ՀԱՅԱՍՏԱՆԻ ՀԱՆՐԱՊԵՏՈՒԹՅԱՆ ԳԻՏՈՒԹՅՈՒՆՆԵՐԻ ԱԶԳԱՅԻՆ ԱԿԱԴԵՄԻԱ
ՄՈԼԵԿՈՒԼԱՅԻՆ ԿԵՆՍԱԲԱՆՈՒԹՅԱՆ ԻՆՍՏԻՏՈՒՏ

ԽԱՉԱՏՐՅԱՆ ՀԱՄԼԵՏ ՆԱՐԵԿԻ

ՎԻՐՈՒՍԱՅԻՆ ՍՊԻՏԱԿՈՒՑՆԵՐԻ ԱԿՏԻՎ ԿԵՆՏՐՈՆՆԵՐԻ ԿՈՆՖՈՐՄԱՑԻՈՆ ՎԵՐԼՈՒԾՈՒԹՅԱՆ ՀԱՐԹԱԿ ՀԱԿԱՎԻՐՈՒՍԱՅԻՆ ՄԻԱՑՈՒԹՅՈՒՆՆԵՐԻ ԿԱՌՈՒՑՎԱԾՔԱՀԵՆ ՀԱՅՏՆԱԲԵՐՄԱՆ ՀԱՄԱՐ

Գ.00.03 – «Մոլեկուլային և բջջային կենսաբանություն» մասնագիտությամբ կենսաբանական գիտությունների թեկնածուի գիտական աստիճանի հայցման ատենախոսության

ՍԵՂՄԱԳԻՐ

ԵՐԵՎԱՆ – 2026

NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF ARMENIA
INSTITUTE OF MOLECULAR BIOLOGY

HAMLET NAREK KHACHATRYAN

ACTIVE SITE CONFORMATIONAL ANALYSIS FRAMEWORK OF VIRAL PROTEINS FOR STRUCTURE-BASED ANTIVIRAL DRUG DISCOVERY

SYNOPSIS

of Dissertation Submitted for the Degree
of Candidate of Biological Sciences (Ph.D.) in the Field of
03.00.03. – “Molecular and Cellular Biology”

YEREVAN – 2026

Ատենախոսության թեման հաստատվել է Մոլեկուլային կենսաբանության ինստիտուտի գիտական խորհրդում:

Գիտական ղեկավար՝	կ.գ.թ. Հովակիմ Սարգսի Զաքարյան
Պաշտոնական ընդդիմախոսներ՝	կ.գ.դ. Կարեն Բաբկենի Նազարյան կ.գ.թ. Բինգ Սիոնգ
Առաջատար կազմակերպություն՝	Ռուս-Հայկական համալսարան

Ատենախոսության պաշտպանությունը տեղի կունենա 2026 թ. սեպտեմբերի 30-ին, ժամը 14:00-ին, Մոլեկուլային կենսաբանության ինստիտուտում, 042 մասնագիտական խորհրդի նիստում (ՀՀ, 0014, ք. Երևան, Է. Հասրաթյան 7):

Ատենախոսությանը կարելի է ծանոթանալ Մոլեկուլային կենսաբանության ինստիտուտի գրադարանում և https://imb.am/ կայքում:

Ատենախոսության սեղմագիրն առաքվել է 2026 թ. օգոստոսի 28-ին:

042 մասնագիտական խորհրդի գիտական քարտուղար,
կենս. գիտ. թեկնածու Ռ.Վ. Զախարյան

Dissertation topic approved at the Scientific Council of the Institute of Molecular Biology.

Scientific supervisor:	Ph.D. Hovakim Sargis Zakaryan
Official opponents:	D.Sc. Karen Babken Nazaryan
	Ph.D. Bing Xiong
Leading organization:	Russian-Armenian University

The dissertation defense will be held on September 30, 2026, at 14:00, at a session of the Specialized Council 042 operating at the Institute of Molecular Biology (E. Hasratyan 7, 0014, Yerevan, RA).

The dissertation is available in the library of the Institute of Molecular Biology and on the official website https://imb.am/.

The synopsis was sent out on August 28, 2026.

