Computational Pipeline for the Discovery of Potential Inhibitors of PARP1

Alberto González Calatayud

1 INTRODUCTION

Oncological research has extensively studied the molecular dynamics of cancer cell perseverance, particularly those governing DNA repair. A pivotal player in this domain is Poly (ADP-ribose) polymerase 1 (PARP1), an enzyme essential for repairing DNA damage. Together with PARP2, PARP1 is a critical agent in cancer-related DNA repair activities (Ko and Ren, 2012),(Schiewer et al., 2012).

This protein is associated with various types of cancer, most commonly with BRCA1 and BRCA2 positive breast and ovarian cancers, some types of prostate and pancreatic cancer (Schiewer et al., 2012).

The expression of PARP1 is upregulated in various cancers, reflecting its sensitivity to DNA-damaging agents and underscoring its value as a therapeutic target (Mateo et al., 2019), (Graziani and Szabó, 2005).

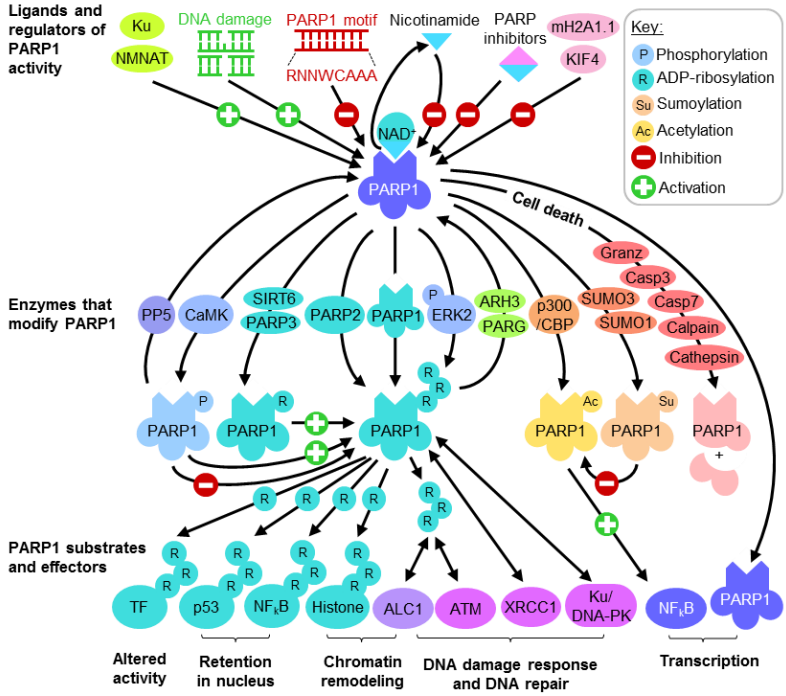


Figure 1: Function and regulation of PARP1

As can be seen in Figure1 from the Article entitled "Functional Aspects of PARP1 in DNA Repair and Transcription" by Yi et al., 2019, the upregulation of this enzyme does not only boost the DNA repair capacity of cancer cells but also engages in a multitude of actions that, in mutated (cancerous) cells, contribute to their malignancy. Beyond maintaining genomic stability, PARP1 also regulates the transcription of genes implicated in the DNA damage response, affirming its multifaceted role in oncology.

Therefore, PARP1 inhibition has emerged as a strategy to induce synthetic lethality in some of these cancers.
These drugs are designed to exploit deficiencies in the DNA repair of cancer cells, causing their death while minimizing damage to normal cells (Yi et al., 2019), (Zuo et al., 2021).

The study's goal of this project is to establish a computational pipeline utilizing three screening techniques: Ligand-based virtual screening, Quantitative Structure-Activity Relationship (QSAR), and affinity docking. This triad aims to identify potential inhibitory compounds for the PARP1 protein, which is the one that has the greatest implication in the mechanism of action of these proteins (Kutuzov et al., 2021).

