



**University of Dundee**

## **Accurate prediction of dynamic protein-ligand binding using P-score ranking**

E. G. F. Ibrahim, Peter; Zuccotto, Fabio; Zachariae, Ulrich; Gilbert, Ian; Bodkin, Mike

*Published in:*  
Journal of Computational Chemistry

*DOI:*  
[10.1002/jcc.27370](https://doi.org/10.1002/jcc.27370)

*Publication date:*  
2024

*Licence:*  
CC BY

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication in Discovery Research Portal](#)

*Citation for published version (APA):*  
E. G. F. Ibrahim, P., Zuccotto, F., Zachariae, U., Gilbert, I., & Bodkin, M. (2024). Accurate prediction of dynamic protein-ligand binding using P-score ranking. *Journal of Computational Chemistry*, 45(20), 1762-1778.  
<https://doi.org/10.1002/jcc.27370>

### **General rights**

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Accurate prediction of dynamic protein–ligand binding using *P*-score ranking

Peter E. G. F. Ibrahim  | Fabio Zuccotto | Ulrich Zachariae | Ian Gilbert | Mike Bodkin

Drug Discovery Unit, Division of Biological Chemistry and Drug Discovery, University of Dundee, Dundee, UK

## Correspondence

Mike Bodkin, Drug Discovery Unit, Division of Biological Chemistry and Drug Discovery, University of Dundee, Dow St, Dundee DD1 5EH, UK.  
Email: [mbodkin001@dundee.ac.uk](mailto:mbodkin001@dundee.ac.uk)

## Funding information

Welcome Centre for Anti-Infectives Research

## Abstract

Protein–ligand binding prediction typically relies on docking methodologies and associated scoring functions to propose the binding mode of a ligand in a biological target. Significant challenges are associated with this approach, including the flexibility of the protein–ligand system, solvent-mediated interactions, and associated entropy changes. In addition, scoring functions are only weakly accurate due to the short time required for calculating enthalpic and entropic binding interactions. The workflow described here attempts to address these limitations by combining supervised molecular dynamics with dynamical averaging quantum mechanics fragment molecular orbital. This combination significantly increased the ability to predict the experimental binding structure of protein–ligand complexes independent from the starting position of the ligands or the binding site conformation. We found that the predictive power could be enhanced by combining the residence time and interaction energies as descriptors in a novel scoring function named the *P*-score. This is illustrated using six different protein–ligand targets as case studies.

## KEYWORDS

binding pose prediction, dynamic average quantum mechanics fragment molecular orbital, *P*-score, supervised molecular dynamics

## 1 | INTRODUCTION

In the absence of any crystallographic data, the accurate prediction of a ligand's binding mode can be critical for success in structure-based drug discovery.<sup>1</sup> Likewise, structure based virtual screening methods, which are widely used for the identification of chemical starting points, rely heavily on accurate pose prediction and scoring.<sup>2–5</sup>

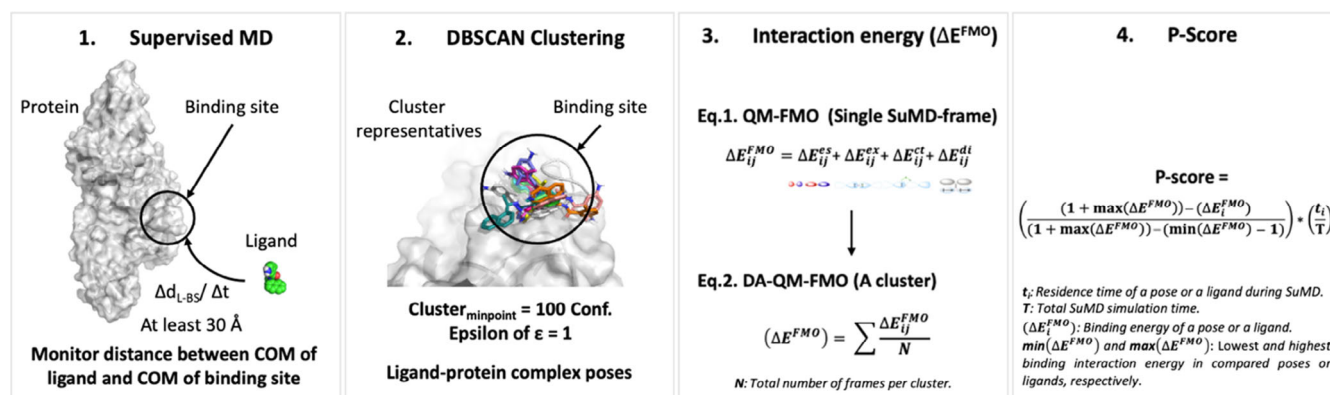
There is widespread application of protein–ligand docking algorithms in molecular design, yet they possess significant drawbacks. To enable docking at speed a variety of approximations are made to derive a scoring function which albeit renders the scores almost qualitative. The scoring system rarely accounts for the conformational

flexibility of either the protein or the ligand associated with induced-fit or conformational-selection binding events. Consequently, water and or ion solvation together with the entropic changes associated on ligand binding are poorly handled. Hence, small changes in the binding site conformation or entropic solvation can lead to drastically different docking results.<sup>6–12</sup>

To address these limitations, molecular mechanics with Poisson-Boltzmann and surface area (MM-PBSA) or molecular mechanics with generalized Born and surface area (MM-GBSA) approaches were introduced to estimate the binding energies, combining molecular mechanics with estimates of solvation energies. These methods were introduced in the late 1990s by the Kollman and Case labs, aiming to

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Journal of Computational Chemistry* published by Wiley Periodicals LLC.



**FIGURE 1** Schematic diagram of the binding pose prediction and scoring, the *P*-score (SuMD X DA-QM-FMO → *P*-score).  $\Delta E$ , interaction energy; COM, centre of mass;  $t$ , residence time.

estimate the binding free energies, or relative binding free energies, of related compounds and protein–ligand binding modes.<sup>13,14</sup>

Despite their potential, MM-PBSA/GBSA have several limitations which include the need for prior knowledge of a protein–ligand bound complex to serve as a starting point, although starting conformations can be taken from prior docked poses. Their accuracy for binding poses prediction and ranking the order of compounds according to their affinities is very system dependent.<sup>15,16</sup>

Recent developments in machine and deep learning approaches have led to a possible paradigm shift in the way protein ligand docking solutions are generated. Programs such as LiGANN, DiffDock, or Pocket2Mol generate both molecule and pose simultaneously.<sup>17–19</sup> However, tools such as PoseBusters or PoseCheck which were developed to assess the geometric and interaction quality of generated ligands have highlighted there is significant room for improvement compared to solutions derived from physics-based approaches.<sup>20,21</sup>

To overcome the issues outlined above, we introduce a novel protocol (Figure 1), which combines supervised molecular dynamics (SuMD) with dynamical averaging of quantum mechanics fragment molecular orbital (DA-QM-FMO). SuMD allows a dynamical sampling of a protein–ligand system in a fully solvated environment with protein flexibility and ligand conformational sampling. DA-QM-FMO estimates the average potential binding interaction energies from dynamically generated protein–ligands binding modes. In this study, DA-QM-FMO was able to outperform simple docking or MM-PBSA/GBSA at predicting the experimental binding pose.

We believe limitations due to the length of the SuMD simulations and the approximations made in energy calculations techniques, accounts for the fact that these calculations on their own did not fully replicate the experimental data. However, by combining the residence time and binding interaction energy the method was able to improve the predictive power. We refer to the combination of descriptors as the *P*-score (Figure 1,4). The *P*-score prioritizes protein–ligand binding poses based on stability and interaction energy of binding modes. We apply the *P*-score theory using six different case studies, with published crystal structures from the Protein Data Bank (PDB); Papain-like protease protein (PL-pro) and Main-protease (Mpro) for

SARS-CoV-2, Heat shock protein 90 (Hsp90), p38 Kinase (p38), Myeloid Cell Leukemia 1 (Mcl-1), and *Pseudomonas. aeruginosa* LpxC (LpxC).

## 2 | MATERIALS AND METHODS

### 2.1 | Protein–ligand sampling phase

SuMD was used to allow a whole ensemble sampling phase between the ligand and its intended target independent from the initial ligand position either in solvent or within the binding site vicinity. SuMD is a novel molecular dynamics approach developed by Moro et al.<sup>22,23</sup> that enables the investigation of the molecular recognition pathway between ligand and receptor. SuMD overcomes the conventional constraints associated with the docking methods in protein–ligand binding pose generation.<sup>24,25</sup>

#### 2.1.1 | System preparation

Protein structures for five systems; Pl-pro, Mpro, p38, Hsp90, and Mcl-1, were prepared with the protein preparation wizard as implemented in Maestro.<sup>26</sup> Hydrogen atoms were added to the complex and missing atoms in protein side chains were built according to the AMBER16 force field topology.<sup>27</sup> The N- and C-terminus were capped. LpxC system was built using Metal Center Parameter Builder (MCPB) from Ambertools for the  $Zn^{+2}$  binding site using Zinc AMBER force field (ZAFF) for 4-coordinated zinc metal centers.<sup>28</sup>

#### 2.1.2 | Ligand parameterization

All ligands were prepared and treated under physiological conditions of pH = 7.4, throughout the present study. For the MD simulations, the AMBER16 force field was used. Geometry minimization, and ligand parameters were derived with GAFF, as implemented in

Ambertools, by using Antechamber and PARMCHK tools. Partial charges were calculated following the procedure suggested by antechamber.<sup>29,30</sup>

### 2.1.3 | Solvated system setup and equilibration

Protein–ligand complexes were assembled with the tleap tool using AMBER16SB, as the force field for the proteins.<sup>31</sup> The systems were explicitly solvated by a cubic water box with cell borders placed at least 13 Å away from any protein or ligand atom using TIP3P as the water model.<sup>32</sup> To neutralize the total charge, Na<sup>+</sup>/Cl<sup>−</sup> counterions were added to a final salt concentration of 0.15 M. The systems were energy minimized for 1000 steps with the conjugate-gradient method and then 50,000 steps of NVE (100 ps) ensemble (micro-canonical ensemble; constant number of particles [N], constant volume [V], constant energy [E]) followed by 1 ns of NPT (isothermal–isobaric ensemble; constant number of particles [N], constant pressure [P], constant temperature [T]) simulation, both using a time step of 2 fs and applying harmonic position constraints on protein and ligand atoms, which were gradually reduced with a scaling factor of 0.1. Pressure was maintained at 1 atm using a Berendsen barostat. The Langevin thermostat was set with a low damping constant of 1 ps<sup>−1</sup>. Bond lengths involving hydrogen atoms were constrained using the M-SHAKE algorithm. The MD production runs were conducted in a NVT ensemble. Long-range Coulomb interactions were treated using the particle mesh Ewald summation method (PME) setting the mesh spacing to 1.0 Å. A nonbonded cutoff distance of 9 Å with a switching distance of 7.5 Å was used.<sup>33–36</sup>

## 2.2 | Protein–ligand binding poses clustering using DBSCAN

After concatenating SuMD trajectories, they were subjected to geometric distinction using DBSCAN. This process helped in identifying clusters of similar conformations in the combined trajectory data. DBSCAN is a density-based clustering algorithm that allows the most populated ligand–protein states to be identified and distinguished from the background noise.<sup>37</sup> Each cluster (protein–ligand binding mode) is composed of a collection of MD snapshots “MD frame.” Binding poses are then evaluated using the energy calculation approaches (MM-PBSA, QM-FMO, and DA-QM-FMO) to estimate the protein–ligand complex binding energy and interaction energy.

## 2.3 | Energy calculation

To evaluate the binding energies from the sampling and clustering phases, we make a comparison between MM-PBSA, Quantum Mechanics fragment molecular orbital (QM-FMO) method and Dynamical averaging of QM-FMO (DA-QM-FMO). For the MM-PBSA, molecular mechanics energies were calculated along with

generalized Born and surface area continuum solvation using the MMPBSA.py scripts from the AmberTools22 suite (<https://ambermd.org/>).<sup>38</sup> QM-FMO is used for calculating the binding interaction energies described by the inter-fragment interaction energy (IFIE) and its pair interaction energy decomposition analysis (PIEDA).<sup>39,40</sup>

Inter-fragment interaction energy (IFIE) analysis estimates protein–ligand interactions with quantum (electronic) effects using a modest level of computational resources. Furthermore, by means of pair interaction energy decomposition analysis (PIEDA), the FMO interaction energy,  $\Delta E^{\text{FMO}}$ , is calculated as the sum of five energy terms: electrostatic ( $\Delta E^{\text{ES}}$ ), exchange repulsion ( $\Delta E^{\text{EX}}$ ), dispersion ( $\Delta E^{\text{DI}}$ ), charge transfer with higher-order mixed terms ( $\Delta E^{\text{CT+mix}}$ ), and solvation energy, as shown in Equation (1). This approach offers a detailed estimation of protein–ligand interactions in an implicit water, polarizable continuum model (PCM) that has been of great value in a range of drug discovery projects.

$$\Delta E^{\text{FMO}} = \Delta E_{ij}^{\text{es}} + \Delta E_{ij}^{\text{ex}} + \Delta E_{ij}^{\text{ct+mix}} + \Delta E_{ij}^{\text{di}} + \Delta E^{\text{Gsol}}, \quad (1)$$

$\Delta E^{\text{FMO}}$ ; interaction energy of a single snapshot (MD frame).

