biology

MDPI

Article

Ligand Design Using Unique Conformations to Preferentially Dock a Specific Site on Collagen-Bound MMP1

Anthony Nash, Chase Harms and Susanta K. Sarkar *

School of Molecular Sciences, Arizona State University, Tempe, AZ 85287, USA; anash19@asu.edu (A.N.)
* Correspondence: susanta.sarkar@asu.edu

Simple Summary

In this paper, we establish a substrate-specific allosteric ligand-design strategy that integrates all-atom molecular dynamics simulations, conformational clustering, machine-learning-based ligand generation, and surface-wide docking-based selectivity screening, using collagen-bound MMP1 as a model system. We show that collagen binding reshapes the conformational ensemble of MMP1, creating unique conformational states that are absent or poorly sampled in the free enzyme. By identifying these substrate-specific conformations, generating candidate ligands based on dynamic fingerprints around the collagen-specific allosteric residue R405, and docking these compounds into binding pockets on the MMP1 surface, we identify ligands predicted to preferentially dock at the collagen-specific R405 region rather than competing pockets. These results establish a generalizable computational framework for prioritizing candidate ligands that may recognize biologically relevant substrate-bound conformations rather than static protein structures alone. Substrate-specific allosteric ligand design may offer new opportunities to develop selective modulators of dynamic protein systems, pending experimental validation of binding and functional effects.

Abstract

Precise site-specific ligand design remains a major challenge in structure-based drug discovery. Most existing approaches screen ligands against binding pockets identified from static protein structures obtained by X-ray crystallography, NMR spectroscopy, cryo-electron microscopy, or AlphaFold predictions. However, protein function is governed by a structure–dynamics–function relationship, and ligand screening that does not account for binding competition across the protein surface or the receptor's dynamic, substrate-dependent conformational states remains incomplete. Substrate-specific conformations are underexplored and may offer new opportunities for selective ligand design, although systematic workflows to identify and exploit such states remain limited. Previously, we showed that collagen alters matrix metalloprotease-1 (MMP1) dynamics and that R405 is a collagen-specific allosteric residue exhibiting strong dynamic correlations with the catalytic site. Here, we present a computational framework for substrate-specific allosteric ligand design using collagen-bound MMP1 as a model system. We characterized the conformational dynamics of free and collagen-bound MMP1 by all-atom molecular dynamics simulations, clustered the resulting conformational ensembles, and identified conformations unique to the collagen-bound state. These conformations were used as structural templates for machine-learning-based generation of approximately 150,000 candidate ligands, which were subsequently docked against both the R405-centered region and all detectable binding pockets on the MMP1 surface. Several candidate ligands were predicted to dock preferentially at the R405 region by at least 0.3 kcal/mol compared with competing surface

Check for updates

Received: 16 May 2026
Revised: 7 July 2026
Accepted: 13 July 2026
Published: 16 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

pockets. Together, these results establish a generalizable computational workflow for identifying candidate ligands predicted to preferentially dock to substrate-specific allosteric conformations and provide a foundation for future experimental validation of selective allosteric modulation.

Keywords: substrate-specific allostery; substrate-specific drug design; collagen; MMP1

1. Introduction

MMPs are zinc-dependent enzymes involved in diverse functions in the human body [1]. MMPs pose a challenge for drug targeting because inhibiting the active site affects both intended and other critical functions. As such, early inhibitors targeting the catalytic zinc-binding sites lacked specificity and produced dose-limiting toxicities, motivating a shift toward exosites and allosteric regulatory regions [2]. Therefore, controlling one specific function of MMPs is necessary to reduce the toxicity of drugs that target MMPs. In this paper, we focus on MMP1, a collagenase that cleaves triple-helical collagen and plays a central role in tissue remodeling and pathological invasion. Structural studies of collagen-bound MMP1 revealed that both the catalytic and hemopexin domains contribute to its function [3], and allosteric communications from the hemopexin domain are essential for cleaving triple-helical type-1 collagen [4].

Previously, we showed that the presence of a bound substrate alters MMP1 dynamics and that substrate-specific allosteric residues distant from the active site exhibit strong dynamic correlations with the active site [5,6]. Identification of these substrate-specific allosteric residues provides an opportunity to selectively modulate one MMP1 function without broadly affecting other functions. The allosteric residues of MMP1 vary with the bound substrate, including the absence of a substrate in free MMP1. For collagen-bound MMP1, we identified R405 as an allosteric residue that shows a strong dynamic correlation with the active-site residue E219. From a biophysical perspective, these observations indicate that substrate binding reshapes the underlying conformational ensemble of MMP1 and establishes long-range dynamic coupling between distal residues. Such ensemble redistribution is a hallmark of allosteric regulation in flexible multidomain proteins [7]. In a previous study, we computationally screened a subset of molecules from the ZINC database against the region around R405 on the MMP1 surface and found that predicted ligand-binding affinity at R405 depends on the presence of collagen, with some molecules predicted to dock more favorably at the allosteric site than at other binding pockets [8]. In the present work, we therefore use machine-learning-based ligand generation around R405 to design candidate ligands predicted to preferentially dock at the collagen-specific allosteric region rather than at other sites on MMP1.

Since the emergence of computational drug discovery in the 1970s, the field has seen remarkable progress enabled by molecular dynamics (MD) simulations [9], high-performance computing facilities [10], on-demand virtual libraries of drug-like small molecules containing billions of ligands [11], an increasing number of available crystal structures [12], structural predictions by AlphaFold [13], and recent rapid progress in artificial intelligence [14]. Advances in MD simulations have been particularly important for capturing time-dependent fluctuations in protein structures and resolving transient conformational states that are not accessible through static structural methods alone [15–17]. We identified conformational fingerprints specific to collagen-bound MMP1. These fingerprints represent physically distinct local conformational states within the broader ensemble and provide a quantitative description of how substrate engagement

modifies accessible MMP1 conformations. We then used a machine-learning approach to generate candidate ligands based on the dynamic fingerprints unique to collagen-bound MMP1 at R405 [18]. We identified candidate ligands predicted by docking to score more favorably at the R405-centered region than at other detected MMP1 surface pockets.

This work establishes a general computational framework for substrate-specific allosteric ligand design that integrates the identification of substrate-specific allosteric sites, dynamic conformational fingerprinting at these sites, machine-learning-based candidate ligand generation, and docking-based comparison across competing surface pockets. By evaluating candidate ligands against dynamic substrate-bound conformations rather than static protein structures alone, this strategy addresses a major limitation of traditional structure-based drug design. The approach provides a quantitative, transferable workflow for prioritizing candidate ligands predicted to engage substrate-dependent allosteric regions, laying the foundation for future experimental validation of binding, selectivity, and functional modulation.

