

Choosing among Competing Pockets in Oligomeric Proteins: An OIPS-Assisted, Traceable Multi-Evidence Analysis

✉ ✉

Nanjing University of Chinese Medicine

Yifan Zhu

Nanjing University of Chinese Medicine

Shilei Zhao

Nanjing University of Chinese Medicine

Xin Zhang

Nanjing University of Chinese Medicine

Haopeng Sun

China Pharmaceutical University

Yao Chen

Nanjing University of Chinese Medicine

Xin Xue

xuexin@njucm.edu.cn

National and Local Collaborative Engineering Center of Chinese Medicinal Resources Industrialization
and Formulae Innovative Medicine

Research Article

Keywords: oligomeric proteins, pocket prioritization, protein-protein interface, allosteric site, evidence integration, molecular dynamics, binding-site prediction

Posted Date: August 10th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10420908/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Choosing among Competing Pockets in Oligomeric Proteins: An OIPS-Assisted, Traceable Multi-Evidence Analysis

Xiao Chen¹, Yifan Zhu¹, Shilei Zhao¹, Xin Zhang¹, Haopeng Sun^{2,*}, Yao Chen^{1,*}, Xin Xue^{3,*}.

¹Nanjing University of Chinese Medicine, Nanjing 210023, China

²School of Pharmacy, China Pharmaceutical University, Nanjing, 211198, China

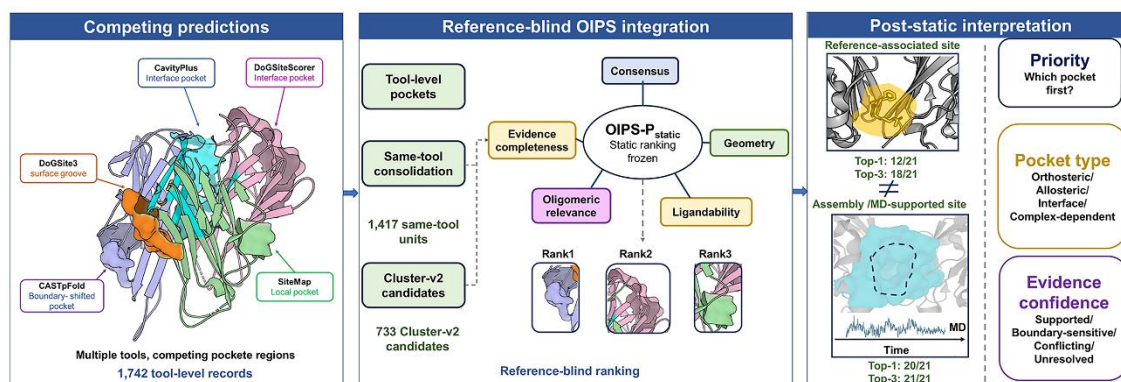
³National and Local Collaborative Engineering Center of Chinese Medicinal Resources Industrialization and Formulae Innovative Medicine, Jiangsu Collaborative Innovation Center of Chinese Medicinal Resources Industrialization, Jiangsu Key Laboratory for High Technology Research of TCM Formulae, Nanjing University of Chinese Medicine, Nanjing 210023, China

* Corresponding author: xuexin@njucm.edu.cn (Xin Xue), 300630@njucm.edu.cn (Yao Chen), sunhaopeng@cpu.edu.cn (Haopeng Sun)

Abstract

Pocket-finding programs often identify several geometrically plausible candidates in oligomeric proteins, creating a practical prioritization problem for docking, simulation, and experimental follow-up. Interchain boundaries, nested subpockets, internal allosteric cavities, complex-dependent sites, and static-dynamic discordance further complicate this choice. We developed OIPS, a traceable framework that integrates heterogeneous evidence to prioritize candidate pockets. Five pocket-finding workflows yielded 1,742 records across a curated case series of 21 oligomeric protein-ligand systems. Reference-blind reconstruction excluded 178 records lacking both a center and residue set, consolidated 1,417 within-tool units, and generated 733 medoid-constrained cross-tool candidates. *OIPS* – P_{static} ranked candidates using prespecified modules for cross-tool consensus, geometry, ligandability, oligomeric relevance, and evidence completeness. Reference association, redocking, molecular dynamics (MD), and literature annotations were added only after static rankings were fixed. Automated quality control and rule-based annotation were applied to all 63 Top-3 candidates without expert adjudication. The first reference-associated candidate ranked first in 12/21 systems, within the Top-3 in 18/21, and within the Top-5 in 19/21 (mean reciprocal rank [MRR] = 0.723). The first candidate supported by reference association or assembly/dynamic evidence ranked first in 20/21 systems and within the Top-3 in all 21 systems (MRR = 0.976). Removing O_{rel} increased reference Top-1 recovery from 0.571 to 0.667, while lowering reference Top-3 recovery from 0.857 to 0.810 and first-supported Top-1

recovery from 0.952 to 0.905. The ablation exposed a trade-off between retrospective ligand-location recovery and preservation of assembly-supported alternatives. Within this case series, OIPS provided an auditable basis for comparing competing pocket hypotheses, identifying evidence conflicts, and selecting targeted follow-up analyses.



Keywords: oligomeric proteins; pocket prioritization; protein-protein interface; allosteric site; evidence integration; molecular dynamics; binding-site prediction

1. Introduction

Binding-site prediction typically focuses on detecting cavities capable of accommodating ligands. In structure-based drug discovery, candidate generation is followed by a consequential prioritization step.[1-3] A single protein structure may contain several plausible clefts, nested subpockets, shallow grooves, interfacial depressions, and local cavities. Researchers must then select the candidate that will define the docking grid, guide fragment growth, focus subsequent simulations, or motivate experimental testing.[1, 3] Choosing the wrong region can yield technically coherent docking or simulation results while directing the project toward a biologically irrelevant hypothesis.

This prioritization problem is amplified in oligomeric proteins. In these systems, pockets can be formed by multiple chains, reshaped by assembly, exposed only in particular complex states, or located adjacent to an interface without being centered on it.[4-6] At the same time, a crystallographic ligand may occupy a compact orthosteric or internal allosteric site far from the dominant interchain surface.[7, 8] An assembly-aware ranking may therefore favor a different candidate from a ranking designed to recover a known ligand position. Reference recovery measures agreement with an experimentally observed ligand position, whereas assembly-aware preservation assesses whether structurally supported alternatives remain available for evaluation.[1, 9] Recent pocket-prediction methods have improved cavity geometry characterization, graph-based structural context, local feature extraction, and integration with blind-docking workflows.[1, 10] Parallel work on

allosteric sites and protein-protein interfaces has emphasized that ligandability, pocket context, and mechanism cannot be inferred from cavity volume alone.[11, 12] These advances improve the quality of candidate generation and annotation, but they do not remove the need to reconcile outputs that use different pocket boundaries, descriptors, and native ranking conventions.[13, 14]

The biological and structural context of a candidate pocket also determines how it should be used in downstream molecular design.[15, 16] Compact orthosteric clefts generally support focused docking grids centered on catalytic or substrate-recognition contacts.[7] Internal allosteric sites require evaluation of conformational state, local interaction networks, and ligand-pose persistence.[15] Interface-associated pockets may depend on hotspot engagement and balanced contacts across chains, while complex-dependent local pockets require the complete assembly to be retained without extending the docking region over the entire interface. Pocket selection consequently affects docking-grid placement, scaffold choice, polarity distribution, fragment-growth direction, and the intervention mechanism pursued in subsequent design.[2, 9]

To organize this decision process, we developed OIPS as a traceable multi-evidence framework.[17, 18] The framework provides three complementary interpretation outputs: OIPS prioritization score (OIPS-P) prioritizes candidate pockets within each structure; OIPS pocket-type interpretation layer (OIPS-T) assigns a working pocket type, including orthosteric, allosteric, interface-associated, and complex-dependent categories; and OIPS confidence interpretation layer (OIPS-C) records the evidence state and any conflicts that qualify that assignment.[19] OIPS-P is a prioritization score and does not estimate ligand affinity, potency, or biological function.[1]

We evaluated OIPS through a mechanistic case-series analysis of candidate-pocket selection. The curated set comprises 21 oligomeric protein-ligand systems selected to represent distinct pocket-choice problems, with several protein families represented more than once.[9] The analysis pursued four objectives: reconstruct cross-tool candidates without single-link chaining or reference leakage; establish a fixed static ranking that incorporates assembly context; apply reproducible structural quality control and rule-based post hoc annotation; and determine where reference, interface, redocking, and dynamic evidence converge or diverge. The resulting output is a traceable map of competing pocket hypotheses, the evidence supporting each interpretation, and the analyses required to resolve remaining uncertainty.

2. Methods

2.1 Study design, structural preparation, and pocket detection

The study comprised 21 curated oligomeric protein-ligand systems, including homodimers, heterodimers, two homotrimers, and multicomponent protein complexes. The systems were selected to represent distinct candidate-pocket selection scenarios and were not intended to estimate population-level prediction accuracy.[14, 20] Inclusion required an interpretable biological assembly, a defined crystallographic ligand or ligand-associated reference region, and sufficient structural information for mapping predictions across multiple pocket-detection workflows.[21] Oligomeric states were assigned according to the biological assemblies used in the structural analysis. Pocket archetypes were classified as orthosteric, allosteric, interface-associated, or complex-dependent according to reference-site location, chain contribution, interface relationship, functional annotation, and assembly dependence.[21, 22] These assignments were used for stratified interpretation after candidate ranking and did not contribute to candidate reconstruction or $OIPS - P_{static}$ calculation.[19, 23] Systems from the same protein family were retained as related cases and treated as clustered units in family-level resampling analyses.[14, 16]

Pocket detection was performed on ligand-depleted structures derived from the corresponding crystallographic holo assemblies. Crystallographic reference ligands, non-target small molecules, and crystallization additives were removed before pocket mapping.[24] Protein chains required to preserve the relevant oligomeric or complex-dependent context were retained, together with structurally required cofactors and metal ions.[25] The same retained assembly was supplied to each available pocket-detection workflow for a given target.[3, 22, 25] Reference-ligand coordinates and reference-site annotations were not provided to the detection or static-ranking procedures. Reference association, redocking, molecular dynamics (MD), functional annotation, and literature-based interpretation were introduced only after the $OIPS - P_{static}$ ranking had been fixed.

Pocket predictions were generated independently using CavityPlus, DoGSiteScorer, DoGSite3, CASTpFold, and SiteMap in Schrödinger Suite 2025. Native or recommended settings were retained unless otherwise specified. Collectively, these workflows represent different cavity definitions and provide partially overlapping sets of geometric, surface, and ligandability-related descriptors. Their native scores were therefore retained as source-specific outputs and were not compared directly across programs before normalization. Pocket outputs were analyzed as candidate-generating records; the native Top-1 prediction of an individual tool was not treated as a definitive site assignment.

For each tool-level prediction, the source program, native pocket identifier, native output order or rank, pocket center, lining-residue set, contributing chains, and available descriptors were retained. When a program did not report a center directly, the center was reconstructed only when spatial coordinates were available in the native output; otherwise, the center remained missing. Residue identifiers were mapped to the corresponding chains and residue numbers in the retained biological assembly. Geometric evidence included available measures of pocket volume, surface area, depth, enclosure, exposure, and mouth or opening characteristics.[26, 27] Ligandability-related evidence included program-specific druggability estimates, predicted - pK_d - related outputs, SiteScore, Dscore, contact, enclosure, exposure, and hydrophobicity-related terms, when available.[10, 14, 28]

Descriptor directionality, clipping limits, normalization rules, and within-module aggregation procedures were prespecified before the final cluster-v2 analysis and were not adjusted according to reference-pocket recovery.[28, 29] Each descriptor was transformed so that a higher normalized value consistently indicated more favorable evidence for the corresponding module. Values outside the prespecified descriptor range were clipped before transformation to the common 0-100 scale. Descriptors not reported by a program remained missing and were not replaced by zero. Module values were calculated from the available descriptors only, treating missing values as unavailable evidence rather than unfavorable evidence.[12, 30]

2.2 Reference-blind cluster-v2 reconstruction

Pocket records were processed in two stages. First, duplicate, nested, or highly overlapping predictions from the same tool were consolidated by deterministic complete-link grouping.[1, 31] Records were grouped when residue IoU was ≥ 0.70 , residue containment was ≥ 0.85 with center distance ≤ 8 Å, or center distance was ≤ 2.5 Å with residue $IoU \geq 0.25$. Center-only records were treated as duplicates within 1.5 Å. Second, same-tool units were grouped across tools when center distance was ≤ 6 Å, center distance was ≤ 10 Å with residue $IoU \geq 0.20$, or residue IoU was ≥ 0.35 . Cross-tool grouping was medoid-constrained: all members were required to remain compatible with the cluster medoid, and the maximum center diameter was limited to 12 Å. Single-link transitive chaining was not permitted.[32]

Each cluster-v2 candidate retained its medoid center, contributing tools and chains, representative pockets, and residue sets. Envelope residues were defined as the union of representative residue sets, whereas core residues required support from at least half of the contributing tools, with a minimum of two tools for multi-tool candidates. Representatives were selected according to within-group similarity, data completeness, residue coverage, and native tool rank.