Dynamical averaging of QM-FMO allows an estimation of the total average interaction energy ( $\Delta E^{\text{FMO}}$ ) of the protein–ligand complex along the MD trajectory instead of relying only on the interaction energy  $\Delta E^{\text{FMO}}$  for a single snapshot (cluster representative).<sup>41</sup>

The total average interaction energy ( $\Delta E^{\text{FMO}}$ ) (DA-QM-FMO) of a binding mode is obtained by calculating the average interaction energy  $\Delta E^{\text{FMO}}$  of all snapshots found per cluster divided by the number of snapshots within the cluster (MD frames) as shown in Equation (2).<sup>41</sup>

$$\left(\Delta E^{\text{FMO}}\right) = \sum \frac{\Delta E_i^{\text{FMO}}}{N}, \quad (2)$$

( $\Delta E^{\text{FMO}}$ ) DA-QM-FMO represents the average  $\Delta E^{\text{FMO}}$  per cluster;  $i$  and  $N$  denote the indices of the MD snapshots and the total number of MD snapshots per cluster, respectively.

Prior to the DA-QM-FMO calculations, all snapshots were treated with a restrained minimization procedure with the OPLS3e force field to alleviate steric clashes, following the protein preparation wizard protocol of Schrödinger's software suite.<sup>42</sup> The heavy atoms of the protein were permitted to move up to 0.3 Å from their original position in the crystal structure, while the heavy atoms of the ligand and water molecules were restrained in position with a force of 1.0 kcal/mol/Å; hydrogen atoms were left unrestrained. The restraint weight was set to ensure that the binding poses of the ligands did not deviate far from their initial positions.

Protein residues within a radius of 5 Å from the ligand atoms were included in the FMO calculations. The C-terminal carboxylic acid of the peptide was capped with N-methylamine, and the N-terminal position acetylated while maintaining the geometry of the neighboring residues. A given amino acid, along with its side chain, C-alpha, and backbone NH, and the carbonyl group of the adjacent amino acid, define each FMO fragment. Fragmentation was carried out according

to a well-established fragmentation strategy inspired from Facio,<sup>43</sup> in a fully automated high-throughput python code. The calculations were performed at second order Møller–Plesset perturbation theory (FMO2-MP2/6-31G\*) and density function tight-binding (FMO2-DFTB) theory level, using GAMESS implementation.<sup>44,45</sup>

All calculations were performed on a hybrid CPU/GPU cluster. SuMD simulations were carried out with the ACEMD engine, on a GPU cluster equipped with 16 GPUs (2 NVIDIA GTX 1080 per node on 8 GPUs).<sup>46</sup>

## 2.4 | *P*-score theory for protein–ligand binding pose scoring

To improve the predictive power for ranking binding poses (cluster representatives), we introduce the *P*-score as a novel scoring function to eliminate false positives in binding pose ranking. The rationale behind the *P*-score for protein–ligand binding prediction is that a ligand binding mode of interest is likely to show a lengthened residence time and high degree of affinity in a biological target binding site. In-silico this is translated into having a prolonged ligand pose stability during the SuMD simulations and a low binding energy or interaction energy ( $\Delta E$ ) with binding site residues. Therefore, the normalization of time ( $t$ ) (stability) and  $\Delta E$  (affinity) can be considered as key descriptors for each binding mode or active hit among a virtual screening library. A binding mode can be distinguished using the *P*-score diagram represented in a quadrant matrix, (Figure 2).

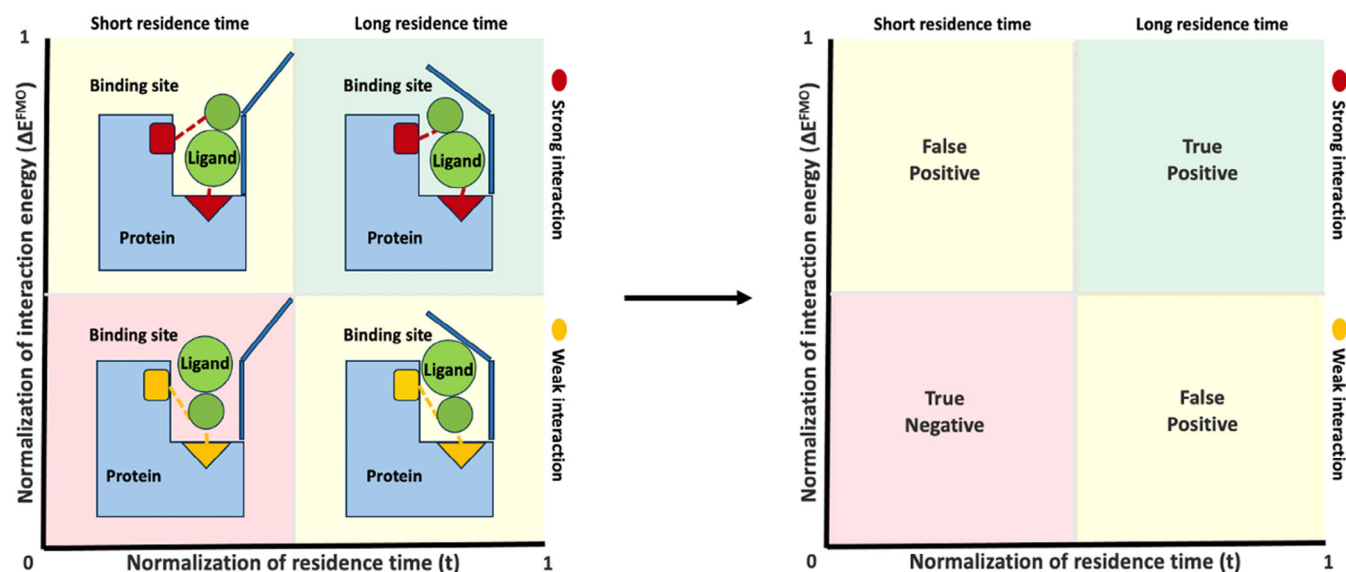
The *P*-score defines a protein–ligand bound pose is a result of dynamic induced-fit binding event, that relies on the residence time ( $t$ )

and total average interaction energy ( $\Delta E^{\text{FMO}}$ ) for each binding mode or hit, using the following Equation (3):

$$P\text{-score} = \left( \frac{(1 + \max(\Delta E^{\text{FMO}})) - (\Delta E_i^{\text{FMO}})}{(1 + \max(\Delta E^{\text{FMO}})) - (\min(\Delta E^{\text{FMO}}) - 1)} \right) \times \left( \frac{t_i}{T} \right), \quad (3)$$

*P*-score equation:  $t_i$  is the residence time of a pose or a ligand.  $T$  is the total SuMD simulation time.  $\Delta E_i$  is the interaction energy of a pose or a ligand.  $\min(\Delta E)$  and  $\max(\Delta E)$  are the lowest and highest interaction energy, in compared poses or ligands, respectively.

The *P*-score equation is inspired from the rescaling normalization equation where the range of all descriptors should be normalized so that each descriptor contributes proportionately to a final score.<sup>47</sup> The *P*-score gives an evaluation of each binding mode by the normalization of the ensemble of all descriptors; time ( $t$ ) and average interaction energy ( $\Delta E^{\text{FMO}}$ ), rather than using a single descriptor (residence time [ $t$ ] or interaction energy [ $\Delta E^{\text{FMO}}$ ]). The *P*-score gives a score between 0 and 1 which is the normalization range of all descriptors included. A *P*-score value towards 1 means that this binding mode is highly prioritized as a stable conformation of the protein–ligand complex, with a long residence time ( $t$ ) and a low interaction energy ( $\Delta E^{\text{FMO}}$ ) (True positive). A pose with a *P*-score value towards 0 is likely to be the least favorable binding mode for the protein–ligand complex, which can be a result of having the shortest residence time, the highest interaction energy, or both combined, (True negative).



**FIGURE 2** *P*-score theory diagram, Norm = normalized value of time ( $t$ ) and interaction energy ( $\Delta E$ ). Top-right green box → protein–ligand binding poses or hits of interest. Bottom-left red box → least interesting protein–ligand binding poses or hits. Top-left yellow box → protein–ligand binding poses or hits with high affinity but low stability and bottom-right yellow-box → protein–ligand binding poses or hits with high stability but low affinity. Vertical line intersecting the x-axis is the mean of normalized time ( $t$ ) values, and horizontal line intersecting the y-axis is the mean of normalized values for interaction energy ( $\Delta E$ ).

### 3 | METHODOLOGY PROTOCOL BENCHMARKING

To benchmark our approach, we perform a comparison between Glide docking, induced-fit docking and SuMD combined with MMPBSA, SuMD combined with QM-FMO and SuMD combined with DA-QM-FMO.

### 4 | DOCKING AND INDUCED-FIT DOCKING

For our first case study, PI-pro, GLIDE and Induced-fit docking modules of the Maestro 12.6.144 software (Schrodinger, LLC, New York, NY, USA)<sup>9</sup> were used to perform the molecular docking studies. Ligprep (Schrodinger, LLC)<sup>48</sup> was employed to optimize the energy of all compounds. The *holo-closed* P-loop conformation of PI-pro (PDB ID: 7JRN) and the *apo-opened* P-loop conformation (PDB ID: 6WUU) were prepared for docking using the 'protein preparation wizard' in Maestro (Schrodinger, LLC).<sup>26</sup> Bond orders and formal charges were added to hetero groups and hydrogens were added to all atoms. Side chains not close to the binding cavity and not participating in salt bridges were neutralized to avoid interference with binding interaction energy, and termini were capped with N-acetyl (ACE) and N-methyl amide (NMA) groups. The OPLSe force field was used to refine the structure and optimize the hydrogen bond network. Docking was carried out using the standard precision (SP) mode on the protein structure grid, and the Glide score was used to evaluate the final binding pose of protein–ligand complexes.

For the Induced-fit docking,<sup>49</sup> the entire receptor molecule was constrained and minimized with an root-mean-square deviation (RMSD) cutoff of 0.18 Å to generate the centroid of the residues, and the box size was automatically generated. Initial Glide docking<sup>26</sup> was carried out for each ligand, and the side chains were automatically trimmed based on the B-factor. The receptor and ligand van der Waals scaling were both set to 0.5, and 20 poses were generated. Prime side-chain prediction and minimization were carried out to refine residues within 5.0 Å of the ligand poses, and side chains were optimized to induce the ligand structure and conformation to fit each pose of the receptor structure. Finally, GLIDE standard precision (SP) redocking was performed on the top-scoring 20 structures overall and structures within 30.0 kcal/mol of the top-scoring structures. The Induced-fit docking (IFD) score was generated for each output pose.

## 5 | RESULTS AND DISCUSSION

### 5.1 | First case study PL-pro target system

To test *P-score* protocol, Papain-like protease protein (PL-pro) for SARS-CoV-2 coronavirus was used as our first case study.<sup>50</sup> PL-pro is expressed during SARS-CoV-2 viral replication and strong evidence in support of its function suggests it as a promising target in combatting

SARS-CoV-2 infection.<sup>51</sup> Multiple crystal structures of SARS-Cov-2 PL-pro, co-crystallized with inhibitors are available on the Protein Data bank (Table 1).

The complexity as well as the flexibility of the PI-pro binding pocket, made this test case particularly challenging. The binding site contains of a highly flexible loop named the “P-loop” that is found in different conformations in the crystal structures. Two crystal structures were chosen for the purpose of including different states of SARS-CoV-2 PL-pro binding site conformation; chain A from the *apo-form* (PDB ID: 6WUU), and the GRL0617-bound *Holo-form* (PDB ID: 7JRN).<sup>52–54</sup> Structural information of the *holo-form* GRL0617-bound PI-pro (PDB ID: 7JRN), revealed that the catalytic site is formed of a P-loop composed of TYR-268 and GLN-269, forming Pi-Pi stacking and hydrogen bond interactions with the bound ligand respectively. This conformation of the P-loop bound to the ligand is named a “*holo-closed* conformation” (Figure 3A).