2 METHODS

2.1 Ligand Based Virtual Screening

Initially, a search was conducted in various databases for the target protein PARP1, with UniProt code: P09874 and CHEMBL3105. Once identified, the Python client for accessing the ChEMBL RESTful API was used to find compounds with inhibitory interaction capacity with PARP1.
Subsequently, to choose the query molecule for the study, different approved drugs (such as Olaparib, with an IC50 of 0.6 nM) as well as the molecule with the lowest IC50 from the database were examined. Their structures were observed in ChEMBL web, and due to their similarity, it was decided to use the molecule CHEMBL4169012, which has an IC50 value of 0.02 nM, as the query molecule for virtual screening.
This decision aimed to identify potential PARP1 inhibitory compounds, starting with a query molecule with high inhibitory capacity and characteristics similar to approved drugs. These compounds adhered to Lipinski's Rule of Five, signifying favorable drug-like properties.

Both 2D and 3D similarity searches were conducted With the PubChem API using different thresholds: a Tanimoto threshold of 0.8 for 2D similarity and a more stringent Tanimoto threshold of 0.95 for 3D similarity.

To ensure the integrity of the screening process, it has been incorporated the query molecule as a standard/positive control and octanoic acid as the negative control due to its structural dissimilarity.

Using Python's RDKit library, the analysis involved the structural analysis of the query molecule, which included the computation of Morgan and MACCS fingerprints. Subsequently, the compounds underwent further evaluation, employing Morgan and MACCS molecular fingerprints, as well as Tanimoto and Dice similarity coefficients.

After a thorough examination of the results obtained from each similarity metric, the Dice MACCS coefficient was chosen as it effectively prioritized compounds with structural similarities to the query molecule. A threshold of 0.91 was applied for this selection.

2.2 QSAR

A Quantitative Structure-Activity Relationship (QSAR) model was developed, trained with a ChEMBL database subset of compounds interacting with the target protein and exhibiting IC50 values. Several models were tested, and the best-performing one, determined through thorough evaluation of cross-validation results, was trained using a random forest algorithm.

This model was then used to evaluate the compounds selected from the Ligand-Based Virtual Screening, predicting their IC50 values. A cutoff of 1.5 nM was applied to filter for the next phase, focusing on those compounds with the highest inhibitory potential and lowering the number of compounds to manage computational complexity in the docking stage.

2.3 ADMET Properties

Following the QSAR analysis, the compounds selected for further investigation underwent an evaluation of their ADMET properties. The aim was to identify compounds that met the criteria outlined by Lipinski's Rule of Five. Only those compounds that satisfied these criteria were retained for the subsequent docking studies. This stringent filtering ensured that the selected compounds exhibited favorable pharmacokinetic and pharmacodynamic properties, enhancing their potential as promising candidates for further exploration.

2.4 Protein-Ligand Docking

The final analysis step involved protein-ligand docking to predict the complex formed by the target protein, PARP1, and the candidate inhibitory compounds. This aimed to evaluate the compounds' affinity energy and interaction capabilities with PARP1.

To facilitate this, the PARP1 catalytic domain complexed with Rucaparib was obtained from the Protein Data Bank (PDB). Preprocessing using Chimera identified the center ligand for orientation and removed crystallized ligands (Rucaparib and sulfate ion).

In Chimera, the measured center ligand was observed to have coordinates at (91.23, -1.28, 122.19). Notably, SeamDock software imposes constraints on coordinate values, permitting only those within the range of -50 to 50. Consequently, to align with SeamDock's requirements, the x and z coordinates were set to 0 for consistency, and a uniform size of 30 was applied to all compounds.

Docking was performed with SeamDock software using the SMILES notations of the candidate compounds and the preprocessed target protein. Initially, all four protein chains (A, B, C, and D) were considered, but computational limitations led to a focus on chains A and B. Docking was performed using chains A and B, yielding improved results in the latter case.

3 RESULTS

In the pursuit of compounds capable of interacting with and inhibiting the target protein Poly (ADP-ribose) polymerase 1 (PARP1), an initial set of 3096 compounds from the CHEMBL database was identified.

As outlined in the methods section, the choice of the query molecule for Ligand-Based Virtual Screening (LBVS) centered on CHEMBL4169012, distinguished by its lowest IC50 value.
This compound showed potent inhibitory activity, adhered to Lipinski's rules, and structurally resembled approved drugs. The strategy involved selecting a query molecule for LBVS that closely resembled established compounds, meeting criteria, and aiming for more potent inhibitors.

