Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Mixed-Solvent MD Suite: Modular Tools for Building and Analyzing Mixed-Solvent Molecular Dynamics Systems

Yue Qiu

sean28299@gmail.com

Jinhua Advanced Research Institute of Technology

Qing Liu

Jian-Wu Xie

Jinhua Advanced Research Institute of Technology

Research Article

Keywords: Mixed-solvent molecular dynamics (MSMD), Solvent probes, Active binding site, Cryptic pockets, GROMACS, Cpptraj

Posted Date: April 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9262224/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Journal of Computer-Aided Molecular Design on May 25th, 2026. See the published version at https://doi.org/10.1007/s10822-026-00829-y.

Abstract

Mixed-solvent molecular dynamics (MSMD) has become a useful simulation strategy for mapping protein interaction hotspots, exploring cryptic or allosteric pockets, and characterizing ligandable surface regions. However, practical MSMD studies still require labor-intensive system preparation, probe bookkeeping, format conversion, and trajectory post-processing across multiple software environments. Here, we present the Mixed-Solvent MD Suite, a modular software platform that automates these operational steps rather than proposing a new physical sampling formalism. The Suite includes (1) a fully web-based Mixed Solvent MD System Builder for protein and probe setup, concentration-controlled solvation (% w/v or mol/L), and automated generation of GROMACS and AMBER compatible input files, and (2) a local Probe GridMap Builder for grid-based probe occupancy analysis of MD trajectories using the AMBER cpptraj backend. Case studies on five benchmark proteins show that the platform reproducibly constructs MSMD systems, supports transferable analysis workflows across simulation outputs, and qualitatively recovers known ligandable regions in representative targets. Overall, the Mixed-Solvent MD Suite lowers technical barriers for MSMD studies and provides an extensible foundation for future benchmarked hotspot analysis and cryptic-site discovery workflows.

Introduction

Mixed-solvent molecular dynamics (MSMD) is a broader class of explicit-solvent simulations in which water is combined with one or more organic probe molecules to map favorable interaction regions on biomolecular surfaces [1]. Within this broader landscape, MixMD refers more specifically to the Carlson group workflow [2, 3], while related families such as SILCS [4], MDmix style solvent-site mapping [5], and more recent MSMD extensions have pursued overlapping goals using different probe sets, scoring strategies, and analysis formalisms. The core concept of MSMD is to allow organic probes, often drug-like fragments, to compete with water molecules during simulations, thereby identifying energetically favorable regions on the protein that may correspond to active or allosteric binding sites [1, 6]. Compared to traditional molecular docking or static structure-based approaches, MSMD offers a dynamic perspective by incorporating protein flexibility and solvent competition into hotspot detection [1, 7]. This is particularly valuable for uncovering cryptic sites (transient or cryptic pockets) that only emerge under specific conformational states of the protein [8]. Numerous studies have demonstrated its effectiveness in mapping active and allosteric sites on kinases, GPCRs, P450s, phosphatases, and viral proteases [8–12]. Together, these applications demonstrate that MSMD is a robust and versatile tool for druggability assessment and structure-based drug design, offering dynamic, probe-driven mapping of biomolecular surfaces that complements traditional docking and pharmacophore modeling approaches [2, 13, 14]. By facilitating the discovery of cryptic sites and providing a more realistic view of protein–ligand interactions in solution [15], MSMD plays a critical role in advancing modern computational drug discovery workflows.

Despite increasing interest in MSMD for identifying cryptic binding sites and assessing druggability [16–18], current tools remain fragmented and technically demanding. First, system construction still requires manual steps such as box generation, probe insertion, and force field setup [19–21], while general-purpose tools like CHARMM-GUI [22–24] and Packmol [25, 26] lack native support for mixing multiple probe types. This complexity raises the risk of error and limits accessibility for non-expert users. Second, trajectory analysis is similarly limited. Although cpptraj [27, 28] enables probe density mapping, it depends on command-line inputs and manual post-processing, offering minimal visualization support. Probeview [6], once a visualization add-on, is now obsolete due to reliance on deprecated Python 2.7 and outdated PyMOL versions. Moreover, tools like cpptraj are AMBER-oriented, while CHARMM-GUI targets CHARMM, with limited interoperability between GROMACS and AMBER workflows [28–31]. Critical features such as automated probe number estimation based on concentration, support for custom probe mixtures, and intuitive hotspot visualization are still lacking or only partially implemented. These limitations highlight the need for a modular, integrated, and user-friendly MSMD platform.

To address these limitations, we developed the Mixed-Solvent MD Suite, a modular and user-friendly framework that automates both system construction and trajectory analysis for MSMD applications. By integrating key functions into a unified platform, the suite simplifies complex setup procedures, reduces manual effort, and enables comprehensive analysis of probe distributions. It comprises two main components: The Mixed Solvent MD System Builder, for automated system generation, and the Probe GridMap Builder, for visualizing probe distributions from MD trajectories. The architecture of the Suite is illustrated in Fig. 2 and Fig. 3, which provides an overview of the software interface and key functionalities.

1.1 Mixed Solvent MD System Builder: Web-Based System Construction Tool

The first component, the Mixed Solvent MD System Builder (Fig. 2A) is a fully web-based tool that allows users to construct MSMD simulation systems by uploading protein structures, selecting solvent probes from a pre-defined library or user-defined molecules, and configuring system boxes. The front-end interface is developed using HTML and JavaScript, while the back-end is powered by a Flask-based Python server that manages file processing, probe insertion, and topology generation tasks. Moreover, a dedicated structure cleaning and repair tools (Figure S1), PDB Structure Preprocessing Tool, also is provided as part of the suite to facilitate this step. These tools help users prepare protein structures by addressing common issues such as missing atoms, repairing PDB files (based on PDBFixer [32]), and cleaning the structure to ensure compatibility with downstream simulation workflows. Alternatively, users may preprocess protein structures using external tools such as PyMOL, Chimera, Schrödinger's Protein Preparation Wizard, or other related software. In addition, the Builder provides two dedicated online calculators: The System Water Count Calculator (Fig. 2B) and The Probe Insertion Calculator (Fig. 2C). it offers water molecule prediction and probe insertion quantity calculation based on user-defined concentrations.

Probe molecule insertion of Mixed Solvent MD System Builder is performed directly within the GROMACS environment, which offers several advantages. GROMACS provides a flexible and efficient platform for handling large simulation systems, ensuring accurate coordinate merging, proper system neutralization, and seamless integration of protein, solvent, and probe molecules [33, 34]. By leveraging GROMACS for probe insertion, users benefit from robust topology handling, compatibility with widely used force fields, and streamlined preparation for subsequent simulation steps. For AMBER users, the system directly generates coordinate and topology files compatible with AMBER-based simulation workflows, enabling seamless construction and execution of mixed-solvent molecular dynamics simulations using AMBER tools. The Builder allows users to directly specify the number of probe molecules to be inserted into the simulation system for each selected probe type, providing fine-grained control over the composition of the mixed-solvent environment. The output includes a ready-to-use GROMACS system package with all necessary input files, including topology and coordinate files. Additionally, the system can automatically generate a combined PDB file containing both the protein and inserted probe molecules.

The Mixed Solvent MD System Builder comes with 13 commonly used probe small molecules pre-built into the system (Table S1), including acetonitrile (ACN), isopropanol (IPA), N-methylacetamide (NMA), phenol (PHN), pyrimidine (PYR), formamide (FMA), dimethyl sulfoxide (DMS), methanol (MEH), guanidinium (GUA), acetate (ACE), benzene (BNZ), toluene (TOL), and indole (IND). Each probe has distinct physicochemical properties that enable recognition of various protein microenvironments, such as hydrophobic pockets, polar or charged regions, hydrogen bonding sites, and aromatic interaction surfaces [5, 15, 35]. All thirteen probe molecules (ACN, IPA, NMA, PHN, PYR, FMA, DMS, MEH, GUA, ACE, BNZ, TOL, and IND) used in this work are fully compatible with both the CHARMM27 and AMBER99SB-ILDN force fields. The system also allows users to upload custom probe molecules for insertion, provided as a set of gro, top, and prm files. To facilitate this process, the force field files for the built-in probes are available for download, enabling users to construct compatible files for their own probe molecules. In addition to probe insertion, the Builder supports the construction of pure protein–water systems without any probes. These systems can be used to estimate the number of water molecules required for solvation, which serves as the basis for calculating the appropriate number of probe molecules to achieve a specific concentration. These integrated features provide users with a convenient and automated workflow for preparing complex mixed-solvent systems, reducing manual intervention and potential sources of error during the setup process.