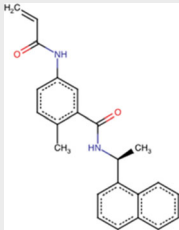
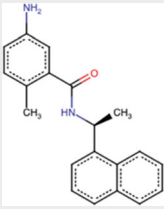
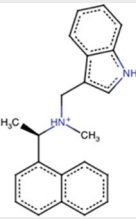
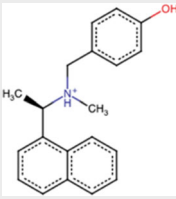
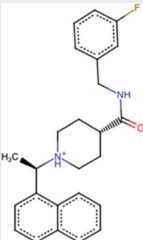
In the *apo-form* of the PL-pro (PDB ID: 6WUU), the P-loop showed an open conformation with about 90° difference in orientation of the side chains of the TYR-268 and GLN-269. This conformation of the P-loop is named an “*apo-opened* conformation” (Figure 3B). This P-loop acts as a one gate-keeper structure of PL-pro inhibitors, having a dynamic motion from opened to closed conformation upon ligand binding. Recent studies by Ismail et al, revealed that upon applying 1 μs of classical molecular dynamics simulation to the PL-pro *apo-form*, the P-loop shows a high degree of flexibility, as shown in (Figure 3B).<sup>55</sup> This creates a critical challenge in the discovery and optimization of new hit compounds against such highly flexible binding sites.

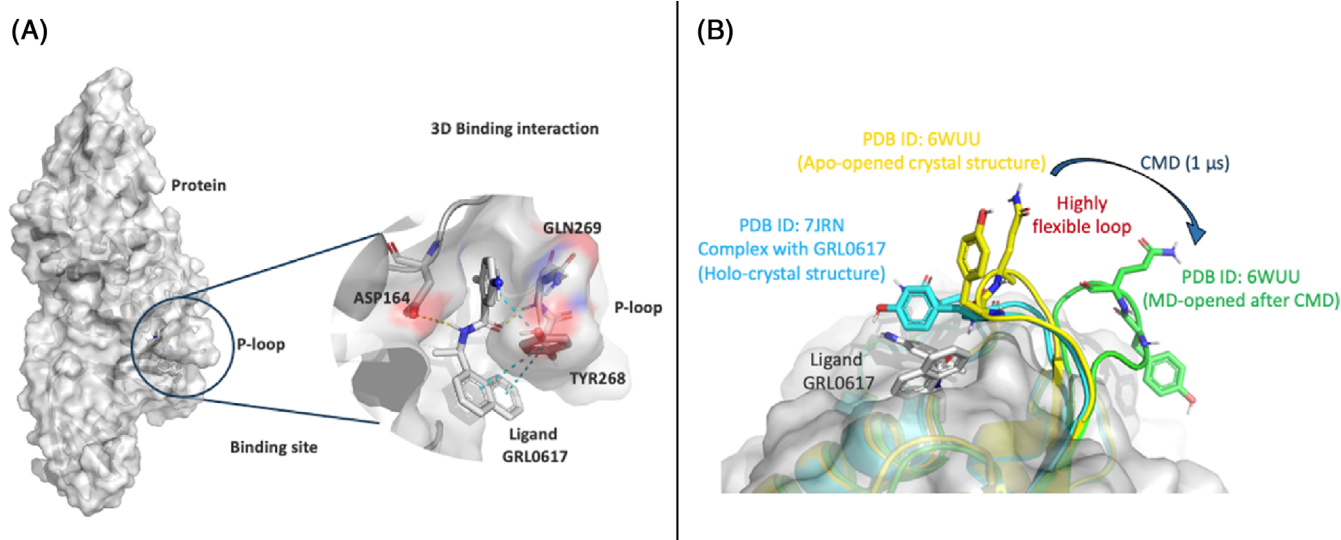
Since accurate prediction of the correct binding-poses during virtual screening is essential for the reliability of the screening process. Our protocol was applied to a set of five different PL-pro inhibitors with a range of binding affinities; as shown in Table 1. These are the ligands from *holo*-structures of PDB IDs: 7JIV, 7JRN, 7RZC, 7SDR, and 7TZJ, targeting the catalytic binding site of PI-pro.<sup>56,57</sup> The ligands were screened against the three binding site conformations formed by the P-loop, the *holo-closed*, *apo-opened* and *CMD-opened* conformations.

The *holo-closed*, *apo-opened* and *CMD-apo-opened* P-loop conformations forming the binding site of the PL-pro were used as starting points for SuMD simulations for the five different PL-pro inhibitors. Each simulation was 200 ns long with three different replicates to ensure sufficient sampling of ligand in the binding site. The SuMD simulations started with the ligands being placed at random coordinates with a distance between 30 and 60 Å from the binding site.

As the ligands reach the vicinity of 9 Å (cut-off) with the putative binding site in each replicate, different stabilities occur between the ligand and the binding site, suggesting that the ligands had achieved a different conformational sampling while forming initial binding interactions with the receptor binding site residues (Appendix S1; Videos). Treating each binding site conformation separately; *holo-closed*, the *apo-opened* and *CMD-opened* P-loop conformations, the three replicates of SuMD simulation were concatenated for each ligand, yielding a total of 600 ns simulation time. Filtration was performed before

**TABLE 1** Five different PL-pro inhibitors targeting the catalytic binding site with different potencies, originally identified by high-quality crystallographic structures were collected from the Protein Data Bank database.

Ligands	Lig-01 (PDB ID: 7JIV)	Lig-02 (PDB ID: 7JRN)	Lig-03 (PDB ID: 7RZC)	Lig-04 (PDB ID: 7SDR)	Lig-05 (PDB ID: 7TZJ)
IC <sub>50</sub> (μM)	6.4 ± 0.6	2.3 ± 0.2	0.67 ± 0.14	0.67 ± 0.08	0.7 ± 0.1
2D Structures					

**FIGURE 3** (A). PDB ID: 7JRN in complex with ligand GRL0617. (B) Binding site conformations observed by the P-loop: *holo-closed* in ligand-protein crystal complex, *apo-opened* in apo crystal structure, and CMD-*apo-opened* conformation.

clustering by removal of all SuMD frames with more than 9 Å distance between the ligand and the catalytic cleft. Specifically, a cluster is a protein-ligand complex (a binding pose) formed of at least 100 similar ligand conformations which differ by no more than 1 Å RMSD. The total number of frames in each cluster yields the total residence time ( $t$ ) for its corresponding protein-ligand binding mode, in which a single SuMD frame is equal to 20 ps of simulation. To calculate the interaction energy between the protein and ligands in the generated binding modes (clusters), snapshots for each cluster were extracted every 2 ns. Subsequently, DA-QM-FMO-based interaction energies ( $\Delta E^{\text{FMO}}$ ) were calculated for each cluster.

## 5.2 | PL-pro system: *Holo-closed* P-loop

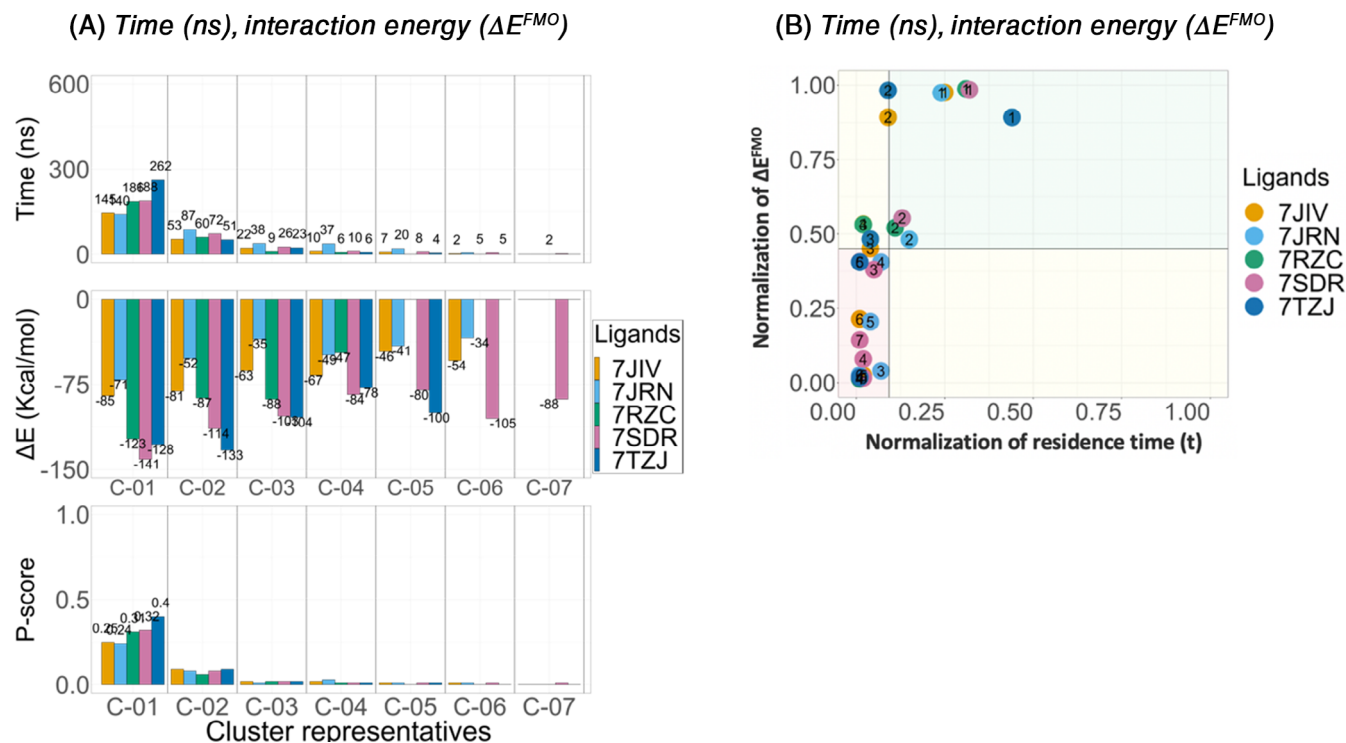
For the *holo-closed* P-loop conformation system, for each cluster of each ligand, the time (ns) the ligand spent stably bound in a certain

pose (residence time) and the total average interaction energy ( $\Delta E^{\text{FMO}}$ ) (kcal/mol) was calculated (Figure 4A).

For the five ligands; Lig-01 (7JIV) (orange), Lig-02 (7JRN) (cyan), Lig-03 (7RZC) (green), Lig-04 (7SDR) (magenta) and Lig-05 (7TZJ) (dark blue), cluster representative C-01 showed the most stable conformation in the binding site, translated by having the longest residence time ( $t$ ), of 145, 140, 186, 188, and 262 ns, respectively out of the 600 ns concatenated trajectories, rendering the most populated binding pose during the SuMD simulations.

Lig-03 (7RZC) (green), Lig-04 (7SDR) (magenta) and Lig-05 (7TZJ) (dark blue) show a lower IC<sub>50</sub> values in comparison to Lig-01 (7JIV) (orange) and Lig-02 (7JRN) (cyan). In addition, ligands Lig-03/04/05 spends a longer residence time in the most stable conformation than Lig-01/02.

The cluster representative C-01 showed the lowest total average interaction energy ( $\Delta E^{\text{FMO}}$ ), of -85 (kcal/mol), -71 (kcal/mol), -123 (kcal/mol), -141 (kcal/mol), respectively, for the four ligands: Lig-01 (7JIV) (orange), Lig-02 (7JRN) (cyan), Lig-03 (7RZC) (green), and Lig-04



**FIGURE 4** Holo-closed P-loop conformation analysis. Clustering of the three replicates for each PL-pro inhibitor, Lig-01 (7JIV), Lig-02 (7JRN), Lig-03 (7RZC), Lig-04 (7SDR), and Lig-05 (7TZJ). For each cluster (C-0\*) the residence time ( $t$ ) (ns), interaction energy ( $\Delta E^{FMO}$ ) (kcal/mol) shown on bar plots and  $P$ -score diagram.

(7SDR) (magenta). Thus, for these four ligands, the most stable conformation was calculated to also display the lowest energy, as would be expected. However, for Lig-05 (7TZJ) (dark blue), although C-01 appears to be the most stable conformation (the longest residence time ( $t$ ) of 262 ns), it was ranked only as the second lowest total average interaction energy ( $\Delta E^{FMO}$ )-127.5 kcal/mol after C-02 with a total average interaction energy ( $\Delta E^{FMO}$ ) of  $-133$  kcal/mol.