Scientific secretary of the specialized council 042
Ph.D. R.V. Zakharyan

INTRODUCTION

Background and Context. The main constraint on developing new antiviral medications is time: the process takes months, sometimes years, and during that interval a new or antigenically drifted strain spreads and evades existing means. Small-molecule antivirals are the first line for containing severe infection and limiting morbidity until immunization becomes available, yet new antigenically distinct and drug-resistant variants keep emerging, and the interval between a strain's appearance and the availability of an active inhibitor is itself part of the clinical burden [Tao et al. 2021]. Shortening the discovery workflow is therefore a primary therapeutic goal, and physics- and machine-learning-based computational methods have become the principal way of addressing it.

Structure-based drug discovery (SBDD) combines techniques such as virtual screening and *de novo* design that rely on a protein's known three-dimensional structure to shorten discovery by searching for or designing new, potentially active compounds *in silico*. Deep-learning-accelerated screening has analyzed libraries of more than a billion compounds against a single viral protein in the time a wet-lab campaign would need for a few hundred [Ton et al. 2020]. The machine learning-driven *de novo* design goes further: rather than ranking members of a fixed library, an agent constructs novel chemical entities and optimizes them against several physicochemical objectives simultaneously [Popova et al. 2018]. Both families rest on a single, understudied premise that they act upon one three-dimensional model of the protein.

That premise matters, because a viral drug target has no single, rigid geometry. For a target of interest, the Protein Data Bank holds tens to hundreds of crystallographic holo- and apo-structures that differ in flap geometry, in how open each subsite is, and in the electrostatic character of the binding cavity [Kneller et al. 2020]. A compound can score well against one conformation and poorly against another of the very same enzyme, and a single conformation often fails to reproduce even the experimentally determined binding pose [Amaro et al. 2018]. The choice of which crystal structure to screen or design against therefore largely determines the outcome [McGovern & Shoichet 2003], yet in everyday practice that choice is made on crystallographic quality parameters like resolution, R-factors, or the number of Ramachandran outliers, which, as this work shows, do not predict a structure's competence for screening or design. Whether a conformation enriches active compounds depends on concrete features of its active site, and this defines the competent-conformation selection problem for structure-based drug design.

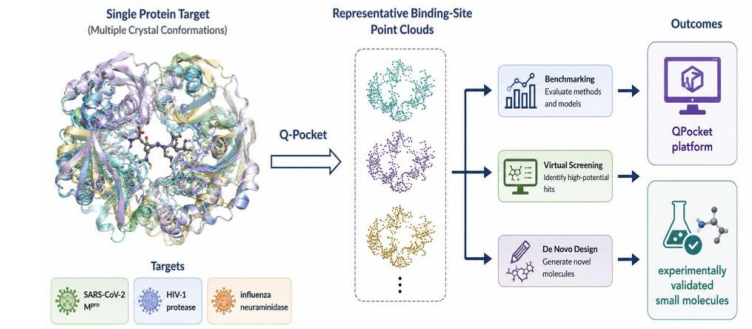

Figure 1. Graphical abstract.
Q-Pocket summarizes the conformational diversity of a protein target by representing multiple crystal structures into a compact set of representative binding-site point clouds. These pocket representations support benchmarking, virtual screening, and *de novo* molecular design across therapeutically relevant targets, facilitating the discovery of small-molecule drug candidates.

This selection problem is the central subject of the dissertation. We developed Q-Pocket, an alignment-free framework that represents each binding site as a physicochemically annotated point cloud and reduces a target's whole crystallographic ensemble to a handful of representative conformational states (Figure 1). We built the framework, and demonstrated its central claim, on the SARS-CoV-2 main protease (Mpro), the HIV-1 protease (HIV-1 PR) active site, a pocket that forms between the enzyme's two chains and offers a more challenging target with more structures deposited in the PDB, and finally applied it to guide the discovery of active inhibitors against Mpro and against influenza neuraminidase (NA), which we confirmed both *in vitro* and *in vivo*.

Research Aim and Objectives. The aim of this work is to establish, and to experimentally validate, a systematic and physicochemically grounded framework for selecting representative protein conformations for structure-based antiviral drug design and discovery. This aim was organised into four objectives.

Objective 1. Establish a benchmarking framework for the selection of representative protein structures for *in silico* antiviral drug discovery: a binding-site representation encoding chemistry and geometry independently of backbone superposition, a distance metric and clustering procedure partitioning an ensemble into representative states, and a reproducible pipeline.