2. Results and Discussion

Collagen-bound MMP1 exhibits restricted conformational dynamics compared to free MMP1. To investigate changes in MMP1 dynamics due to collagen binding, we compared MD simulations of free and collagen-bound MMP1. We started from the crystal structure (PDB ID: 4AUO) of collagen-bound inactive MMP1 [3]. Then, we computationally restored the catalytic glutamate by reverting the inactive E219A crystallographic mutation to E219 and then constructed one model for free MMP1 and another for collagen-bound MMP1. Both models were equilibrated and simulated using three sets of randomly chosen starting velocities. Each production trajectory ran for 500 ns, yielding a cumulative simulation time of 1.5 μs for each system (see Methods for details).

The catalytic domain exhibited greater intrinsic mobility than the hemopexin domain in the free systems, consistent with prior experimental and computational observations. In collagen-bound systems, global domain fluctuations were significantly reduced, indicating that substrate interactions stabilize specific conformational states. This effect is supported by residue-level root-mean-squared fluctuation (RMSF) analysis, which shows decreased flexibility across multiple regions in the collagen-bound systems relative to the free systems. Residue flexibility was mapped onto the structures using RMSF values calculated every 20 frames. System structures were captured throughout the simulation trajectory and superimposed to provide a visual reference for the extent of flexibility across all residues (Figure 1).

These results show that collagen binding restricts conformational variability and stabilizes distinct interdomain configurations in MMP1 relative to the free enzyme. The reduced flexibility defines substrate-dependent conformational states that provide a structural basis for identifying collagen-specific dynamic fingerprints. These fingerprints serve as the structural templates for the computational design and screening of ligands predicted to preferentially dock to the collagen-specific allosteric region in the subsequent analyses.

Collagen-bound MMP1 has unique conformations distinct from those of free MMP1. To quantify substrate-dependent structural divergence, we compared representative conformations from clustered MD simulation frames of free and collagen-bound MMP1. Figure 2A presents a map of structural similarity scores computed between clusters derived from the dynamics of free and collagen-bound MMP1. Similarity measurements of conformation clusters were performed using conformations around the collagen-specific "fingerprint" of R405, within a sphere of radius 12 Å centered on its alpha-carbon atom. Structural similarity scores were calculated using Deeply Tough, a deep learning-based method for quantifying pocket and local structural similarity across protein conformations [19]. Scores ranged

from −1 to 0, indicating maximal structural dissimilarity and identical local structural environments, respectively. Figure 2B shows the total similarity score across all free MMP1 clusters for each cluster of collagen-bound MMP1 conformations.

We identified three conformational clusters with the lowest mean similarity scores relative to free-state conformations: -0.40 ± 0.02 (mean ± SEM, ID# 370), -0.55 ± 0.03 (mean ± SEM, ID# 375), and -0.46 ± 0.02 (mean ± SEM, ID# 405). These clusters, therefore, represent structural states uniquely enriched in the collagen-bound system and minimally sampled by free MMP1. We expected to generate distinct local geometries and physicochemical environments that influence ligand recognition. Consequently, structural features governing substrate-dependent drug binding are most likely to be captured within these uniquely collagen-associated conformational clusters, making them prime candidates for targeted ligand design.

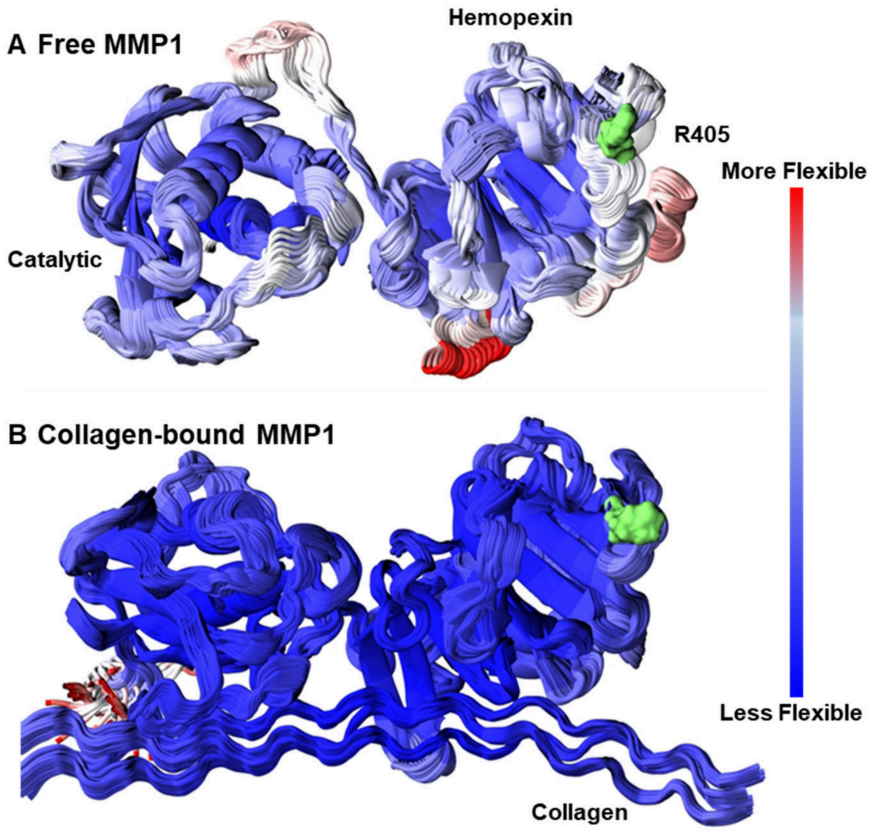


Figure 1. Dynamic flexibility of free and collagen-bound MMP1. Residue flexibility of (**A**) free MMP1 and (**B**) collagen-bound MMP1. The catalytic domain (F100–Y260) is on the left, the hemopexin domain (D279–C466) is on the right, and the two domains are connected by the linker region (G261–C278). Residue flexibility is mapped onto the structures using RMSF values calculated every 20 frames, with blue indicating the least mobile residues. Structures show the superposition of frames throughout the MD simulation. Collagen binding significantly reduces overall MMP1 flexibility. Residue R405 (in green) is a collagen-specific allosteric residue having strong correlations with the active site residue E219. We use the unique collagen-specific conformation around R405 to generate candidate ligands predicted to preferentially dock at sites around R405.