1.2 Probe GridMap Builder: Local Trajectory Analysis Tool

The second component, the Probe GridMap Builder (Fig. 3), is a standalone graphical user interface (GUI) application for trajectory analysis, built on a Conda-based backend and leveraging CPPTRAJ for data processing. It supports molecular dynamics trajectories generated by GROMACS (TRR format) and AMBER (NetCDF format), enabling the generation of grid-based probe density maps for visualizing probe distributions and identifying dynamic binding hotspots. The tool automates probe occupancy calculations and generates grid-based density files (e.g., nonortho.pdb), which can be visualized in

external molecular viewers such as PyMOL or Chimera. This enables researchers to map probe distributions onto protein structures and identify dynamic binding hotspots that may correspond to cryptic or allosteric sites. In addition, the tool produces a ranked file listing probe density clusters, identified using the DBSCAN clustering algorithm. This enables the identification and prioritization of potential binding hotspots and cryptic sites based on probe accumulation data.

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [36] is a clustering algorithm that groups data points based on density connectivity. A point *p* is considered a core point if its ε-neighborhood contains at least MinPts points:

$$Core\,(p) \iff |N\epsilon\ \ (p)| \geq\ MinPts$$

Where $N_{\varepsilon(p)}$ is the ε−neighborhood of *p*:

$$N\epsilon\ \ (p) = \{\ q \in\ D||dist,(p,q) \leq\ \epsilon\ \}$$

A point *q* is directly density-reachable from *p* if $q \in N_{\varepsilon(p)}$ and *p* is a core point. A cluster is defined as a maximal set of density−connected points:

$$C = \{p \in\ D \mid\ \exists\ \text{a core point } o \text{ such that } p \text{ is density} - \text{reachable from } o\}$$

Points not belonging to any cluster are classified as noise.

Together, the Mixed-Solvent MD Suite provides an end-to-end solution for MSMD workflows, from system construction and probe placement to trajectory analysis and hotspot visualization. The distinctive features of the suite's two core components, the Mixed Solvent MD System Builder and the Probe GridMap Builder, are summarized in Table 1. By combining these tools, empowers researchers to efficiently construct, simulate, and analyze MSMD systems, facilitating the identification of cryptic pockets and dynamic interaction hotspots that may guide structure-based drug design efforts.

Table 1
Summary of the features and functionalities of the two core components in the Mixed-Solvent MD Suite.

Components	Platform	Features
Mixed Solvent MD System Builder	Web-Online Tool	1. Supports insertion of 13 pre-defined probe molecules and custom user-uploaded molecules for flexible system setup. 2. Integrated utilities for system water molecule count, probe molecule insertion quantity calculation, and basic PDB preprocessing. 3. Supports customizable parameters for field and box setup, energy minimization settings, and ion/salt addition. 4. Automatic generation of topology and coordinate files compatible with GROMACS and AMBER simulation workflows.
Probe GridMap Builder	Local-GUI Tool	1. Supports trajectory formats generated by GROMACS (TRR) and AMBER (NetCDF). 2. Utilizes a cpptraj-powered backend for automated grid-based probe occupancy mapping. 3. Generates output compatible with PyMOL for visualizing probe distributions and hotspot regions. 4. Includes probe density sorting and hotspot ranking using the DBSCAN clustering algorithm, with direct export of sorted results for downstream analysis and visualization.

Methods

The Mixed-Solvent MD System Builder is an online platform designed as a web-based application, enabling users to build mixed-solvent MD systems directly in the browser without the need for local installation. To evaluate the functionality and robustness of the Mixed-Solvent MD Suite, we selected five representative proteins as benchmark targets (Table 2). These include the Mouse Double Minute 2 homolog (MDM2) [37], PDB ID: 4HG7, a key regulator of p53 and a classical protein–protein interaction target; B-cell lymphoma 2 (Bcl-2) [38], PDB ID: 6O0L, an anti-apoptotic protein that contains hydrophobic surface grooves often targeted by small molecules; Interleukin-2 (IL-2) [39], PDB ID: 1M47, a cytokine whose dynamic surface topology harbors transient ligand-binding regions; Protein Tyrosine Phosphatase 1B (PTP1B) [40], PDB ID: 1Q1M, a model phosphatase known for its allosteric regulatory site; and the SARS-CoV-2 papain-like protease (PLpro) [41], PDB ID: 6WUU, a promising target for the development of antiviral drugs. Together these targets cover a broad range of biological functions and structural complexities.

Table 2
Structural Dataset of Target Proteins

No.	PDB ID	Resol.	Water molecules	Probe molecules	Total atoms	Released	PubMed ID
1	4HG7	1.60 Å	11020	200	34231	2013-07-31	23897459
2	6O0L	2.20 Å	12893	234	40357	2019-05-22	31160589
3	1M47	1.99 Å	14852	270	45942	2002-07-31	12582206
4	1Q1M	2.60 Å	28004	508	87808	2003-09-16	13678400
5	6WUU	2.79 Å	65364	1187	199096	2020-05-20	33067239

For each target, we used the System Water Count Calculator to estimate the number of water molecules required to solvate the protein within a 1.4 nm buffer box. An example output of the water distribution estimation is shown in Fig. 2B, illustrating the single-water system generated for the 1M47 protein. Based on this solvation volume, the Probe Insertion Calculator was employed to calculate the number of ACN probe molecules corresponding to a 5% (w/v) concentration (Fig. 2C). Subsequently, we used the MSMD System Builder to generate mixed-solvent systems for all five proteins, each solvated with TIP3P water and 5% (w/v) acetonitrile (ACN) as the selected probe molecule. In this particular study, the CHARMM27 force field was selected for system construction and simulation, ensuring compatibility with the mixed-solvent environment. Subsequently, production molecular dynamics simulations were carried out with GROMACS, probe-density grids were generated and analyzed using the Probe GridMap Builder, and the resulting hotspots were visualized in PyMOL. A schematic overview of the data flow in the Mixed-Solvent MD System Builder and Probe GridMap Builder are shown in Fig. 4.

2.1 Construction of Mixed-Solvent Systems

The Mixed-Solvent MD System Builder was employed to construct five distinct mixed-solvent systems for the selected target proteins (1M47, 1Q1M, 4HG7, 6O0L, 6WUU). Protein structures for each system were pre-processed using the PDB Structure Preprocessing Tool (Figure S1) prior to system construction. Force field selection was set to CHARMM27, with TIP3P as the solvent water model. Simulation boxes were configured using an automatic buffer mode with a 1.4 nm padding applied uniformly around the protein. For each target, the System Water Count Calculator was used to estimate the total number of water molecules required to fully solvate the protein structure. These values served as the basis for computing the number of probe molecules to be inserted at a concentration of 5% (w/v), using the Probe Insertion Calculator. The calculation results are shown in Table 2. For this study, ACN was selected for the benchmark simulations due to its well-established usage in MSMD studies and balanced physicochemical properties; however, all other probe molecules provided by the Suite were

parameterized and validated under both CHARMM27 and AMBER99SB-ILDN force fields, and can be used interchangeably depending on the target system. Each system was optionally subjected to ion addition and energy minimization. Users specified standard GROMACS mdp options including the minimization algorithm, step size, convergence threshold, and cutoffs for electrostatic and van der Waals interactions. Ions (e.g., Na^+, Cl^-) were introduced to neutralize the system or achieve target salt concentrations. Finally, by clicking the “Build System” button, the Builder generates a complete set of GROMACS-compatible input files, including topology, coordinate, and parameter files required for simulation. These prepared system files were subsequently used for production simulations as described in Section 2.2.

2.2 Production MD Simulation

Following the successful construction of five distinct mixed-solvent systems using the Mixed-Solvent MD System Builder (Section 2.1), molecular dynamics (MD) simulations were carried out to evaluate probe behavior and solvent–protein interactions under realistic dynamic conditions. Each protein system, composed of TIP3P water and 5% (w/v) acetonitrile (ACN), was subjected to production MD simulation using GROMACS 2022.4. After energy minimization, NVT and NPT equilibration, each system underwent a 100 ns production simulation. The resulting trajectories were saved in GROMACS-compatible formats, including trajectory files (.trr), associated topology (.tpr) and coordinate files (.gro), as input for subsequent GridMap-based probe distribution analysis as described in Section 2.3.