The  $P$ -score was calculated for each cluster of each ligand (Figure 4A,B). The  $P$ -score diagram shows that for the five PL-pro inhibitors Lig-01 (7JIV), Lig-02 (7JRN), Lig-03 (7RZC), Lig-04 (7SDR), and Lig-05 (7TZJ), C-01 falls in the top right corner (green box) formed by the mean of normalization of interaction energy ( $\Delta E^{FMO}$ ) and time ( $t$ ) for all the simulations. The  $P$ -score bar-chart shows that C-01 for Lig-01 (7JIV), Lig-02 (7JRN), Lig-03 (7RZC), Lig-04 (7SDR), and Lig-05 (7TZJ), has the highest values among other poses for each ligand separately (Appendix S1). This indicates that C-01 is predicted to be the highest-ranking pose for each ligand.

For the first four ligands, Lig-01 (7JIV), Lig-02 (7JRN), Lig-03 (7RZC), and Lig-04 (7SDR), C-01 showed the most stable conformation with the longest residence time (ns) and the lowest total average interaction energy ( $\Delta E^{FMO}$ ) (Figure 4A). However, looking for Lig-05 (7TZJ) (dark blue), although C-02 showed the lowest binding interaction energy ( $\Delta E^{FMO}$ ), yet it shows a low residence time ( $t$ ) in comparison to C-01, with 51 ns compared to 262 ns for C-01. The  $P$ -score ranks the C-01 over C-02 for the lig-05 (7TZJ) (dark blue).

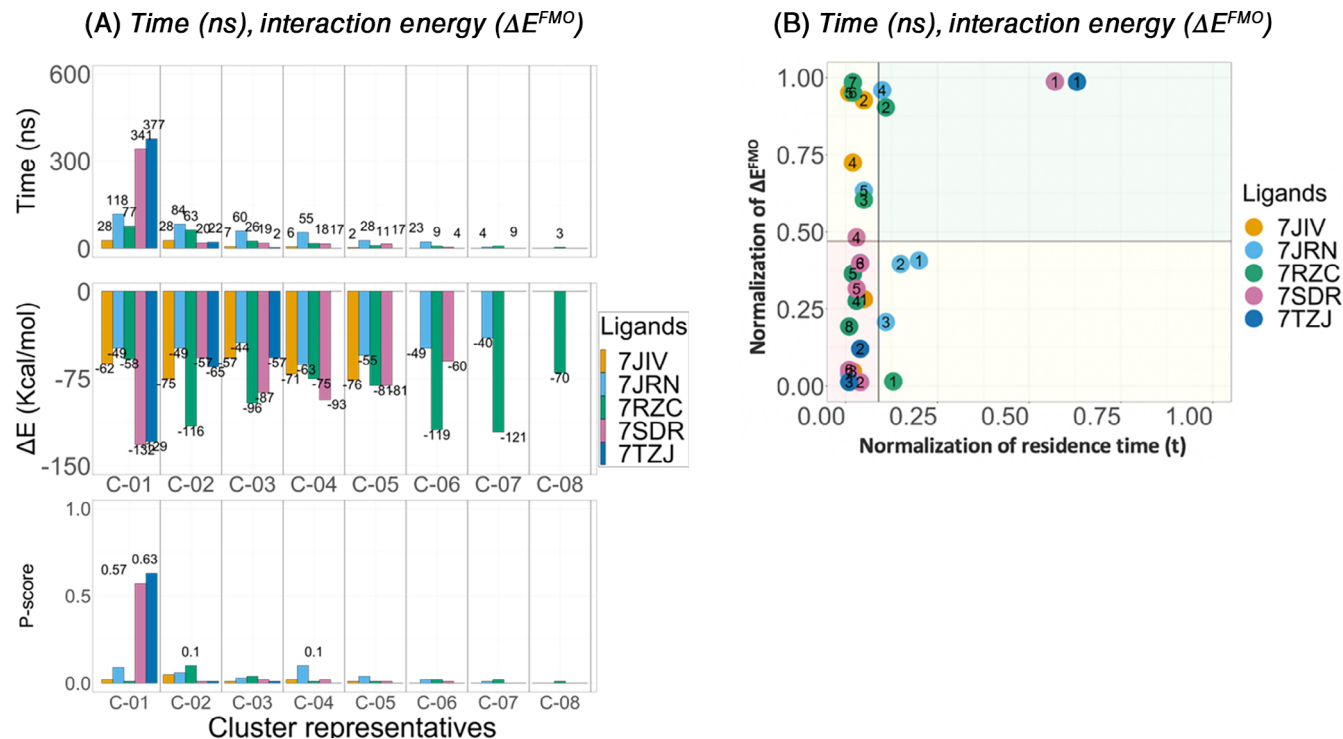
### 5.3 | PL-pro system: Apo-opened P-loop

The same procedures of calculating the time (ns), and interaction energy ( $\Delta E^{FMO}$ ) (kcal/mol), for clusters of each ligand, were performed on the *apo-opened* P-loop conformation system, (Figure 5A).

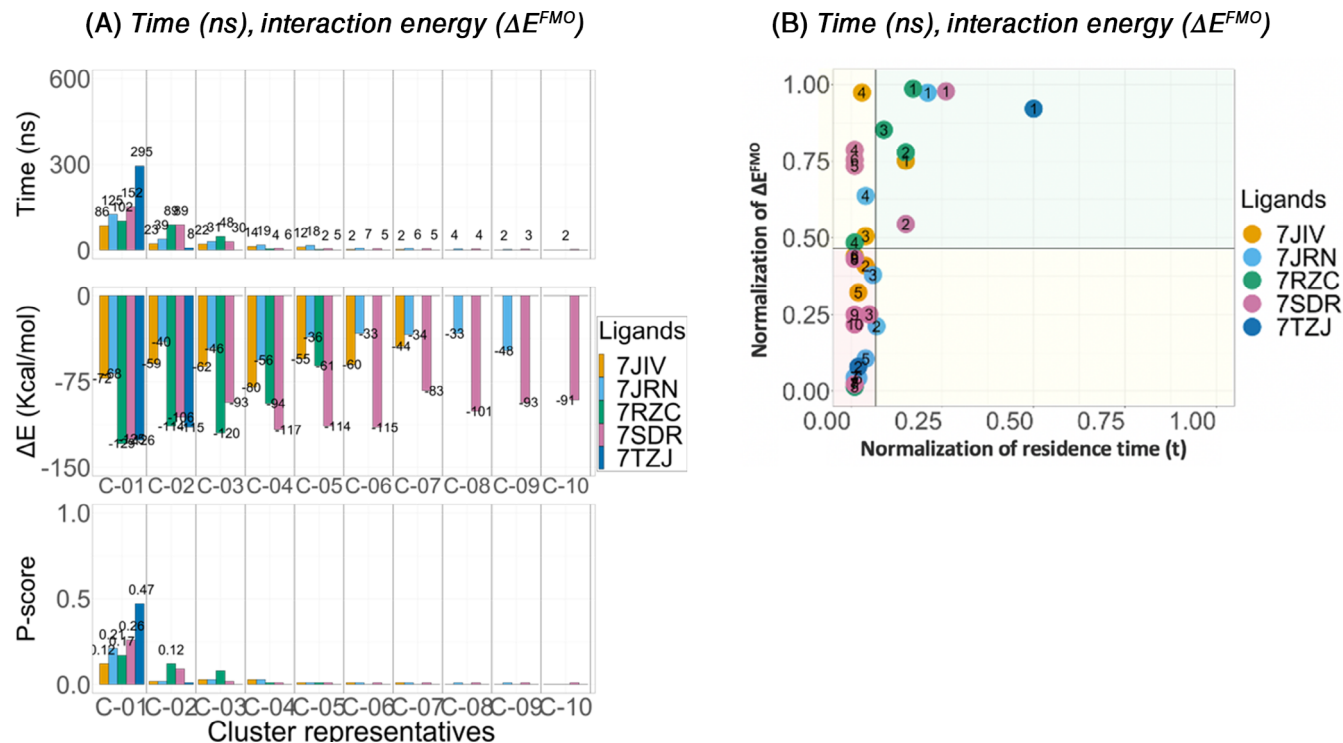
For Lig-01 (7JIV) (orange), the maximum residence times for cluster representatives show a very short residence time of 28 ns found for both C-01 and C-02 from a total simulation time of 600 ns. This could be due to the relatively low affinity of  $6.4 \pm 0.6 \mu\text{M}$  (Table 1) or indicate that the simulation was not long enough to achieve a stable pose, when starting from the *apo-opened* P-loop. When considering the total interaction energy for the ligand, C-05 showed the lowest interaction energy of  $-76$  kcal/mol.

For Lig-02 (7JRN) (cyan) although C-01 showed the longest residence time in the binding site with 118 ns, C-04 showed the lowest interaction energy with  $-63$  kcal/mol and ranked 4th according to the residence time with 55 ns. For Lig-03 (7RZC) (green), C-01 showed the longest residence in the binding site with 77 ns; however, cluster C-07 showed the lowest interaction energy with a value of  $-121$  kcal/mol. Cluster representative C-01 for ligands, Lig-04 (7SDR) (magenta) and Lig-05 (7TZJ) (dark blue), showed the most stable conformation in the binding site by having the longest residence time ( $t$ ) of 377s and 341 ns respectively. In addition, C-01 for Lig-04 (7SDR) and Lig-05 (7TZJ), showed the lowest total average interaction energy ( $\Delta E^{FMO}$ ) of  $-132$  and  $-129$  kcal/mol, respectively.





**FIGURE 5** Apo-opened P-loop conformation analysis. Clustering of the three replicates for each PL-pro inhibitor, Lig-01 (7JIV), Lig-02 (7JRN), Lig-03 (7RZC), Lig-04 (7SDR), and Lig-05 (7TZJ). For each cluster (C-0\*) the residence time (t) (ns), interaction energy ( $\Delta E^{FMO}$ ) (kcal/mol) shown on bar plots and P-score diagram.



**FIGURE 6** CMD-opened P-loop conformation analysis. Clustering of the three replicates for each PL-pro inhibitor, Lig-01 (7JIV), Lig-02 (7JRN), Lig-03 (7RZC), Lig-04 (7SDR), and Lig-05 (7TZJ). For each cluster (C-0\*) the residence time (t) (ns), interaction energy ( $\Delta E^{FMO}$ ) (kcal/mol) shown on bar plots and P-score diagram.

By calculating the *P*-score diagram and bar-chart, indicates that for ligands, Lig-04 (7SDR) and Lig-05 (7TZJ), C-01 is predicted to be the highest-ranking binding pose of interest with a *P*-score values of 0.57 and 0.63 respectively, (Figure 5B).

For ligand Lig-03 (7RZC), the *P*-score ranks C-02 as the highest-ranking pose with a value of 0.1. Although C-02 did not show the longest residence time in the binding site, with only 63 ns when compared to C-01 of 77 ns, nor the lowest interaction energy  $-115$  kcal/mol compared to C-07 with  $-121$  kcal/mol, yet the *P*-score predicts that the highest-ranking pose is C-02.

For Lig-02 (7JRN), although C-01 shows the longest residence time (*t*) of 118 ns out of the 600 ns trajectories) (Figure 5A), *P*-score prioritizes C-04 as the pose of interest which shows much lower residence time (*t*) of 55 ns. Also, *P*-score can identify that C-02 for Lig-03 (7RZC) and C-04 for Lig-02 (7JRN) yet it also says that both predicted poses are on the border line of residence time (normalization of residence time (Figure 5B), which means, these poses can be significant but prolongation of SuMD simulations is recommended to increase the percentage of confidence in selected poses.

However, for Lig-01 (7JIV) (orange), the *P*-score suggests that none of the poses are likely to be prioritized. In the *P*-score diagram, no points for the Lig-01 (7JIV) (orange) were found in the top right corner (green box).

## 5.4 | PL-pro system: CMD-opened P-loop

Starting from the CMD-opened P-loop conformation system, (Figure 6). For Lig-01 (7JIV) (orange), C-01 showed the longest residence time of 86 ns found. This could also be due to the relatively low affinity of  $6.4 \pm 0.6$   $\mu$ M (Table 1). When considering the total interaction energy for the ligand, C-04 showed the lowest interaction energy of  $-80$  (kcal/mol). For Lig-02 (7JRN) (cyan) although C-01 showed the longest residence time in the binding site with 125 ns, with the lowest interaction energy with  $-68$  kcal/mol. For Lig-03 (7RZC) (green), Lig-04 (7SDR) (magenta) and Lig-05 (7TZJ) (dark blue), C-01 showed the longest residence in the binding site with 102s, 152, and 295 ns respectively. Also, C-01 for Lig-03 (7RZC) (green), Lig-04 (7SDR) and Lig-05 (7TZJ), showed the lowest total average interaction energy ( $\Delta E^{\text{FMO}}$ ) of  $-129$ ,  $-125$  kcal/mol and  $-126$  kcal/mol, respectively.