Objective 2. Systematically evaluate all available crystallographic holo-structures of a defined target and identify the binding-site conformational determinants that govern virtual-screening performance, including whether crystallographic quality predicts that performance.

Objective 3. Apply the selected representative conformations within advanced virtual screening (AVS) and reinforcement-learning (RL) *de novo* design campaigns and confirm the computational predictions through experimental measurement.

Objective 4. Showcase the framework for targets of moderate sequence homology through a pan-influenza neuraminidase inhibitor discovery campaign spanning influenza A and B.

Scientific Novelty. A validated framework has been developed to represent target's crystallographic holo-ensemble as a set of physicochemically annotated binding-site point clouds with further reduction to a small number of representative conformational states. The procedure is alignment-free and uses side-chain chemistry and cavity electrostatics rather than backbone geometry.

Crystallographic quality and virtual-screening competence are shown to be mutually independent properties of a structure: across 342 Mpro holo-structures certain conformations underperform on enrichment, ROC-AUC, and pose prediction, and they do so reproducibly across three independent docking engines, so the deficiency is a property of the structure itself and not of the method. Because the underperforming structures have well-defined structural features, this gives a computable rather than statistical criterion for including or excluding a conformation.

The framework's applicability has been shown on the HIV-1 PR active site, a pocket formed within the homodimer interface and well characterised in the PDB (684 structures retrieved and featurized). Eight populated conformational states were identified, four of which account for 85% of all structures. The framework was tested experimentally against two targets, the final validation being *de novo*-designed neuraminidase inhibitors that show activity against both influenza A and B viruses (IAV, IBV), retained potency against the oseltamivir-resistant H275Y variant. Compounds showed high survival rates of mice in an animal model, meanwhile being dissimilar to oseltamivir (a mean Tanimoto coefficient of 0.13).

Scientific and Practical Significance. The main contribution of this work is the developed methodology and platform to turn conformation selection from an informal checkbox guided by

crystallographic quality into a defined, benchmarkable stage of the structure-based discovery workflow. A target's holo-ensemble is finite and collapses to a small number of representative states, so the correct input to virtual screening and *de novo* design is that representative ensemble rather than a single structure chosen based on statistical quality metrics. Selecting on demonstrated structural competence rather than on resolution should lower the false-negative rate of screening campaigns aimed at viral proteins and comparable targets. The practical value is demonstrated through *in silico*, *in vitro*, and *in vivo* validation: representative-conformation selection helped identify micromolar Mpro inhibitors confirmed in an enzymatic assay and defined the representative structures for a *de novo* neuraminidase campaign whose leads inhibited the neuraminidases of IAV and IBV, including the oseltamivir-resistant H275Y mutant.

Approbation of the Work. The main results of the research were presented at "COST Actions BEST-CSP workshop" 9-11 September 2024, Warsaw, Poland and "European Bioinformatics Institute" 27-28 October 2022, Cambridge, UK, and at the meetings of the Scientific Council of the Institute of Molecular Biology NAS RA.

Volume and Structure. The dissertation comprises 103 pages of computer-formatted English text, including 7 tables, 35 figures, 10 equations and 174 references, and consists of the following sections: Introduction, Literature Review, Materials and Methods, Results and Discussion, Conclusion, Inferences, and References.

Publications. The main results of the dissertation are published in four peer-reviewed papers.

MATERIALS AND METHODS

Structural datasets and curation. We assembled crystallographic datasets for each studied target directly from the Protein Data Bank [Berman et al. 2000] using the Q-Pocket retrieval and curation utility. For SARS-CoV-2 Mpro we collected holo-structures at $\leq$ 2.5 Å resolution and filtered out single-atom and oligopeptide ligands, ligands bound outside the catalytic site, and structures with missing binding-site residues. Water and biologically not relevant heteroatoms were stripped and the A chain kept. For the retrieval of HIV-1 PR dataset, 1HSG was used as the reference structure. The resolution threshold and a sequence-identity threshold were set $\leq$ 2.50 Å and $\geq$ 90% correspondingly. Further, all structures with missing homodimer chain were discarded. Retrieved structures were renumbered to match the numbering of the UniProt P04585 entry. For neuraminidase data retrieval, we used the reference structure 6G01 (influenza A H1N1) with a $\leq$ 2.50 Å cut-off and a relaxed homology threshold ($\geq$ 20%) to match the influenza A and B subtypes, then applied biological filters to retain wild-type influenza A HxN1 and influenza B entries.