2.3 Hotspot Analysis via Probe GridMap Builder

To perform probe-based spatial analysis following production MD simulations, we employed the Probe GridMap Builder to identify and rank potential binding hotspots. After generating the five mixed-solvent systems using the Mixed-Solvent MD System Builder and conducting 100 ns production simulations with GROMACS, the trajectory files (in TRR format) and corresponding topology files (in PDB format, with water molecules removed) were extracted. These de-solvated files were then uploaded to the GridMap Builder for analysis.

Probe occupancy mapping was performed by selecting predefined probe types from the internal library or by specifying user-defined custom probes. Key parameters included grid spacing (dx, default 0.5 nm), normalization by frame count (typically 1000 frames), and a maximum cutoff distance (default 0.35 nm) to define spatial proximity. Upon execution, the tool computed frame-normalized probe densities over a three-dimensional grid surrounding the protein structure. The resulting occupancy maps, output as grid-aligned PDB files (e.g., nonortho.pdb), were visualized using PyMOL to locate recurrent probe-enriched regions indicative of cryptic or allosteric binding sites. Additionally, the tool provided ranked density values for each region, facilitating downstream hotspot comparison and prioritization.

2.4 System Requirements and Availability

The Mixed-Solvent MD Suite is designed to be platform-flexible, with components optimized for different usage scenarios. The MSMD System Builder is deployed as a fully web-based application and can be

accessed through modern web browsers on any operating system, including Windows, macOS, and Linux, without the need for local installation. This ensures broad accessibility for users regardless of their computational environment. The Probe GridMap Builder is a local application that requires installation of a compatible conda environment. All source code, installation instructions, and demonstration data for the Mixed-Solvent MD Suite are freely available on GitHub at: https://sean28.github.io/MixMD/. These resources provide users with comprehensive documentation, usage examples, and tutorial materials to facilitate adoption and reproducibility in MSMD research.

Results and Discussion

3.1 Result of Construction of Mixed-Solvent Systems

The resulting five systems (Fig. 5) represent structurally diverse targets solvated in a consistent mixed-solvent environment, demonstrating the capability of the Suite to construct high-quality MSMD systems with concentration control, topology automation, and reproducibility.

The total number of atoms in the constructed systems ranges from approximately 34,000 to nearly 200,000 (Table 2), covering a wide spectrum of system sizes and highlighting the scalability and flexibility of the platform. Each system was built using a 1.4 nm buffer and solvated with TIP3P water along with 5% (w/v) acetonitrile (ACN) as the probe molecule. The number of inserted ACN molecules scaled accordingly from 200 to 1187 (Table 2), as calculated by the Probe Insertion Calculator. This demonstrates the accuracy and adaptability of the Suite's concentration control mechanism with respect to box volume and solute geometry. The systems also reflect a wide distribution in structural resolution and biological relevance. Resolutions of the selected PDB entries range from 1.60 Å (4HG7) to 2.79 Å (6WUU), ensuring the reliability of structural features used in simulation.

In terms of individual system characteristics, the constructed systems for the five benchmark proteins varied in both box dimensions and molecular complexity (Table 2). For 4HG7 (MDM2), the system contained 11,020 water molecules and 200 ACN probes, totaling 34,231 atoms, representing a relatively compact MSMD setup. 600L (Bcl-2) yielded a moderately larger system with 12,893 water molecules and 234 ACN molecules, resulting in 40,357 total atoms, reflecting its more extended surface topology. 1M47 (IL-2) required 14,852 water molecules and 270 ACN molecules, producing a 45,942-atom system, consistent with its more solvent-accessible and flexible structure. The 1Q1M (PTP1B) system involved 28,004 water molecules and 508 ACN molecules, resulting in a larger system of 87,808 atoms, likely due to its bulky domain and deeper box configuration. Notably, 6WUU (PLpro) generated the largest system, comprising 65,364 water molecules and 1,187 ACN probes, with a total of 199,096 atoms, demonstrating the platform's scalability to handle large, multi-domain targets. Together, these results validate the robustness of the Mixed-Solvent MD System Builder in handling both compact and large-scale protein systems, accurately computing solvent/probe quantities, and producing simulation-ready input files across varying protein classes and structural resolutions.

The Mixed-Solvent MD Suite introduces a streamlined, modular framework that addresses key limitations in traditional MSMD workflows. By integrating both web-based system construction and local trajectory analysis, the suite provides an end-to-end solution that significantly lowers the technical barrier for setting up and interpreting mixed-solvent molecular dynamics simulations. Its dual-mode architecture, comprising the Mixed Solvent MD System Builder and the Probe GridMap Builder, caters to diverse user needs while maximizing accessibility and platform compatibility. Key innovations include flexible probe molecule support, customizable probe concentrations in % w/v or mol/L units, and force field compatibility with GROMACS and AMBER, all of which contribute to enhanced reproducibility and ease of use. Importantly, the suite eliminates the need for fragmented manual operations typically required in conventional MSMD workflows, replacing them with intuitive graphical interfaces and automated calculators. These features collectively enhance user efficiency, ensure consistency across systems, and empower researchers to visualize and interpret probe distributions with greater clarity and confidence.

3.2 Result of GridMap Analysis

To investigate probe accumulation patterns and identify active binding site (include potential cryptic sites), we applied the Probe GridMap Builder to analyze the MSMD trajectories of the five benchmark proteins. This tool processed the simulation outputs and generated three-dimensional grid-based occupancy maps based on the spatial distribution of acetonitrile (ACN) probe atoms over the entire trajectory. The resulting density maps were visualized in PyMOL, where high-density regions indicate potential binding hotspots (Fig. 6). As illustrated in Fig. 6, each protein exhibited multiple distinct probe-enriched regions, shown from two opposing views (180° rotation) to capture surface accessibility. The right-hand bar plots display the quantitative probe occupancy rankings for the top five sites on each protein. These values, derived from DBSCAN-based clustering of grid points, reflect the relative strength of probe accumulation at different surface positions. The ranking provides a reliable metric for prioritizing binding pockets for further exploration. A full summary of occupancy scores for all systems is listed in Table S2.

For 4HG7 (MDM2), the GridMap analysis identified five distinct probe-accumulating sites with occupancy values ranging from 81 (Site 1) to 2 (Site 5). The two top-ranked sites (81 and 79) were symmetrically distributed near shallow surface grooves, suggesting a degree of convergence in probe binding behavior. Sites 3 and 4 (37 and 36) remained moderately populated, while Site 5 showed significantly reduced occupancy, indicating limited probe accessibility at that position. Importantly, Sites 1, 2, and 3 spatially coincide with the native ligand-binding pocket of the 4HG7 structure, where a small-molecule inhibitor is bound [37]. Although the original cleft contains hydrophobic residues, the observed ACN enrichment suggests that this region remains solvent-accessible and provides sufficient polar or partially polar interactions to favor ACN accumulation. This highlights the ability of ACN-based MSMD simulations to identify functionally relevant pockets, even when dominated by shallow or semi-polar microenvironments.

600L (Bcl-2) exhibited the highest absolute occupancy values among all proteins analyzed, with the top two sites reaching 138 and 133, respectively. These regions clustered around concave surface pockets, forming clear probe enrichment zones. Notably, Sites 1 and 3 spatially overlapped with the native ligand-binding pocket observed in the crystal structure [38], indicating that ACN probes can effectively sample pharmacologically relevant regions. Site 3 showed a moderate occupancy of 50, while Sites 4 and 5 dropped to 17 and 4, respectively. The sharp decline in probe density beyond the top three sites suggests a high degree of spatial selectivity in probe accumulation, likely reflecting a well-defined and solvent-accessible binding groove.

For 1M47 (IL-2), probe enrichment was more evenly distributed across the surface. Although Site 1 showed a peak occupancy of 91, subsequent sites declined gradually—22, 10, 7, and 3 for Sites 2 through 5—suggesting broader, less sharply defined probe–protein interactions. This distribution pattern may correspond to the dynamic and partially disordered regions characteristic of cytokine surfaces. Importantly, no probe hotspots were observed to coincide with the known ligand-binding site in this structure, which may reflect the limited ability of hydrophilic probes such as ACN to access deep or hydrophobic clefts in IL-2.