2.3 OIPS-P_{static} calculation and ranking

All cluster-level evidence modules were normalized to a common 0-100 scale before integration. Five static modules were included: cross-tool consensus (C_{cons}), geometric evidence (G_{geo}), ligandability-related evidence (P_{lig}), oligomeric relevance (O_{rel}), and evidence completeness (Q_{evidence}). Higher values represented stronger support within the scope of the corresponding module.

C_{cons} was calculated from the weighted support of the contributing tools. Tool weights were 1.00 for CavityPlus, DoGSiteScorer, and DoGSite3, 1.15 for SiteMap, and 0.75 for CASTpFold. The summed weight was divided by the maximum possible support of 4.90 and converted to a 0-100 scale. A factor of 0.5 was applied to candidates supported by only one program. G_{geo} was calculated as the mean of up to the three highest available source-level geometry scores, and P_{lig} as the mean of up to the two highest available ligandability scores. Evidence completeness for candidate i was calculated as:

$$Q_{\text{evidence},i} = \text{clamp}(30 + 9n_i + 18f_{\text{center},i} + 14f_{\text{residue},i} + 4I_{\text{SiteMap},i}, 0, 100)$$

Where n_i is the number of contributing tools, $f_{\text{center},i}$ and $f_{\text{assembly},t}$ are the fractions of representatives with an available center and residue set, respectively; and $I_{\text{SiteMap},i} = 1$ when SiteMap contributed to the candidate and 0 otherwise. The clamp function was defined as $\text{clamp}(x, 0, 100) = \min[100, \max(0, x)]$.

The target-level assembly-context component, $S_{\text{assembly},t}$, was derived from the number of retained protein chains and the global fraction of retained protein residues involved in an interchain contact:

$$S_{\text{assembly},t} = \text{clamp}(35 + 10K_t + 80F_{\text{interface},t}, 0, 100)$$

$$F_{\text{interface},t} = \frac{N_{\text{interface},t}}{N_{\text{envelope},t}}$$

where K_t is the number of retained protein chains, $N_{\text{interface},t}$ is the number of unique residues satisfying the 5.0 Å inter-chain contact definition, and $N_{\text{envelope},t}$ is the total number of retained protein residues. The same $S_{\text{assembly},t}$ value was assigned to all candidates within target t . When an evaluable interchain interface was absent, $S_{\text{assembly},t}$ was assigned the prespecified fallback value of 35.

Candidate-specific interface overlap was calculated as:

$$f_{\text{interface},i} = \frac{N_{\text{interface},i}}{N_{\text{envelope},i}}, S_{\text{overlap},i} = \text{clamp}(25 + 130f_{\text{interface},i}, 0, 100)$$

Where $N_{\text{interface},i}$ is the number of envelope residues satisfying the 5.0 Å interchain-contact definition and $N_{\text{envelope},i}$ is the total number of candidate-envelope residues. For a candidate containing residues from $K_i \geq 2$ chains, normalized chain entropy was calculated as:

$$H_{\text{chain},i} = -\frac{\sum_{c=1}^{K_i} p_{i,c} \ln p_{i,c}}{\ln K_i}$$

Where $p_{i,c}$ is the proportion of candidate-envelope residues contributed by chain c . The chain component was defined as:

$$S_{\text{chain},i} = \begin{cases} 75 + 25H_{\text{chain},i}, K_i \geq 2 \\ 65, K_i = 1 \text{ and } f_{\text{interface},i} \geq 0.35 \\ 30, \text{otherwise} \end{cases}$$

The center-to-interface distance d_i , was calculated as the minimum Euclidean distance between the candidate medoid center and any heavy atom belonging to an inter-chain interface residue. The distance component was defined as

$$S_{\text{distance},i} = \begin{cases} 100, d_i \leq 4 \\ 90 - 5(d_i - 4), 4 < d_i \leq 8 \\ 70 - 6(d_i - 8), 8 < d_i \leq 16 \\ \max[10, 22 - 1.2(d_i - 16)], d_i > 16 \end{cases}$$

The four components were combined as

$$O_{\text{rel},i} = 0.42S_{\text{overlap},i} + 0.25S_{\text{chain},i} + 0.20S_{\text{distance},i} + 0.13S_{\text{assembly},t}$$

Higher O_{rel} values indicated stronger structural association with the retained oligomeric interface or assembled complex. The module did not include PISA-derived interaction energies, buried surface area, hydrogen bonds, salt bridges, hotspot annotations, evolutionary conservation, or mutational evidence because these data were not available consistently across the case series. Accordingly, O_{rel} was interpreted as an assembly-aware structural relevance measure.

For candidate cluster, the static pocket-priority score was calculated as:

$$\text{OIPS-P}_{\text{static},i} = \frac{0.22I_{C,i}C_{\text{cons},i} + 0.18I_{G,i}G_{\text{geo},i} + 0.24I_{P,i}P_{\text{lig},i} + 0.24I_{O,i}O_{\text{rel},i} + 0.12I_{Q,i}Q_{\text{evidence},i}}{0.22I_{C,i} + 0.18I_{G,i} + 0.24I_{P,i} + 0.24I_{O,i} + 0.12I_{Q,i}}$$

Where $I_{i,m}=1$ when module m was available for candidate i , and $I_{i,m}=0$ otherwise. Module weights were fixed before the final cluster-v2 analysis and were not optimized against reference-ligand recovery, redocking, molecular dynamics, or candidate-ranking results. Missing modules were excluded from both the numerator and denominator, and the remaining weights were renormalized rather than treating missing evidence as zero. Reference-ligand location, center-to-ligand distance, ligand-contact recovery, reference-residue overlap, redocking, molecular dynamics, literature annotation, and D_{dyn} were excluded from $\text{OIPS} - \text{P}_{\text{static}}$. These evidence layers were introduced only after the static ranking had been frozen. Candidates were ranked independently within each PDB structure using unrounded $\text{OIPS} - \text{P}_{\text{static}}$ values. Scores were not compared directly across unrelated protein systems. Exact ties were assigned the mean of the tied rank positions.

Local weight sensitivity was evaluated by independently decreasing or increasing each of the five static module weights by 20%, producing ten one-at-a-time perturbation scenarios. After each perturbation, the weights were renormalized, while candidate definitions, module values, missing-data treatment, and post-static annotations remained unchanged. Ranking stability was assessed using baseline Top-1 retention, Top-3 Jaccard overlap, within-target Spearman rank correlation, and changes in the first reference-associated and first-supported candidate ranks. Reference-ligand, redocking, and dynamic evidence remained excluded from all perturbed static rankings.

2.4 Automated structural quality control and post hoc evidence annotation

After the within-target OIPS – P_{static} ranking had been frozen, the Top-3 candidates from each system underwent automated structural quality control. The assessment considered mappability, tool support, cluster diameter, center dispersion, spatial continuity, pairwise residue *IoU*, and the ratio of core to envelope residues. A candidate was flagged as $QC_{possible\ overmerge}$ when it lacked spatial continuity, when its diameter exceeded 10.5 Å with a core/envelope ratio below 0.25, or when extensive same-tool secondary predictions were combined with low cross-tool residue agreement. $QC_{possible\ split}$ required support from at least two tools together with center dispersion >4 Å, median residue *IoU* <0.15, or a core/envelope ratio <0.20. $QC_{boundary\ sensitive}$ was assigned to multi-tool candidates with diameter >9 Å or comparable boundary disagreement. $QC_{insufficient\ evidence}$ was assigned to center-only candidates, unsupported single-tool candidates lacking ligandability evidence, or low-scoring surface-associated predictions. Primary states followed the order $QC_{unmappable}$, $QC_{possible\ overmerge}$, $QC_{possible\ split}$, $QC_{boundary\ sensitive}$, $QC_{insufficient\ evidence}$, and QC_{pass} . QC labels did not alter static scores or ranks.

Reference and orthogonal evidence were appended only after static ranking. A candidate was labeled R_{auto} when it satisfied at least one reference-association rule: $DCC \leq 6$ Å with contact recall ≥ 0.10 ; $DCC \leq 10$ Å with contact recall ≥ 0.20 or residue *IoU* ≥ 0.15 ; or $DCC \leq 12$ Å with contact recall ≥ 0.50 and residue *IoU* ≥ 0.15 . Distance to crystallographic ligand center (DCC) was the distance between the candidate medoid center and the crystallographic ligand centroid. Contact recall was the fraction of crystallographic ligand-contact residues recovered in the candidate envelope, and reference-residue IoU compared the candidate envelope with the crystallographic ligand-contact residue set.

Candidates not assigned R_{auto} were labeled A_{auto} when they remained structurally interpretable and were supported by either an assembly-interface relationship or a persistent MD contact region. Assembly-interface support required an interface-residue fraction ≥ 0.35 and a center-to-interface distance ≤ 8 Å. Candidates with

acceptable structural quality but insufficient support were labeled U_{auto} , whereas X_{auto} was reserved for unmappable candidates or clear integration artifacts.

Reference ligands were identified using curated residue name, chain identifier, and residue number, and ambiguous multi-ligand structures were resolved before mapping. Boundary-sensitive candidates retained an unresolved flag where appropriate. These automated labels represent reproducible evidence states rather than experimentally validated pocket functions.

2.5 Candidate-prioritization metrics and uncertainty

Reference-associated Top-1, Top-3, Top-5, mean reciprocal rank (MRR), median rank, and rank percentile were calculated using the first R_{auto} candidate in each evaluable target. When multiple R_{auto} candidates represented different boundaries of the same reference region, all were retained, but only the highest-ranked candidate defined the endpoint. The first-supported-site endpoint was defined separately using the highest-ranked R_{auto} or A_{auto} candidate. Unresolved targets were excluded from the corresponding metric denominator.

Confidence intervals were estimated from 10,000 bootstrap resamples using a fixed random seed. Target-level resampling treated each PDB structure as one unit, whereas family-clustered resampling sampled protein families and retained all associated structures. Leave-one-family-out analysis was used to assess the influence of repeated protein series.[23, 33] All resampling procedures used the same prespecified random seed.

For the O_{rel} ablation, candidate definitions and the four remaining module values were held fixed. O_{rel} was removed, the available module weights were renormalized, and candidates were reranked independently within each PDB structure. The analysis evaluated changes in the first reference-associated and first-supported ranks, Top-3 inclusion, within-target Spearman rank correlation, and Top-1 candidate identity. Reference-location recovery and preservation of assembly-supported alternatives were reported as separate endpoints. Because both O_{rel} and the assembly-supported component of the first-supported endpoint used interface-residue fraction and center-to-interface distance, the first-supported endpoint was interpreted as a measure of assembly-aware internal consistency rather than external validation.

2.6 Comparison with native single-tool priorities

To evaluate whether multi-tool integration provided information beyond the native ranking of individual pocket detectors, the outputs of CavityPlus, DoGSiteScorer, DoGSite3, CASTpFold, and SiteMap were compared with the frozen OIPS-P ranking.[19] Within each PDB-tool pair, raw tool records were first consolidated into the same-tool units used during cluster-v2 reconstruction. Same-tool units were ordered

according to the retained native output order. When multiple units from the same tool mapped to the same cluster-v2 candidate, only the earliest occurrence was retained, preventing parent pockets, subpockets, or alternative boundary definitions of the same spatial region from occupying multiple regional ranks.

A single-tool region was considered reference-associated when its mapped cluster-v2 candidate satisfied the prespecified R_{auto} criteria. The rank of the first such region was used to calculate Top-1, Top-3, Top-5, and MRR. A no-hit state was assigned when a program produced at least one mappable region but none was reference-associated. Outputs containing records without any mappable region were distinguished from unavailable tool outputs. Because tool coverage differed across the 21 systems, quantitative across-method comparisons were based primarily on the complete-case subset of 14 systems for which all five programs produced at least one mappable region. Native Top-1 regions were also mapped to cluster-v2 candidates to determine whether the programs prioritized the same spatial region within each target.

2.7 Redocking, molecular dynamics, and representative-case analysis

Redocking was performed using Glide in Schrödinger Suite 2025 in standard-precision (SP) mode. [34, 35] The docking-grid center was defined from the centroid of the corresponding crystallographic ligand, and the grid box was generated using the Absolute size option with the default dimensions. Receptor and ligand preparation, pose generation, and all other docking parameters followed the default Schrödinger workflow. The first-ranked SP pose was retained for Protein root-mean-square deviation (RMSD) calculation and subsequent structural interpretation.[36] GlideScore was recorded as a docking output and was not interpreted as binding affinity or biological potency.

Molecular dynamics simulations were performed using Desmond in Schrödinger Suite 2025. Each trajectory was run for 100 ns using the default Desmond simulation protocol under an NPT ensemble at 300 K. System preparation, solvation, ion placement, minimization, equilibration, and trajectory analysis followed the default workflow.[37, 38] Trajectory coordinates were recorded every 1000 ps, and the complete 0-100 ns interval was used for analysis. Twenty-three trajectories covering 15 targets were analyzed, including native-ligand, redocked-ligand, and apo systems.