According to the *P*-score diagram and bar-chart, indicates that for all ligands, C-01 is predicted to be the highest-ranking binding pose of interest with a *P*-score values of 0.12, 0.21, 0.17, 0.26, and 0.47 respectively. Although C-01 did not show the lowest total average interaction energy, for ligands Lig-01 (7JIV) (orange), and Lig-02 (7JRN) (cyan).

## 6 | VALIDATION OF PREDICTED POSES BY *P*-score

The accuracy of the *P*-score predictions was then assessed by comparing the highest *P*-score structure for each ligand with its respective

crystallographic reference, computing the RMSD of non-hydrogen atomic coordinates. Selected binding poses for each ligand in each SuMD simulation system, starting from both *holo-closed* or *apo-opened* P-loop conformation, shows an RMSD from the reference structure that falls below 2 Å, with the lowest being 0.6 Å and the highest of 1.6 Å, (Figure 7).

This indicates a good prediction of the correct binding pose for each ligand. The selected pose for each ligand also established nearly the same interaction energy ( $\Delta E^{\text{FMO}}$ ) against the binding site residues of its corresponding crystalized pose (reference pose). To have a further investigation on the predicted binding poses, detailed protein–ligand binding analysis is performed on the interaction with the binding site residues (Appendix S1 (3–8)).

## 7 | PROTOCOL BENCHMARKING

Simple docking and induced-fit docking were evaluated as a basis of comparison for the binding pose prediction of the five ligands in terms of different scoring functions, against the three P-loop conformation, the *holo-closed*, *apo-opened* and *CMD-opened* conformations.

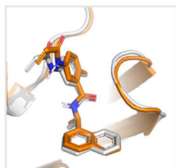
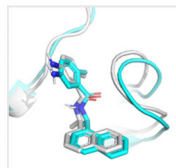
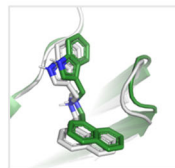
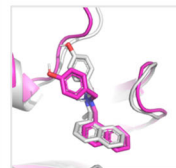
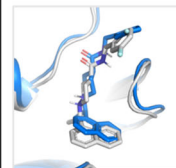
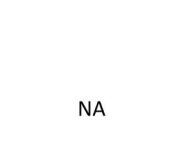
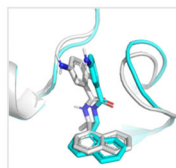
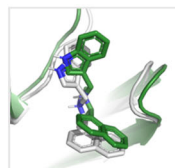


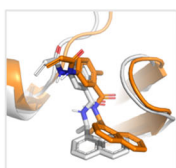
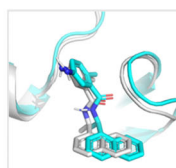
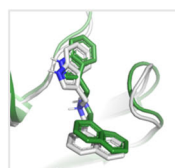
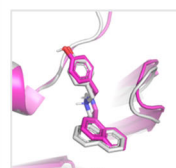
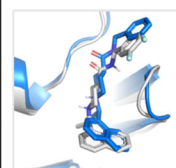
Simple docking using Glide and its docking score function was successful in generating the correct binding pose (RMSD  $<2$  Å) of only Lig-01 (7JIV) and Lig-02 (7JRN) in the *holo-closed* P-loop conformations. However, simple docking as well as induced-fit docking failed to predict the correct binding pose for the five ligands in both *apo-opened* and *CMD-opened* P-loop conformations (RMSD  $>2$  Å) (Figure 8).

However, by utilizing the SuMD approach, we were able to generate the correct binding pose independent from the binding site conformation except for the ligand Lig-01 (7JIV) in the *apo-opened* simulation. These latter results may be due to the ligand's low affinity with the open conformation of the binding site, probably indicating the need for longer simulation times. By contrast, when MM-PBSA ( $\Delta E$ ) was applied on the clusters from the SuMD, it showed low accuracy in predicting the correct binding pose with 50% correct predictions.

Taking the residence time (*t*) or the interaction energy DA-FMO ( $\Delta E$ ) alone as a single descriptor to account for the differentiation between binding poses, the most stable conformation as predicted from the SuMD was also the lowest energy conformation as predicted by DA-QM-FMO or FMO (on a single snapshot) in most cases.

However, in some cases, either one or both metrics did not predict the experimental binding pose. This is likely due to the limited simulation time during SuMD, and the fact that FMO accounts for the interaction energy between ligand and binding site residues but does not calculate the binding free energy of the system on atomic level. Also, FMO calculations rely on a basis set of atomic orbitals to describe the electronic structure of the system. However, if the basis set is not sufficiently large, important electronic interactions between

**FIGURE 7** Root-mean-square deviations of clusters with the highest *P*-score value for each PL-pro inhibitor starting from the three different binding sites conformations, Lig-01 (7JIV), Lig-02 (7JRN), Lig-03 (7RZC), Lig-04 (7SDR), and Lig-05 (7TZJ), against its correspondent reference crystal structure (white). NA, not available, since *P*-score did not prioritize any pose during the simulation.

Lig-01 (7JIV)	Lig-02 (7JRN)	Lig-03 (7RZC)	Lig-04 (7SDR)	Lig-05 (7TZJ)
<b>Holo-closed P-Loop conformation system analysis</b>				
C-01	C-01	C-01	C-01	C-01
				
RMSD = 0.6 Å <i>P</i> -score = 0.25	RMSD = 0.7 Å <i>P</i> -score = 0.24	RMSD = 0.7 Å <i>P</i> -score = 0.31	RMSD = 1.6 Å <i>P</i> -score = 0.32	RMSD = 1.0 Å <i>P</i> -score = 0.40
<b>Apo-opened P-Loop conformation system analysis</b>				
NA	C-04	C-02	C-01	C-01
				
NA	RMSD = 1.3 Å <i>P</i> -score = 0.10	RMSD = 1.7 Å <i>P</i> -score = 0.10	RMSD = 1.3 Å <i>P</i> -score = 0.57	RMSD = 0.8 Å <i>P</i> -score = 0.63
<b>CMD-opened P-Loop conformation system analysis</b>				
C-01	C-01	C-01	C-01	C-01
				
RMSD = 0.8 Å <i>P</i> -score = 0.12	RMSD = 0.7 Å <i>P</i> -score = 0.21	RMSD = 1.5 Å <i>P</i> -score = 0.17	RMSD = 1.6 Å <i>P</i> -score = 0.26	RMSD = 1.6 Å <i>P</i> -score = 0.47

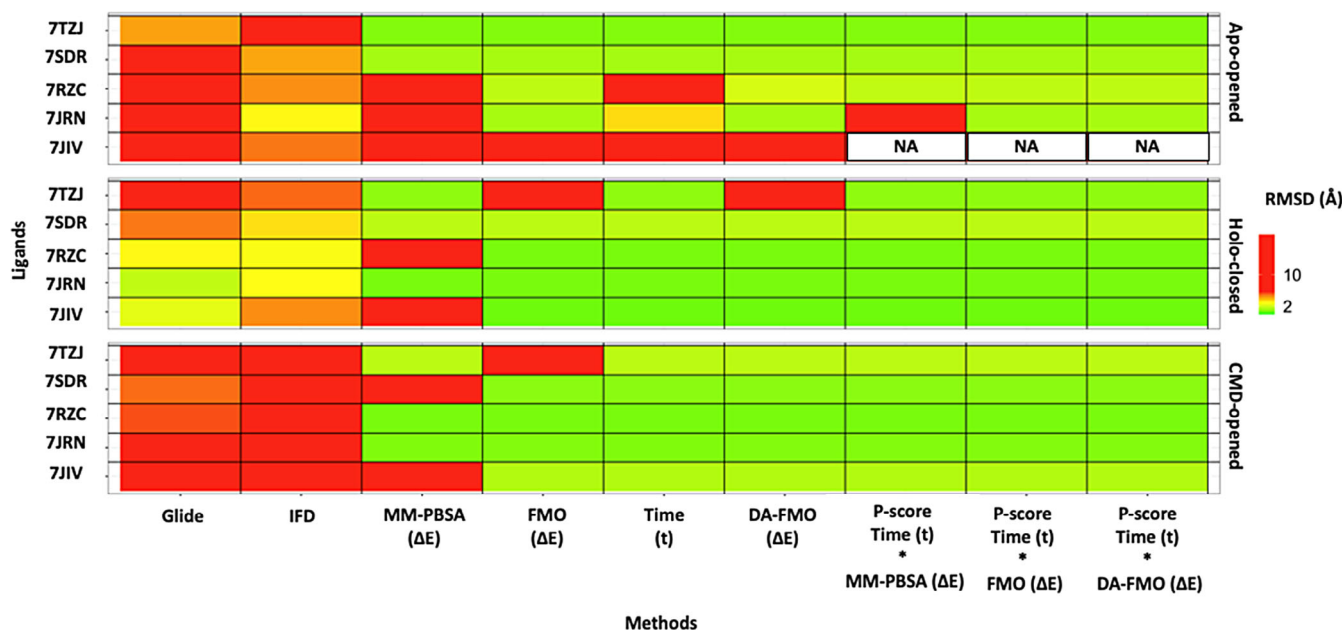
fragments may be missed, leading to errors in the predicted properties.

Applying the *P*-score matrix, using the residence time and the binding energy from MM-PBSA, FMO or DA-FMO, was applied to accurately rank order generated poses for all ligands independent from the starting point of the binding site. Also, in the case of 7JIV in the apo-opened conformation simulation, the *P*-score was able to identify it as “NA” (Not available) which refers to no poses were prioritized in these simulations, eliminating the false positive predictions. A single false prediction by the *P*-score is found with the ligand 7JRN in the Apo-opened, however this is because MM-PBSA as energy calculation is based on molecular mechanics which is less accurate than quantum-level FMO, therefore using less accurate methods to provide descriptors for the *P*-score matrix can result in false positives.

## 7.1 | M-pro target system

A second SARS-CoV-2 protease protein is Mpro which is one of two cysteine viral proteases essential for viral replication. We apply the same procedures mentioned for PL-pro on six different ligands co-crystallized with the Mpro protein with PDB IDs: Lig-01 (5R81), Lig-02 (5R83), Lig-03 (5R84), Lig-04 (5RE4), Lig-05 (5RF1), and Lig-06 (5RF3).<sup>58</sup> In this case, each ligand was subjected to different sampling time using SuMD, with a total simulation time of 257 ns all-combined.

*P*-score quadrant matrix was performed using the residence time and QM-FMO, to predict the binding pose for each ligand, (Figure 9A). In relative comparison between all simulations combined, *P*-score prioritized a binding pose for ligands; Lig-03 (5R84), Lig-04 (5RE4), and Lig-06 (5RF3), represented in clusters numbers C-01, and for Lig-05 (5RF1) cluster number C-02, were identified as the poses of



**FIGURE 8** Heatmap of root-mean-square deviation (RMSD) with a threshold of 2 Å, showing lowest RMSD reach by each ligand, Lig-01 (7JIV), Lig-02 (7JRN), Lig-03(7RZC), Lig-04 (7SDR), and Lig-05 (7TZJ) (y-axis), against different method used, Glide (simple docking using glide and docking scoring function), IFD (induced fit docking, scoring using docking scoring function), MM-PBSA (ΔE) (MM-PBSA (ΔE) binding energy used for prediction), time (t) (using time [t] for residence), DA-FMO (ΔE) (using interaction energy of DA-QM-FMO as a single descriptor for affinity). NA, not applicable.

interest. Poses were accurately predicted in the four prioritized ligands shows a value of (RMSD <2 Å) against the reference crystal structure. However, for the ligands Lig-01 (5R81), and Lig-02 (5R83), *P-score* did not prioritize any of the poses that have been identified in relative comparison to the rest of the ligands. By treating each of the screened molecules separately, (Figure 9C; Appendix S1), *P-score* identify for ligands, Lig-01 (5R81), cluster number C-02 as potential pose of interest, while for Lig-02 (5R83), both C-02 and C-03 falls on the mean line that intersect the residence time.

Accordingly, *P-score* in these three cases is prioritizing the mentioned poses per ligand but with lower confidence than the first four ligands. Therefore, *P-score* is suggesting that these are the potential poses of interest, but more simulation time is needed to assure the prediction. However, assessing the RMSD in respect to the co-crystallized reference, we can find that *P-score* prioritized poses even with low confidence, have a value of (RMSD <2.5 Å), where the ligands bind in a well-defined orientation in respect to the reference pose, where these poses are found to be not the lowest binding interaction energy nor the longest residence time, yet *P-score* prioritized the nearest poses to their corresponding reference structure.