In the case of 1Q1M (PTP1B), GridMap identified five relatively well-populated sites with occupancy values of 58, 54, 35, 33, and 18. The first two positions were close in ranking, suggesting potential cooperative or neighboring interactions. The remaining sites also retained appreciable probe occupancy, indicating a diverse set of probe-accessible surface patches. However, similar to 1M47, none of the identified hotspots overlapped with the native ligand-binding site in the crystal structure. This observation may again be attributed to the limitations of ACN as a polar probe, which may be suboptimal for detecting certain hydrophobic or buried allosteric sites. Future applications may benefit from alternative probe selections from the Suite's internal library of 13 solvent types to better match the physicochemical environment of specific targets.

Finally, 6WUU (PLpro) yielded five probe-enriched sites with occupancy values ranging from 36 to 22, showing a consistent and gradual distribution across the surface. The relatively uniform probe presence across multiple sites may reflect the high surface flexibility and solvent-exposed loops characteristic of PLpro's catalytic and ubiquitin-binding regions. Notably, the top-ranked site (Position 1) coincides with a cryptic binding pocket previously reported in multiple studies [9, 42, 43], and this region is flanked by conformationally dynamic residues associated with allosteric regulation, further supporting the functional relevance of this probe-identified hotspot.

Across all five target proteins, the occupancy values derived from GridMap analysis provide a semi-quantitative basis for identifying frequently visited probe positions. This data-driven ranking, enabled by the integrated DBSCAN clustering algorithm, supports systematic identification and prioritization of potential binding pockets. It is important to note that the current analysis serves primarily as a demonstration of the Suite's functionality, and does not attempt to comprehensively validate the pharmacological relevance of each identified site. Nevertheless, the occupancy-based ranking enables

systematic pocket prioritization, and the generated data can be directly exported for follow-up applications such as fragment docking, pharmacophore alignment, or site-directed mutagenesis studies. These results demonstrate the utility of GridMap Builder in translating dynamic solvent behavior into interpretable spatial features, offering both visual and numerical insight into cryptic pocket detection across diverse protein systems. By lowering the technical barrier through its integrated web-based and local GUI tools, the Suite enables broader adoption of MSMD techniques by users without extensive simulation experience. This democratization of MSMD workflows makes it feasible for researchers in fields such as structural biology, protein engineering, and molecular design to systematically explore protein dynamics and identify cryptic or allosteric pockets that may otherwise remain inaccessible in traditional apo-state simulations.

3.3 Limitations and Prospects

Despite the notable advantages of the Mixed-Solvent MD Suite in automating and simplifying MSMD workflows, some limitations remain. Currently, while it currently supports direct generation of topology and coordinate files for both GROMACS and AMBER workflows, several areas remain for future enhancement. The trajectory analysis module, although effective in detecting potential binding hotspot regions, is limited in analytical scope and lacks advanced features such as probe–protein interactions, inter-solvent interaction analysis or AI-driven hotspot prediction. Moreover, as the suite is built on GROMACS and AMBER engines, its functionality is inherently constrained by these platforms' capabilities, highlighting the need for broader engine compatibility and future modular enhancements.

In the future, advanced machine learning techniques will be introduced to enhance the interpretation of probe-derived occupancy data. One key direction is the development of diffusion-based binding site prediction models, which treat sparse probe distributions as partial observations and reconstruct complete binding surfaces using generative diffusion models (e.g., DDPMs). This could improve the resolution and coverage of cryptic or transient site identification. We also plan to implement a binding interaction scoring network based on graph neural networks (e.g., EGNN, SchNet), which will evaluate DBSCAN-detected hotspots by integrating local topology and probe density features, providing affinity scores and interaction type labels. These enhancements will evolve the Mixed-Solvent MD Suite into an AI-driven probe-guided binding site discovery platform, enabling tighter integration with virtual screening and drug design workflows.

Conclusions

In this work, we present the Mixed-Solvent MD Suite, an integrated and modular framework designed to streamline the construction and analysis of MSMD simulations. The suite combines the Mixed Solvent MD System Builder, a fully web-based tool for automated system generation, and the Probe GridMap Builder, a local interface for trajectory-based probe interaction mapping. This integrated solution addresses key limitations of current MSMD workflows, including the lack of multi-probe support, complex setup procedures, and fragmented analysis pipelines. The Mixed-Solvent MD Suite lowers

technical barriers for researchers, enabling flexible probe selection, automated topology generation, concentration-based probe calculation, and grid-based hotspot visualization. Moreover, the suite's compatibility with both GROMACS and AMBER engines ensures broader applicability across diverse molecular dynamics simulation pipelines. Through benchmark studies, we demonstrate its ability to efficiently identify cryptic binding sites and dynamic probe distributions in biomolecular systems. We believe the Mixed-Solvent MD Suite will serve as a valuable resource for the broader molecular dynamics and drug discovery communities, facilitating both routine simulations and advanced structural biology research.

Declarations

Conflict of Interest

The authors declare that patent applications have been filed related to the software described in this work. These patent filings do not affect the authors' commitment to open academic research and reproducibility.

Author Contribution

Y.Q. conceived the study and developed the methodology and software. Y.Q. also performed data curation and wrote the original draft of the manuscript. Q.L. contributed to visualization and investigation. J.-W.X. contributed to project administration and funding acquisition.All authors reviewed the manuscript and approved the final version.

Acknowledgement

We acknowledge financial support from the Scientific Research Projects of the Jinhua Advanced Research Institute of Technology (G202509 and G202510), and from the Jinhua Municipal Health Commission and the Jinhua Municipal Science and Technology Bureau (No. 2025ZD08).

Data Availability Statement

All data and software required to reproduce the results reported in this study are publicly available. Representative input structures, probe parameter files, example trajectories, and analysis scripts used in this work are provided in the same repository.

The Mixed-Solvent MD Suite, including the web-based Mixed-Solvent MD System Builder and the local Probe GridMap Builder, together with source code, documentation, and example input files, is available at: https://sean28.github.io/MixMD/

References

1. Ghanakota P, Carlson HA (2016) Moving Beyond Active-Site Detection: MixMD Applied to Allosteric Systems. J Phys Chem B 120(33):8685–8695. 10.1021/acs.jpcb.6b03515
2. Smith RD, Carlson HA (2021) Identification of Cryptic Binding Sites Using MixMD with Standard and Accelerated Molecular Dynamics. J Chem Inf Model 61(3):1287–1299. 10.1021/acs.jcim.0c01002
3. Wakefield AE, Kozakov D, Vajda S (2022) Mapping the binding sites of challenging drug targets. Curr Opin Struct Biol 75:102396. 10.1016/j.sbi.2022.102396
4. MacKerell AD Jr., Jo S, Lakkaraju SK, Lind C, Yu W (2020) Identification and characterization of fragment binding sites for allosteric ligand design using the site identification by ligand competitive saturation hotspots approach (SILCS-Hotspots). Biochim Biophys Acta Gen Subj 1864(4):129519. 10.1016/j.bbagen.2020.129519
5. Yanagisawa K, Yoshino R, Kudo G, Hirokawa T (2022) Inverse Mixed-Solvent Molecular Dynamics for Visualization of the Residue Interaction Profile of Molecular Probes. Int J Mol Sci 23(9). 10.3390/ijms23094749
6. Graham SE, Leja N, Carlson HA (2018) MixMD Probeview: Robust Binding Site Prediction from Cosolvent Simulations. J Chem Inf Model 58(7):1426–1433. 10.1021/acs.jcim.8b00265
7. Ghanakota P, DasGupta D, Carlson HA (2019) Free Energies and Entropies of Binding Sites Identified by MixMD Cosolvent Simulations. J Chem Inf Model 59(5):2035–2045. 10.1021/acs.jcim.8b00925
8. Chan WKB, Carlson HA, Traynor JR (2023) Application of Mixed-Solvent Molecular Dynamics Simulations for Prediction of Allosteric Sites on G Protein-Coupled Receptors. Mol Pharmacol 103(5):274–285. 10.1124/molpharm.122.000612
9. Qiu Y, Liu Q, Tu G, Yao X-J (2022) Discovery of the cryptic sites of SARS-CoV-2 papain-like protease and analysis of its druggability. Int J Mol Sci 23(19):11265
10. Feng Y, Gong C, Zhu J, Liu G, Tang Y, Li W (2024) Unraveling the Ligand-Binding Sites of CYP3A4 by Molecular Dynamics Simulations with Solvent Probes. J Chem Inf Model 64(8):3451–3464. 10.1021/acs.jcim.4c00089
11. DasGupta D, Mehrani R, Carlson HA, Sharma S (2024) Identifying Potential Ligand Binding Sites on Glycogen Synthase Kinase 3 Using Atomistic Cosolvent Simulations. ACS Appl Bio Mater 7(2):588–595. 10.1021/acsabm.2c01079
12. Gopinath K, Jokinen EM, Kurkinen ST, Pentikainen OT (2020) Screening of Natural Products Targeting SARS-CoV-2-ACE2 Receptor Interface - A MixMD Based HTVS Pipeline. Front Chem 8:589769. 10.3389/fchem.2020.589769
13. Ung PM, Ghanakota P, Graham SE, Lexa KW, Carlson HA (2016) Identifying binding hot spots on protein surfaces by mixed-solvent molecular dynamics: HIV-1 protease as a test case. Biopolymers 105(1):21–34. 10.1002/bip.22742
14. Makley LN, Johnson OT, Ghanakota P, Rauch JN, Osborn D, Wu TS, Cierpicki T, Carlson HA, Gestwicki JE (2021) Chemical validation of a druggable site on Hsp27/HSPB1 using in silico solvent mapping