Protein RMSD and root-mean-square fluctuation (RMSF), protein-aligned ligand RMSD, protein-ligand contacts, and pocket-residue flexibility were extracted where available. Ligand-contact occupancy was calculated over the complete trajectory, and contacts present for more than 30% of the analyzed simulation time were treated as persistent. Persistent contact regions were mapped to cluster-v2 candidates using residue overlap, residue *IoU*, and center-distance measures. MD evidence was classified as concordant, partially

concordant or boundary-shifted, conflicting, insufficient, or apo-only. Dynamic evidence was used only for post hoc OIPS-C interpretation and did not alter OIPS-Pstatic scores or ranks.[39]

2.8 Reproducibility, code, and data availability

All analyses were implemented in a version-controlled Python workflow. The accompanying public repository contains the processed pocket-prediction records, redistribution-permitted prepared structures, fixed configuration files, cluster-v2 membership tables, frozen OIPS-P rankings, post-static reference, redocking, and MD evidence tables, automated QC outputs, statistical analyses, tests, and figure source data. Static candidate reconstruction and OIPS-P scoring were completed before reference-ligand, redocking, and MD evidence were loaded; these post-static evidence layers could not modify candidate membership, module values, or within-target ranking. The repository is publicly available at <https://github.com/xiaochen3316/oips-reproducibility> (commit: 2516d10ff8d7bd72a4afcf7e68174f859b8c7a36). Redistribution-restricted third-party software outputs and large raw MD trajectories are not included; their provenance and availability are documented in the repository.

3. Results

3.1 Reference-blind reconstruction defined the formal candidate landscape

Across the five pocket-finding workflows, 1,742 tool-level records were collected, comprising 792 CASTpFold, 443 DoGSiteScorer, 207 CavityPlus, 197 DoGSite3, and 103 SiteMap records. Of these, 178 records lacked both a pocket center and a residue set and were retained as unmappable audit records. The remaining 1,564 records were consolidated into 1,417 same-tool units and subsequently reconstructed into 733 formal cluster-v2 candidates across the 21 systems. Candidate counts ranged from 17 to 96 per target, with a median of 24 (**Figure 1A**) .

Comparison with the previous working clusters showed that 112 clusters mapped one-to-one to a cluster-v2 candidate, whereas 79 mappable working clusters were separated into two or more candidates. This increase reflected the separation of spatially heterogeneous or boundary-distinct regions rather than additional pocket-detection runs. No cluster-v2 candidate exceeded the 12 Å center-diameter limit, and no tool contributed more than one formal vote to a candidate. The previous 369-cluster set was therefore retained only for reconstruction audit and was not used as the formal evaluation set.

The first reference-associated candidate ranked first in 12 of 21 systems, within the Top-3 in 18 systems, and within the Top-5 in 19 systems. When the endpoint was expanded to the first candidate supported by either reference association or independent assembly/MD evidence, the corresponding proportions increased to 20

of 21 for Top-1 and 21 of 21 for both Top-3 and Top-5. The largest differences between the two endpoints occurred in 4W9H, 6WKQ, and 1TCO, where the first reference-associated candidates ranked tenth, sixth, and fourth, respectively, whereas the first-supported candidate ranked first (**Figure 1B,C**).

Target-level and family-clustered bootstrap analyses produced identical point estimates but different interval widths. Family-clustered intervals were generally wider, particularly for reference-associated Top-1, reflecting the repeated protein families and limited effective sample size. In contrast, first-supported Top-3 remained 1.00 under both resampling schemes (**Figure 1D**).

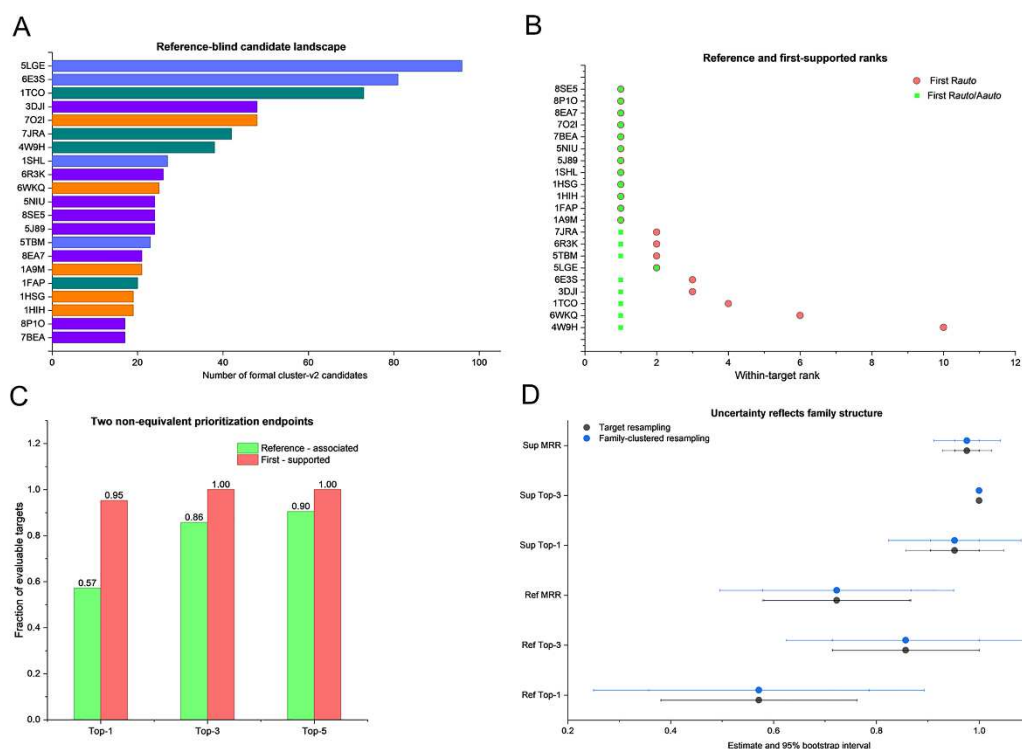


Figure 1. Reference-blind candidate reconstruction and prioritization. (A) Number of formal cluster-v2 candidates per target after exclusion of unmappable records; colors indicate pocket archetypes. (B) Within-target ranks of the first reference-associated candidate and the first-supported candidate; the shaded region denotes ranks 1-3. (C) Overall Top-1, Top-3, and Top-5 proportions for the reference-associated and first-supported endpoints. (D) Point estimates and 95% bootstrap intervals obtained by target-level and family-clustered resampling.

3.2 Individual tools produced divergent native regional priorities

The rank of the first reference-associated region varied substantially across methods and targets. Several systems showed nearly identical prioritization across the individual tools and OIPS-P, whereas others showed broad rank dispersion. For example, the first reference-associated rank ranged from 1 to 10 in 1SHL, from 2

to 10 in 4W9H, and from 1 to 13 in 6E3S. These differences indicate that the constituent methods varied not only in their highest-ranked output but also in how deeply the reference-associated region was placed within the native candidate list (**Figure 2A**).

Mapping the native Top-1 outputs of the individual detectors to cluster-v2 candidates revealed additional tool-dependent variation at the spatial level. Complete agreement among all available single-tool Top-1 regions was observed in only 6 of the 21 systems. The remaining 15 systems contained at least two distinct Top-1 spatial regions, and eight systems contained three or more. Thus, disagreement among the constituent methods frequently reflected selection of different structural regions rather than minor numerical differences in the ranking of the same pocket (**Figure 2B**).

In the 14-system complete-case subset, DoGSite3 showed the highest retrospective reference-associated Top-1 recovery, ranking the reference region first in 12 of 14 systems. OIPS-P ranked the reference-associated region first in 7 of 14 systems and therefore did not uniformly maximize retrospective Top-1 recovery. However, OIPS-P retained the first reference-associated region within the Top-3 in 12 of 14 systems, equaling the highest Top-3 retention observed among the compared methods, and within the Top-5 in 13 of 14 systems. The complete method-level results are summarized in **Table 1** and **Figure 2C**.

Table 1. Reference-associated prioritization of individual tools and OIPS-P in the 14-system complete-case subset.

Method	N	Top-1	Top-3	Top-5	MRR
CavityPlus	14	6/14 (0.429)	10/14 (0.714)	13/14 (0.929)	0.613
DoGSiteScorer	14	9/14 (0.643)	12/14 (0.857)	13/14 (0.929)	0.761
DoGSite3	14	12/14 (0.857)	12/14 (0.857)	13/14 (0.929)	0.880
CASTpFold	14	6/14 (0.429)	10/14 (0.714)	11/14 (0.786)	0.590
SiteMap	14	7/14 (0.500)	9/14 (0.643)	11/14 (0.786)	0.592
OIPS-P	14	7/14 (0.500)	12/14 (0.857)	13/14 (0.929)	0.680

The contribution of OIPS was therefore not to outperform every individual detector on retrospective Top-1 reference recovery. OIPS-P selected one of the modal single-tool Top-1 regions in 14 of the 21 systems. In the remaining seven systems, the integrated Top-1 differed from the modal single-tool region, consistent with the additional contribution of geometry, ligandability, evidence completeness, and oligomeric relevance to the final priority order. In 5TBM, four of the five individual tools prioritized the internal reference-associated region V2C002, whereas OIPS-P prioritized the assembly-associated candidate V2C001. Conversely, all five individual tools and OIPS-P prioritized V2C003 in 4W9H, indicating that the separation from the local reference region was already present in the detector outputs rather than being introduced solely by OIPS

weighting (Figure 2A,B). Together, these results support the use of OIPS as a common candidate-reconstruction and decision layer rather than as a claim of universal detector superiority.

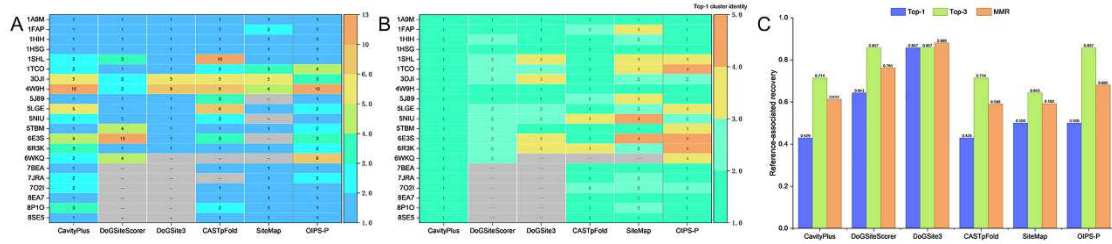


Figure 2. Comparison of native single-tool priorities and OIPS-P. (A) Rank of the first reference-associated region across the 21 systems. Gray cells indicate unavailable, unmappable, or no-hit outputs. (B) Native Top-1 cluster-v2 region selected by each method. Identical numbers within a target indicate the same spatial region. Complete agreement among available tools occurred in 6 of 21 systems. (C) Top-1, Top-3, and MRR in the 14-system complete-case subset.

3.3 Automated structural quality control and post hoc evidence annotation

Automated structural quality control (QC) was applied to the 63 static Top-3 candidates. Thirty-eight candidates were classified as QC_{pass} , 12 as $QC_{boundary\ sensitive}$, 11 as $QC_{possible\ split}$, one as $QC_{possible\ overmerge}$, and one as $QC_{insufficient\ evidence}$; no candidate was classified as $QC_{unmappable}$ (Figure 3A). Thus, most high-ranking candidates were structurally interpretable, whereas boundary- and split-related warnings accounted for the majority of retained uncertainties.

Boundary-related warnings were concentrated in multi-tool candidates with differences in pocket centers or residue envelopes. In 5J89, the first- and third-ranked candidates were classified as $QC_{boundary\ sensitive}$, whereas the second-ranked candidate passed QC. In 5TBM, the assembly-associated first-ranked candidate was classified as $QC_{possible\ split}$, while the second-ranked reference-associated internal site passed QC. These candidates remained in the ranked set, and their QC states did not alter the frozen static scores or ranks.

Post hoc annotation assigned 24 Top-3 candidates to R_{auto} , 26 to A_{auto} , and 13 to U_{auto} , with no X_{auto} assignments (Figure 3B). R_{auto} candidates were most frequent at rank 1, whereas U_{auto} candidates became more frequent toward rank 3; A_{auto} candidates were distributed across all three rank positions. Together, the QC states and post hoc evidence labels provided the operational OIPS-C annotation while preserving unresolved candidates as explicit uncertainties rather than excluding them from the candidate set.

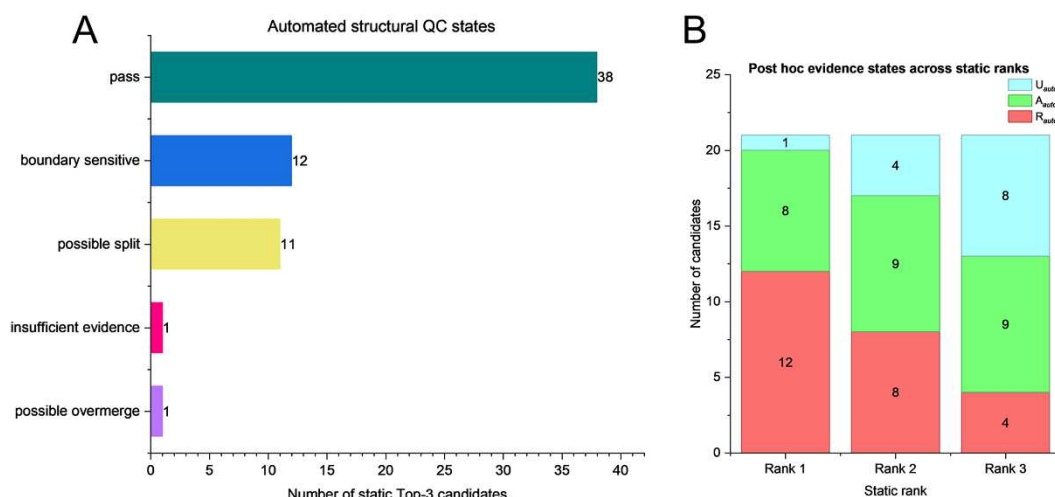


Figure 3. Automated structural quality control and post hoc evidence annotation. (A) Distribution of automated QC states among the 63 static Top-3 candidates. (B) Distribution of R_{auto} , A_{auto} , and U_{auto} evidence states across static ranks 1-3. Each rank contained one candidate from each of the 21 systems. No Top-3 candidate was classified as $QC_{unmappable}$ or assigned X_{auto} .

