, 601–608. doi:10.1002/embr.201338369
- Hambarde, S., Tsai, C. L., Pandita, R. K., Bacolla, A., Maitra, A., Charaka, V., et al. (2021). EXO5-DNA structure and BLM interactions direct DNA resection critical for ATR-dependent replication restart. *Mol. Cell* 81, 2989–3006 e9. doi:10.1016/j.molcel.2021.05.027
- Hammel, M., Yu, Y., Fang, S., Lees-Miller, S. P., and Tainer, J. A. (2010). XLF regulates filament architecture of the XRCC4-ligase IV complex. *Structure* 18, 1431–1442. doi:10.1016/j.str.2010.09.009
- Han, S., and Tainer, J. A. (2002). The ARTT motif and a unified structural understanding of substrate recognition in ADP-ribosylating bacterial toxins and eukaryotic ADP-ribosyltransferases. *Int. J. Med. Microbiol.* 291, 523–529. doi:10.1078/1438-4221-00162

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2024.1442267/full#supplementary-material>

- Helleday, T., and Rudd, S. G. (2022). Targeting the DNA damage response and repair in cancer through nucleotide metabolism. *Mol. Oncol.* 16, 3792–3810. doi:10.1002/1878-0261.13227
- Hitomi, K., Iwai, S., and Tainer, J. A. (2007). The intricate structural chemistry of base excision repair machinery: implications for DNA damage recognition, removal, and repair. *DNA Repair (Amst)* 6, 410–428. doi:10.1016/j.dnarep.2006.10.004
- Hopkins, J. L., Lan, L., and Zou, L. (2022). DNA repair defects in cancer and therapeutic opportunities. *Genes Dev.* 36, 278–293. doi:10.1101/gad.349431.122
- Hou, J., Zhao, R., Xia, W., Chang, C. W., You, Y., Hsu, J. M., et al. (2020). PD-L1-mediated gasdermin C expression switches apoptosis to pyroptosis in cancer cells and facilitates tumour necrosis. *Nat. Cell Biol.* 22, 1264–1275. doi:10.1038/s41556-020-0575-z
- Houl, J. H., Ye, Z., Brosey, C. A., Balapiti-Modarage, L. P. F., Namjoshi, S., Bacolla, A., et al. (2019). Selective small molecule PARP inhibitor causes replication fork stalling and cancer cell death. *Nat. Commun.* 10, 5654. doi:10.1038/s41467-019-13508-4
- Hura, G. L., Hodge, C. D., Rosenberg, D., Guzenko, D., Duarte, J. M., Monastyrskyy, B., et al. (2019). Small angle X-ray scattering-assisted protein structure prediction in CASP13 and emergence of solution structure differences. *Proteins* 87, 1298–1314. doi:10.1002/prot.25827
- Johannes, J. W., Balazs, A., Barratt, D., Bista, M., Chuba, M. D., Cosulich, S., et al. (2021). Discovery of 5-[4-[(7-Ethyl-6-oxo-5,6-dihydro-1,5-naphthyridin-3-yl)methyl]piperazin-1-yl]-N-methylpyridine-2-carboxamide (AZD5305): a PARP1-DNA trapper with high selectivity for PARP1 over PARP2 and other PARPs. *J. Med. Chem.* 64, 14498–14512. doi:10.1021/acs.jmedchem.1c01012
- Kocic, G., Chernev, A., Tegunov, D., Dienemann, C., Urlaub, H., and Cramer, P. (2019). Structural basis of TFIH activation for nucleotide excision repair. *Nat. Commun.* 10, 2885. doi:10.1038/s41467-019-10745-5
- Korshunov, K. S., Blakemore, L. J., and Trombley, P. Q. (2017). Dopamine: a modulator of circadian rhythms in the central nervous system. *Front. Cell Neurosci.* 11, 91. doi:10.3389/fncel.2017.00091