Collagen-specific allosteric fingerprints on MMP1 are dynamic. Traditional structure-based drug discovery commonly relies on screening compounds against static protein structures, implicitly assuming that binding pockets remain structurally stable and uniquely targetable across functional states [20]. However, proteins are dynamic and adopt an ensemble of conformations, and ligand or substrate binding can alter the distribution of conformations. The top three unique clusters (IDs 370, 375, and 405) in Figure 2 differ in

their three-dimensional arrangements, as shown in Figure 3. Separate enlarged views of the three fingerprints are provided in Figure S1. As such, collagen-specific allosteric fingerprints that surround defined residues represent dynamic, substrate-dependent screening targets. By incorporating conformational variability around a selected site, we aimed to generate candidate ligands predicted to preferentially dock to these biologically meaningful regions rather than nonspecifically interact with multiple surface pockets. Such dynamic, site-selective docking strategies may support future efforts to develop more selective allosteric modulators, pending experimental validation.

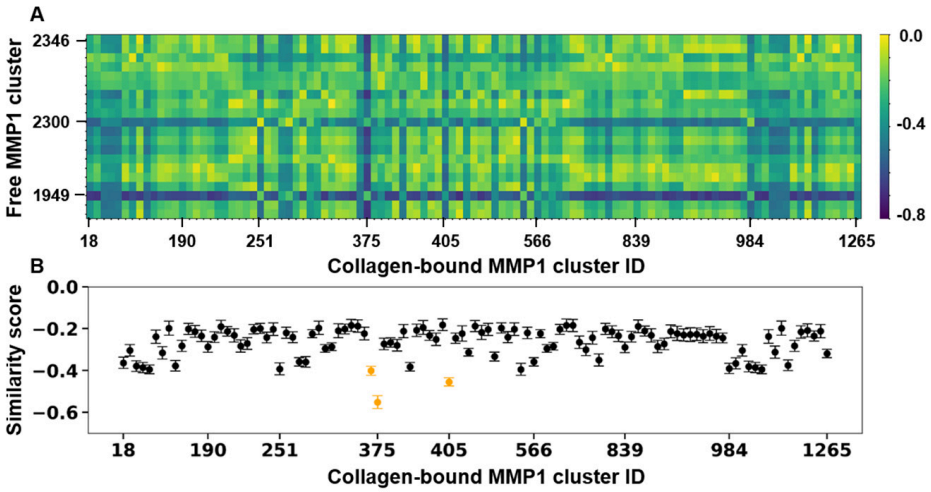


Figure 2. Identification of unique conformations of collagen−bound MMP1. (**A**) Pairwise similarity map between conformation clusters of free and collagen-bound MMP1. (**B**) The mean and standard error of the mean of similarity scores between each collagen-bound MMP1 conformation cluster and all clusters of free MMP1. The three clusters for collagen-bound MMP1 are shown in yellow.

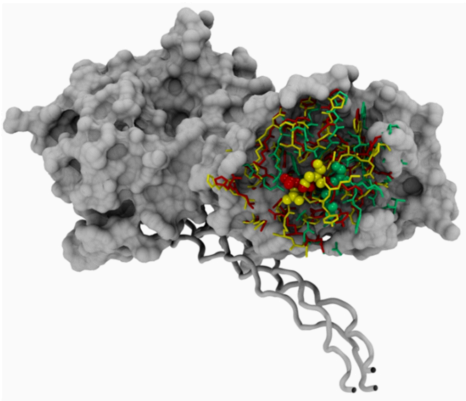


Figure 3. Dynamic collagen-specific allosteric fingerprints on MMP1. Three collagen-specific unique fingerprints are projected onto the protein surface, with each color representing one of three structurally distinct states. The position of R405 is indicated by spheres.

Candidate ligands generated using dynamic allosteric fingerprints yield favorable docking scores. For each of the three dynamic fingerprints in Figure 3, we generated candidate ligands using an approach called the Geometry-Complete Diffusion Model [18]. For each fingerprint, residues within 12 Å of R405 were defined as the local protein environment, and this region was used as a constraint for generating drug-like molecules. An iterative design and curation process was applied with a target of 50,000 unique candidate ligands per fingerprint, while enforcing uniqueness both within each fingerprint-specific set and across all three sets. Molecules were subsequently filtered using RDKit-based criteria [21] and PoseBusters validation [22], with a threshold of 0.80 applied to retain

ligands satisfying at least 80% of drug-likeness checks. Molecules failing these criteria were excluded, while accepted candidate ligands were accumulated into fingerprint-specific datasets, yielding three structurally conditioned and chemically distinct ligand libraries.

The resulting ligand sets were evaluated with AutoDock GNINA (version 1.3.2) docking [23], with five poses generated per ligand and the search space centered within 12 Å of R405 to fully encapsulate the fingerprint-defined region. This redocking framework enabled direct comparison of predicted docking behavior across all generated compounds. The distribution of minimized docking scores (kcal/mol) demonstrates that ligands generated from each fingerprint exhibit favorable docking within the targeted region (Figure 4). A substantial proportion of ligands fall within the range of approximately −2 to −5 kcal/mol, with a subset showing even more favorable values. As shown in Figure 4, the three fingerprint-derived ligand sets follow closely aligned score distributions, indicating that conditioning ligand generation on dynamic allosteric fingerprints yields compounds that consistently interact with the same local environment and have comparable docking scores.

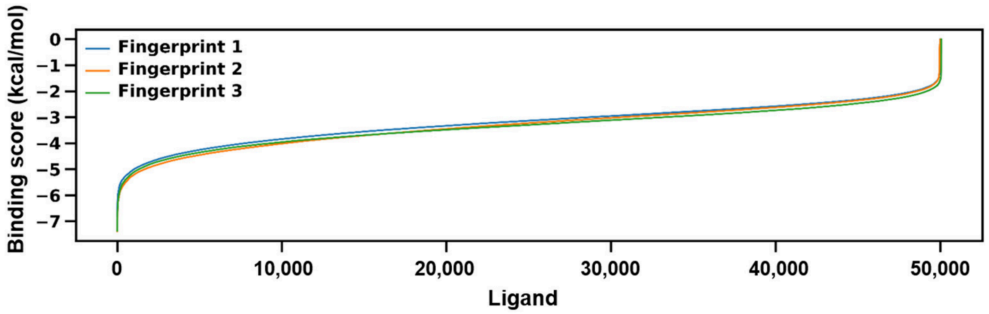


Figure 4. Docking scores of generated ligands. Distribution of docking scores (kcal/mol) for three compound libraries derived from distinct fingerprints. All compounds exhibit favorable binding, with scores below 0 kcal/mol. A small subset of compounds lies near 0 kcal/mol or extends to more favorable negative energies. The three libraries display similar distributions of docking scores.