Following this, a PubChem **similarity search**, as elaborated in the 2, produced 7985 compounds. Among these, 718 surpassed the Dice similarity coefficient threshold of 0.91 with the MACCS fingerprints, advancing to the subsequent stage.

These 718 compounds underwent IC50 value prediction using the trained **QSAR** model. Only those with predicted IC50 values below 1.5 nM were retained, resulting in a final selection of 13 molecules. The lowest predicted IC50 value among these compounds was 0.514 nM.

For the query molecule (which was removed from the training dataset to be able to evaluate it) a significant difference was obtained between the predicted (2) and actual (0.02) IC50

values. However, an IC50 value of 2 is still considered favorable. This deviation is primarily due to the limited presence of very low IC50 values in the training dataset, while most predictions aligned closely with actual values during model testing.

Of these 13 compounds, adherence to **ADMET** properties was evaluated, leading to the retention of only 7 compounds. The most prevalent "limiting" factors among these compounds, approaching the violation of the Rule of Five (RO5), were molecular weight and log P value. This implies that the candidates selected for docking exhibit favorable pharmacokinetic properties.

The initial **docking** was performed with A and B chains, but subsequently, it was decided to retain only chain A of the PARP1 protein, as it exhibited the most favorable affinity with the potential molecules. This selection streamlined the analysis and ensured a more targeted examination of interaction possibilities.

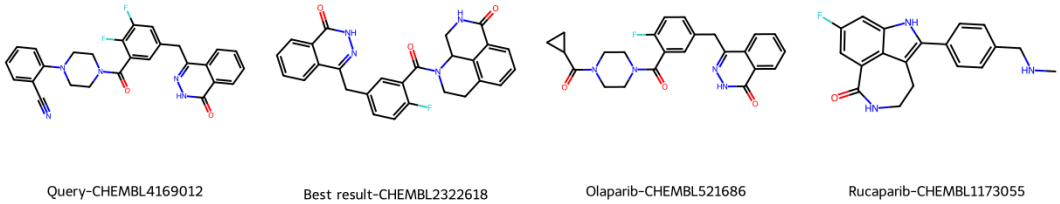
Table 1: Docking Results for Candidate Compounds and query molecule

Compound CID	**Affinity (kcal/mol)**	**Compound CID**	**Affinity (kcal/mol)**
71606888	-12.4	59535475	-11.0
154276338	-11.2	145960349	-11.4
145952366	-11.6	68498505	-11.1
59535477	-10.9	query	-12

As observed in Table 1, all evaluated compounds have demonstrated notably high affinity towards the Poly (ADP-ribose) polymerase 1 (PARP1) protein, with all recorded affinities falling below -10 kcal/mol. This high affinity suggests that these compounds hold significant potential as candidates for inhibiting the activity of the PARP1 protein, particularly in its chain A. Docking of the query molecule was also carried out to compare the result, since its structure is very similar to the candidate molecules and it has a very low real IC50 and a very good affinity was observed.

4 DISCUSSION

Among the ultimate candidates, CID:71606888 exhibited the highest affinity in the docking simulations, making it a strong contender as a PARP1 inhibitor.
The initial hypothesis, which involved selecting a query molecule with the highest inhibitory potency, despite not being an approved drug, proved successful. The chosen query molecule, CHEMBL4169012, demonstrated both structural similarity to approved drugs like Olaparib and Rucaparib (known to act on PARP1 and PARP2) and adherence to Lipinski's rules. This strategic approach resulted in the discovery of compounds with significant inhibitory potential.
It is important to highlight that these compounds exhibit their highest affinity for the A chain of PARP1. PARP1 is composed of four chains, with the A and B chains being the most crucial for its mechanism of action (Kutuzov et al., 2021), (Lin et al., 2022). These chains bear structural similarities to PARP2, leading to the expectation that these compounds could also influence PARP2. However, the focus was intentionally placed on the A chain of PARP1, given its significant role in the protein's mechanism and its higher abundance. In fact, PARP1 is implicated in approximately 80% of DNA damage-induced poly ADP-ribosylation in mammalian cells, making it a compelling target (Lin et al., 2022), (Zuo et al., 2021).