3.4 Reference-associated and first-supported prioritization showed distinct patterns

Every target contained at least one evaluable R_{auto} candidate. The first reference-associated candidate ranked first in 12 of 21 systems (0.571), within the Top-3 in 18 systems (0.857), and within the Top-5 in 19 systems (0.905). The corresponding mean reciprocal rank (MRR) was 0.723, and the median first-reference rank was 1. Target-level and family-clustered bootstrap analyses produced the same point estimates, whereas the family-clustered intervals were generally wider, particularly for reference-associated Top-1 prioritization, reflecting the repeated protein families in the curated case series (**Figure 1B-D**). Target-level static outputs, including the top-ranked cluster, its $OIPS - P_{static}$ score, evidence state, and the two ranking endpoints, are summarized in **Table 2**.

When the endpoint was expanded to the first candidate supported by either reference association or independent assembly/MD evidence, Top-1 prioritization increased to 20 of 21 systems (0.952), Top-3 and Top-5 prioritization reached 21 of 21 systems (1.000), and MRR increased to 0.976. The only system without a first-supported candidate at rank 1 was 5LGE, in which the static top candidate remained U_{auto} , whereas the first reference-associated and first-supported region ranked second (**Figure 1B,C; Table 2**). The first-supported endpoint was treated as an assembly-aware interpretive endpoint rather than an independent biological ground truth.

The largest separations between the two endpoints occurred in 1TCO, 6WKQ, and 4W9H. Their first reference-associated candidates ranked fourth, sixth, and tenth, respectively, whereas an A_{auto} -supported candidate ranked first in each system. At the category level, interface/ protein-protein interaction (PPI) systems showed reference-associated Top-1 and Top-3 proportions of 0.750 and 1.000, respectively, indicating relatively strong agreement between reference location and assembly-aware prioritization. Complex-dependent systems showed the largest divergence, with reference-associated Top-1 and Top-3 proportions of 0.250 and 0.500, whereas both first-supported proportions were 1.000. These category-level comparisons were descriptive because of the small group sizes.

Table 2. Target-level OIPS – P_{static} prioritization, post hoc evidence, and local weight sensitivity.

PDB	Pocket archetype	Top-ranked cluster	Top OIPS – P_{static}	Top evidence state	First R_{auto} rank ¹	First-supported rank ¹	Top-1 retention ²
1A9M	Orthosteric	V2C001	95.26	R_{auto}	1	1	10/10
1FAP	Complex-dependent	V2C002	87.49	R_{auto}	1	1	10/10
1HIH	Orthosteric	V2C004	84.85	R_{auto}	1	1	10/10
1HSG	Orthosteric	V2C001	95.24	R_{auto}	1	1	10/10
1SHL	Allosteric	V2C002	72.85	R_{auto}	1	1	10/10
1TCO	Complex-dependent	V2C004	91.00	A_{auto}	4 (3-4)	1	10/10
3DJI	Interface/PPI	V2C004	89.59	A_{auto}	3	1	10/10
4W9H	Complex-dependent	V2C003	87.55	A_{auto}	10 (9-10)	1	10/10
5J89	Interface/PPI	V2C003	83.57	R_{auto}	1	1	10/10
5LGE	Allosteric	V2C006	81.76	U_{auto}	2 (2-3)	2 (2-3)	10/10
5NIU	Interface/PPI	V2C002	79.82	R_{auto}	1	1	10/10
5TBM	Allosteric	V2C001	84.10	A_{auto}	2 (2-3)	1	10/10
6E3S	Allosteric	V2C003	83.31	A_{auto}	3 (3-4)	1	10/10
6R3K	Interface/PPI	V2C002	81.10	A_{auto}	2	1	10/10
6WKQ	Orthosteric	V2C002	77.72	A_{auto}	6 (6-7)	1	10/10
7BEA	Interface/PPI	V2C003	79.48	R_{auto}	1	1	10/10
7JRA	Complex-dependent	V2C006	83.51	A_{auto}	2	1	10/10
7O2I	Orthosteric	V2C001	77.29	R_{auto}	1 (1-2)	1	9/10
8EA7	Interface/PPI	V2C003	79.79	R_{auto}	1	1	10/10
8P1O	Interface/PPI	V2C001	73.70	R_{auto}	1	1	10/10
8SE5	Interface/PPI	V2C003	80.07	R_{auto}	1	1	10/10

¹ Values outside parentheses indicate the baseline rank. Parenthetical values indicate the minimum-maximum rank across ten one-at-a-time module-weight perturbations, in which each of the five static module weights was independently decreased or increased by 20% followed by weight renormalization. Parenthetical ranges are omitted when the rank remained unchanged.

² Number of the ten perturbation scenarios in which the baseline Top-1 cluster was retained. Candidate definitions, module values, missing-value handling, and cluster-v2 assignments were held fixed. Reference-ligand, redocking, and dynamic evidence were not included in the recalculated static rankings.

OIPS – P_{static} prioritization was locally robust to $\pm 20\%$ perturbation of the five module weights. Twenty of the 21 systems retained the baseline Top-1 candidate in all ten scenarios, whereas 7O2I retained it in nine scenarios. Reference-associated ranks shifted by no more than one position in seven systems, and the first-supported rank remained unchanged except in 5LGE, where it varied between ranks 2 and 3 (**Table 2**).

Candidate prioritization was largely stable under moderate perturbation of the prespecified module weights. Across the ten one-at-a-time $\pm 20\%$ perturbation scenarios, the baseline Top-1 candidate was retained in 20-21 of the 21 systems. Nine scenarios retained the baseline Top-1 candidate in all systems, and the only Top-1 change occurred in 7O2I after a 20% reduction in Plig, where V2C003 replaced the baseline V2C001 candidate. Mean Top-3 Jaccard overlap ranged from 0.905 to 1.000, and the median within-target Spearman rank correlation ranged from 0.995 to 1.000. Reference-associated Top-1 prioritization varied from 11/21 to 12/21, whereas reference-associated Top-3 prioritization varied from 17/21 to 19/21. In contrast, the first-supported endpoint remained unchanged at 20/21 for Top-1 and 21/21 for Top-3 across all perturbation scenarios. Target-level rank ranges and Top-1 retention are summarized in Table 2. These results indicate that the principal candidate-prioritization pattern was locally robust to moderate weight changes, although shortlist composition and reference-associated rank remained sensitive in selected systems. Because Orel represents a biologically distinct assembly-aware component, its contribution was additionally assessed by complete module removal rather than by local weight perturbation alone.

3.5 Effect of oligomeric relevance on candidate prioritization

Removal of Orel slightly increased reference-associated Top-1 recovery from 12/21 (0.571) to 14/21 (0.667) and increased reference MRR from 0.723 to 0.752, but reduced Top-3 retention from 18/21 (0.857) to 17/21 (0.810). In contrast, the first-supported endpoint consistently decreased after Orel removal, with Top-1, Top-3, and MRR changing from 0.952, 1.000, and 0.976 to 0.905, 0.905, and 0.925, respectively (**Figure 4A**). These results indicate that Orel does not simply optimize crystallographic reference recovery but contributes to preservation of broader assembly-supported candidate prioritization.

Target-level effects were heterogeneous. Reference-associated ranks improved after removing Orel in several systems, including 4W9H (rank 10 to 3), 5TBM (rank 2 to 1), and 6E3S (rank 3 to 1). Conversely, reference-associated ranking decreased in 5LGE (rank 2 to 6), 6R3K (rank 2 to 5), and 7JRA (rank 2 to 3) (**Figure 4B**). These changes reflect the different balance between reference-site recovery and oligomerization-related structural relevance.

Changes in first-supported ranking were generally smaller than reference-only changes, with the largest shift observed for 6R3K, where the first-supported candidate moved from rank 1 to rank 4 after Orel removal (**Figure 4C**). Across the 21 systems, only four targets changed top-candidate identity after ablation (5LGE, 5TBM, 6E3S, and 6R3K). The mean within-target Spearman correlation between full and ablated rankings was 0.777, indicating that overall candidate ordering was largely preserved despite local redistribution in a subset of systems (**Figure 4D**).

Together, these results demonstrate that Orel introduces an assembly-aware constraint that may slightly reduce reference-only optimization while improving preservation of structurally supported oligomeric candidates.

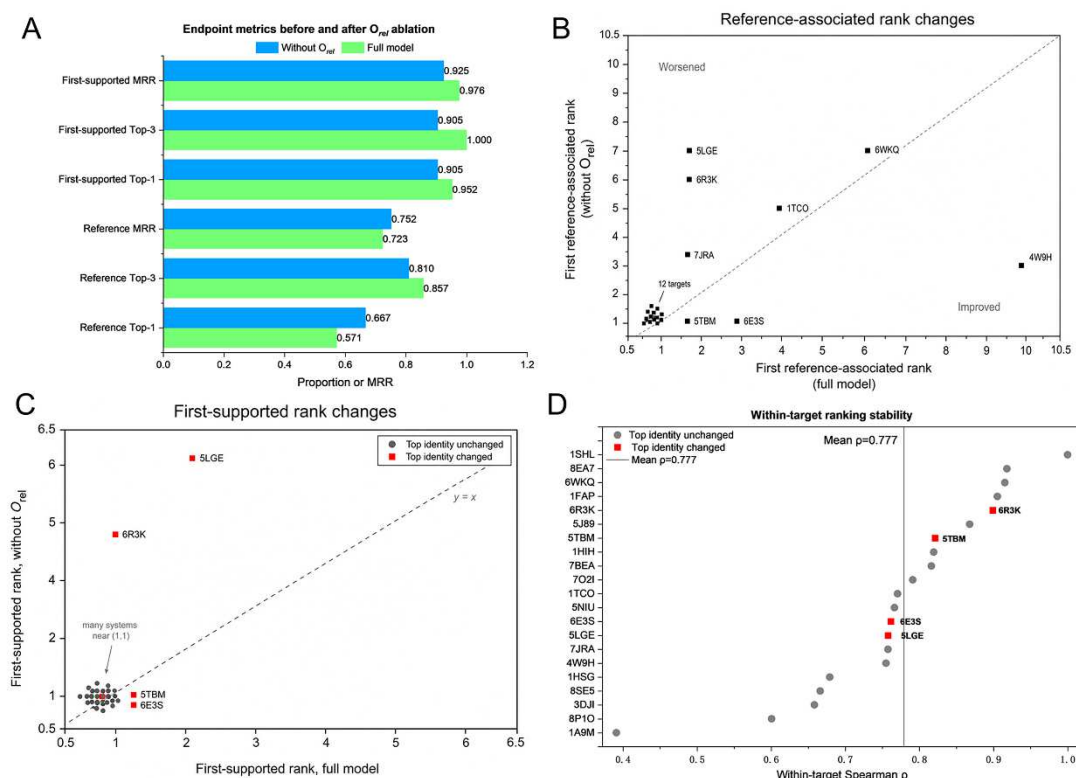


Figure 4. Ablation analysis reveals the contribution of oligomeric relevance to evidence-integrated pocket prioritization.(A) Endpoint comparison between the complete OIPS model and the model lacking Orel. Reference-associated recovery and first-supported candidate preservation were evaluated using Top-1, Top-3, and MRR metrics.(B) Changes in reference-associated

ranking after Orel removal. Each point corresponds to one target, comparing the rank of the first reference-associated candidate between the full and ablated models.(C) Changes in first-supported ranking after Orel removal. This endpoint integrates reference-associated and assembly-supported candidates to evaluate preservation of biologically interpretable candidate pockets.(D) Stability of within-target candidate ordering after Orel ablation. Spearman correlation coefficients quantify the similarity between full-model and ablated rankings across the 21 oligomeric systems.

3.6 MD and redocking qualified static candidate prioritization

Molecular dynamics and redocking were analyzed as post-static evidence and did not alter the frozen $OPIS - P_{static}$ scores or within-target rankings. The MD dataset comprised 23 trajectories across 15 of the 21 systems, including native-ligand, redocked-ligand, and apo simulations. Six systems had no MD data. Seven trajectories were classified as concordant, three as partially concordant or boundary-shifted, ten as conflicting, and three as apo-only. **Figure 5A** shows the static ranks of the candidates most strongly supported by the 20 ligand-containing trajectories; the three apo-only trajectories were classified separately and were not included in this panel. These results indicate that dynamic evidence frequently refined the spatial interpretation of a static candidate rather than simply confirming or rejecting the top-ranked region.