- Kuhlbrandt, W. (2014). Biochemistry. The resolution revolution. *Science* 343, 1443–1444. doi:10.1126/science.1251652
- Kuhn, L. A., Siani, M. A., Pique, M. E., Fisher, C. L., Getzoff, E. D., and Tainer, J. A. (1992). The interdependence of protein surface topography and bound water molecules revealed by surface accessibility and fractal density measures. *J. Mol. Biol.* 228, 13–22. doi:10.1016/0022-2836(92)90487-5
- Lee, D., Huntoon, K., Wang, Y., Kang, M., Lu, Y., Jeong, S. D., et al. (2024). Synthetic cationic helical polypeptides for the stimulation of antitumour innate immune pathways in antigen-presenting cells. *Nat. Biomed. Eng.* 8, 593–610. doi:10.1038/s41551-024-01194-7
- Longo, M. A., Roy, S., Chen, Y., Tomaszowski, K. H., Arvai, A. S., Pepper, J. T., et al. (2023). RAD51C-XRCC3 structure and cancer patient mutations define DNA replication roles. *Nat. Commun.* 14, 4445. doi:10.1038/s41467-023-40096-1
- Lu, C., Wu, C., Ghoreishi, D., Chen, W., Wang, L., Damm, W., et al. (2021). OPLS4: improving force field accuracy on challenging regimes of chemical space. *J. Chem. Theory Comput.* 17, 4291–4300. doi:10.1021/acs.jctc.1c00302
- Lu, H., Zhou, Q., He, J., Jiang, Z., Peng, C., Tong, R., et al. (2020). Recent advances in the development of protein-protein interactions modulators: mechanisms and clinical trials. *Signal Transduct. Target Ther.* 5, 213. doi:10.1038/s41392-020-00315-3
- Lyu, J., Irwin, J. J., and Shoichet, B. K. (2023). Modeling the expansion of virtual screening libraries. *Nat. Chem. Biol.* 19, 712–718. doi:10.1038/s41589-022-01234-w
- Maignan, S., Guilloteau, J. P., Fromage, N., Arnoux, B., Becquart, J., and Ducruix, A. (1995). Crystal structure of the mammalian Grb2 adaptor. *Science* 268, 291–293. doi:10.1126/science.7716522
- Moiani, D., Link, T. M., Brosey, C. A., Katsonis, P., Lichtarge, O., Kim, Y., et al. (2021). An efficient chemical screening method for structure-based inhibitors to nucleic acid enzymes targeting the DNA repair-replication interface and SARS CoV-2. *Methods Enzymol.* 661, 407–431. doi:10.1016/bs.mie.2021.09.003
- Moiani, D., Ronato, D. A., Brosey, C. A., Arvai, A. S., Syed, A., Masson, J. Y., et al. (2018). Targeting allostery with avatars to design inhibitors assessed by cell activity: dissecting MRE11 endo- and exonuclease activities. *Methods Enzymol.* 601, 205–241. doi:10.1016/bs.mie.2017.11.030
- Moiani, D., Salvalaglio, M., Cavallotti, C., Bujacz, A., Redzynia, I., Bujacz, G., et al. (2009). Structural characterization of a Protein A mimetic peptide dendrimer bound to human IgG. *J. Phys. Chem. B* 113, 16268–16275. doi:10.1021/jp909405b
- Mol, C. D., Arvai, A. S., Sanderson, R. J., Slupphaug, G., Kavli, B., Krokan, H. E., et al. (1995a). Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell* 82, 701–708. doi:10.1016/0092-8674(95)90467-0
- Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E., et al. (1995b). Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell* 80, 869–878. doi:10.1016/0092-8674(95)90290-2
- Mol, C. D., Izumi, T., Mitra, S., and Tainer, J. A. (2000). DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination [corrected]. *Nature* 403, 451–456. doi:10.1038/35000249
- Morin, L. P. (1999). Serotonin and the regulation of mammalian circadian rhythmicity. *Ann. Med.* 31, 12–33. doi:10.3109/07853899909019259
- Morris, M. C., Gondeau, C., Tainer, J. A., and Divita, G. (2002). Kinetic mechanism of activation of the Cdk2/cyclin A complex. Key role of the C-lobe of the Cdk. *J. Biol. Chem.* 277, 23847–23853. doi:10.1074/jbc.M107890200
- Mui, T. P., Fuss, J. O., Ishida, J. P., Tainer, J. A., and Barton, J. K. (2011). ATP-stimulated, DNA-mediated redox signaling by XPD, a DNA repair and transcription helicase. *J. Am. Chem. Soc.* 133, 16378–16381. doi:10.1021/ja207222t
- Newman, J. A., Cooper, C. D. O., Aitkenhead, H., and Gileadi, O. (2015). Structure of the helicase domain of DNA polymerase theta reveals a possible role in the microhomology-mediated end-joining pathway. *Structure* 23, 2319–2330. doi:10.1016/j.str.2015.10.014
- Nguyen, M. T., Moiani, D., Ahmed, Z., Arvai, A. S., Namjoshi, S., Shin, D. S., et al. (2021). An effective human uracil-DNA glycosylase inhibitor targets the open pre-catalytic active site conformation. *Prog. Biophys. Mol. Biol.* 163, 143–159. doi:10.1016/j.pbiomolbio.2021.02.004
- Noordermeer, S. M., and Van Attikum, H. (2019). PARP inhibitor resistance: a tug-of-war in BRCA-mutated cells. *Trends Cell Biol.* 29, 820–834. doi:10.1016/j.tcb.2019.07.008
- Nussinov, R., and Tsai, C. J. (2013). Allostery in disease and in drug discovery. *Cell* 153, 293–305. doi:10.1016/j.cell.2013.03.034
- O'Connor, M. J. (2015). Targeting the DNA damage response in cancer. *Mol. Cell* 60, 547–560. doi:10.1016/j.molcel.2015.10.040
- Parge, H. E., Arvai, A. S., Murtari, D. J., Reed, S. I., and Tainer, J. A. (1993). Human CksHs2 atomic structure: a role for its hexameric assembly in cell cycle control. *Science* 262, 387–395. doi:10.1126/science.8211159
- Parikh, S. S., Putnam, C. D., and Tainer, J. A. (2000a). Lessons learned from structural results on uracil-DNA glycosylase. *Mutat. Res.* 460, 183–199. doi:10.1016/s0921-8777(00)00026-4
- Parikh, S. S., Walcher, G., Jones, G. D., Slupphaug, G., Krokan, H. E., Blackburn, G. M., et al. (2000b). Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5083–5088. doi:10.1073/pnas.97.10.5083
- Pilie, P. G., Tang, C., Mills, G. B., and Yap, T. A. (2019). State-of-the-art strategies for targeting the DNA damage response in cancer. *Nat. Rev. Clin. Oncol.* 16, 81–104. doi:10.1038/s41571-018-0114-z
- Provasek, V. E., Bacolla, A., Rangaswamy, S., Mitra, J., Kodavati, M., Yusuf, I. O., et al. (2024). RNA/DNA binding protein TDP43 regulates DNA mismatch repair genes with implications for genome stability. bioRxiv.