and biophysical methods. Bioorg Med Chem 34:115990. 10.1016/j.bmc.2020.115990

15. Lexa KW, Goh GB, Carlson HA (2014) Parameter Choice Matters: Validating Probe Parameters for Use in Mixed-Solvent Simulations. J Chem Inf Model 54(8):2190–2199. 10.1021/ci400741u
16. Skrhak V, Novotny M, Feidakis CP, Krivak R, Hoksza D (2024) CryptoBench: cryptic protein-ligand binding sites dataset and benchmark. Bioinformatics 41(1). 10.1093/bioinformatics/btae745
17. Kapitonova AA, Perfilova KV, Cooley RB (2024) Sluchanko, N. N. Phosphorylation code of human nucleophosmin includes four cryptic sites for hierarchical binding of 14-3-3 proteins. *bioRxiv* 10.1101/2024.02.13.580064
18. Yang WC, Gong DH, Hong W, Gao YY, Hao GF (2023) Grasping cryptic binding sites to neutralize drug resistance in the field of anticancer. Drug Discov Today 28(9):103705. 10.1016/j.drudis.2023.103705
19. Berendsen HJ, van der Spoel D, van Drunen R (1995) GROMACS: a message-passing parallel molecular dynamics implementation. Comput Phys Commun 91(1–3):43–56
20. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free. J Comput Chem 26(16):1701–1718. 10.1002/jcc.20291
21. Case DA, Aktulga HM, Belfon K, Ben-Shalom I, Brozell SR, Cerutti D, Cheatham T, Cruzeiro VWD, Darden T, Duke RE (2021) Amber : Reference Manual. 2021
22. Kong L, Park SJ, Im W, CHARMM-GUI PDB (2024) Reader and Manipulator: Covalent Ligand Modeling and Simulation. J Mol Biol 436(17):168554. 10.1016/j.jmb.2024.168554
23. Gee S, Glover KJ, Wittenberg NJ, Im W (2024) CHARMM-GUI Membrane Builder for Lipid Droplet Modeling and Simulation. ChemPlusChem 89(8):e202400013. 10.1002/cplu.202400013
24. Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem 29(11):1859–1865. 10.1002/jcc.20945
25. Huang J, Wu C, Yang X, Yang Z, Liu S, Yu G (2025) PACKMOL-GUI: An All-In-One VMD Interface for Efficient Molecular Packing. J Chem Inf Model 65(2):778–784. 10.1021/acs.jcim.4c01639
26. Martinez L, Andrade R, Birgin EG, Martinez JM (2009) PACKMOL: a package for building initial configurations for molecular dynamics simulations. J Comput Chem 30(13):2157–2164. 10.1002/jcc.21224
27. Roe DR, Cheatham TE 3 (2018) Parallelization of CPPTRAJ enables large scale analysis of molecular dynamics trajectory data. J Comput Chem 39(25):2110–2117. 10.1002/jcc.25382
28. Roe DR, Cheatham TE 3 (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. J Chem Theory Comput 9(7):3084–3095. 10.1021/ct400341p
29. Suh D, Arattu Thodika AR, Kim S, Nam K, Im W (2024) CHARMM-GUI QM/MM Interfacer for a Quantum Mechanical and Molecular Mechanical (QM/MM) Simulation Setup: 1. Semiempirical Methods. J Chem Theory Comput 20(12):5337–5351. 10.1021/acs.jctc.4c00439
30. Guterres H, Im W (2023) CHARMM-GUI-Based Induced Fit Docking Workflow to Generate Reliable Protein-Ligand Binding Modes. J Chem Inf Model 63(15):4772–4779. 10.1021/acs.jcim.3c00416

31. Park SJ, Kern N, Brown T, Lee J, Im W (2023) CHARMM-GUI PDB Manipulator: Various PDB Structural Modifications for Biomolecular Modeling and Simulation. J Mol Biol 435(14):167995. 10.1016/j.jmb.2023.167995
32. Eastman P, Swails J, Chodera JD, McGibbon RT, Zhao Y, Beauchamp KA, Wang LP, Simmonett AC, Harrigan MP, Stern CD et al (2017) OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. PLoS Comput Biol 13(7):e1005659. 10.1371/journal.pcbi.1005659
33. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D et al (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29(7):845–854. 10.1093/bioinformatics/btt055
34. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comput 4(3):435–447. 10.1021/ct700301q
35. Yanagisawa K, Moriwaki Y, Terada T, Shimizu KEXPRORER (2021) Rational Cosolvent Set Construction Method for Cosolvent Molecular Dynamics Using Large-Scale Computation. J Chem Inf Model 61(6):2744–2753. 10.1021/acs.jcim.1c00134
36. Ester M, Kriegel H-P, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, ; Vol. 96, pp 226–231
37. Anil B, Riedinger C, Endicott JA, Noble ME (2013) The structure of an MDM2-Nutlin-3a complex solved by the use of a validated MDM2 surface-entropy reduction mutant. Acta Crystallogr D Biol Crystallogr 69(Pt 8):1358–1366. 10.1107/S0907444913004459
38. Birkinshaw RW, Gong JN, Luo CS, Lio D, White CA, Anderson MA, Blombery P, Lessene G, Majewski IJ, Thijssen R et al (2019) Structures of BCL-2 in complex with venetoclax reveal the molecular basis of resistance mutations. Nat Commun 10(1):2385. 10.1038/s41467-019-10363-1
39. Arkin MR, Randal M, DeLano WL, Hyde J, Luong TN, Oslob JD, Raphael DR, Taylor L, Wang J, McDowell RS et al (2003) Binding of small molecules to an adaptive protein-protein interface. Proc Natl Acad Sci U S A 100(4):1603–1608. 10.1073/pnas.252756299
40. Liu G, Xin Z, Pei Z, Hajduk PJ, Abad-Zapatero C, Hutchins CW, Zhao H, Lubben TH, Ballaron SJ, Haasch DL et al (2003) Fragment screening and assembly: a highly efficient approach to a selective and cell active protein tyrosine phosphatase 1B inhibitor. J Med Chem 46(20):4232–4235. 10.1021/jm034122o
41. Rut W, Lv Z, Zmudzinski M, Patchett S, Nayak D, Snipas SJ, El Oualid F, Huang TT, Bekes M, Drag M et al (2020) Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: A framework for anti-COVID-19 drug design. Sci Adv 6(42). 10.1126/sciadv.abd4596
42. Ferreira JC, Adem KA, Fadl S, Villanueva AJ, Alzyoud L, Ghattas MA, Rabeh WM Discovery of Novel Allosteric Sites of SARS-CoV-2 Papain-Like Protease (PLpro). *bioRxiv* 2023, 2023.2005. 2016.540953.
43. Yue Q, Hua-Juan J, Yu-Shun Y, Xiao-Qin H, Xue-Wen Z (2025) Review of the Crystallized Structures of the SARS-CoV-2 papain-like protease. J Mol Struct 141730