Detected binding pockets on the MMP1 surface are not centered at the collagen-specific allosteric residue R405. We identified potential binding pockets with solvent-exposed regions on the MMP1 surface using Fpocket analysis [24]. To this end, we considered whether alpha carbon atoms were solvent-exposed, ensuring that pocket identification was confined to regions physically accessible to ligands. Collagen-bound MMP1 yielded more detected pockets than free MMP1. More pockets for collagen-bound MMP1 are likely due to the presence of collagen, which contributes additional surface features and interfacial geometries that Fpocket recognizes as candidate binding regions. The statistical distribution of Fpocket's Drug Scores further differentiates the two systems. Collagen-bound MMP1 exhibits a higher Drug Score of 0.07 ± 0.20 (mean ± SD) across 107 identified pockets. In contrast, free MMP1 displays a lower score of 0.02 ± 0.07 (mean ± SD) across 90 identified pockets.

The highest-scoring pocket for free MMP1, with a Drug Score of 0.48, is near the active site in the catalytic domain, indicating that this region remains the most favorable predicted binding site in the absence of a substrate (Figure 5A). In the collagen-bound system, the maximum Drug Score increases to 0.93, indicating the presence of a highly druggable predicted pocket, a larger structural region that extends from the linker region to near the catalytic domain (Figure 5B). The minimum Drug Score in both systems is 0.00, with multiple instances observed in each case, indicating that a subset of the detected pockets is not predicted to be druggable under the scoring scheme used by Fpocket. We used all identified pockets as docking targets, irrespective of Drug Scores, for subsequent candidate ligand screening. This approach ensures that the full spectrum of surface-accessible regions is investigated, allowing downstream analyses to evaluate predicted docking behavior

across both high-scoring and low-scoring pockets in a consistent and unbiased manner. Note that there are no high-scoring binding pockets near R405 (Figure 5), which would normally indicate that R405 is not a conventionally druggable site. Nevertheless, as shown next, some candidate ligands generated from dynamic allosteric fingerprints around R405 are predicted to dock more favorably at R405 than at the high-scoring binding pockets shown in Figure 5.

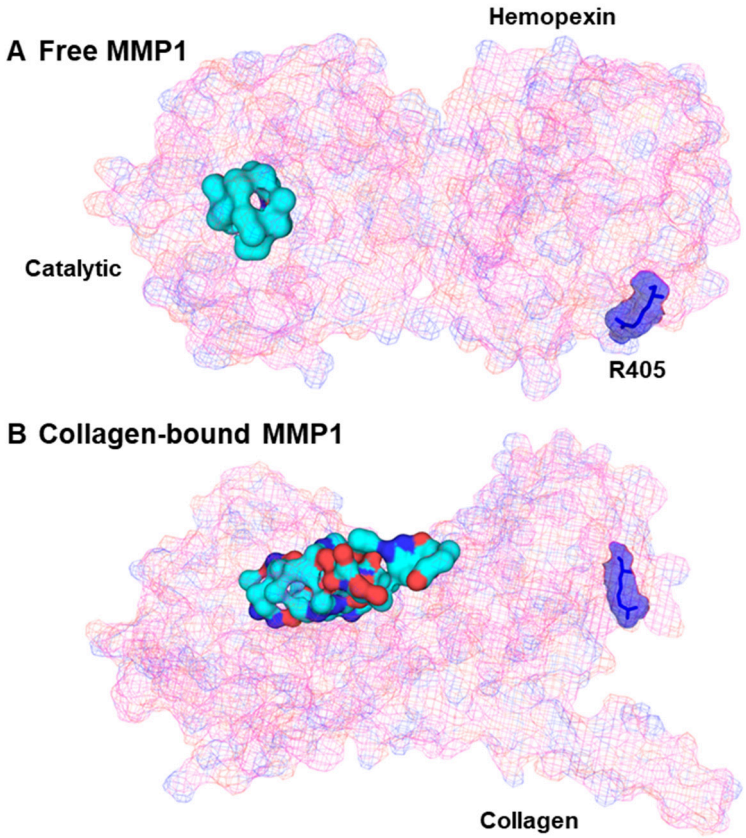


Figure 5. Binding pockets on MMP1. Locations of binding pockets for (**A**) free MMP1 and (**B**) collagen-bound MMP1. In both panels, the full protein is shown in wireframe, with maximal binding pockets rendered as surface representations. The position of R405, corresponding to the allosteric fingerprint site, is highlighted as a dark blue semi-transparent surface. Cyan, blue, and red represent carbon, nitrogen, and oxygen atoms.

Fingerprint-conditioned candidate ligands are predicted to dock preferentially to the R405 allosteric region compared to other sites on MMP1. We generated 50,000 ligands per fingerprint (Figure 4) and selected the top 5000 based on docking scores. This reduction left three curated libraries of high-scoring, fingerprint-conditioned candidate ligands. For the three fingerprints shown in Figure 3, we had a total of 15,000 designed candidate ligands. To compare predicted docking behavior at R405 and across all identified pockets, we used AutoDock GNINA docking [23] within an area with a 12 Å radius centered on R405. The docking search space was centered on the geometric center of each pocket, and the number of poses per docking run increased to twenty. For each ligand–pocket pair, a representative pose was determined using the median medoid structure. Mean docking scores were calculated across all twenty poses for each pair. These values were compared to the corresponding docking score obtained at the R405-centered fingerprint for the same candidate ligand. Candidate ligands were retained when their mean docking score at the R405-centered site was more favorable than at any alternative pocket, thereby identifying

compounds predicted to preferentially dock at the allosteric region around R405, as defined by the collagen-bound state.

Figure 6 shows that the top five candidate ligands are predicted to dock near R405 with scores at least 0.3 kcal/mol more favorable than those at competing binding pockets on the MMP1 surface. At body temperature (310 K), this energetic difference corresponds to an estimated ~1.6-fold preference for the R405-centered site under idealized Arrhenius and Eyring assumptions with comparable prefactors. Representative 2D ligand–receptor interaction maps and compound structures for the top-ranked candidates are shown in Figure S2. Because docking scores are approximate computational ranking metrics rather than experimentally measured binding free energies, kinetic barriers, or dissociation rates, this estimate should be interpreted only as a qualitative indication of preferential docking. These results identify candidate ligands predicted to preferentially dock at the collagen-specific R405 region rather than at other detected surface pockets.