- Putnam, C. D., Shroyer, M. J., Lundquist, A. J., Mol, C. D., Arvai, A. S., Mosbaugh, D. W., et al. (1999). Protein mimicry of DNA from crystal structures of the uracil-DNA glycosylase inhibitor protein and its complex with *Escherichia coli* uracil-DNA glycosylase. *J. Mol. Biol.* 287, 331–346. doi:10.1006/jmbi.1999.2605
- Revillo Imbernon, J., Chiesa, L., and Kellenberger, E. (2023). Mining the Protein Data Bank to inspire fragment library design. *Front. Chem.* 11, 1089714. doi:10.3389/fchem.2023.1089714
- Rosenberg, D. J., Syed, A., Tainer, J. A., and Hura, G. L. (2022). Monitoring nuclease activity by X-ray scattering interferometry using gold nanoparticle-conjugated DNA. *Methods Mol. Biol.* 2444, 183–205. doi:10.1007/978-1-0716-2063-2_12
- Rotheneder, M., Staktye, K., Van De Logt, E., Bartho, J. D., Lammens, K., Fan, Y., et al. (2023). Cryo-EM structure of the Mre11-Rad50-Nbs1 complex reveals the molecular mechanism of scaffolding functions. *Mol. Cell* 83, 167–185 e9. doi:10.1016/j.molcel.2022.12.003
- Ryan, K., Bolanos, B., Smith, M., Palde, P. B., Cuenca, P. D., Vanarsdale, T. L., et al. (2021). Dissecting the molecular determinants of clinical PARP1 inhibitor selectivity for tankyrase1. *J. Biol. Chem.* 296, 100251. doi:10.1074/jbc.RA120.016573
- Sadybekov, A. V., and Katritch, V. (2023). Computational approaches streamlining drug discovery. *Nature* 616, 673–685. doi:10.1038/s41586-023-05905-z
- Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des.* 27, 221–234. doi:10.1007/s10822-013-9644-8
- Saxena, S., Nabel, C. S., Seay, T. W., Patel, P. S., Kawale, A. S., Crosby, C. R., et al. (2024). Unprocessed genomic uracil as a source of DNA replication stress in cancer cells. *Mol. Cell* 84, 2036–2052.e7. doi:10.1016/j.molcel.2024.04.004
- SCHRODINGER (2017). *Small-molecule drug discovery suite*. 2017-2.
- Scott, D. E., Bayly, A. R., Abell, C., and Skidmore, J. (2016). Small molecules, big targets: drug discovery faces the protein-protein interaction challenge. *Nat. Rev. Drug Discov.* 15, 533–550. doi:10.1038/nrd.2016.29
- Setton, J., Zinda, M., Riaz, N., Durocher, D., Zimmermann, M., Koehler, M., et al. (2021). Synthetic lethality in cancer therapeutics: the next generation. *Cancer Discov.* 11, 1626–1635. doi:10.1158/2159-8290.CD-20-1503

- Shibata, A., Moiani, D., Arvai, A. S., Perry, J., Harding, S. M., Genois, M. M., et al. (2014). DNA double-strand break repair pathway choice is directed by distinct MRE11 nuclease activities. *Mol. Cell* 53, 7–18. doi:10.1016/j.molcel.2013.11.003
- Singh, M., Bacolla, A., Chaudhary, S., Hunt, C. R., Pandita, S., Chauhan, R., et al. (2020). Histone acetyltransferase MOF orchestrates outcomes at the crossroad of oncogenesis, DNA damage response, proliferation, and stem cell development. *Mol. Cell Biol.* 40, 002322–e320. doi:10.1128/MCB.00232-20
- Slupphaug, G., Mol, C. D., Kavli, B., Arvai, A. S., Krokan, H. E., and Tainer, J. A. (1996). A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* 384, 87–92. doi:10.1038/384087a0
- Syed, A., Filandr, F., Patterson-Fortin, J., Bacolla, A., Ravindranathan, R., Zhou, J., et al. (2023). Novobiocin blocks nucleic acid binding to Polθ and inhibits stimulation of its ATPase activity. *Nucleic Acids Res.* 51, 9920–9937. doi:10.1093/nar/gkad727