(a) Query Molecule (b) Best Result (c) Olaparib (d) Rucatinib

Figure 2: Chemical Structures

In terms of challenges and potential improvements, enhancing computational complexity to include shape similarity-based filtering for exploring diverse compounds is one avenue. Adjusting thresholds or emphasizing different characteristics during screening could yield compounds with unique inhibitory actions, offering potential for alternative drug formulations. Further refinement of the QSAR model for improved prediction accuracy is also a promising area for enhancement.
Hence, in future applications of this approach, it could be explored compounds with greater dissimilarity from the standard inhibitors of this protein. This could lead to the discovery of novel drug candidates, expanding treatment options and possibilities.

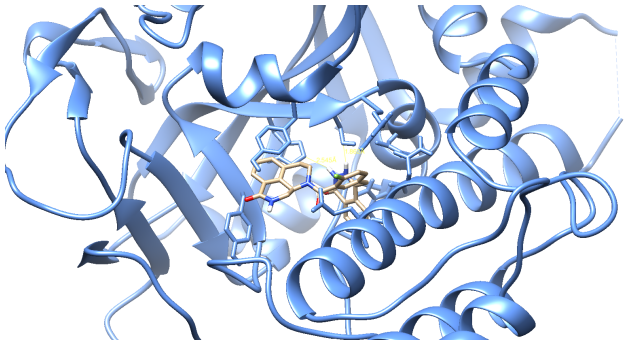


Figure 3: Interaction of the best compound with chain A of the PARP1 protein in Chimera.

References

Graziani, G., & Szabó, C. (2005). Clinical perspectives of parp inhibitors. *Pharmacological Research*, *52*(1), 109–118. https://doi.org/10.1016/j.phrs.2005.02.013 (cit. on p. 1).

Ko, H. L., & Ren, E. C. (2012). Functional aspects of parp1 in dna repair and transcription. *Biomolecules*, *2*(4), 524–548. https://doi.org/10.3390/biom2040524 (cit. on p. 1).

Kutuzov, M. M., Belousova, E. A., Kurgina, T. A., Ukraintsev, A. A., Vasil'eva, I. A., Khodyreva, S. N., & Lavrik, O. I. (2021). The contribution of parp1, parp2 and poly(adp-ribosyl)ation to base excision repair in the nucleosomal context. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-84351-1 (cit. on pp. 2, 4).

Lin, X., Jiang, W., Rudolph, J., Lee, B. J., Luger, K., & Zha, S. (2022). Parp inhibitors trap parp2 and alter the mode of recruitment of parp2 at dna damage sites. *Nucleic Acids Research*, *50*(7), 3958–3973. https://doi.org/10.1093/nar/gkac188 (cit. on p. 4).

Mateo, J., Lord, C., Serra, V., Tutt, A., Balmaña, J., Castroviejo-Bermejo, M., Cruz, C., Oaknin, A., Kaye, S., & de Bono, J. (2019). A decade of clinical development of parp inhibitors in perspective. *Annals of Oncology*, *30*(9), 1437–1447. https://doi.org/10.1093/annonc/mdz192 (cit. on p. 1).

Schiewer, M. J., Goodwin, J. F., Han, S., Brenner, J. C., Augello, M. A., Dean, J. L., Liu, F., Planck, J. L., Ravindranathan, P., Chinnaiyan, A. M., McCue, P., Gomella, L. G., Raj, G. V., Dicker, A. P., Brody, J. R., Pascal, J. M., Centenera, M. M., Butler, L. M., Tilley, W. D., ... Knudsen, K. E. (2012). Dual roles of parp-1 promote cancer growth and progression. *Cancer Discovery*, *2*(12), 1134–1149. https://doi.org/10.1158/2159-8290.cd-12-0120 (cit. on p. 1).

Yi, M., Dong, B., Qin, S., Chu, Q., Wu, K., & Luo, S. (2019). Advances and perspectives of parp inhibitors. *Experimental Hematology amp; Oncology*, *8*(1). https://doi.org/10.1186/s40164-019-0154-9 (cit. on p. 1).

Zuo, X., Zhao, H., & Li, D. (2021). Systematic inhibitor selectivity between parp1 and parp2 enzymes: Molecular implications for ovarian cancer personalized therapy. *Journal of Molecular Recognition*, *34*(7). https://doi.org/10.1002/jmr.2891 (cit. on pp. 1, 4).