Static-dynamic conflicts were concentrated in cases where cluster-v2 separated a localized ligand-contact region from a broader assembly-associated candidate. In 4W9H, the native trajectory mapped predominantly to the local reference-associated cluster V2C001 rather than the assembly-associated static Top-1 candidate V2C003. In 5LGE, the native and redocked trajectories mapped most strongly to V2C016, whereas the intended 6VN-associated region was represented by neighboring clusters and the static Top-1 candidate V2C006 remained U_{auto} . In 6WKQ, both native and redocked trajectories supported the reference-associated V2C011 region, whereas V2C002 remained the static Top-1 candidate. A conflict designation therefore did not necessarily indicate failure of the static model, but instead identified cases in which assembly-aware prioritization and localized dynamic contacts supported different structural hypotheses.

Redocking assessed local pose recoverability within the reference-defined region. Eighteen systems showed ligand heavy-atom RMSD values of ≤ 2 Å, two were within the 2-3 Å warning range, and 4W9H showed poor pose reproduction with an RMSD of 3.544 Å (**Figure 5B**). These results supported local reference-pose recovery in most systems but reinforced the uncertainty of 4W9H, where the static Top-1 candidate, reference-associated region, MD-supported region, and redocked pose were not mutually concordant.

Residue-level contact occupancies provided direct dynamic context for three representative systems (**Figure 5C**). Persistent contacts were defined as interactions present in at least 30% of the analyzed frames. In 5J89,

persistent contacts involved residues from both protomers, supporting retention of an interface-associated ligand-contact region while also illustrating uncertainty in its local boundary. In 5TBM, persistent interactions were concentrated within a compact set of residues, consistent with localized binding in the internal allosteric pocket. In 4W9H, extensive persistent contacts were retained within the local MD-supported region despite its disagreement with the broader assembly-associated static Top-1 candidate. Contact persistence therefore supported local ligand retention but did not by itself resolve competition between ligand-centered and assembly-centered candidates.

MD and redocking were most informative when the static Top-1 candidate was boundary-sensitive, when the reference-associated region ranked outside the Top-3, or when the assembly-associated candidate differed from the ligand-centered region. These post-static evidence layers were therefore incorporated into OIPS-C to qualify candidate interpretation rather than to retrospectively rescore $OPIS - P_{static}$.

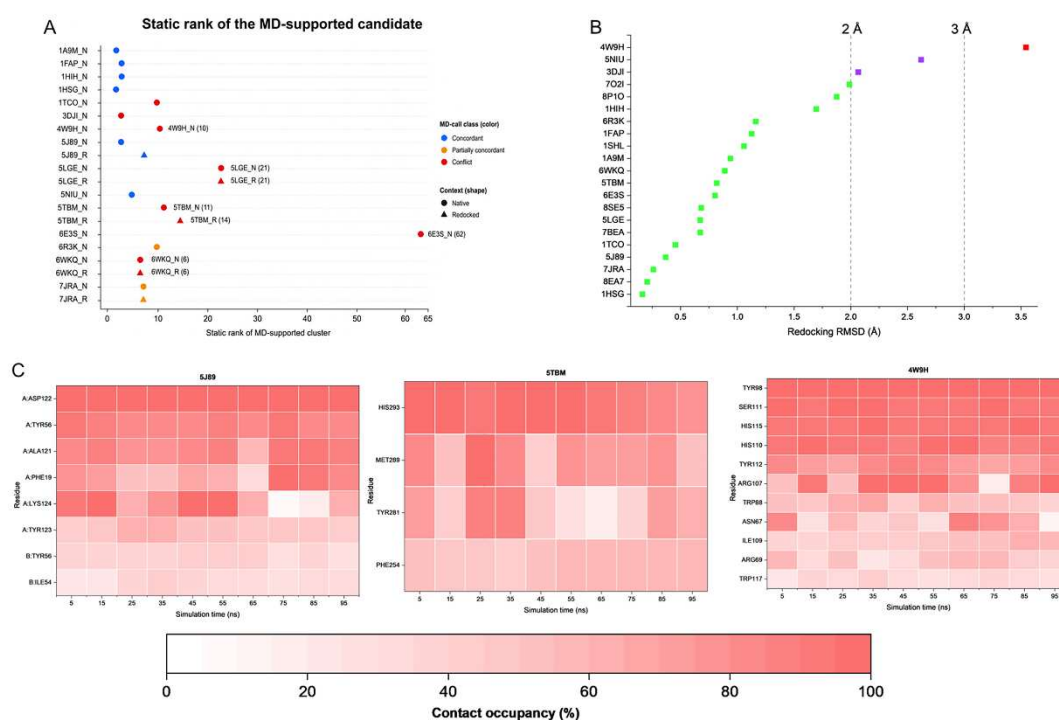


Figure 5. Dynamic and structural validation of OIPS-selected pocket interpretations. (A) Static ranks of MD-supported candidates across ligand-containing trajectories. Colors indicate MD interpretation classes, and symbols distinguish native and redocked simulations. (B) Redocking RMSD of OIPS-selected Top-1 candidates across 21 systems. Dashed lines indicate 2 Å and 3 Å reference thresholds. (C) Residue-level ligand-contact occupancy during 100 ns MD simulations of representative systems (5J89, 5TBM, and 4W9H). Heatmaps show the percentage of frames containing at least one ligand contact per residue, with persistent contacts defined as occupancy $\geq 30\%$.

3.7 Structural atlas of representative OIPS-prioritized pocket archetypes

To illustrate the structural breadth of the OIPS framework beyond target-level ranking metrics, five representative systems were selected to capture distinct pocket archetypes observed in the curated case series (**Figure 6**). These examples included an orthosteric reference pocket, an interface-associated pocket, an internal allosteric pocket, a trimeric assembly-associated pocket, and a complex-dependent local pocket. For each system, the OIPS-prioritized region was visualized in the context of the full oligomeric assembly together with a local zoomed view showing the pocket surface, crystallographic reference ligand, and representative lining residues.

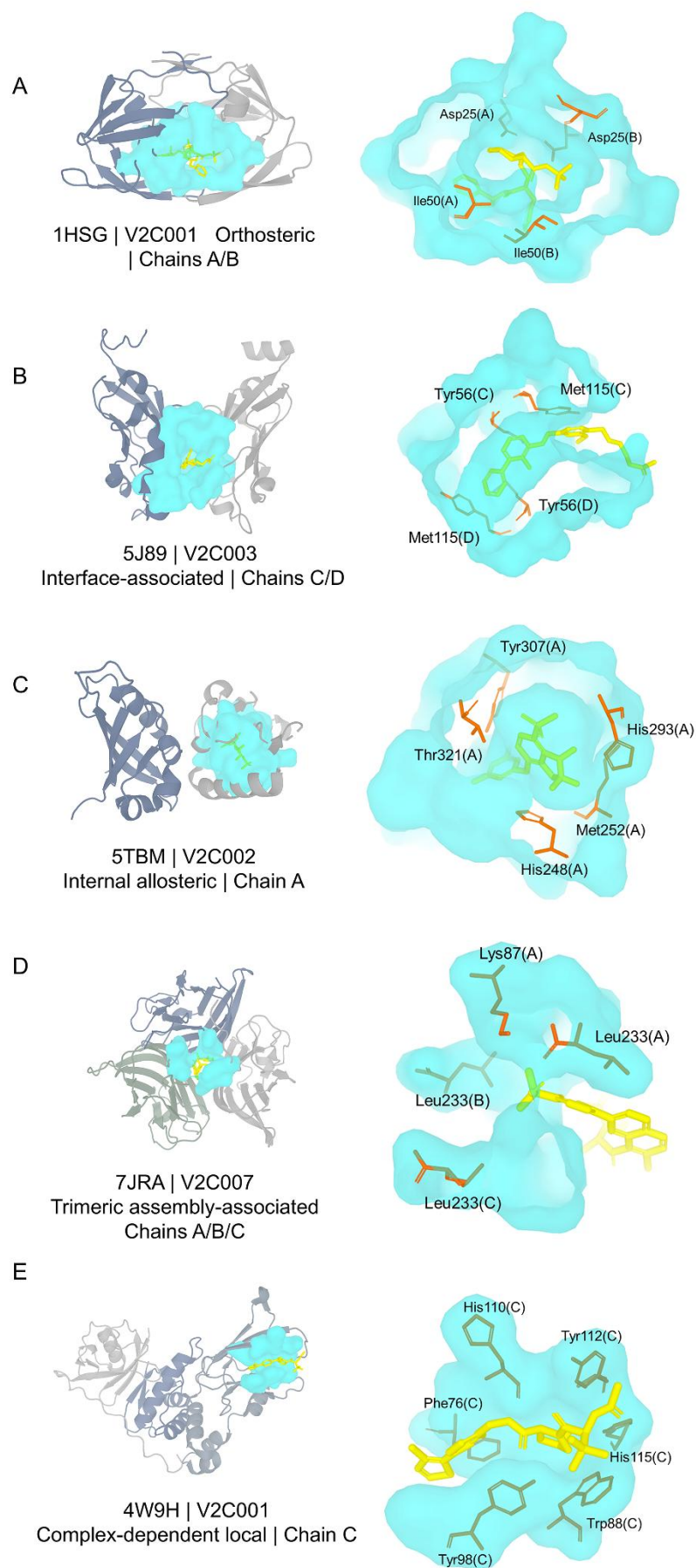


Figure 6. Structural atlas of representative OIPS-prioritized pocket archetypes. (A) Orthosteric reference pocket in 1HSG. (B) Interface-associated pocket in 5J89. (C) Internal allosteric pocket in 5TBM. (D) Trimeric assembly-associated pocket in 7JRA. (E) Complex-dependent local pocket in 4W9H.

For each system, the left panel shows the full oligomeric assembly and the right panel shows a local enlarged view of the OIPS Top-1 pocket. Pocket surfaces indicate the prioritized candidate, yellow sticks indicate the crystallographic reference ligand, and selected lining residues and chain identifiers illustrate the local structural context.

The structural atlas showed that the prioritized candidates were not restricted to one pocket class. Instead, OIPS captured compact ligand-centered clefts, inter-chain interface pockets, assembly-shaped local cavities, and pockets whose interpretation depended on preservation of the full biological complex. These representative structures therefore provide a direct visual overview of the pocket categories addressed by the present framework and complement the ranking and evidence analyses summarized above.

Table 3. Quantitative structural and evidence characteristics of the representative OIPS candidates.

PDB	Candidate	Pocket interpretation	Rank	OIPS-P	Tools	Chains	Core /envelope residues	Interface fraction	Cluster diameter (Å)	DCC (Å)	QC _{state}	Redocking RMSD (Å)
1HSG	V2C001	Orthosteric	1	95.24	5	A/B	33/57	0.702	2.417	0.613	QC _{pass}	0.164
5J89	V2C003	Interface-associated	1	83.57	4	C/D	35/65	0.369	11.084	3.337	QC _{boundary}	0.368
5TBM	V2C002	Internal allosteric	2	82.13	5	A	25/48	0.104	4.468	0.302	QC _{pass}	0.819
7JRA	V2C007	Trimeric assembly-associated	2	78.11	2	A/B/C	9/77	0.506	2.252	5.022	QC _{possible split}	0.260
4W9H	V2C001	Complex-dependent local	10	70.24	5	C	10/22	0.000	10.023	4.232	Not evaluated	3.544

^a Automated structural QC was formally applied only to the static Top-3 candidates. Therefore, 4W9H V2C001, which ranked tenth, was not assigned a formal Top-3 QC state. Evidence labels and redocking results were appended post hoc and did not alter static ranks.

DCC, distance between the cluster-v2 candidate center and the crystallographic reference-ligand center. The interface fraction represents the fraction of candidate envelope residues classified as inter-chain interface residues. Core residues were supported by at least half of the contributing tools, whereas envelope residues represent the union of the representative tool-level residue sets.

Quantitative characteristics of the five illustrated candidates are summarized in Table 3. The selected regions differed substantially in static rank, cross-tool support, oligomeric chain contribution, interface association, cluster boundary, and proximity to the crystallographic reference ligand. The representative candidates were not restricted to static Top-1 regions, because the purpose of the structural atlas was to illustrate distinct pocket interpretations rather than only the highest-ranked candidate in each target.

The atlas also helps distinguish structurally different decision contexts. In the orthosteric reference example, the prioritized candidate closely recapitulated the known ligand-binding region. In contrast, the interface-associated, trimeric, and complex-dependent examples showed stronger contributions from assembly context and multi-chain structural organization. More detailed evidence conflicts and boundary-related ambiguities in selected systems are examined in the following section.

3.8 Representative cases illustrate three distinct candidate-selection problems

Building on the broader structural overview in **Figure 6**, three systems were further examined as detailed representative case studies to illustrate boundary uncertainty, competing mechanistic hypotheses, and cross-layer evidence conflict.

5J89 represented boundary uncertainty around a consistently supported interface pocket(**Figure7A**). The reference-associated V2C003 candidate ranked first with an $OIPS - P_{static}$ score of 83.571. It was supported by four pocket-detection tools, contained residues from both chains C and D, and showed an interface-residue fraction of 0.369. However, its 11.084 Å cluster diameter resulted in a $QC_{boundary-sensitive}$ assignment. Removal of Orel did not alter either the top-candidate identity or the reference-associated rank. Native and redocked MD mapped most strongly to the neighboring V2C001 and V2C021 boundaries, respectively, but both trajectories were classified as concordant at the regional decision level. Redocking reproduced the crystallographic pose with an RMSD of 0.368 Å. Thus, the evidence consistently identified the same interface region, whereas the remaining uncertainty concerned the precise boundary used to define the actionable pocket.