Figures

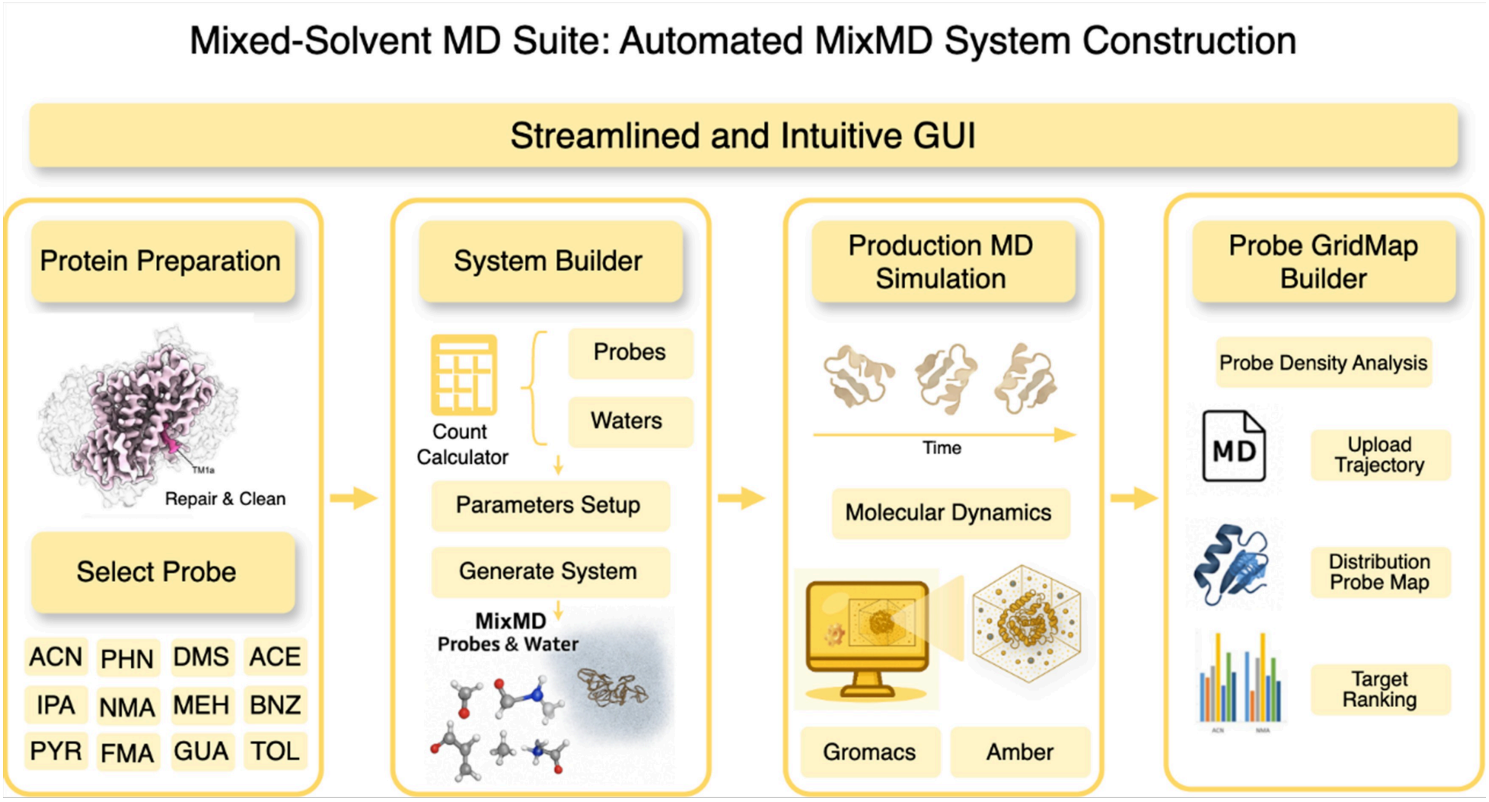


Figure 1

Mixed-Solvent MD Suite Overview. The workflow illustrates an integrated pipeline for mixed-solvent molecular dynamics simulations, consisting of a web-based Mixed-solvent molecular dynamics System Builder for automated system construction and a local Probe GridMap Builder for trajectory analysis and probe density visualization.

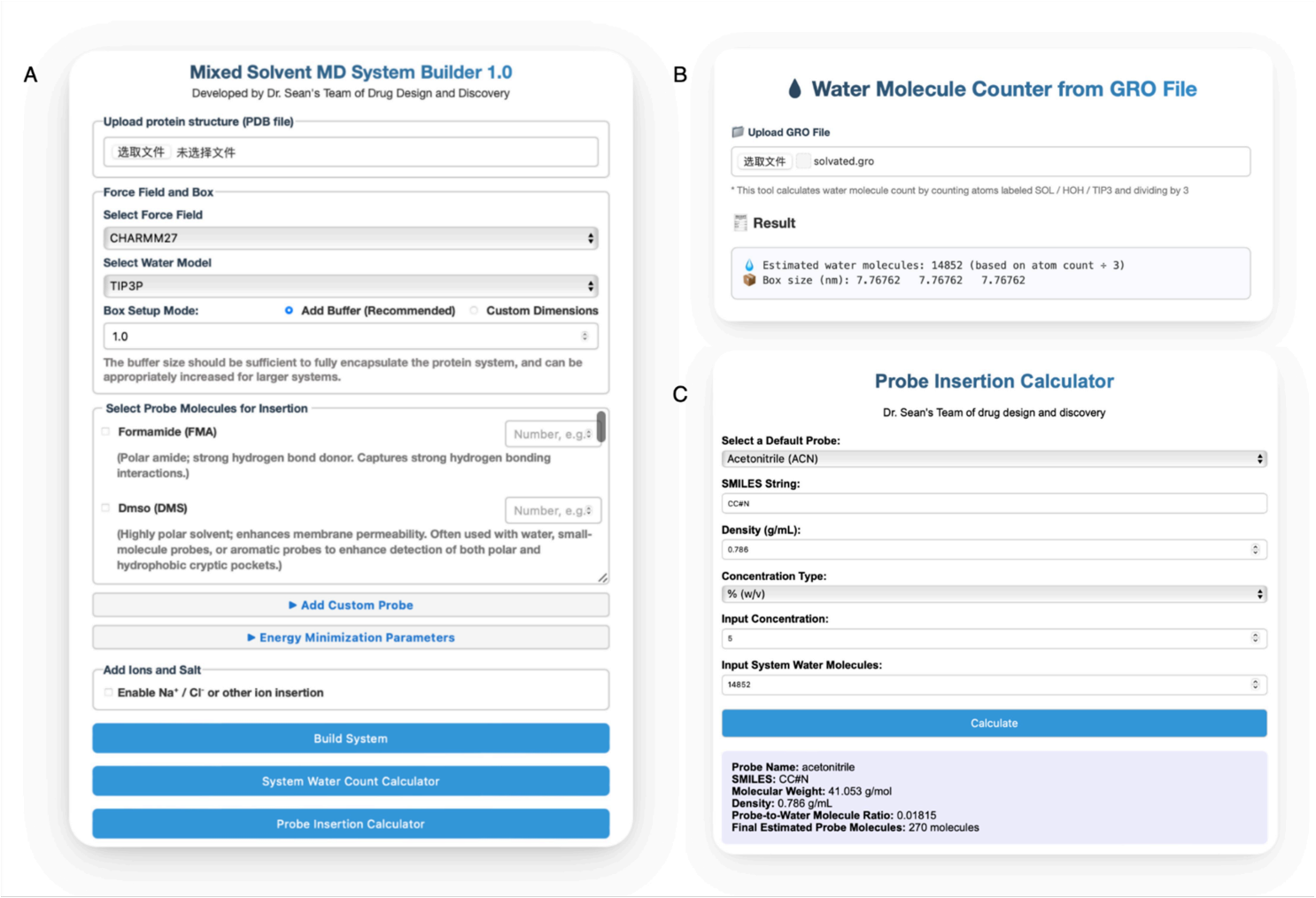


Figure 2

Interface of Mixed Solvent MD System Builder. (A) The Mixed Solvent MD System Builder allows users to upload a protein structure, select a force field and water model, define box parameters, and insert probes into the system. (B) The Water Molecule Counter estimates the number of water molecules from a GRO file. (C) Probe Insertion Calculator calculates probe molecule counts based on system properties and desired concentration.