The Q-Pocket framework. Q-Pocket represents a binding site not by its backbone coordinates but by the chemistry and geometry of the cavity itself. Each site is a point cloud of grid points at 0.70 Å spacing filling the solvent-accessible cavity (CavitOmiX v.1.0, Innophore GmbH), every point carrying six physicochemical descriptors (Figure 2): electrostatic potential, obtained by solving the linearized Poisson–Boltzmann equation with APBS [Baker et al. 2001; Jurrus et al. 2018] using PARSE partial charges [Sitkoff et al. 1994], a Coulomb potential, a lipophilic potential, hydrogen-bond donor and acceptor densities, and an aromatic bonding density. Because these fields describe the chemistry of the cavity, two structures with an identical global fold but chemically distinct cavities receive distinct descriptors, independently of backbone RMSD. The Q-Pocket is implemented as an open-source Python package built on SciPy, scikit-learn [Pedregosa et al. 2011] and RDKit [Landrum 2023].

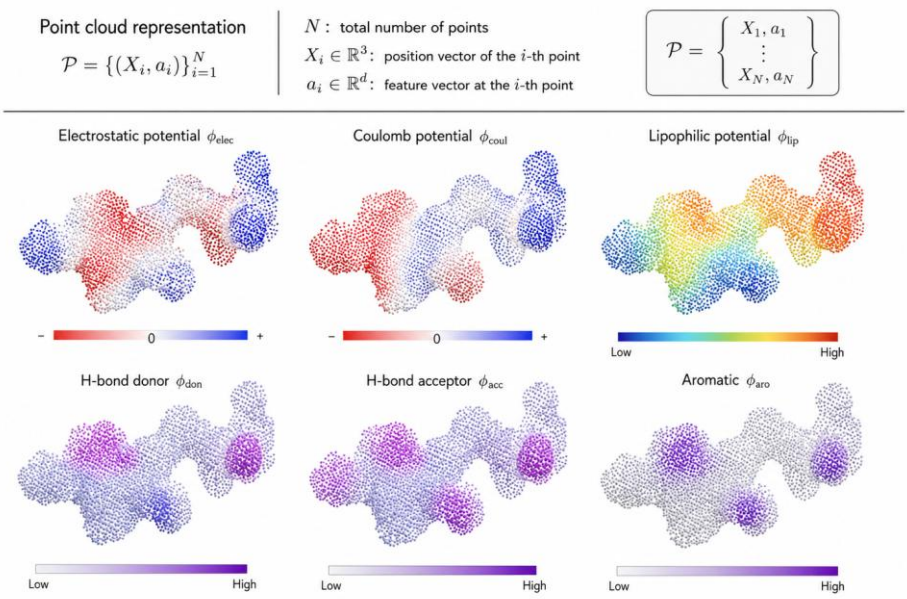

Figure 2. Point cloud representation and featurization of binding pocket within Q-Pocket framework.
The cavity volume is defined as a set of N points, where each point has a three-dimensional coordinate X_i and an associated feature vector a_i. The mapped physicochemical features include electrostatic, Coulomb, lipophilic, and interaction potentials, with color scales indicating their relative magnitudes across the volume.

To respect the geometry of the cavity surface, we used geodesic graph representations via connecting the points of each cloud into an ε-ball graph and computed all point-to-point geodesic distances with Dijkstra's algorithm [Dijkstra 1959]. The geodesic distances were converted to affinities through a Gaussian heat kernel and reduced to the symmetric normalized graph Laplacian [von Luxburg 2007; Belkin & Niyogi 2003], its low-frequency eigenvalues form the pocket's shape signature, while the graph-Fourier power spectrum [Shuman et al. 2013] of the six physicochemical fields forms its chemistry signature. For every pair of pockets these fingerprints were combined into a single spectral distance summing a geometric and a chemical term, with fixed weights applied unchanged across every target so that no result depends on target-specific tuning (Table 1). We clustered the pairwise distance matrix by affinity propagation [Frey & Dueck 2007], which returns exemplar structures directly and reports singletons as noise points. Alternative clustering and representative selection algorithms were also implemented [Rousseeuw 1987; Kaufman & Rousseeuw 1990].