## 7.2 | *P. aeruginosa* LpxC—*P-score* using MMPBSA

*P. aeruginosa* LpxC, is a zinc metalloenzyme that catalyzes the first step in the biosynthesis of Lipid A, an essential component of the cell envelope of Gram-negative bacteria.<sup>59</sup> Its binding site is formed of a 4-coordinated zinc metal centers with 3 residues; ASP-241, HIS-237,

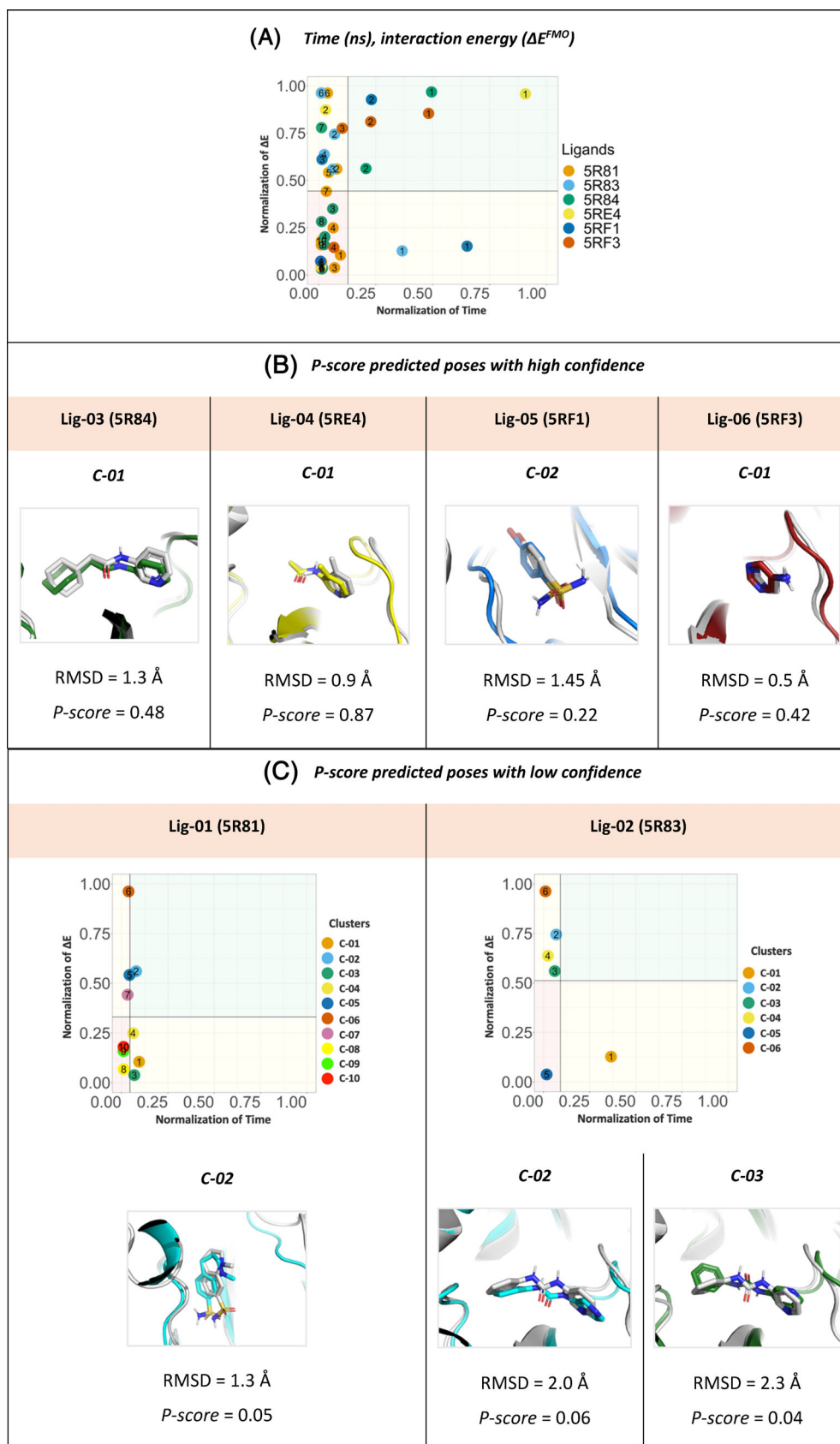
and HIS-78, (Figure 10A). In this example, the ligand is sampled within the binding site, however for energy calculations MM-PBSA was used instead of DA-QM-FMO, due to the positive charge of the Zn<sup>2+</sup>-ion, that can be challenging for QM-approaches. Therefore, the *P-score* matrix, is built around residence time (t) and the binding energy calculation from MM-PBSA. This protein is found in a complex structure with PDB ID; 7CI8, where there have been many modifications to the core scaffold and R-groups of this molecule reported in the literature.<sup>56</sup> We apply the *P-score* protocol on the crystalized ligand and one of its core scaffold modifications (referred as 7CI8-R).<sup>59</sup>

In both ligands, the 7CI8 and 7CI8-R, the *P-score* has a high confidence in the binding pose predicted, represented in cluster number C-01. The 7CI8 ligand has a well aligned conformation with Zn<sup>2+</sup> ion binding interaction, reflected in the RMSD of 1.2 Å with the reference molecule. On the other hand, for the 7CI8-R as there is no crystal structure, however they share the same common head that is binding to the Zn<sup>2+</sup> ion, RMSD was calculated only on the atom pair with the reference molecule, which shows also a well aligned Zn<sup>2+</sup>-ion binding and an RMSD of 1.3 Å. Figure 10B.

## 7.3 | Further study cases; heat shock protein 90, p38 kinase, and myeloid cell leukemia 1

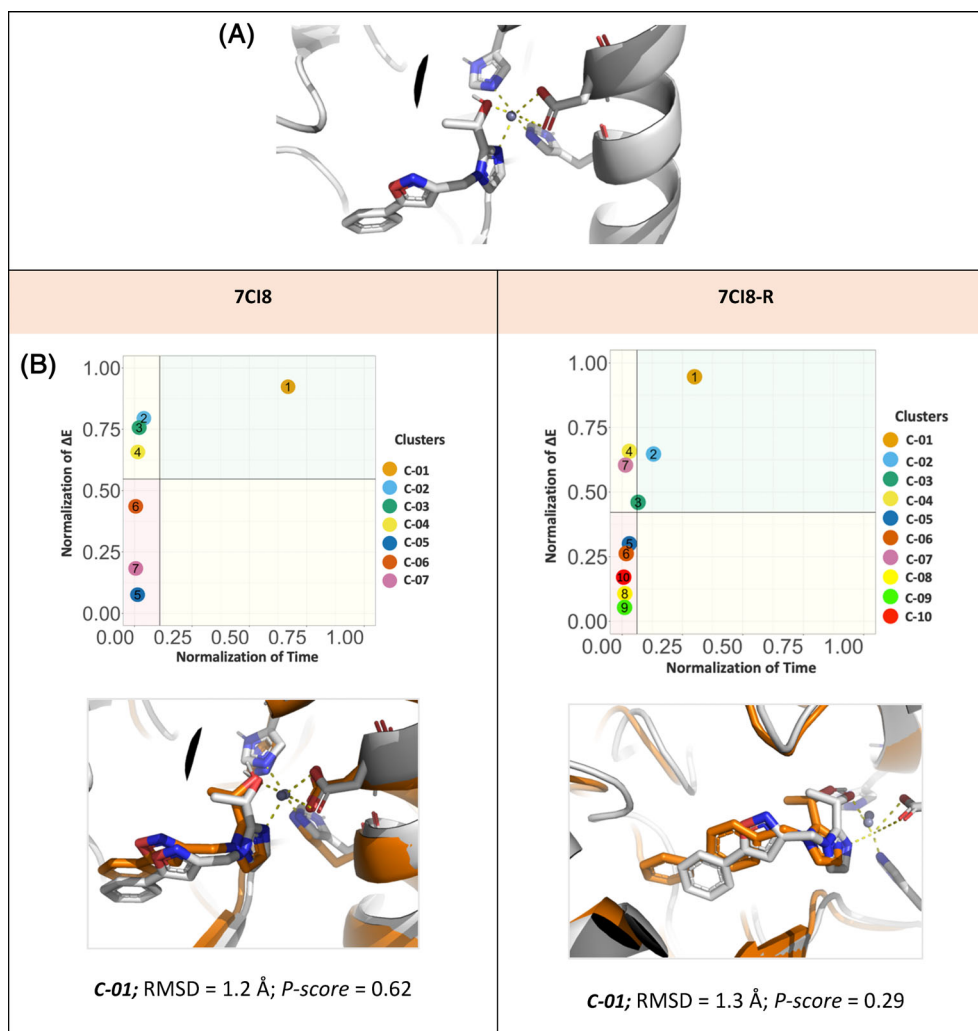
Heat shock protein 90 (Hsp90) plays a key role in stress response and protection of the cell against the effects of mutation, p38 Kinase (p38), and Myeloid Cell Leukemia 1 (Mcl-1) a member of the Bcl-2

**FIGURE 9** (A) *P*-score diagram for Mpro target screening. Clusters representative for PDB IDs: 5R81, 5R83, 5R84, 5RE4, 5RF1, and 5RF3. For each cluster numbered, and the residence time (*t*) (ns), interaction energy ( $\Delta E^{FM0}$ ) (kcal/mol) shown on bar *P*-score diagram *x* and *y*-axis respectively. (B) Root-mean-square deviation (RMSD) of clusters with the highest *P*-score value for each M-pro binders, Lig-03 (5R84), Lig-04 (5RE4), Lig-05 (5RF1), and Lig-06 (5RF3), against its correspondent reference crystal structure (white). (C) *P*-score diagram for Mpro target screening. RMSD of clusters with the low *P*-score confidence value for each M-pro binders, Lig-01 (5R81), Lig-02 (5R83), against its correspondent reference crystal structure (white).



family of proteins, is overexpressed and amplified in various cancers and promotes the aberrant survival of tumor cells that otherwise would undergo apoptosis. These targets are for treatment of many

diseases. For the Hsp90, p38 and Mcl-1 the PDB IDs; 3FT5, 1W7H and 4HW3 were used as representative targets with their co-crystallized ligands.<sup>60–62</sup>



**FIGURE 10** (A) *Pseudomonas aeruginosa* LpxC Zinc binding site,  $Zn^{2+}$ , between ASP-241, HIS-237 and HIS-78. (B) *P*-score diagram for LpxC target. RMSD of Clusters with the low *P*-score confidence value for each M-pro binders, 7C18, 7C18-R, against its correspondent reference crystal structure (white).

As shown in (Figure 11A,B), *P*-score prioritized a pose for both targets the Hsp90 and p38, where cluster number C-01 showed high confidence, that is reflected in a high *P*-score value of 0.74 and 0.81, respectively. Compared to the RMSD value of the co-crystallized poses, both ligands; Hsp90 and p38 shows a well-established superposition on the crystalized reference with RMSD 0.95 Å and 1.70 Å, respectively. For the case of Mcl-1, *P*-score matrix shows two different clusters C-02 and C-03 on the mean value of residence time (*t*) and C-01 on the mean value of interaction energy (Figure 11C). Although the pose represented by C-03 is with acceptable range of RMSD value 2.7 Å, it is recommended to increase the sampling phase to allow more exploration of local minima around the protein-ligand binding.