- Syed, A., and Tainer, J. A. (2018). The MRE11-RAD50-NBS1 complex conducts the orchestration of damage signaling and outcomes to stress in DNA replication and repair. *Annu. Rev. Biochem.* 87, 263–294. doi:10.1146/annurev-biochem-062917-012415
- Tainer, J. A., Getzoff, E. D., Alexander, H., Houghten, R. A., Olson, A. J., Lerner, R. A., et al. (1984). The reactivity of anti-peptide antibodies is a function of the atomic mobility of sites in a protein. *Nature* 312, 127–134. doi:10.1038/312127a0
- Tsuihiji, K., Honda, E., Kojoh, K., Katoh, S., Taguri, T., Yoshimori, A., et al. (2022). Rational strategy for designing peptidomimetic small molecules based on cyclic peptides targeting protein-protein interaction between CTLA-4 and B7-1. *Pharm. (Basel)* 15, 1506. doi:10.3390/ph15121506
- Tsutakawa, S. E., Bacolla, A., Katsonis, P., Bralic, A., Hamdan, S. M., Lichtarge, O., et al. (2021). Decoding cancer variants of unknown significance for helicase-nuclease-RPA complexes orchestrating DNA repair during transcription and replication. *Front. Mol. Biosci.* 8, 791792. doi:10.3389/fmolb.2021.791792
- Tsutakawa, S. E., Classen, S., Chapados, B. R., Arvai, A. S., Finger, L. D., Guenther, G., et al. (2011). Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell* 145, 198–211. doi:10.1016/j.cell.2011.03.004
- Tsutakawa, S. E., Sarker, A. H., Ng, C., Arvai, A. S., Shin, D. S., Shih, B., et al. (2020a). Human XPG nuclease structure, assembly, and activities with insights for neurodegeneration and cancer from pathogenic mutations. *Proc. Natl. Acad. Sci. U. S. A.* 117, 14127–14138. doi:10.1073/pnas.1921311117
- Tsutakawa, S. E., Shin, D. S., Mol, C. D., Izumi, T., Arvai, A. S., Mantha, A. K., et al. (2013). Conserved structural chemistry for incision activity in structurally non-homologous apurinic/apyrimidinic endonuclease APE1 and endonuclease IV DNA repair enzymes. *J. Biol. Chem.* 288, 8445–8455. doi:10.1074/jbc.M112.422774
- Tsutakawa, S. E., Thompson, M. J., Arvai, A. S., Neil, A. J., Shaw, S. J., Algasai, S. I., et al. (2017). Phosphate steering by Flap Endonuclease 1 promotes 5'-flap specificity and incision to prevent genome instability. *Nat. Commun.* 8, 15855. doi:10.1038/ncomms15855
- Tsutakawa, S. E., Tsai, C. L., Yan, C., Bralic, A., Chazin, W. J., Hamdan, S. M., et al. (2020b). Envisioning how the prototypic molecular machine TFIIH functions in transcription initiation and DNA repair. *DNA Repair (Amst)* 96, 102972. doi:10.1016/j.dnarep.2020.102972
- Tubbs, J. L., Latypov, V., Kanugula, S., Butt, A., Melikishvili, M., Kraehenbuehl, R., et al. (2009). Flipping of alkylated DNA damage bridges base and nucleotide excision repair. *Nature* 459, 808–813. doi:10.1038/nature08076
- Voutsadakis, I. A., and Digkila, A. (2023). The landscape and therapeutic targeting of BRCA1, BRCA2 and other DNA damage response genes in pancreatic cancer. *Curr. Issues Mol. Biol.* 45, 2105–2120. doi:10.3390/cimb45030135
- Wang, M., Chen, S., and Ao, D. (2021). Targeting DNA repair pathway in cancer: mechanisms and clinical application. *MedComm* 2, 654–691. doi:10.1002/mco2.103
- Weth, F. R., Hoggarth, G. B., Weth, A. F., Paterson, E., White, M. P. J., Tan, S. T., et al. (2024). Unlocking hidden potential: advancements, approaches, and obstacles in repurposing drugs for cancer therapy. *Br. J. Cancer* 130, 703–715. doi:10.1038/s41416-023-02502-9
- Williams, R. S., Moncalian, G., Williams, J. S., Yamada, Y., Limbo, O., Shin, D. S., et al. (2008). Mre11 dimers coordinate DNA end bridging and nuclease processing in double-strand-break repair. *Cell* 135, 97–109. doi:10.1016/j.cell.2008.08.017