Interaction and chemical-space profiling. We analyzed protein–ligand interactions across all evidenced small-molecule binders with PLIP [Salentin et al. 2015; Adasme et al. 2021] and reported them by interaction type and occurring frequency. To analyze the chemical space of the unique inhibitors we calculated Morgan fingerprints (radius 2, 2048 bits) for each and used the pairwise Tanimoto distance matrix (computed with RDKit [Landrum 2023]) for diversity analysis. For unique ligands additional properties like molecular weight, logP by the Wildman–Crippen method [Wildman & Crippen 1999], TPSA, and hydrogen-bond and rotatable-bond counts were computed.

Table 1. Default parameters of the Q-Pocket spectral-distance framework.

Parameter	Symbol	Default
Point-cloud grid spacing	—	0.70 Å
Geodesic-graph neighbor radius	r	2.0 Å
Heat-kernel bandwidth	σ	10.0
Retained Laplacian eigenmodes	K	100
Feature-normalization clip	—	±5 (standard deviations)
Shape weight	α	0.85
Chemical-term scale	β	12.5
Maximum pocket size (larger excluded)	—	5000 points
Affinity-propagation damping	—	0.9
Affinity-propagation iterations	—	300
Medoid clustering: maximum clusters searched	$k_{\max}$	10

Structural clustering and backbone flexibility. For the Mpro benchmark, we superimposed and hierarchically clustered the Cα coordinates of the interacting binding-site residues with the Bio3D R package (1.2 Å cut-off) [Grant et al. 2006; Skjærven et al. 2014], giving a backbone-geometry grouping to set against the chemistry-based Q-Pocket clustering. We quantified backbone flexibility by atomic fluctuation and ensemble normal-mode analysis with an elastic-network model [Kolossváry 2024].

Docking, benchmarking, and advanced virtual screening. We used UCSF Chimera [Pettersen et al. 2004] and the Schrödinger Protein Preparation Wizard under the OPLS force field [Schrödinger LLC 2020; Harder et al. 2016; Jorgensen & Tirado-Rives 1988] to prepare the receptor structures for further studies. To assess docking sensitivity to receptor conformation, we docked a benchmarking library of the active ligands plus property-matched decoys generated with DUD-E [Mysinger et al. 2012] using three independent programs that apply different search and scoring algorithms: the open-source AutoDock Vina [Trott & Olson 2010] and rDock [Ruiz-Carmona et al. 2014], and the commercial Glide [Friesner et al. 2004]. We computed the enrichment factor, ROC-AUC, and BEDROC ($\alpha = 20$) with scikit-learn [Pedregosa et al. 2011] and scored a 2 Å pose-prediction success rate separately. For the prospective Mpro screen, we clustered known inhibitors from ChEMBL and BindingDB [Gaulton et al. 2012; Gilson et al. 2016] on topological fingerprints and used them to build ligand-based pharmacophores in LigandScout [Wolber & Langer 2005]. We screened the selected model against roughly 200 million commercial compounds through the Pharmit server [Sunseri & Koes 2016], docked the matches into eight representative Mpro structures with two engines, the Monte-Carlo-based ICM-Pro and AutoDock Vina [Abagyan et al. 1994; Trott & Olson 2010], and summed the ranks across structures and engines (ensemble and consensus docking).

Reinforcement-learning *de novo* design. We generated novel neuraminidase inhibitors *de novo* with a deep reinforcement-learning platform [Matevosyan et al. 2022] formulated as a Markov decision process that begins from a single benzene ring and assembles novel scaffolds without external training data, scoring each modification to shape the reward. We then filtered the generated population by synthetic accessibility [Ertl & Schuffenhauer 2009], quantitative estimate of drug-likeness [Bickerton et al. 2012], structural alerts, and vendor availability, and ranked it by MM/PBSA binding free energy against the influenza A and B neuraminidases.