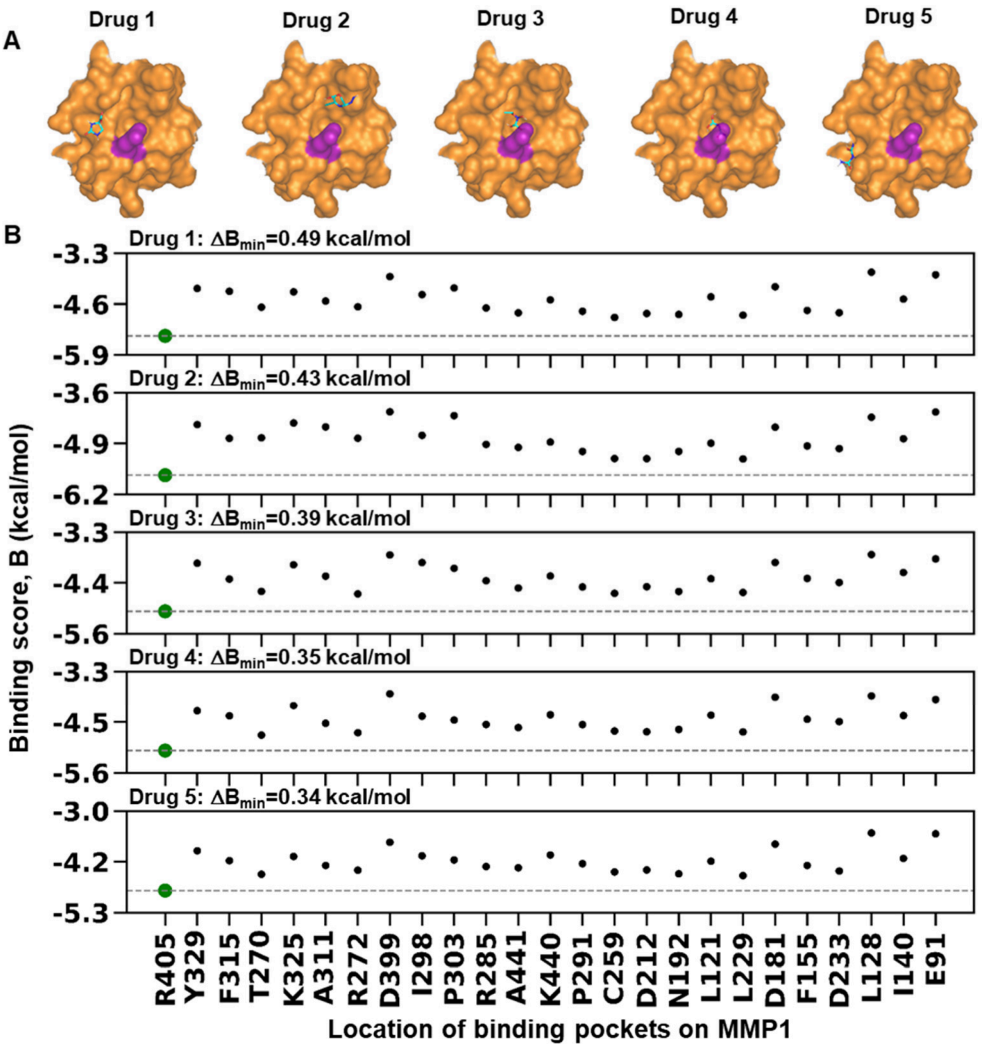


Figure 6. Identification of designed candidate ligands predicted to preferentially dock at the targeted site. (**A**) Top five candidate ligands (green sticks) docked near R405 (purple). (**B**) Docking scores of the top five candidate ligands at R405 (green solid circles) and other binding pockets across the MMP1 surface. These candidate ligands are predicted to dock at R405 with at least 0.3 kcal/mol more favorable scores.

3. Conclusions

This study establishes a general computational strategy for identifying candidate ligands predicted to preferentially dock at specific protein sites within a broader substrate-specific allosteric ligand-design framework (Figure 7). We integrated experimentally supported molecular dynamics simulations, substrate-specific dynamic allosteric fingerprint identification, machine-learning-based ligand generation, and surface-wide docking-based selectivity screening to prioritize candidate ligands predicted to favor specific regions on protein surfaces. Using collagen-bound MMP1 as a model, we identified the top three dynamic allosteric fingerprints near R405 that were structurally distinct from those of free MMP1. These findings support the idea that substrate binding can create functional allosteric states that are not apparent from the unbound enzyme alone.

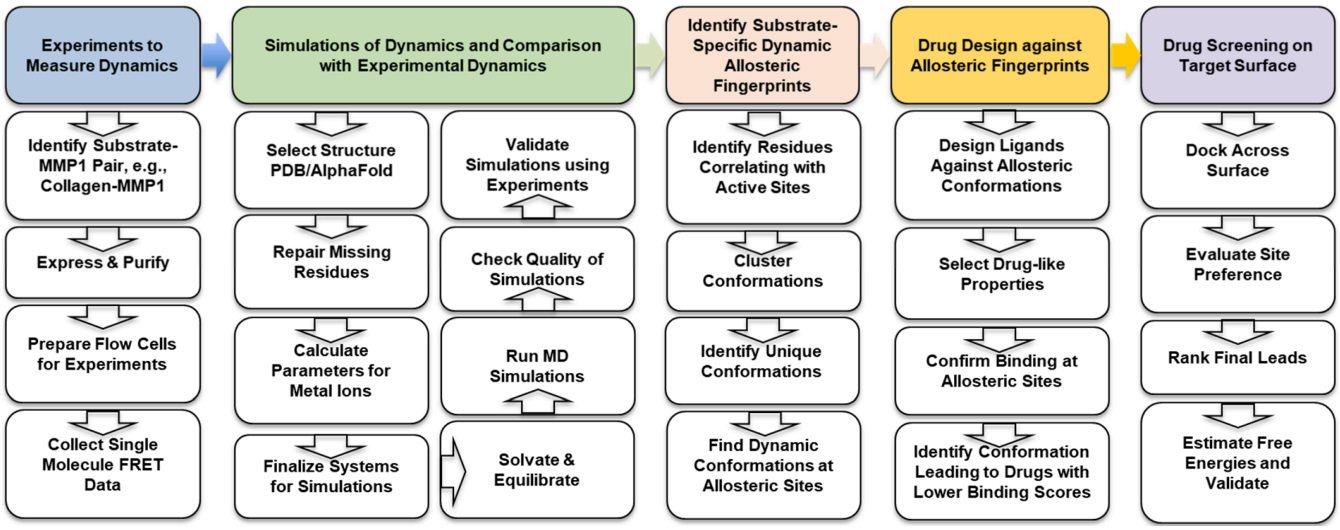


Figure 7. A framework for designing candidate ligands predicted to preferentially dock at a specific site on MMP1. Two-way feedback between experiments and simulations enables validation of simulations. We used validated simulations to identify substrate-specific allosteric fingerprints and to design ligands based on dynamic conformational arrangements at the targeted site. We identified candidate ligands predicted to dock more favorably at the target site than at other sites on the protein surface.