#### 7.4 | Correlation between interaction energy (kcal/mol) and residence time (*t*)

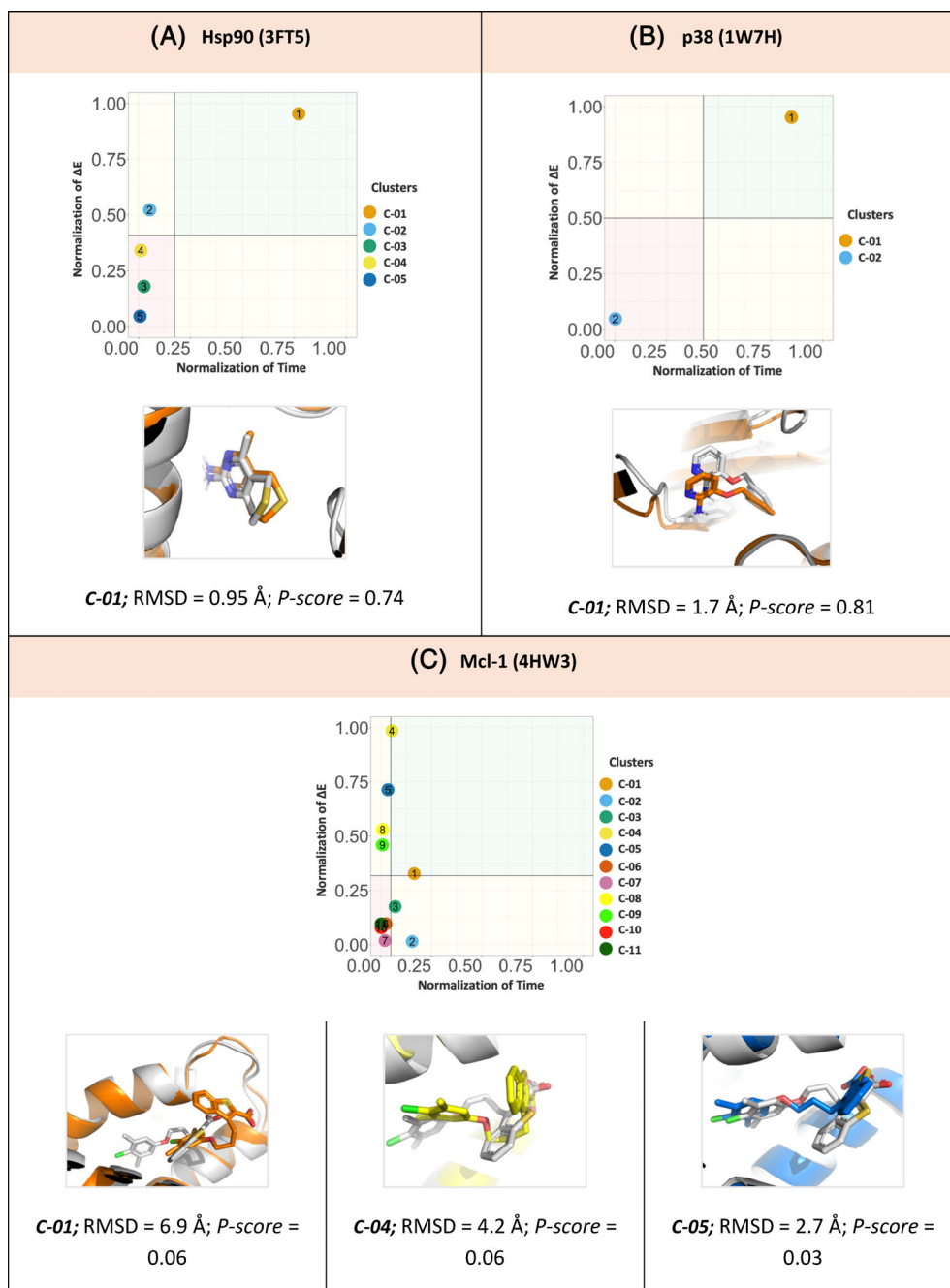
Pearson correlation coefficient (*r*) was calculated for each ligand simulation independently based on the simulation data, (as detailed in the Appendix S1; 9).<sup>63</sup> Our analysis reveals a consistent trend; when

the sampling time is long enough to allow the ligand to adequately explore the binding site and assume stable conformations across various local minima, a negative correlation emerges between prolonged residence time (*t*) and lower interaction energy (kcal/mol). This correlation suggests that; as the energy of interaction decreases, the residence time of the ligand tends to increase, indicating a more stable binding. This observation is supported by a high confidence level in the prediction, as indicated by the associated p-values. However, in simulations with insufficient dynamic sampling time, correlation is barely predicted and mostly not found, yet the *P*-score can accurately prioritize the correct binding pose by combining both descriptors in a normalization equation. These findings underline the significance of both descriptors; residence time and interaction energy, in characterizing protein-ligand binding dynamics, and thus affirming the principles outlined in the *P*-score theory.

#### 7.5 | *P*-score for hit identification

The experimental Free energy ( $\Delta G_{exp}$ ) for the IC50, as shown in Equation (4):

**FIGURE 11** (A,B) *P*-score diagram for Hsp90 and p38 targets. root-mean-square deviation (RMSD) of clusters with the low *P*-score confidence value for both, Hsp90 (3FT5), p38 (1W7H), against its correspondent reference crystal structure (white). (C) *P*-score diagram for Mcl-1 target. RMSD of Clusters with the low *P*-score confidence value for Mcl-1 (4HW3), against its correspondent reference crystal structure (white).



$$\Delta G_{\text{exp}} = R \times T \times \ln(\text{IC50}). \quad (4)$$

Experimental free energy equation,  $R$  and  $T$  represent the gas constant and temperature, respectively.

The *P*-score for the experimental value can be calculated by considering the normalization of the simulation time;  $t_i/T = 1$ , since the crystal structure represents a static reference pose.

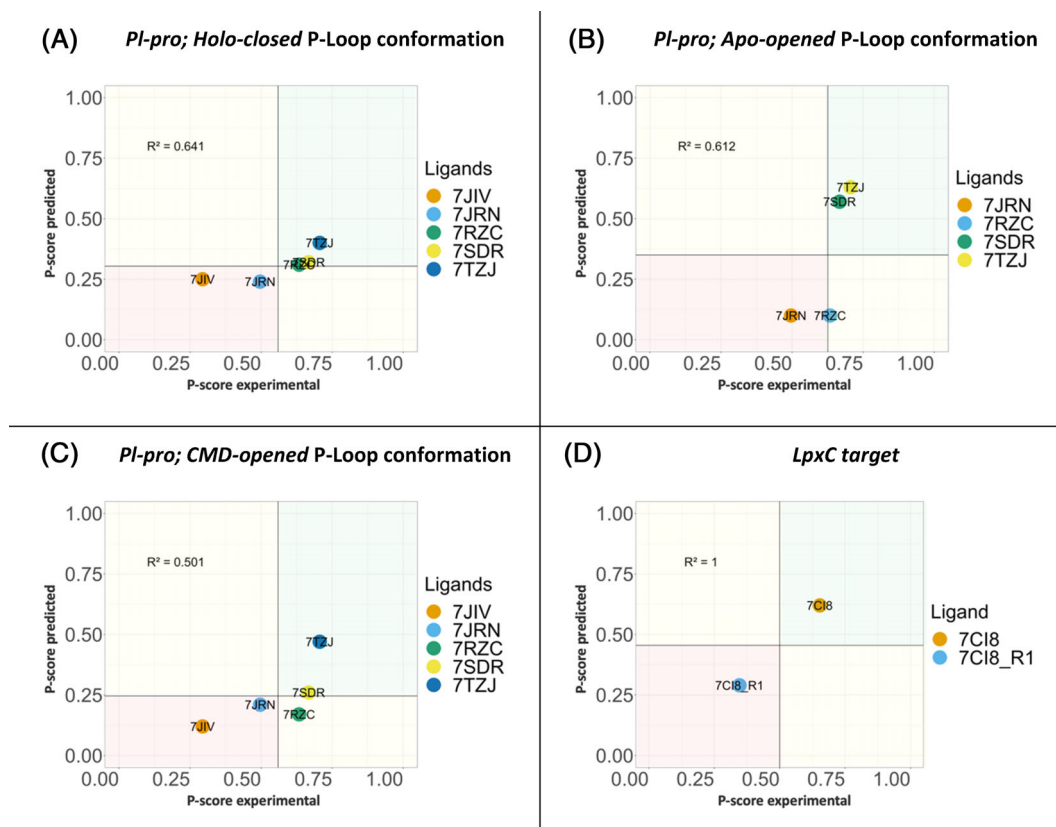
## 7.6 | PL-pro system

Starting from the *holo-closed* P-loop conformation system, the *P*-score predicted against *P*-score experimental, identify Lig-03 (7RZC), Lig-04

(7SDR), and Lig-05 (7TZJ) as the most active hits when compared to the Lig-01 (7JIV) and Lig-02 (7JRN) (Figure 12A).

Starting from the *apo-opened* P-loop conformation system the *P*-score identifies the Lig-04 (7SDR) and Lig-05 (7TZJ), as the most active hits. For Lig-03 (7RZC), as mentioned earlier, prolongation of the simulation is needed to increase the confidence of the binding pose. Lig-02 (7JRN) falls in the less potent (red box) left bottom corner of the graph, and for Lig-01 (7JIV) no poses were identified by the *P*-score, (Figure 12B).

Starting from the *CMD-opened* P-loop conformation *P*-score ranks Lig-04 (7SDR), and Lig-05 (7TZJ) as the most active hits when compared to the Lig-01 (7JIV) and Lig-02 (7JRN) and Lig-03 (7RZC), same as in *apo-opened* P-loop conformation system (Figure 12C).



**FIGURE 12** *P*-score against normalization of delta G experimental for the five ligands of PL-pro inhibitors, Lig-01 (7JIV), Lig-02 (7JRN), Lig-03 (7RZC), Lig-04 (7SDR), and Lig-05 (7TZJ). Starting from, (A) Holo-closed. (B) Apo-opened. (C) CMD-opened P-loop conformation systems. (D) LpxC target.

These results correlated with the experimental  $IC_{50}$  values showing that the difference in potency between molecules Lig-03 (7RZC), Lig-04 (7SDR), and Lig-05 (7TZJ) and molecules Lig-01 (7JIV), and Lig-02 (7JRN) can be identified as more potent hits in comparison to molecules Lig-01 (7JIV), and Lig-02 (7JRN).

Also, there you can see a correlation between the *P*-score predicted and the *P*-score experimental, in the three cases, showing that the holo-closed conformation has the highest correlation of 0.64  $R^2$  compared to the Apo-opened and CMD-opened P-loop systems conformations. This can conclude that the more stable the binding conformation is the better the correlation of the *P*-score predicted with the *P*-score experimental, as in a wider open binding site conformation, the system needs more sampling time to stabilize a local minima with a binding interaction.

## 7.7 | LpxC system

Both molecules in the protocol mentioned above have reported  $IC_{50}$  values (7C18;  $IC_{50} = 0.88$  nM and 7C18-R;  $IC_{50} = 3.94$  mM). Although MM-PBSA energy calculation showed that the 7C18-R binding energy is  $-23.9$  (kcal/mol), and 7C18 binding energy is  $-21.7$  (kcal/mol), which is not reflecting the correct correlation with the experimental  $IC_{50}$ s. However, combining the residence time ( $t$ ) using

the *P*-score predicted shows a very good correlation to distinguish between the  $\mu$ M and nM range molecules, showing that 7C18 falls in the green box while the  $\mu$ M range 7C18-R is in the red box. (Figure 12D).

## 8 | CONCLUSION

The Identification of a suitably representative binding pose for a ligand in a protein pocket is challenging as it is a complex process. Here we introduce an approach using supervised molecular dynamics, coupled with scoring the binding poses using energy calculation methods; Dynamical Averaging of Quantum Mechanics Fragment Molecular Orbital (DA-QM-FMO), FMO or MM-PBSA.

Our studies highlight the approach could be further enhanced by combining the trajectories of the supervised molecular dynamics with the calculated binding energy using DA-QM-FMO as the *P*-score. When using both the time of the ligand in any binding pose and the calculated binding energy are considered as descriptors and normalized, giving a score between 0 and 1. The *P*-score was able to predict correctly the experimental binding poses with good accuracy, independent from the ligands starting point or the binding site conformation. In addition, it was able to predict when none of the correct binding poses were found to be the experimental binding pose.



Although each of the mentioned case studies have their complexity, PL-pro was a particularly challenging target, due to the flexibility of the active site and the very flexible P-loop which closes over the ligand.

Although there is an increase in time and cost using SuMD compared to Glide and docking strategies. Considering the flexible nature of protein targets is crucial for predicting the binding events hence the introduction of dynamics is a necessity, even though the cost can be a limitation that needs careful consideration for future perspectives.

## ACKNOWLEDGMENTS

Acknowledgement to the Wellcome Centre for Anti-Infectives Research for funding this scientific project. Thanks to Mattia Sturlese, for sharing the scripts for the Supervised Molecular dynamics method.

## DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

## ORCID

Peter E. G. F. Ibrahim  <https://orcid.org/0009-0001-8774-1068>

## REFERENCES

- [1] W. L. Jorgensen, *Science* **2004**, 303, 1813.
- [2] E. Gallicchio, M. Lapelosa, R. M. Levy, *J. Chem. Theory Comput.* **2010**, 6, 2961.
- [3] N. M. Lim, L. Wang, R. Abel, D. L. Mobley, *J. Chem. Theory Comput.* **2016**, 12, 4620.
- [4] M. Araki, N. Kamiya, M. Sato, M. Nakatsui, T. Hirokawa, Y. Okuno, *J. Chem. Inf. Model.* **2016**, 56, 2445.
- [5] D. L. Mobley, K. A. Dill, *Structure* **2009**, 17, 489.
- [6] G. Jones, P. Willett, R. C. Glen, A. R. Leach, R. Taylor, *J. Mol. Biol.* **1997**, 267, 727.
- [7] A. J. A. Docking, A. J. Autodock, G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, *J. Comput. Chem.* **1998**, 19, 1639.
- [8] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, P. S. Shenkin, *J. Med. Chem.* **2004**, 47, 1739.
- [9] R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin, D. T. Mainz, *J. Med. Chem.* **2006**, 49, 6177.
- [10] O. Trott, A. J. Olson, *J. Comput. Chem.* **2010**, 31, 455.
- [11] W. J. Allen, T. E. Balius, S. Mukherjee, S. R. Brozell, D. T. Moustakas, P. T. Lang, D. A. Case, I. D. Kuntz, R. C. Rizzo, *J. Comput. Chem.* **2015**, 36, 1132.
- [12] M. L. Verdonk, I. Giangreco, R. J. Hall, O. Korb, P. N. Mortenson, C. W. Murray, *J. Med. Chem.* **2011**, 54, 5422.
- [13] A. P. Graves, D. M. Shivakumar, S. E. Boyce, M. P. Jacobson, D. A. Case, B. K. Shoichet, *J. Mol. Biol.* **2008**, 377, 914.
- [14] T. E. Cheatham 3rd, J. Srinivasan, D. A. Case, P. A. Kollman, *J. Biomol. Struct. Dyn.* **1998**, 16, 265.
- [15] D. A. Pearlman, *J. Med. Chem.* **2005**, 48, 7796.
- [16] T. Steinbrecher, D. A. Case, A. Labahn, *J. Med. Chem.* **2006**, 49, 1837.
- [17] M. Ragoza, T. Masuda, D. R. Koes, *Chem. Sci.* **2022**, 13, 2701.
- [18] X. Peng, S. Luo, J. Guan, Q. Xie, J. Peng, J. Ma, *Proc. 39th Int. Conf. Mach. Learn. Proc. Mach. Learn. Res.*, Vol. 162, PMLR, Baltimore **2022**, p. 17644.
- [19] G. Corso, H. Stark, B. Jing, R. Barzilay, T. Jaakkola, *Int. Conf. Learn. Representations (ICLR)*. **2023**.
- [20] M. Buttenschoen, G. Morris, C. Deane, *Chem. Sci.* **2024**, 15, 3130.
- [21] C. Harris, K. Didi, A. Jamasb, C. Joshi, S. Mathis, P. Lio, T. Blundell, Benchmarking Generated Poses: How Rational is Structure-based Drug Design with Generative Models. arXiv preprint arXiv: 2308.07413. **2023**.
- [22] V. Salmasso, S. Moro, *Front. Pharmacol.* **2018**, 9, 923.
- [23] V. Salmasso, M. Sturlese, A. Cuzzolin, S. Moro, *Structure* **2017**, 25, 655.
- [24] M. Bissaro, G. Bolcato, M. Pavan, D. Bassani, M. Sturlese, S. Moro, *ChemMedChem* **2021**, 16, 2075.
- [25] M. Bissaro, M. Sturlese, S. Moro, *Drug Discov. Today* **2020**, 25, 1693.
- [26] G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, *J. Comput. Aided Mol. Des.* **2013**, 27, 221.
- [27] M. B. Peters, Y. Yang, B. Wang, L. Füstí-Molnár, M. N. Weaver, K. M. Merz Jr., *J. Chem. Theory Comput.* **2010**, 6, 2935.
- [28] C. Dickson, L. Rosso, R. Betz, R. Walker, I. Gould, *Soft Matter* **2012**, 8, 9617.
- [29] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, 25, 1157.
- [30] D. A. Case, H. M. Aktulga, K. Belfon, I. Y. Ben-Shalom, J. T. Berryman, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, G. A. Cisneros, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, G. Giambasu, M. K. Gilson, H. Gohlke, A. W. Goetz, R. Harris, S. Izadi, S. A. Izmailov, K. Kasavajhala, M. C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K. M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K. A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C. L. Simmerling, N. R. Skrynnikov, J. Smith, J. Swails, R. C. Walker, J. Wang, J. Wang, H. Wei, R. M. Wolf, X. Wu, Y. Xiong, Y. Xue, D. M. York, S. Zhao, P. A. Kollman, *Amber 2022*, University of California, San Francisco **2022**.
- [31] C. Tian, K. Kasavajhala, K. A. A. Belfon, L. Raguette, H. Huang, A. N. Miguels, J. Bickel, Y. Wang, J. Pincay, Q. Wu, C. Simmerling, *J. Chem. Theory Comput.* **2020**, 16, 528.
- [32] P. Mark, L. Nilsson, *J. Comput. Chem.* **2002**, 23, 1211.
- [33] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, J. R. Haak, *J. Chem. Phys.* **1984**, 81, 3684.
- [34] R. Elber, A. P. Ruymgaart, B. Hess, *Eur. Phys. J. Spec. Top.* **2011**, 200, 211.
- [35] V. Kräutler, W. van Gunsteren, P. Hünenberger, *J. Comput. Chem.* **2001**, 22, 501.
- [36] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **1995**, 103, 8577.
- [37] Pei, J., Tseng, V. S., Conference, P., & Goebel, R. (2013). *Advances in Knowledge Discovery (Issue April)*, Gold Coast.
- [38] B. R. Miller 3rd, T. D. McGee Jr., J. M. Swails, N. Homeyer, H. Gohlke, A. E. Roitberg, *J. Chem. Theory Comput.* **2012**, 8, 3314.
- [39] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, M. Uebayasi, *Chem. Phys. Lett.* **1999**, 313, 701.
- [40] K. Kitaura, T. Sawai, T. Asada, T. Nakano, M. Uebayasi, *Chem. Phys. Lett.* **1999**, 312, 319.
- [41] K. Takaba, C. Watanabe, A. Tokuhisa, Y. Akinaga, B. Ma, R. Kanada, M. Araki, Y. Okuno, Y. Kawashima, H. Moriwaki, N. Kawashita, T. Honma, K. Fukuzawa, S. Tanaka, *J. Comput. Chem.* **2022**, 43, 1362.
- [42] Schrödinger, LLC, *Protein Preparation Wizard; Epik*, Schrödinger, LLC, New York, NY **2021**.
- [43] M. Suenaga, *J. Comput. Chem., Jpn.* **2005**, 4, 25.
- [44] M. Suenaga, *J. Comput. Chem., Jpn.* **2008**, 7, 33.
- [45] A. Heifetz, M. Aldeghi, E. I. Chudyk, D. G. Fedorov, M. J. Bodkin, P. C. Biggin, *Biochem. Soc. Trans.* **2016**, 44, 574.
- [46] M. J. Harvey, G. Giupponi, G. D. Fabritiis, *J. Chem. Theory Comput.* **2009**, 5, 1632.

- [47] S. Aksoy, R. M. Haralick, *Pattern Recognit. Lett.* **2001**, 22, 563.
- [48] Schrödinger, LLC, *Schrödinger Release 2023-1. LigPrep*, Schrödinger, LLC, New York, NY **2021**.
- [49] W. Sherman, H. S. Beard, R. Farid, *Chem. Biol. Drug Des.* **2006**, 67, 83.
- [50] S. A. Amin, T. Jha, *Eur. J. Med. Chem.* **2020**, 201, 112559.
- [51] V. Thiel, K. A. Ivanov, Á. Putics, T. Hertzog, B. Schelle, S. Bayer, B. Weißbrich, E. J. Snijder, H. Rabenau, H. W. Doerr, A. E. Gorbalenya, J. Ziebuhr, *J. Gen. Virol.* **2003**, 84, 2305.
- [52] W. Rut, Z. Lv, M. Zmudzinski, S. Patchett, D. Nayak, S. J. Snipas, F. El Oualid, T. T. Huang, M. Bekes, M. Drag, S. K. Olsen, *Sci. Adv.* **2020**, 6, eabd4596.
- [53] C. Ma, M. D. Sacco, Z. Xia, G. Lambrinidis, J. A. Townsend, Y. Hu, X. Meng, T. Szeto, M. Ba, X. Zhang, M. Gongora, F. Zhang, M. T. Marty, Y. Xiang, A. Kolocouris, Y. Chen, J. Wang, *ACS Cent. Sci.* **2021**, 7, 1245.
- [54] C. Ma, J. Wang, *ACS Pharmacol. Transl. Sci.* **2022**, 5, 102.
- [55] M. I. Ismail, H. M. Ragab, A. A. Bekhit, T. M. Ibrahim, *Comput. Biol. Med.* **2021**, 131, 104295.
- [56] J. Osipiuk, S. A. Azizi, S. Dvorkin, M. Endres, R. Jedrzejczak, K. A. Jones, S. Kang, R. S. Kathayat, Y. Kim, V. G. Lisnyak, S. L. Maki, V. Nicolaescu, C. A. Taylor, C. Tesar, Y. A. Zhang, Z. Zhou, G. Randall, K. Michalska, S. A. Snyder, B. C. Dickinson, et al., *Nat. Commun.* **2021**, 12, 743.
- [57] D. J. Calleja, N. Kuchel, B. G. C. Lu, R. W. Birkinshaw, T. Klemm, M. Doerflinger, J. P. Cooney, L. Mackiewicz, A. E. Au, Y. Q. Yap, T. R. Blackmore, K. Katneni, E. Crighton, J. Newman, K. E. Jarman, M. J. Call, B. C. Lechtenberg, P. E. Czabotar, M. Pellegrini, S. A. Charman, K. N. Lowes, J. P. Mitchell, U. Nachbur, G. Lessene, D. Komander, *Front. Chem.* **2022**, 10, 861209.
- [58] A. Douangamath, D. Fearon, P. Gehrtz, T. Krojer, P. Lukacik, C. D. Owen, E. Resnick, C. Strain-Damerell, A. Aimon, P. Ábrányi-Balogh, J. Brandão-Neto, A. Carbery, G. Davison, A. Dias, T. D. Downes, L. Dunnett, M. Fairhead, J. D. Firth, S. P. Jones, A. Keeley, G. M. Keserü, H. F. Klein, M. P. Martin, M. E. M. Noble, P. O'Brien, A. Powell, R. N. Reddi, R. Skyner, M. Snee, M. J. Waring, C. Wild, N. London, F. von Delft, M. A. Walsh, *Nat. Commun.* **2020**, 11, 1.
- [59] H. Kanazawa, T. Yamaguchi-sasaki, N. Tanaka-yamamoto, J. Yamagishi, R. Kurimoto-tsuruta, Y. Ogata, N. Ohtake, H. Angove, L. Baker, R. Harris, A. Macias, A. Robertson, A. Surgenor, H. Watanabe, K. Nakano, M. Mima, K. Iwamoto, A. Okada, I. Takata, K. Hitaka, A. Tanaka, K. Fujita, H. Sugiyama, R. E. Hubbard, *J. Med. Chem.* **2020**, 14805.
- [60] C. W. V. Gmbh, C. Kga, J. J. Barker, O. R. Barker, R. Boggio, V. Chauhan, R. K. Y. Cheng, V. Corden, S. M. Courtney, V. M. Falque, F. Fusar, M. Gardiner, M. Estelle, N. Hamelin, T. Hesterkamp, O. Ichihara, R. S. Jones, O. Mather, C. Mercurio, S. Minucci, C. A. G. N. Montalbetti, A. Müller, D. Patel, B. G. Phillips, M. Varasi, M. Whittaker, D. Winkler, C. J. Yarnold, B. G. Phillips, *ChemMedChem* **2009**, 4, 963.
- [61] M. J. Hartshorn, C. W. Murray, A. Cleasby, M. Frederickson, I. J. Tickle, *J. Med. Chem.* **2005**, 48, 403.
- [62] A. Friberg, D. Vigil, B. Zhao, R. N. Daniels, J. P. Burke, P. M. Garcia-barrantes, D. Camper, B. A. Chauder, T. Lee, E. T. Olejniczak, S. W. Fesik, *J. Med. Chem.* **2013**, 56, 15.
- [63] J. Benesty, J. Chen, Y. Huang, I. Cohen, *Noise Reduction in Speech Processing*, Springer, Heidelberg **2009**, p. 37.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** P. E. G. F. Ibrahim, F. Zuccotto, U. Zachariae, I. Gilbert, M. Bodkin, *J. Comput. Chem.* **2024**, 45(20), 1762. <https://doi.org/10.1002/jcc.27370>