Molecular dynamics. We assessed binding-mode stability and end-state free energies by molecular dynamics, using the Uni-GBSA workflow (Amber03/GAFF2) for the Mpro screen [Sahakyan 2021; Yang et al. 2023] and GROMACS 2023.1 (ff14SB/GAFF) for the neuraminidase leads [Abraham et al. 2015; Maier et al. 2015; Wang et al. 2004], under SHAKE constraints [Ryckaert et al. 1977], and computed MM/PBSA energies with gmx_MMPBSA [Valdés-Tresanco et al. 2021].

Experimental methods. We measured neuraminidase inhibition with the NA-Fluor MUNANA assay and Mpro inhibition in a FRET-based protease assay (25 nM enzyme, 25 μM substrate), determined antiviral activity and cytotoxicity in MDCK cells (EC50, CC50), and assessed *in vivo* efficacy in lethally infected BALB/c mice (approval #14062023/1). We analyzed the data in GraphPad Prism and plotted the computational results with Matplotlib [Hunter 2007].

The framework is openly available at https://github.com/hamlet-khachatryan/QPocket under the MIT license.

RESULTS AND DISCUSSION

The Q-Pocket framework and the receptor-selection problem on SARS-CoV-2 Mpro

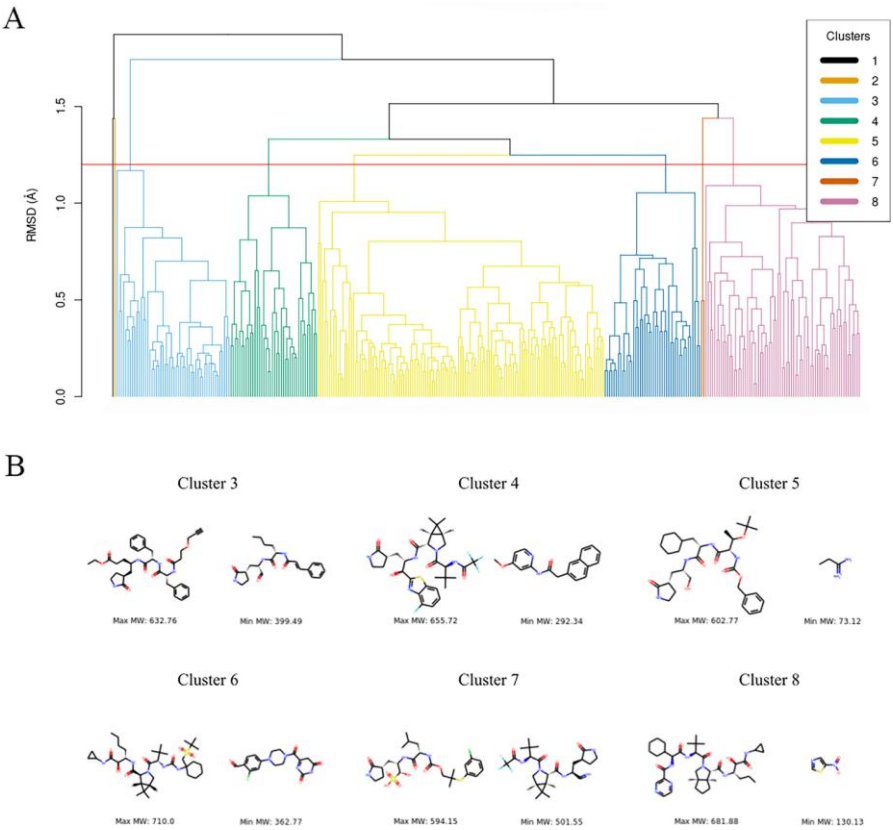

Figure 3. (A) Dendrogram clustering 342 Mpro holo-structures by RMSD of the alpha-carbon coordinates of interacting amino acids (B) Chemical diversity of the ligands within each cluster.