- Wilson, D. M., Deacon, A. M., Duncton, M. A. J., Pellicena, P., Georgiadis, M. M., Yeh, A. P., et al. (2021). Fragment- and structure-based drug discovery for developing therapeutic agents targeting the DNA Damage Response. *Prog. Biophys. Mol. Biol.* 163, 130–142. doi:10.1016/j.pbiomolbio.2020.10.005
- Yan, C., Dodd, T., He, Y., Tainer, J. A., Tsutakawa, S. E., and Ivanov, I. (2019). Transcription preinitiation complex structure and dynamics provide insight into genetic diseases. *Nat. Struct. Mol. Biol.* 26, 397–406. doi:10.1038/s41594-019-0220-3
- Yang, T., Huang, W., Ma, T., Yin, X., Zhang, J., Huo, M., et al. (2023). The PRMT6/PARP1/CRL4B complex regulates the circadian clock and promotes breast tumorigenesis. *Adv. Sci. (Weinh)* 10, e2202737. doi:10.1002/adv.202202737
- Yang, Y., Yao, K., Repasky, M. P., Leswing, K., Abel, R., Shoichet, B. K., et al. (2021). Efficient exploration of chemical space with docking and deep learning. *J. Chem. Theory Comput.* 17, 7106–7119. doi:10.1021/acs.jctc.1c00810
- Ye, Z., Shi, Y., Lees-Miller, S. P., and Tainer, J. A. (2021a). Function and molecular mechanism of the DNA damage response in immunity and cancer immunotherapy. *Front. Immunol.* 12, 797880. doi:10.3389/fimmu.2021.797880
- Ye, Z., Xu, S., Shi, Y., Bacolla, A., Syed, A., Moiani, D., et al. (2021b). GRB2 enforces homology-directed repair initiation by MRE11. *Sci. Adv.* 7, eabe9254. doi:10.1126/sciadv.abe9254
- Ye, Z., Xu, S., Shi, Y., Cheng, X., Zhang, Y., Roy, S., et al. (2024). GRB2 stabilizes RAD51 at reversed replication forks suppressing genomic instability and innate immunity against cancer. *Nat. Commun.* 15, 2132. doi:10.1038/s41467-024-46283-y
- Yuan, X., Duan, Y., Xiao, Y., Sun, K., Qi, Y., Zhang, Y., et al. (2022). Vitamin E enhances cancer immunotherapy by reinvigorating dendritic cells via targeting checkpoint SHP1. *Cancer Discov.* 12, 1742–1759. doi:10.1158/2159-8290.CD-21-0900
- Yu, J., Yan, C., Dodd, T., Tsai, C. L., Tainer, J. A., Tsutakawa, S. E., et al. (2023). Dynamic conformational switching underlies TFIIH function in transcription and DNA repair and impacts genetic diseases. *Nat. Commun.* 14, 2758. doi:10.1038/s41467-023-38416-6
- Zandarashvili, L., Langelier, M. F., Velagapudi, U. K., Hancock, M. A., Steffen, J. D., Billur, R., et al. (2020). Structural basis for allosteric PARP-1 retention on DNA breaks. *Science* 368, eaax6367. doi:10.1126/science.aax6367
- Zhang, X., Blundell, T., and Tainer, J. A. (2021). The renaissance in biophysics and molecular biology enabled by the interface of DNA repair and replication with cancer. *Prog. Biophys. Mol. Biol.* 163, 1–4. doi:10.1016/j.pbiomolbio.2021.03.006
- Zhao, H. (2024). The science and art of structure-based virtual screening. *ACS Med. Chem. Lett.* 15, 436–440. doi:10.1021/acsmedchemlett.4c00093
- Zhou, J., Gelot, C., Pantelidou, C., Li, A., Yucel, H., Davis, R. E., et al. (2021). A first-in-class polymerase theta inhibitor selectively targets homologous-recombination-deficient tumors. *Nat. Cancer* 2, 598–610. doi:10.1038/s43018-021-00203-x
- Zhu, J., Wang, J., Wang, X., Gao, M., Guo, B., Gao, M., et al. (2021). Prediction of drug efficacy from transcriptional profiles with deep learning. *Nat. Biotechnol.* 39, 1444–1452. doi:10.1038/s41587-021-00946-z