We then used these collagen-specific fingerprints to generate candidate ligand libraries and screened candidate molecules not only against the R405-centered site but also across alternative pockets on the MMP1 surface. This surface-wide comparison is important because earlier MMP inhibitor programs were limited by poor selectivity and musculoskeletal toxicity, motivating approaches that avoid broad catalytic-site inhibition [25]. Our results show that fingerprint-conditioned candidate ligands are predicted to dock more favorably at the collagen-specific R405 region than at competing pockets, providing a quantitative computational route for prioritizing site-selective allosteric ligand candidates.

Overall, this work addresses a central limitation of conventional in silico structure-based drug design: the tendency to optimize ligands against static, predefined pockets without testing whether the same ligands are predicted to dock more favorably elsewhere on the protein. By combining substrate-dependent conformational selection with full-surface docking-based screening, the framework provides a general strategy for identifying candidate ligands predicted to preferentially dock at specific protein sites. Future experimental studies will be required to test whether these candidates bind to the R405-centered region, modulate MMP1 activity allosterically, and produce substrate-selective functional effects.

4. Materials and Methods

4.1. Structural Preparation of Free and Collagen-Bound MMP1 Systems for Simulations

We prepared two systems for simulation: a collagen-bound MMP1 system and a free MMP1 system in which the collagen substrate was removed. Both systems were modeled in the catalytically active form with glutamate at residue 219 (E219). The free system was generated by removing the collagen from the complex and subsequently equilibrating the protein to reduce structural bias associated with collagen binding. We obtained the starting structure from PDB ID 4AUO [3]. The inactive crystallographic mutation E219A present in PDB ID 4AUO was computationally reverted to the wild-type catalytic glutamate (A219E) before equilibration and production molecular dynamics simulations. No additional active-site remodeling was performed. We identified missing atoms and residues from the PDB metadata and reconstructed these using Modeller (version 10.8, r13157) [26]. No prodomain was included. We capped the termini of MMP1 using ACE at the *N*-terminus and NME at the *C*-terminus.

We constructed the collagen model as a periodic molecule spanning the simulation box by forming covalent bonds between the termini of each individual polypeptide chain, such that the molecule crosses one boundary of the unit cell and continues through the opposite boundary. This configuration was used to approximate the rigidity of a collagen fibril, rather than allowing short collagen-like molecules to diffuse in the solvent while bound to MMP1. The collagen triple helix was built to undergo sufficient turns to maintain collagen-like structural behavior, and the system was configured such that MMP1 did not interact with its periodic images. All systems were placed in cubic simulation boxes. For both the collagen-bound and free systems, box dimensions were 7.35 nm $\times$ 9.65 nm $\times$ 8.93 nm. Periodic boundary conditions were applied in three dimensions. We solvated all systems using the TIP3P water model [27], and added chloride ions to neutralize the system's charge.

We described the protein using the AMBER99SB-ILDN force field [28]. We parameterized the catalytic and structural Zn^{2+} and Cu^{2+} ions using values derived from prior quantum-chemical calculations, including force constants, equilibrium bond lengths and angles, and torsional parameters for coordinating ligands.

4.2. MD Simulation Setup

We performed molecular dynamics simulations using OpenMM (version 8.2.0) [29]. Simulations were executed on the CUDA platform using mixed precision. We maintained the temperature at 295.15 K and the pressure at 1 bar using a Monte Carlo barostat with an update frequency of 25 steps, corresponding to isotropic pressure coupling. We used a Langevin integrator [30] with a time step of 2 fs and a friction coefficient of 1 ps^{-1}. Initial velocities were assigned at 295.15 K. We treated long-range electrostatics using the particle-mesh Ewald method with a 0.8 nm cutoff [31]. Short-range nonbonded interactions were truncated at the same cutoff distance. All hydrogen-containing bonds were constrained using the HBonds constraint scheme. Periodic boundary conditions were applied to all force terms within the system. We applied harmonic positional restraints to collagen alpha carbon atoms using their initial coordinates as reference positions, with a force constant of 1000 kJ mol^{-1} nm^{-2}. Positional restraints were implemented using CustomExternalForce within OpenMM.

4.3. Energy Minimization, Equilibration, and Production Simulations

We performed energy minimization using OpenMM's minimizeEnergy() routine prior to dynamical propagation. We then carried out pre-equilibration to relax structural distortions and stabilize the system. Systems were heated from 0 K to 295.15 K over 1.0×10^6 time steps of 2 fs, corresponding to 100 ps of equilibration. From the minimized structure, three

independent simulations were initiated through separate heating procedures, each followed by 40 ns of molecular dynamics equilibration to allow relaxation of the system, including adjustment following collagen removal and equilibration of the unit cell under the applied pressure. Each simulation was then propagated for 500 ns, yielding three independent trajectories per system and a total sampling time of 1.5 µs. Coordinates were written in DCD format using the OpenMM DCDReporter at 50,000-step intervals (100 ps). Energies and simulation statistics were recorded at regular intervals during the simulation. Simulation checkpoints were written periodically to enable restart and reproducibility.

Previous single-molecule FRET experiments and all-atom molecular dynamics simulations showed that restoration of the catalytic glutamate recovers conformational dynamics characteristic of active MMP1 [4]. These prior experimental and computational results support the use of the restored E219 model as a catalytically active-state model for the present simulations.