5TBM illustrated competition between an assembly-associated candidate and an established internal allosteric site(**Figure7B**). The assembly-associated V2C001 candidate ranked first with an $OIPS - P_{static}$ score of 84.102, involved both chains, and showed an interface-residue fraction of 0.405. In contrast, the PT2385-associated internal pocket V2C002 ranked second, passed structural QC, and was centered within 0.302 Å of the reference region. Removing Orel promoted V2C002 to rank 1 and changed the top-candidate identity, demonstrating that the preference for V2C001 depended on assembly-aware evidence. Redocking supported

the local reference geometry with an RMSD of 0.819 Å. However, native and redocked MD mapped to lower-ranked local candidates V2C023 and V2C020 rather than directly to either static Top-1 or V2C002. This pattern indicates that the dynamic ligand-contact region was divided across narrower cluster-v2 boundaries. Accordingly, V2C001 represents an assembly-associated structural hypothesis, whereas V2C002 remains the chemically established internal allosteric hypothesis; the appropriate candidate therefore depends on the intended mechanism rather than on a single universal definition of correctness.

4W9H retained the strongest cross-layer evidence conflict(**Figure7C**). The assembly-associated V2C003 candidate ranked first with an $OIPS - P_{static}$ score of 87.551, support from all five pocket-detection tools, contributions from two chains, and an interface-residue fraction of 0.396. The local reference-associated V2C001 candidate ranked tenth, despite substantial overlap with the crystallographic 3JF ligand region. Removal of Orel improved the reference-associated rank from 10 to 3 but did not displace V2C003 from the top position. In contrast, native MD mapped directly to V2C001, with complete recall of the persistent MD-contact residues, whereas redocking failed to reproduce the crystallographic pose satisfactorily RMSD = 3.544Å. The evidence therefore supported two distinct hypotheses: static assembly-aware evidence favored V2C003, whereas the crystallographic ligand and MD favored the more localized V2C001 region. Redocking added a pose-recovery warning rather than resolving the disagreement. Both candidates should therefore remain explicit until a targeted structural or experimental analysis establishes which mechanism is relevant. Together, these cases demonstrate that evidence disagreement can arise from pocket-boundary variation, competition between assembly-associated and ligand-centered sites, or genuine cross-layer conflict. OIPS preserves these distinctions through separate priority, pocket-type, and evidence-state outputs rather than collapsing them into a single retrospective validation label.

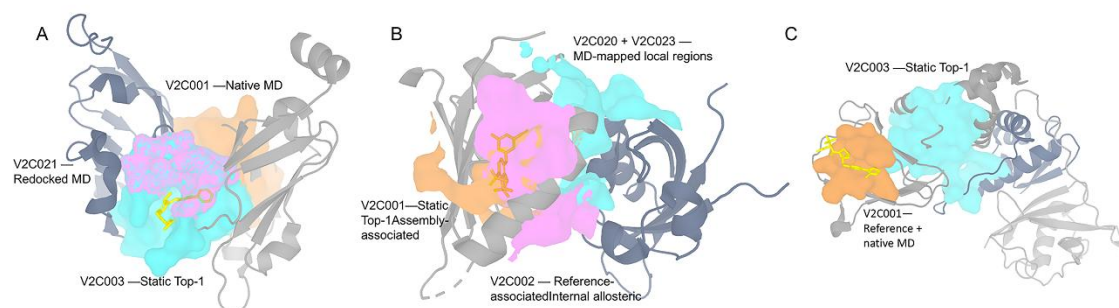


Figure 7. Representative structural interpretation of three candidate-selection scenarios. (A) 5J89 showing a reference-associated interface pocket with native- and redocked-MD-mapped boundary regions. (B) 5TBM showing competition between the assembly-associated static Top-1 candidate and the internal allosteric reference pocket, together with MD-mapped local regions. (C) 4W9H

showing spatial separation between the assembly-associated static Top-1 candidate and the local pocket supported by the reference ligand and native MD. Yellow sticks indicate the crystallographic reference ligand.

4. Discussion

4.1 Principal findings and scope of OIPS

Rather than demonstrating that a single scoring scheme can identify a universally correct binding pocket, this study shows that pocket selection in oligomeric proteins is a problem of defining and comparing competing structural hypotheses. Reference-blind cluster-v2 reconstruction separated predictions that were spatially adjacent but operationally distinct, demonstrating that candidate identity depends on how parent cavities, nested subpockets, and inter-chain grooves are consolidated. Within this reconstructed candidate space, *OIPS* – Pretained reference-associated regions within the high-priority shortlist in most systems but did not uniformly maximize retrospective Top-1 reference recovery. Instead, its assembly-aware component preserved alternative candidates whose relevance depended on the oligomeric structural context. Post-static molecular dynamics and redocking further distinguished agreement in site identity from disagreement in pocket boundary, local ligand-contact geometry, or mechanistic interpretation.

These findings support a deliberately restricted interpretation of OIPS. Its contribution is not a new cavity detector, an affinity predictor, or a universal measure of biological correctness. Rather, OIPS provides a traceable decision layer that separates candidate priority (*OIPS* – *P*), working pocket type (*OIPS* – *T*), and evidence state (*OIPS* – *C*). This separation is particularly important in oligomeric systems, where a compact orthosteric or internal allosteric site may coexist with a broader assembly-associated region. Accordingly, reference-associated recovery and the first-supported endpoint should be regarded as complementary interpretive measures rather than interchangeable definitions of success. Because the assembly-supported component partially overlaps with the interface information encoded by O_{rel} , the first-supported endpoint represents internal interpretive consistency rather than independent external validation.

The central advance is therefore both methodological and conceptual: OIPS retains disagreement among candidate pockets as actionable information. Boundary ambiguity, alternative mechanistic hypotheses, and cross-layer conflicts are not forced into a single retrospective accuracy value but are recorded explicitly to guide the next docking calculation, simulation, structural analysis, or experiment.

4.2 Candidate reconstruction is part of pocket interpretation

Pocket integration is often treated as preprocessing, but the present results indicate that candidate reconstruction is itself part of pocket interpretation because it defines the structural units that are subsequently

ranked. Pocket-detection methods use different surface representations and segmentation procedures, including sphere-, grid-, tessellation-, and surface-based approaches, and may therefore assign different centers, boundaries, or levels of granularity to the same region.[40] Some methods additionally distinguish broad parent cavities from localized subpockets, which may better represent the portion of a cavity occupied by a ligand.[41]

These observations are consistent with our finding that cross-tool disagreement may arise from candidate definition rather than only from ranking. Pocket generation and pocket ranking are separate analytical steps, and method comparison is sensitive to both the ranking protocol and the operational definition of a correct prediction. Partial spatial or residue overlap is therefore insufficient to establish that two outputs represent the same actionable pocket.

Cluster-v2 addressed this problem in two stages. Same-tool consolidation prevented parent pockets, nested subpockets, or alternative outputs from one detector from contributing duplicate votes. Medoid-constrained cross-tool grouping then limited transitive chaining, through which mutually incompatible predictions could otherwise be merged via an intermediate overlapping record. This issue is particularly relevant in oligomeric proteins, where an inter-chain groove may lie adjacent to a compact internal cavity or contain several locally distinct depressions. The reconstruction separated 79 previous working clusters into two or more candidates and generated 733 traceable candidates across the 21 systems.

Candidate definition also affects apparent performance and downstream use. Overmerged candidates may recover a reference ligand through only one part of an oversized envelope, while combining regions with different centers, geometry, or assembly relationships. Conversely, excessive fragmentation may divide one functional site into several nearly equivalent candidates and reduce apparent consensus. Previous subcavity studies have similarly identified pocket overspanning as an important limitation of broad cavity definitions.[41] Cluster-v2 was therefore designed to reduce both overmerging and overfragmentation while retaining boundary uncertainty.

The 5J89 case illustrates this distinction. Multiple tools supported the same broader inter-chain region but differed in the local center and residue envelope. The uncertainty therefore concerned the actionable boundary rather than the existence of the site. The core-and-envelope representation preserved both the shared regional interpretation and the unresolved boundary variation.

Candidate reconstruction thus defines the objects to which subsequent scores, ranks, and evidence labels are assigned. Pocket-prediction evaluation should accordingly distinguish failure to detect a region, failure to prioritize an available candidate, and disagreement over candidate boundary or granularity.

4.3 Oligomeric relevance creates competing ranking objectives

The ablation analysis indicates that pocket prioritization in oligomeric proteins is a multi-objective problem. Removing O_{rel} modestly improved retrospective Top-1 recovery of the crystallographic reference region, but reduced reference-associated Top-3 retention and weakened preservation of the first-supported candidate. Assembly awareness therefore does not simply add another descriptor to a ligand-centered score; it preserves candidates whose relevance depends on the oligomeric context, even when they do not coincide with the observed ligand position.

This trade-off is consistent with the broader distinction between compact enclosed pockets and larger interface-associated regions. Previous cavity analyses have noted that inter-chain cavities may receive high geometric ranks because they are larger than neighboring enclosed ligand-binding sites, even though the two regions represent different structural contexts.[41] A crystallographic ligand consequently supports one experimentally observed binding mode, but does not establish that this site is the only actionable region in the assembled protein. Orthosteric, internal allosteric, and interface-associated candidates may coexist and imply different mechanisms, including competitive inhibition, allosteric modulation, interface disruption, or complex stabilization.

The 5TBM system illustrates this distinction. Most individual detectors prioritized the established internal allosteric reference pocket, whereas the complete OIPS model ranked an assembly-associated candidate first; removing O_{rel} restored the reference-associated site to the first position. This change should not be interpreted simply as correction of an erroneous ranking. The candidates represent different intervention hypotheses, and their relevance depends on the intended mechanism and downstream design strategy. The direction of rank changes after ablation varied among targets, indicating that O_{rel} acts primarily as a local discriminator where ligand-centered and assembly-centered evidence diverge rather than imposing a uniform interface bias. O_{rel} is an assembly-awareness term, not a measure of interface energy or biological function. It summarizes candidate-interface overlap, chain contribution, center-to-interface distance, and assembly context, but does not include buried surface area, interaction energetics, hotspot importance, evolutionary conservation, or experimental mutational evidence. A high value therefore indicates structural association with the retained assembly, not proof of ligandability or therapeutic relevance.

The first-supported endpoint requires the same caution because its assembly-supported component partially overlaps with the interface information encoded by O_{rel} . Its decline after O_{rel} removal should therefore be interpreted as reduced internal assembly-aware consistency rather than loss of independent external validation. Reference association and oligomeric relevance should consequently remain separate dimensions: the former asks whether a candidate recovers a known ligand-binding region, whereas the latter asks whether it is structurally connected to the assembled protein context.

4.4 Static and post-static evidence resolve different levels of pocket uncertainty

Static pocket prioritization, redocking, and molecular dynamics address different levels of structural evidence. The frozen OIPS-P ranking asks which candidate is most strongly supported by the ligand-depleted oligomeric structure. Redocking tests whether a known ligand pose can be reproduced within a predefined local region, whereas MD evaluates the persistence and spatial distribution of interactions under a specific starting state and simulation protocol. Agreement among these layers strengthens a common interpretation, but disagreement does not necessarily identify one method as incorrect.

Redocking should therefore be interpreted as a test of local pose recoverability rather than blind validation of pocket selection. Because the docking region was defined from the crystallographic ligand, a low RMSD supports the compatibility of the local geometry with the known pose but does not establish that the region would be selected over an assembly-associated alternative. Conversely, poor pose recovery may reflect receptor conformation, ligand preparation, or scoring limitations rather than the absence of a meaningful pocket.

MD provides complementary but similarly conditional evidence. Persistent contacts indicate that a local interaction network remains stable within the sampled trajectory, but the outcome depends on the initial structure, ligand state, force field, and accessible timescale. Previous work has likewise shown that static pocket predictors can identify regions for focused MD or enhanced sampling, while emphasizing that their native rankings are not necessarily optimized for cryptic-pocket identification and that apo-to-holo conformational transitions may remain difficult to sample.[42] This supports using static and dynamic analyses as complementary evidence rather than treating either as definitive ground truth.

The present results support a conceptual distinction among three patterns of disagreement. Boundary variation occurs when static and dynamic evidence identify the same broader region but differ in center position or residue extent. Candidate competition occurs when ligand-centered and assembly-associated regions remain distinct but mechanistically plausible. Cross-layer conflict occurs when static ranking, reference association,

redocking, and MD favor spatially different candidates. These patterns require different responses rather than a common designation of prediction failure.

The representative systems illustrate this distinction. In 5J89, the principal uncertainty concerned the local boundary of a consistently supported inter-chain region. Post-static evidence therefore helps refine the actionable residue envelope rather than determine whether the broader site exists. In contrast, 4W9H showed spatial separation between the assembly-associated static Top-1 candidate and the local reference- and MD-supported region, together with the poorest redocking recovery. The available evidence therefore remained insufficient for a single confident assignment.

For this reason, MD and redocking were recorded in OIPS-C rather than used to retrospectively rescore OIPS-P. Maintaining the frozen static ranking preserves the analytical audit trail and distinguishes initial prioritization from subsequent evidence qualification. Post-static evidence should guide the next discriminating analysis, such as alternative docking grids, replicate simulations, receptor-state comparison, or experimental validation rather than automatically overwrite the original candidate order.