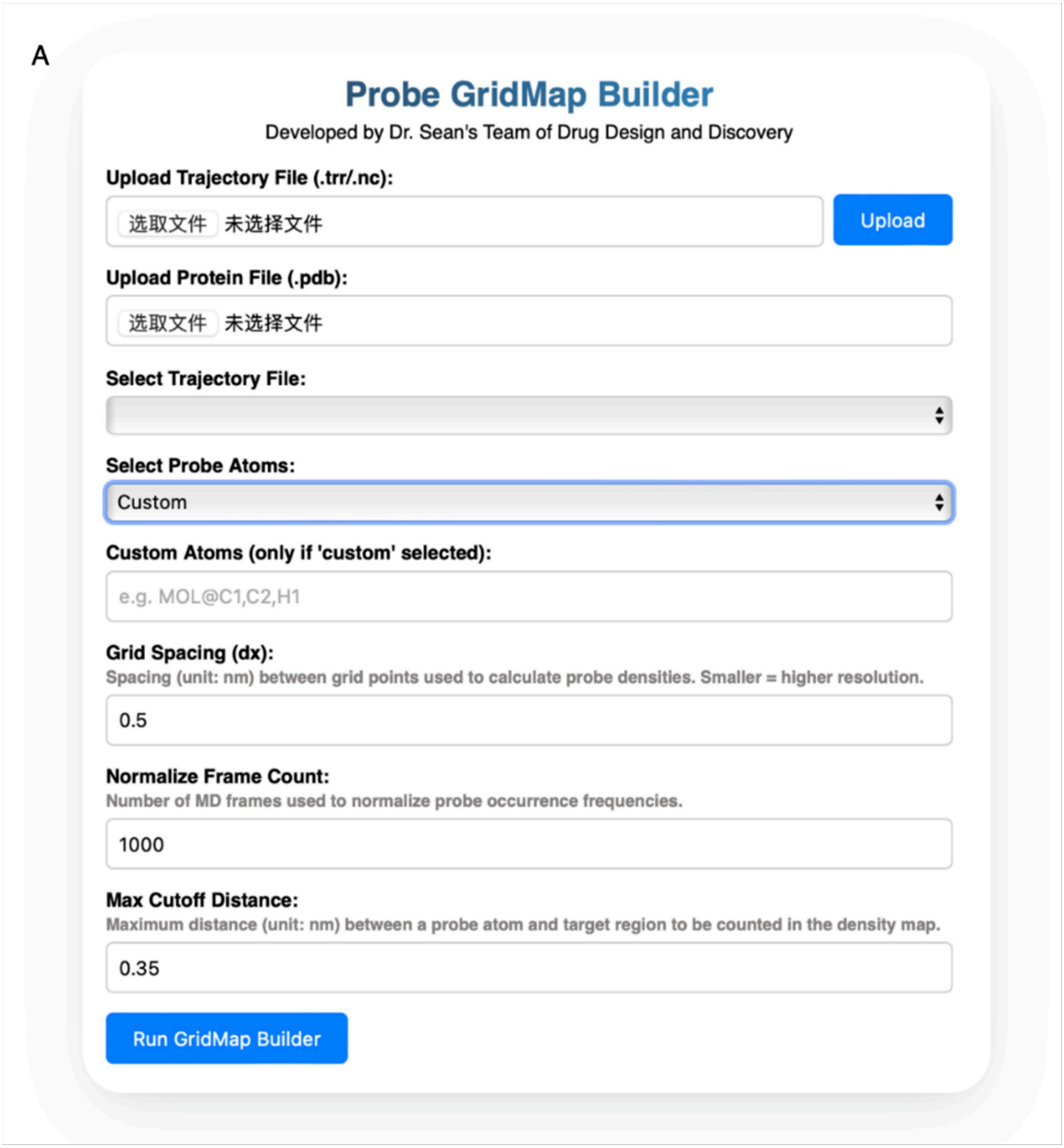


Figure 3

Interface of Probe GridMap Builder. The Probe GridMap Builder allows users to upload trajectory and protein structure files, select probe atoms (pre-defined or custom), set grid spacing, frame normalization, and cutoff distance, and run the density map calculation.

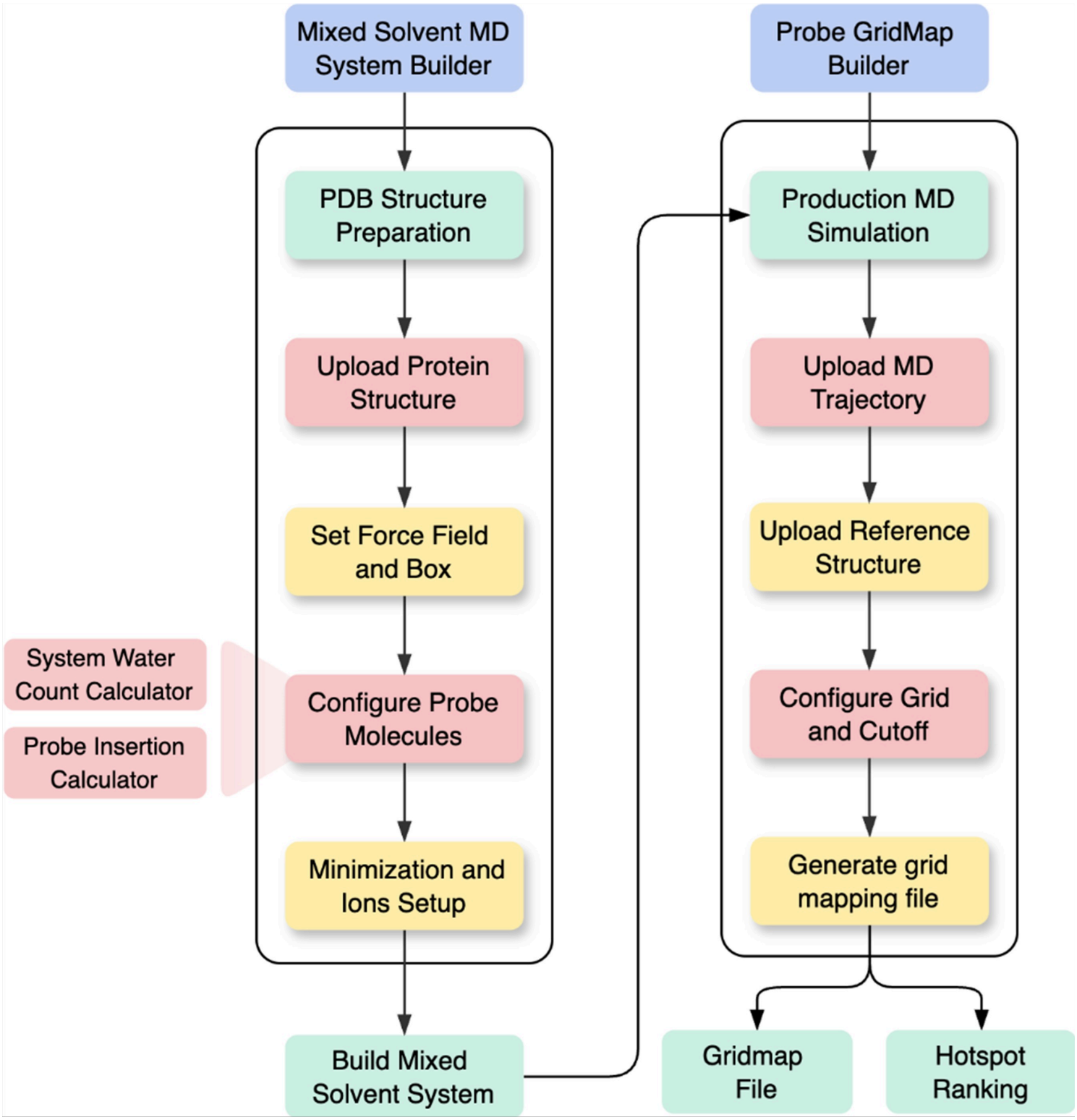


Figure 4

The overall workflow of Mixed-Solvent MD Suite.