4.4. Quality Check of Simulated Conformational Dynamics

We analyzed the final 100 ns of each trajectory. We calculated the root-mean-squared deviation using alpha-carbon atoms as the reference selection across this interval for all repeats of both the collagen-bound and free systems. The initial frame of each trajectory was used as the reference structure. We calculated the solvent accessible surface area using FreeSASA (version 2.1.2) [32]. SASA was evaluated across the entire structure for all frames in the final 100 ns of each trajectory. Hydrogens were included in the calculation. We used a probe radius of 1.40 Å and applied the default Shrake–Rupley algorithm implemented in FreeSASA.

4.5. Classification of MMP1 Conformational Dynamics Based on Solvent-Accessibility of R405

We classified frames based on the solvent-accessible surface area near R405. Frames with a surface area less than 5 $Å^2$ were classified as buried and excluded. Frames with a surface area between 5 and 30 $Å^2$ were classified as partially exposed and excluded. Frames with a surface area greater than or equal to 30 $Å^2$ were classified as fully exposed and retained. These thresholds were chosen to distinguish buried, partially exposed, and solvent-accessible states. We used retained frames to construct structurally consistent snapshot ensembles across all trajectories and systems. For each retained frame, we recorded the residue identity, chain designation, residue index, and alpha-carbon Cartesian coordinates for residues that met the surface-exposure criteria. We calculated the fraction of solvent-accessible alpha carbon atoms per frame and collated frame-wise lists of surface-exposed residues to quantify the prevalence and spatial distribution of solvent exposure across systems.

4.6. Identification of Dynamic Allosteric Fingerprints

We identified candidate binding site fingerprints from the subset of trajectory frames in which R405 was fully solvent-exposed. For the collagen-bound MMP1 ensemble, SASA filtering retained 20,000 structures for clustering. Pairwise backbone RMSD values were calculated after centering and rigid-body superposition using alpha-carbon atoms. The resulting condensed RMSD distance matrix was clustered using average-linkage hierarchical clustering, and discrete clusters were assigned using a 1.2 Å backbone RMSD cutoff. This procedure resolved the collagen-bound trajectory into 1286 conformational clusters. We defined local binding sites by selecting residues within a 12 Å radius centered on the alpha carbon of R405. Clusters containing fewer than 20 frames were excluded. Among the representative clusters selected for downstream analysis, cluster 405 contained 210 frames, cluster 375 contained 67 frames, and cluster 370 contained 62 frames. For each retained

cluster, we extracted a representative structure and isolated the local binding site. These extracted sites were stored as PDB and SDF files.

4.7. Identification of Binding Pockets Using DeeplyTough

We obtained representative structures by clustering molecular dynamics trajectories for both the free and collagen-bound systems. For each system, we retained the three dominant clusters, yielding three candidate conformations centered on R405. We compared these binding pockets using DeeplyTough (GitHub commit 7536c5b) [19]. We performed all-against-all comparisons between fingerprints to generate similarity matrices [33]. We ranked fingerprints by mean similarity score and selected the three lowest-ranking fingerprints per system.

4.8. Ligand Generation Based on Collagen-Specific Allosteric Fingerprints

We performed automated ligand generation using a diffusion-based structure-based design workflow implemented in the GCDM-SBDD framework [18]. Ligand generation was performed only for the collagen-bound system, using the three identified fingerprints. For each fingerprint, we identified binding-site residues by combining solvent accessibility and spatial proximity criteria, using R405 as the reference. Residues within 12 Å of the reference residue and exhibiting solvent accessibility were retained and used to define the binding environment.

We generated ligands iteratively, producing 500 samples per iteration with incrementally random seeds, and targeted 50,000 unique ligands per fingerprint. We enforced uniqueness using global SMILES-based deduplication. We filtered generated ligands using RDKit-based criteria to ensure chemical validity and removed duplicates. We evaluated ligands using PoseBusters (version 0.6.5) [22], retaining molecules that satisfied at least 80% of validation checks assessing chemical validity, structural integrity, protein–ligand steric compatibility, geometric strain, and topological consistency. Ligands passing all filters were aggregated into a master dataset for each fingerprint.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/biology15141169/s1. Figure S1. Dynamic collagen-specific allosteric fingerprints on MMP1. Representative conformations of fingerprint clusters corresponding to (A) #370, (B) #375, and (C) #405. Collagen-bound MMP1 is shown as a gray surface, with R405 highlighted as blue spheres. Residues within the local fingerprint environment are shown as sticks and colored according to side-chain chemistry: acidic (red), basic (blue), aliphatic (green), sulfur-containing (yellow), heterocyclic (purple), hydroxyl-containing (cyan), aromatic (orange), and amide (pink). Figure S2. Chemical structures and docked ligand-receptor interaction maps. Chemical structures of the five top-ranked compounds and their two-dimensional docked ligand-receptor interaction maps are shown. Each panel shows a docked compound together with neighboring receptor residues, labeled and colored according to side-chain chemistry. Interaction distances are shown in Å. Contact lines indicate the predicted interaction type, including polar, hydrophobic, ionic, aromatic, cation–aromatic, and van der Waals interactions. Residue colors denote the classes of acidic, basic, aliphatic, sulfur-containing, heterocyclic, hydroxyl-containing, aromatic, and amide side chains.

Author Contributions: S.K.S. conceived and designed the overall project. A.N. conceived an approach to drug design based on molecular dynamics and machine learning. A.N. ran MD simulations, implemented machine learning-based drug design, and screened drugs. A.N., C.H., and S.K.S. analyzed the results. S.K.S., A.N., and C.H. wrote the manuscript. All authors edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Institutes of Health grant number GM145210.

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest or competing financial interests.