To characterize the conformational landscape of the SARS-CoV-2 main protease, we retrieved all deposited holo-structures at ≤ 2.5 Å resolution (resulting in 454 structures), which filtering on ligand type and binding-site criteria reduced to 342. Hierarchical clustering of the Cα coordinates of the interacting binding-site residues (Bio3D, 1.2 Å cut-off) [Grant et al. 2006; Skjærven et al. 2014] yielded eight clusters (Figure 3), and we selected one representative structure per cluster by proximity to the cluster centroid. 297 unique co-crystallized ligands were identified. Analysis of unique ligands showed that they span a wide range of chemotypes, the largest being narlaprevir (709.44 g/mol). The PLIP interaction analysis revealed that more than the half of unique ligands interacted with His41, Gly143, Cys145, Glu166, and Gln189, the residues forming the catalytic dyad, oxyanion hole, and S1/S3–S4 subsites [Salentin et al. 2015; Adasme et al. 2021]. Of the 297 ligands, 187 (63%) were covalent binders targeting the Cys145 thiol through the reversible thioimidate mechanism used by nirmatrelvir-class inhibitors [Owen et al. 2021]. Elastic-network normal-mode analysis [Kolossváry 2024] found only moderate fluctuations (0.6–1 Å) confined to six regions, with the catalytic His41, Gly143, and Cys145 among the least mobile, so the screening-performance differences below cannot be attributed to backbone flexibility.

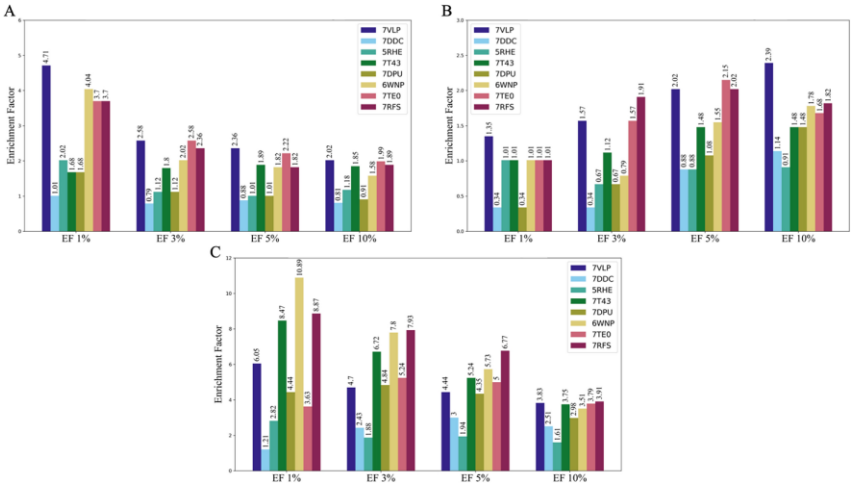

Figure 4. Virtual screening performance evaluation
Virtual screening performance reported as enrichment factors at 1, 3, 5 and 10 %-s are for AutoDock Vina **(A)**, rDock **(B)**, and Glide **(C)** on selected representative Mpro structures.

To determine how the choice of conformation affects virtual-screening performance, we screened a benchmark library of 297 active ligands and 8,910 property-matched decoys (9,207 compounds) against each representative structure with AutoDock Vina [Trott & Olson 2010], rDock [Ruiz-Carmona et al. 2014], and Glide [Friesner et al. 2004]. Performance varied significantly with the structure choice, and the performance pattern held across all three metrics and all three engines (Figure 4): although rDock gave the lowest enrichment overall and Glide the highest, the ranking of structures was preserved between engines. Structure 7DDC consistently returned the lowest EF across every program, and 7DDC and 5RHE the lowest ROC-AUC. The same pattern governed pose prediction: for 7DDC no program recovered a single accurate pose, whereas the highest success rate (29.02%) was obtained on 7RFS with Glide. Because the underperformance of 5RHE, 7DDC, and 7DPU reproduces across three independently developed scoring functions, it is a property of the structures themselves, and it does not track crystallographic quality: the underperforming

structures are not distinguished by poorer resolution or R-factors, so screening competence must be treated as a separate, measurable property.

Reducing each representative site to its Q-Pocket point cloud made the origin of the performance gap apparent (Figure 5).