4.5 From ranked candidates to modality-aware structural decisions

The practical output of OIPS is not an isolated numerical rank, but a candidate interpreted together with its pocket type, structural QC state, supporting evidence, and unresolved alternatives. This distinction is necessary because candidates occupying the same rank may represent compact ligand-centered cavities, interface-associated clefts, assembly-dependent pockets, or complex-induced local regions. Their downstream treatment should therefore depend on the structural hypothesis they represent rather than on rank alone.

Pocket type directly affects computational setup. Compact orthosteric or internal allosteric candidates can generally be investigated using a focused local docking region. Interface-associated or complex-dependent candidates require preservation of the complete biological assembly because removal of a partner chain may alter or eliminate the pocket. However, the retained docking region should remain restricted to a localized cleft rather than span the entire protein-protein interface. Interface cavities can be substantially larger than neighboring enclosed ligand-binding pockets, making the whole interface an excessively broad and chemically heterogeneous search space.

Boundary uncertainty also requires a distinct strategy. When several candidates describe different local envelopes of the same broader region, merging them into one oversized pocket may conceal meaningful differences. Candidate-specific grids can instead be compared through pose convergence, interaction consistency, and targeted simulations. By contrast, an internal allosteric cavity and an assembly-associated

pocket should not be treated as alternative boundaries of one site, because they imply different mechanisms, ligand requirements, and experimental assays.

An additional implication is that candidate interpretation may inform therapeutic-modality selection rather than only docking-site selection. Oligomeric protein surfaces can contain structural features at multiple spatial scales, including a compact local core, a broader interaction envelope, and assembly-dependent surfaces contributed by one or more subunits. These features may favor different intervention formats. Compact and enclosed cores are generally more compatible with fragments or conventional small molecules, whereas broader but geometrically delimited regions may provide starting points for constrained peptides, cyclic peptides, or peptidomimetics. Extended or conformationally dependent surfaces may instead be more appropriate for antibody-based recognition. Interface-associated candidates may also support stabilization rather than disruption of an oligomeric assembly. Small molecules have been shown to bind an allosteric pocket at the TLR8 dimer interface and stabilize an inactive receptor state, while dual-site molecular glues have been designed to engage both a primary binding region and its surrounding protein-protein interaction context.[43, 44]

The core-and-envelope representation provides a potential bridge between small-molecule and peptide design. Core residues may define an anchor interaction for a fragment, compact pharmacophore, or noncanonical amino acid, whereas the surrounding envelope defines the space available for peptide extension, cyclization, or peptidomimetic growth. Structure-guided anchor-extension strategies have demonstrated that a known binding group can be incorporated into a cyclic peptide and supplemented with additional interactions around the target surface. Computational peptide-design workflows have similarly combined sequence generation, peptide docking, and molecular dynamics to refine target-specific binders. OIPS candidates could therefore support a graded design path from fragment or small molecule, through peptidomimetic and macrocyclic peptide formats, rather than assigning each structural region to a single predefined modality.[45, 46]

Assembly-aware interpretation may also have implications for monoclonal and multispecific antibody engineering. OIPS is not an antibody-epitope predictor, but its chain-contribution, boundary, and assembly-context information may help distinguish monomeric surface regions from quaternary or inter-protomer sites. For bispecific antibodies, the relative positions and orientations of two candidate regions may help determine whether simultaneous engagement is geometrically plausible and whether intra-subunit or inter-subunit crosslinking should be considered. Structure-guided multispecific-antibody studies have used epitope proximity and domain-terminal distances to select crosslinking topologies, and constrained antibody geometry

has been shown to alter receptor agonism. Extending OIPS toward this application would require additional descriptors for surface accessibility, glycosylation, epitope conservation, conformational stability, linker reach, and antibody developability.[47]

OIPS-T and OIPS-C are therefore essential complements to OIPS-P. OIPS-T records the working pocket hypothesis, whereas OIPS-C indicates whether that interpretation is supported, boundary-sensitive, competing, or unresolved. Static pocket predictions have similarly been proposed as starting points for focused docking, MD, or enhanced-sampling calculations rather than as definitive site assignments. The next calculation should consequently be selected according to the source of uncertainty, without retrospectively collapsing all evidence into a revised score.

OIPS changes the output of pocket analysis from a single "predicted binding site" to a ranked set of testable structural hypotheses. Concordant evidence can support progression to focused ligand design, whereas boundary variation or cross-layer conflict identifies the additional analysis required before substantial resources are committed. Retaining these alternatives is therefore not a weakness, but the mechanism by which structural uncertainty is converted into an explicit decision pathway.

4.6 Limitations and prospective validation

Several limitations define the scope of the present conclusions. The 21 systems constitute a curated, family-structured case series rather than an external benchmark. Family-clustered resampling and leave-one-family-out analysis reduced the influence of repeated protein families, but cannot establish population-level prediction accuracy or generalizability across the broader structural space of oligomeric proteins. Comparisons among pocket-prediction methods are also sensitive to the ranking protocol and the operational definition of a correct prediction. The present results should therefore be interpreted as a mechanistic and methodological evaluation of candidate selection rather than as a universal performance estimate.

The reconstruction thresholds and module weights were intentionally not optimized against reference-pocket recovery, reducing the risk of retrospective information leakage. Their stability under moderate weight perturbation supports the local robustness of the principal within-target rankings, but does not demonstrate global optimality. External validation should apply the locked reconstruction rules, normalization procedures, missing-data treatment, and weights without target-specific adjustment.

The current implementation of O_{rel} provides a structural approximation of assembly relevance rather than a complete measure of interface function. It incorporates candidate-interface overlap, chain contribution, center-to-interface distance, and assembly context, but does not explicitly include buried surface area, interaction

energetics, residue hotspots, evolutionary conservation, or mutational evidence. These descriptors may improve discrimination between geometrically interface-associated depressions and functionally actionable pockets, although their contribution must be evaluated without systematically favoring large or highly connected interfaces.

Post-static evidence was incomplete and heterogeneous across the case series. MD data were available for 15 of the 21 systems, with different combinations of native-ligand, redocked-ligand, and apo trajectories. The sampled trajectories can identify persistent local contacts or marked instability, but cannot exhaustively characterize slow conformational transitions, cryptic-pocket opening, or ligand-binding kinetics. Conventional MD is similarly limited by accessible timescales, whereas focused or enhanced sampling generally requires prior information about the region of interest.[42] Redocking was performed within crystallographically defined regions and should therefore be interpreted as evidence of local pose recoverability rather than as an independent blind test of pocket selection.

The proposed extension toward peptide, peptidomimetic, and antibody-based modalities remains hypothesis-generating. The current case series was constructed around cavity and ligand-binding-site interpretation and was not designed to benchmark peptide-binding surfaces, antibody epitopes, or multispecific engagement geometry. Dedicated modality-specific descriptors and prospective validation would therefore be required before OIPS could be applied as a design framework for these therapeutic formats.

Automated QC and evidence-state assignments improved reproducibility but were not independently adjudicated by multiple experts or validated experimentally. Reference definitions may remain sensitive in systems containing multiple ligands, neighboring subpockets, or ambiguous functional annotations. Such ambiguity does not alter the frozen candidate ranking, but it limits the certainty with which post hoc evidence labels can be translated into biological pocket assignments. Structured expert review and inter-rater assessment would help distinguish uncertainty intrinsic to the structure from uncertainty introduced by annotation rules. Prospective validation should apply the locked framework to an independent oligomeric collection containing ligand-bound, allosteric, and experimentally characterized interface sites. Boundary-sensitive and conflicting cases should be examined using candidate-specific docking grids, replicate or enhanced-sampling simulations, standardized interface descriptors, and predefined expert-review criteria. A blinded prospective study would provide the most stringent assessment by requiring OIPS to prioritize a previously uncharacterized candidate before its functional relevance is known, followed by ligand screening, mutagenesis, biophysical

measurements, or structural determination. Until such evidence is available, OIPS should be regarded as a traceable decision-support framework rather than a definitive predictor of biological pocket function.

5. Conclusions

Selecting a binding pocket in an oligomeric protein is not simply a matter of identifying the largest or highest-scoring cavity. It requires distinguishing among competing structural hypotheses that may represent ligand-centered, allosteric, interface-associated, or assembly-dependent regions. This study developed OIPS as a reference-blind and traceable framework for reconstructing, prioritizing, and interpreting such candidates across heterogeneous pocket-prediction outputs.

The analysis yielded three principal conclusions. Candidate reconstruction materially determines the objects being ranked: the cluster-v2 procedure converted heterogeneous tool outputs into distinct, auditable candidates while preserving meaningful boundary uncertainty. Oligomeric relevance introduced a complementary ranking objective that retained structurally plausible assembly-associated regions, although these candidates did not always coincide with the crystallographic reference site. Redocking and molecular dynamics further showed that local pose recoverability, dynamic contact persistence, pocket boundary, and static candidate priority are related but non-equivalent forms of evidence. Maintaining a frozen static ranking while recording post-static agreement or conflict preserved these distinctions rather than obscuring them through retrospective rescoreing.

The practical contribution of OIPS is therefore not the assignment of a universally correct pocket, but the transformation of pocket prediction into a ranked set of testable structural hypotheses. By separating candidate priority, working pocket type, structural QC, and evidence state, the framework can support more defensible decisions about biological-assembly preparation, docking-grid definition, simulation design, and experimental validation. It also provides an explicit audit trail for cases in which ligand-centered and assembly-centered interpretations compete.

The results suggest that uncertainty in pocket selection should be retained and classified rather than eliminated prematurely. By distinguishing compact cores, broader interaction envelopes, and assembly-dependent structural contexts, the framework may also provide a basis for modality-aware progression from small molecules to peptidomimetics, cyclic peptides, and antibody-based formats. External evaluation with locked parameters and blinded prospective testing will be required to determine generalizability and experimental utility. Within its present scope, OIPS provides a structured decision-support framework for converting

heterogeneous pocket predictions into actionable and experimentally discriminable hypotheses for oligomeric protein drug discovery.

Abbreviations

OIPS: oligomerization-associated Pocket Interpretation Score;

OIPS-P: OIPS prioritization score;

OIPS-T: OIPS pocket-type interpretation layer;

OIPS-C: OIPS evidence-confidence interpretation layer;

MD: molecular dynamics;

MRR: mean reciprocal rank;

RMSD: root-mean-square deviation;

RMSF: root-mean-square fluctuation;

QC: quality control;

PPI: protein-protein interaction;

DCC: distance to crystallographic ligand center.

Data and Code Availability

The processed data, source code, fixed configuration files, automated tests, and documentation required to reproduce the reported OIPS analyses are publicly available at <https://github.com/xiaochen3316/oips-reproducibility> (commit: 2516d10ff8d7bd72a4afcf7e68174f859b8c7a36). The repository includes input manifests and checksums, redistribution-permitted prepared structures, cluster-v2 reconstruction and post hoc analysis scripts, record-level mappings, frozen reference tables, figure source data, numeric crosschecks, and the unresolved-case register. Redistribution-restricted third-party software outputs and large raw molecular dynamics trajectories are not included; their provenance and availability conditions are documented in the repository.

Author Contributions

Authors

Xiao Chen - School of Pharmacy, Nanjing University of Chinese Medicine, Nanjing 210023, China; Email: 202410632@njucm.edu.cn.

Yifan Zhu - School of Pharmacy, Nanjing University of Chinese Medicine, Nanjing 210023, China; Email: 042123123@njucm.edu.cn.

Shilei Zhao - School of Pharmacy, Nanjing University of Chinese Medicine, Nanjing 210023, China; Email: 042024127@njucm.edu.cn.

Xin Zhang - School of Pharmacy, Nanjing University of Chinese Medicine, Nanjing 210023, China; Email: 042024109@njucm.edu.cn.

Corresponding Authors

Xin Xue - National and Local Collaborative Engineering Center of Chinese Medicinal Resources Industrialization and Formulae Innovative Medicine, Jiangsu Collaborative Innovation Center of Chinese Medicinal Resources Industrialization, Jiangsu Key Laboratory for High Technology Research of TCM Formulae, Nanjing University of Chinese Medicine, Nanjing 210023, China; ORCID: 0000-0001-6143-2921; Phone: +86-25-85811916; Email: xuexin@njucm.edu.cn.

Yao Chen - Nanjing University of Chinese Medicine, Nanjing 210023, China; ORCID: 0000-0002-6009-3208; Phone: +86-13921442470; Email: 300630@njucm.edu.cn.

Haopeng Sun - Department of Medicinal Chemistry, China Pharmaceutical University, Nanjing 210009, China; Email: sunhaopeng@cpu.edu.cn.

Author Contributions

Xiao Chen: Writing - original draft, Visualization. Yifan Zhu: Software, Formal analysis, Computational investigation. Shilei Zhao: Validation, Writing - review and editing, Literature review and reference curation. Xin Zhang: Validation, Writing - review and editing, Literature review and reference curation. Haopeng Sun: Conceptualization, Methodology, Investigation, Resources, Supervision. Yao Chen: Conceptualization, Methodology, Investigation, Resources, Supervision. Xin Xue: Conceptualization, Methodology, Investigation, Resources, Supervision, Funding acquisition, Writing - review and editing. All authors reviewed and approved the final version of the manuscript.