References

1. Rodríguez, D.; Morrison, C.J.; Overall, C.M. Matrix metalloproteinases: What do they not do? New substrates and biological roles identified by murine models and proteomics. *Biochim. Biophys. Acta Mol. Cell Res.* **2010**, *1803*, 39–54. [CrossRef]
2. Vandenbroucke, R.E.; Libert, C. Is there new hope for therapeutic matrix metalloproteinase inhibition? *Nat. Rev. Drug Discov.* **2014**, *13*, 904–927. [CrossRef] [PubMed]
3. Manka, S.W.; Carafoli, F.; Visse, R.; Bihan, D.; Raynal, N.; Farndale, R.W.; Murphy, G.; Enghild, J.J.; Hohenester, E.; Nagase, H. Structural insights into triple-helical collagen cleavage by matrix metalloproteinase 1. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 12461–12466. [CrossRef] [PubMed]
4. Kumar, L.; Nash, A.; Harms, C.; Planas-Iglesias, J.; Wright, D.; Klein-Seetharaman, J.; Sarkar, S.K. Allosteric communications between domains modulate the activity of matrix metalloprotease-1. *Biophys. J.* **2020**, *119*, 360–374. [CrossRef] [PubMed]
5. Kumar, L.; Planas-Iglesias, J.; Harms, C.; Kamboj, S.; Wright, D.; Klein-Seetharaman, J.; Sarkar, S.K. Activity-dependent interdomain dynamics of matrix metalloprotease-1 on fibrin. *Sci. Rep.* **2020**, *10*, 20615. [CrossRef] [PubMed]
6. Kamboj, S.; Harms, C.; Wright, D.; Nash, A.; Kumar, L.; Klein-Seetharaman, J.; Sarkar, S.K. Identification of allosteric fingerprints of alpha-synuclein aggregates in matrix metalloprotease-1 and substrate-specific virtual screening with single molecule insights. *Sci. Rep.* **2022**, *12*, 5764. [CrossRef] [PubMed]
7. Nussinov, R.; Tsai, C.J. Allostery in disease and in drug discovery. *Cell* **2013**, *153*, 293–305. [CrossRef] [PubMed]
8. Harms, C.; Nash, A.; Sarkar, S.K. The substrate changes matrix metalloprotease-1 dynamics, allostery, and ligand binding affinity. *bioRxiv* **2025**. [CrossRef]
9. Durrant, J.D.; McCammon, J.A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **2011**, *9*, 71. [CrossRef] [PubMed]
10. Ge, H.; Wang, Y.; Li, C.; Chen, N.; Xie, Y.; Xu, M.; He, Y.; Gu, X.; Wu, R.; Gu, Q. Molecular dynamics-based virtual screening: Accelerating the drug discovery process by high-performance computing. *J. Chem. Inf. Model.* **2013**, *53*, 2757–2764. [CrossRef] [PubMed]
11. Irwin, J.J.; Tang, K.G.; Young, J.; Dandarchuluun, C.; Wong, B.R.; Khurelbaatar, M.; Moroz, Y.S.; Mayfield, J.; Sayle, R.A. ZINC20—A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *J. Chem. Inf. Model.* **2020**, *60*, 6065–6073. [CrossRef] [PubMed]
12. Burley, S.K.; Bhatt, R.; Bhikadiya, C.; Bi, C.; Biester, A.; Biswas, P.; Bittrich, S.; Blaumann, S.; Brown, R.; Chao, H.; et al. Updated resources for exploring experimentally-determined PDB structures and Computed Structure Models at the RCSB Protein Data Bank. *Nucleic Acids Res.* **2025**, *53*, D564–D574. [CrossRef] [PubMed]
13. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef] [PubMed]
14. Chan, H.S.; Shan, H.; Dahoun, T.; Vogel, H.; Yuan, S. Advancing drug discovery via artificial intelligence. *Trends Pharmacol. Sci.* **2019**, *40*, 592–604. [CrossRef] [PubMed]
15. Shaw, D.E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R.O.; Eastwood, M.P.; Bank, J.A.; Jumper, J.M.; Salmon, J.K.; Shan, Y. Atomic-level characterization of the structural dynamics of proteins. *Science* **2010**, *330*, 341–346. [CrossRef] [PubMed]
16. Hollingsworth, S.A.; Dror, R.O. Molecular dynamics simulation for all. *Neuron* **2018**, *99*, 1129–1143. [CrossRef] [PubMed]
17. Dror, R.O.; Dirks, R.M.; Grossman, J.; Xu, H.; Shaw, D.E. Biomolecular simulation: A computational microscope for molecular biology. *Annu. Rev. Biophys.* **2012**, *41*, 429–452. [CrossRef] [PubMed]
18. Morehead, A.; Cheng, J. Geometry-complete diffusion for 3D molecule generation and optimization. *Commun. Chem.* **2024**, *7*, 150. [CrossRef] [PubMed]
19. Simonovsky, M.; Meyers, J. DeeplyTough: Learning structural comparison of protein binding sites. *J. Chem. Inf. Model.* **2020**, *60*, 2356–2366. [CrossRef] [PubMed]
20. Boehr, D.D.; Nussinov, R.; Wright, P.E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **2009**, *5*, 789–796. [CrossRef] [PubMed]
21. Bento, A.P.; Hersey, A.; Félix, E.; Landrum, G.; Gaulton, A.; Atkinson, F.; Bellis, L.J.; De Veij, M.; Leach, A.R. An open source chemical structure curation pipeline using RDKit. *J. Cheminform.* **2020**, *12*, 51. [CrossRef] [PubMed]
22. Buttenschoen, M.; Morris, G.M.; Deane, C.M. PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chem. Sci.* **2024**, *15*, 3130–3139. [CrossRef] [PubMed]
23. McNutt, A.T.; Francoeur, P.; Aggarwal, R.; Masuda, T.; Meli, R.; Ragoza, M.; Sunseri, J.; Koes, D.R. GNINA 1.0: Molecular docking with deep learning. *J. Cheminform.* **2021**, *13*, 43. [CrossRef] [PubMed]

24. Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: An open source platform for ligand pocket detection. *BMC Bioinform.* **2009**, *10*, 168. [CrossRef]
25. Coussens, L.M.; Fingleton, B.; Matrisian, L.M. Matrix metalloproteinase inhibitors and cancer—Trials and tribulations. *Science* **2002**, *295*, 2387–2392. [CrossRef] [PubMed]
26. Šali, A.; Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779–815. [CrossRef] [PubMed]
27. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935. [CrossRef]
28. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J.L.; Dror, R.O.; Shaw, D.E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct. Funct. Bioinform.* **2010**, *78*, 1950–1958. [CrossRef]
29. Eastman, P.; Swails, J.; Chodera, J.D.; McGibbon, R.T.; Zhao, Y.; Beauchamp, K.A.; Wang, L.-P.; Simmonett, A.C.; Harrigan, M.P.; Stern, C.D. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017**, *13*, e1005659. [CrossRef] [PubMed]
30. Schneider, T.; Stoll, E. Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Phys. Rev. B* **1978**, *17*, 1302. [CrossRef]
31. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092. [CrossRef]
32. Mitternacht, S. FreeSASA: An open source C library for solvent accessible surface area calculations. *F1000Research* **2016**, *5*, 189. [CrossRef] [PubMed]
33. Eguida, M.; Rognan, D. Estimating the similarity between protein pockets. *Int. J. Mol. Sci.* **2022**, *23*, 12462. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.