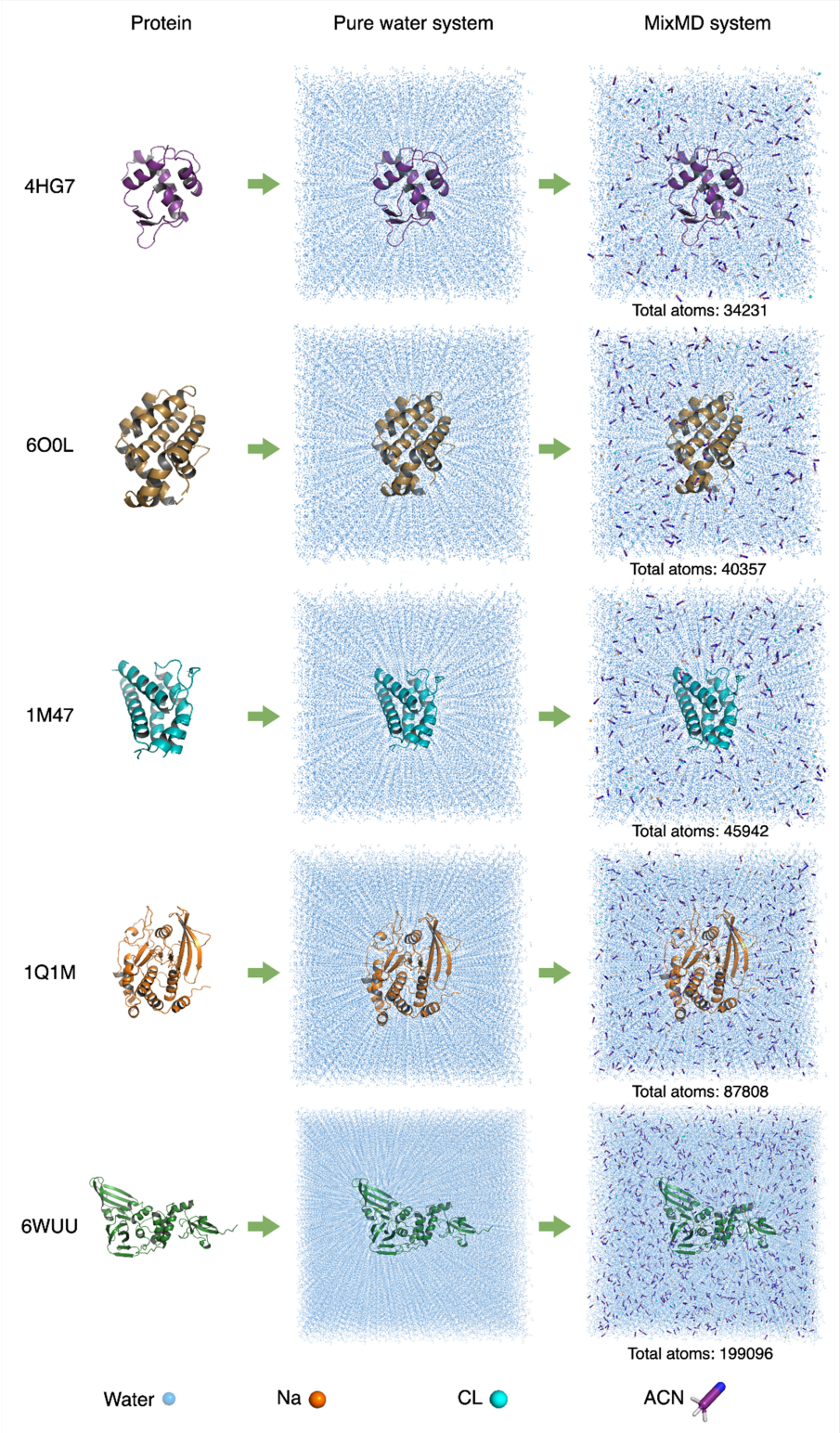


Figure 5

Construction of mixed-solvent systems for five representative protein targets. From left to right: initial protein structures (PDB IDs: 1M47, 1Q1M, 4HG7, 6O0L, and 6WUU), pure water solvated systems, and mixed-solvent systems containing 5% (w/v) acetonitrile (ACN) and TIP3P water. Color scheme: Water (light blue), Na^+ (orange), Cl^- (cyan), ACN (purple).

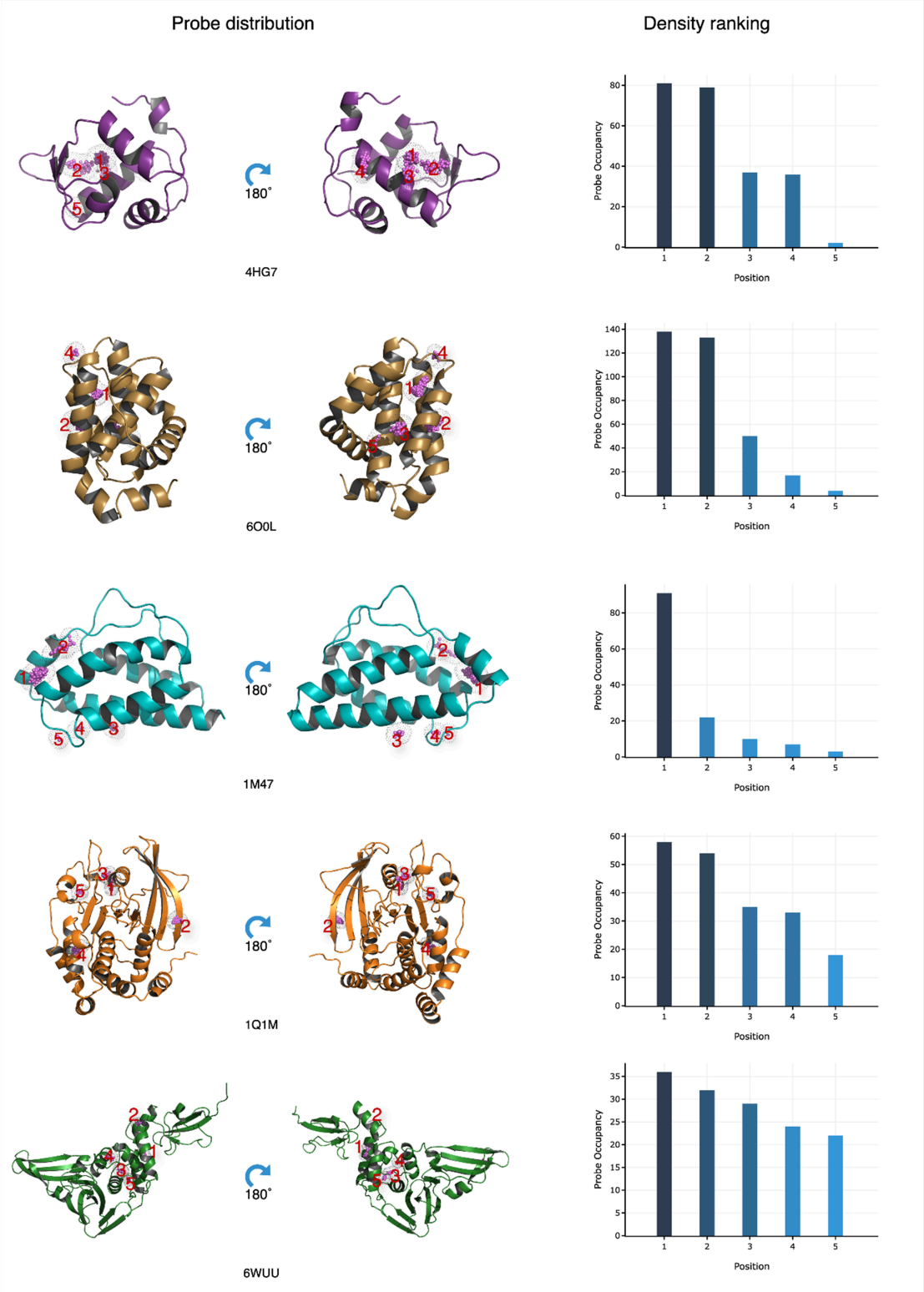


Figure 6

Visualization of probe distribution and hotspot density ranking for five protein systems. For each target, the left panels show two opposite views (180° rotation) of the protein structure overlaid with the top five probe-enriched regions identified by the Probe GridMap Builder. The right panels display bar plots of probe occupancy values for each ranked position, highlighting the relative strength of probe clustering across sites.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupportingInformation20260309.docx