Competing Interests

The authors declare no competing interests.

Acknowledgments

We are grateful for financial support from the National Natural Science Foundation of China (Grant Nos. 82473781 to H.P.S. and 81502986 to X.X.), the Natural Science Foundation of Jiangsu Province (Grant No. BK20221522), and the Jiangsu “333 High-Level Talents Cultivation” Leading Talents Program (Grant Nos. 2022-3-16-203 to H.P.S. and 2024-3-0230 to Y.C.).

References

- (1) Balytskyi Y, Hubenko I, Balytska A, Kelly CV: RAPID-Net: Accurate Pocket Identification for Binding-Site-Agnostic Docking. *J Chem Inf Model* 2025, 65(22):12221-12239.
- (2) Xu Y, Wang S, Hu Q, Gao S, Ma X, Zhang W, Shen Y, Chen F, Lai L, Pei J: CavityPlus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction. *Nucleic Acids Res* 2018, 46(W1):W374-w379.
- (3) Graef J, Ehrt C, Rarey M: Binding Site Detection Remastered: Enabling Fast, Robust, and Reliable Binding Site Detection and Descriptor Calculation with DoGSite3. *J Chem Inf Model* 2023, 63(10):3128-3137.
- (4) Moine-Franel A, Mareuil F, Nilges M, Ciambur CB, Sperandio O: A comprehensive dataset of protein-protein interactions and ligand binding pockets for advancing drug discovery. *Sci Data* 2024, 11(1):402.
- (5) Cao P, Zhang J, Shen Y, Xing Z, Han F, Chen X, Jiang N: MechMemDyn: Coupling Frustration Analysis with Membrane Dynamics to Target the TREM2-DAP12 Complex Interface. *J Chem Inf Model* 2026, 66(9):5382-5392.
- (6) Jiang M, Zhang Z, Ma T, Hang H, Fan Y, Pang S, Zhou W, Zhang Y: ProtGeoNet-Pocket: A Binding Site Prediction Approach Integrating Sequence, Geometry, and Graph Structure. *J Chem Inf Model* 2025, 65(19):10736-10753.
- (7) Evteev SA, Ereshchenko AV, Ivanenkov YA: SiteRadar: Utilizing Graph Machine Learning for Precise Mapping of Protein-Ligand-Binding Sites. *J Chem Inf Model* 2023, 63(4):1124-1132.
- (8) Yáñez O, Alegría-Arcos M, Suardiaz R, Morales-Quintana L, Castro RI, Palma-Olate J, Galarza C, Catagua-González Á, Rojas-Pérez V, Urra G *et al*: Calcium-Alginate-Chitosan Nanoparticle as a Potential Solution for Pesticide Removal, a Computational Approach. *Polymers (Basel)* 2023, 15(14).
- (9) Peter S, Siragusa L, Thomas M, Palomba T, Cross S, O'Boyle NM, Bajusz D, Ferenczy GG, Keserü GM, Bottegoni G *et al*: Comparative Study of Allosteric GPCR Binding Sites and Their Ligandability Potential. *J Chem Inf Model* 2024, 64(21):8176-8192.
- (10) Utgés JS, MacGowan SA, Barton GJ: LIGYSIS-web: a resource for the analysis of protein-ligand binding sites. *Nucleic Acids Res* 2025, 53(W1):W351-w360.
- (11) Shen A, Yuan M, Ma Y, Du J, Wang M: PGBind: pocket-guided explicit attention learning for protein-ligand docking. *Brief Bioinform* 2024, 25(5).
- (12) Prajapati C, Saini LS, Gaur A, Kumar P, Uddin J, Ghazi RM: GC-MS characterization and in-silico evaluation of TNF- α binding efficacy of leaf extract from *Abies spectabilis* (D. Don) Spach. *Sci Rep* 2026, 16(1):3563.
- (13) Khibech O, Bouammali H, Hammouti Y, Bouhrim M, Merzouki M, Abadi S, Al-Zharani M, Nasr FA, Qurtam AA, Bouammali B *et al*: Elucidating the isorhamnetin-3-O-glucoside-iNOS interaction via molecular dynamics and Hirshfeld surface analyses. *PLoS One* 2025, 20(12):e0339357.
- (14) Halgren TA: Identifying and characterizing binding sites and assessing druggability. *J Chem Inf Model* 2009, 49(2):377-389.
- (15) Cankara F, Senyuz S, Sayin AZ, Gursay A, Keskin O: DiPPI: A Curated Data Set for Drug-like Molecules in Protein-Protein Interfaces. *J Chem Inf Model* 2024, 64(13):5041-5051.
- (16) Martin-Garcia JM, Botha S, Hu H, Jernigan R, Castellví A, Lisova S, Gil F, Calisto B, Crespo I, Roy-Chowdhury S *et al*: Serial macromolecular crystallography at ALBA Synchrotron Light Source. Erratum. *J Synchrotron Radiat* 2022, 29(Pt 4):1130.

- (17) Patanè GT, Lombardo L, Putaggio S, Tellone E, Ficarra S, Barreca D, Laganà G, De Luca L, Calderaro A: Anti-Aggregative and Protective Effects of Vicenin-2 on Heat and Oxidative Stress-Induced Damage on Protein Structures. *Int J Mol Sci* 2023, 24(24).
- (18) Plushchetskaya P, Künze G: Evaluation of Small-Molecule Binding Site Prediction Methods on Membrane-Embedded Protein Interfaces. *J Chem Inf Model* 2025, 65(13):6949-6967.
- (19) Galli U, Moro M, Balestrero FC, Colombo G, Kotten M, Roncaglio B, Genazzani AA, Aprile S, Massarotti A, Orsomando G *et al*: Development of PROTACs Targeting the Moonlighting Enzyme Nicotinamide Phosphoribosyltransferase (NAMPT) for Breast Cancer Therapy. *J Med Chem* 2026, 69(5):5259-5290.
- (20) Ibrahim M, Zacharias M: Protein-Protein Interaction Stabilizers from MD Simulation-Derived Pharmacophores. *J Chem Inf Model* 2026, 66(10):6131-6140.
- (21) Koseki J, Motono C, Yanagisawa K, Kudo G, Yoshino R, Hirokawa T, Imai K: CryptToth: Cryptic Pocket Detection through Mixed-Solvent Molecular Dynamics Simulations-Based Topological Data Analysis. *J Chem Inf Model* 2025, 65(11):5567-5575.
- (22) Zhuo C, Zeng C, Liu H, Wang H, Peng Y, Zhao Y: Advances and Mechanisms of RNA-Ligand Interaction Predictions. *Life (Basel)* 2025, 15(1).
- (23) González-Moles M, Warnakulasuriya S, Ramos-García P: Prognosis Parameters of Oral Carcinomas Developed in Proliferative Verrucous Leukoplakia: A Systematic Review and Meta-Analysis. *Cancers (Basel)* 2021, 13(19).
- (24) El-Sayed Ebead E, Aboelnaga A, Nassar E, Naguib MM, Ismail MF: Ultrasonic-induced synthesis of novel diverse arylidenes via Knoevenagel condensation reaction. Antitumor, QSAR, docking and DFT assessment. *RSC advances* 2023, 13(42):29749-29767.
- (25) Yang X, Zhang W, Luo C, Deng Z, Gu D, Yu DJ, Hu S, Zhong Y: SGLEPocket: A Spatial Gating and Local Feature Enhancement Network for Protein-Ligand Binding Pocket Prediction. *J Chem Inf Model* 2026, 66(7):3621-3642.
- (26) Xu Q, Dunbrack RL, Jr.: Principles and characteristics of biological assemblies in experimentally determined protein structures. *Curr Opin Struct Biol* 2019, 55:34-49.
- (27) Wu Q, Ellis H, Siravegna G, Michel AG, Norden BL, Fece de la Cruz F, Balasooriya ER, Zhen Y, Silveira VS, Che J *et al*: Landscape of Clinical Resistance Mechanisms to FGFR Inhibitors in FGFR2-Altered Cholangiocarcinoma. *Clin Cancer Res* 2024, 30(1):198-208.
- (28) Volkamer A, Kuhn D, Rippmann F, Rarey M: DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics* 2012, 28(15):2074-2075.
- (29) Sun YA, Kalpakavadi S, Prior S, Thrift AG, Waddingham S, Phan H, Gall SL: Socioeconomic status and health-related quality of life after stroke: a systematic review and meta-analysis. *Health Qual Life Outcomes* 2023, 21(1):115.
- (30) Tian W, Chen C, Lei X, Zhao J, Liang J: CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res* 2018, 46(W1):W363-w367.
- (31) Park J, Lim H, Lee Y, Kim T, Song H, Lee S, Heo Y, Kim J, Namkung W, Han G: Discovery of PAR2 antagonists through fragment molecular orbital-based hotspot analysis and allosteric sodium site dynamics simulations. *Bioorg Chem* 2025, 167:109285.
- (32) Gainza P, Sverrisson F, Monti F, Rodolà E, Boscaini D, Bronstein MM, Correia BE: Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat Methods* 2020, 17(2):184-192.

- (33) Haas J, Barbato A, Behringer D, Studer G, Roth S, Bertoni M, Mostaguir K, Gumienny R, Schwede T: Continuous Automated Model EvaluatiOn (CAMEO) complementing the critical assessment of structure prediction in CASP12. *Proteins* 2018, 86 Suppl 1(Suppl 1):387-398.
- (34) Bosch-Fernández AM, Menacho W, Pérez R, H VG: Glycans Modulate the Adsorption of RBD Glycoproteins on Polarizable Surfaces. *J Chem Inf Model* 2026, 66(5):2719-2734.
- (35) Wadsworth PA, Singh AK, Nguyen N, Dvorak NM, Tapia CM, Russell WK, Stephan C, Laezza F: JAK2 regulates Nav1.6 channel function via FGF14(Y158) phosphorylation. *Biochim Biophys Acta Mol Cell Res* 2020, 1867(10):118786.
- (36) Mustali J, Yasuda I, Hirano Y, Yasuoka K, Gautieri A, Arai N: Unsupervised deep learning for molecular dynamics simulations: a novel analysis of protein-ligand interactions in SARS-CoV-2 M(pro). *RSC Adv* 2023, 13(48):34249-34261.
- (37) Aranda-García D, Stepniewski TM, Torrens-Fontanals M, García-Recio A, Lopez-Balastegui M, Medel-Lacruz B, Morales-Pastor A, Peralta-García A, Dieguez-Eceolaza M, Sotillo-Nuñez D *et al*: Large scale investigation of GPCR molecular dynamics data uncovers allosteric sites and lateral gateways. *Nat Commun* 2025, 16(1):2020.
- (38) Menchon G, Bombarde O, Trivedi M, Négrel A, Inard C, Giudetti B, Baltas M, Milon A, Modesti M, Czaplicki G *et al*: Structure-Based Virtual Ligand Screening on the XRCC4/DNA Ligase IV Interface. *Sci Rep* 2016, 6:22878.
- (39) Capra JA, Laskowski RA, Thornton JM, Singh M, Funkhouser TA: Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure. *PLoS Comput Biol* 2009, 5(12):e1000585.
- (40) Gagliardi L, Rocchia W: SiteFerret: Beyond Simple Pocket Identification in Proteins. *J. Chem. Theory Comput* 2023, 19(15):5242-5259.
- (41) Marchand JR, Pirard B, Ertl P, Sirockin F: CAVIAR: a method for automatic cavity detection, description and decomposition into subcavities. *Journal of computer-aided molecular design* 2021, 35(6):737-750.
- (42) Ge Y, Pande V, Seierstad MJ, Damm-Ganamet KL: Exploring the Application of SiteMap and Site Finder for Focused Cryptic Pocket Identification. *The journal of physical chemistry B* 2024, 128(26):6233-6245.
- (43) Zhang S, Hu Z, Tanji H, Jiang S, Das N, Li J, Sakaniwa K, Jin J, Bian Y, Ohto U *et al*: Small-molecule inhibition of TLR8 through stabilization of its resting state. *Nature chemical biology* 2018, 14(1):58-64.
- (44) Zhang Z, Li Y, Yang J, Li J, Lin X, Liu T, Yang S, Lin J, Xue S, Yu J *et al*: Dual-site molecular glues for enhancing protein-protein interactions of the CDK12-DDB1 complex. *Nat. Commun* 2024, 15(1):6477.
- (45) Hosseinzadeh P, Watson PR, Craven TW, Li X, Rettie S, Pardo-Avila F, Bera AK, Mulligan VK, Lu P, Ford AS *et al*: Anchor extension: a structure-guided approach to design cyclic peptides targeting enzyme active sites. *Nat. Commun* 2021, 12(1):3384.
- (46) Chen S, Lin T, Basu R, Ritchey J, Wang S, Luo Y, Li X, Pei D, Kara LB, Cheng X: Design of target specific peptide inhibitors using generative deep learning and molecular dynamics simulations. *Nat. Commun* 2024, 15(1):1611.
- (47) Steinhardt JJ, Guenaga J, Turner HL, McKee K, Louder MK, O'Dell S, Chiang CI, Lei L, Galkin A, Andrianov AK *et al*: Rational design of a trispecific antibody targeting the HIV-1 Env with elevated anti-viral activity. *Nat. Commun* 2018, 9(1):877.