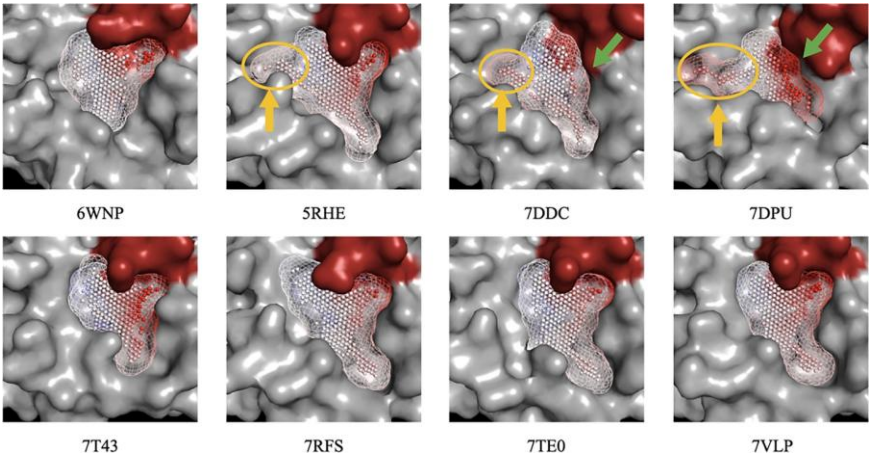

Figure 5. Q-Pocket point cloud representations of selected Mpro structures.
Binding sites are depicted as point clouds withing the binding cavity, each point annotated by its electrostatic potential. The red surface marks Cys145. Green and yellow arrows indicate the absence of the S1 subsite in 7DDC and 7DPU and the emergence of an S2 subpocket in 7DDC, 7DPU, and 5RHE respectively.

Two structural differences set the worst structures apart: in 7DDC and 7DPU the S1 subsite was missing, and in 7DDC, 7DPU, and 5RHE an extra subpocket had opened inside the S2 subsite, and neither feature is present in the better conformations. To quantify the effect of the missing S1 subsite, we analyzed the shortest distances between the active ligands and the Cα of Ser139, which lines the S1 subsite, this linked the collapse of S1 to poor recognition of the actives. The deficiency is therefore structural, not statistical: a conformation that has lost the S1 subsite, or opened an aberrant S2 subpocket, cannot present the interaction surface the true actives require. Three conclusions follow. First, crystallographic quality and screening competence are independent properties, so benchmarking must be treated as a mandatory, separate stage. Second, the verdict is reproducible across engines, so it travels with the structure and can be reused. Third, the discriminating features are chemical and geometric properties of the cavity, which is exactly what the point-cloud representation encodes and a backbone comparison would miss.

Generalization to the HIV-1 protease conformational space

To test whether Q-Pocket generalizes to a harder case, we applied it to the HIV-1 protease, whose binding site forms between the two homodimer chains rather than within a single domain. We retrieved 690 crystal structures at ≥ 90% sequence identity and ≤ 2.50 Å resolution (687 by X-ray, three by neutron diffraction), extracted 452 unique inhibitors, and successfully featurized 684 of them. Sequence-guided superposition confirmed a near-invariant global fold (the mean Cα RMSD to the reference structure 1HSG was 0.41 ± 0.09 Å), so any variation the framework detects is a property of the binding site. Cavity volumes of retrieved structures ranged from 328 to 2,236 Å^3 (mean 792 Å^3; Figure 6).

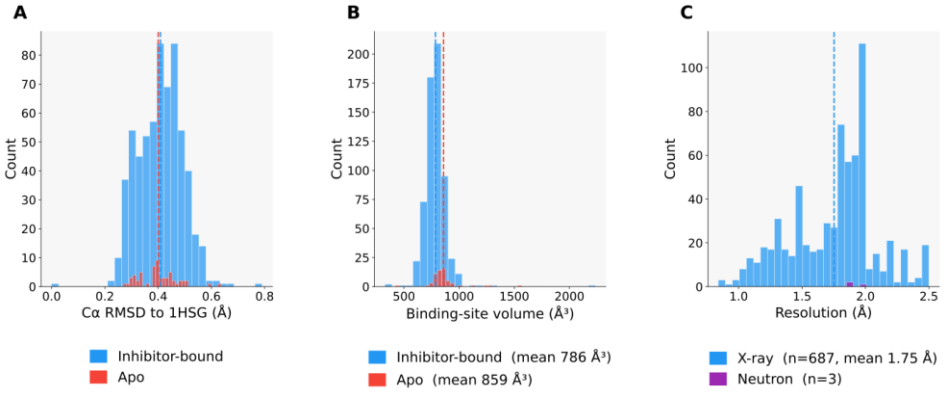

Figure 6. HIV-1 PR dataset overview.
(A) Cα RMSD to 1HSG across the holo and apo states, with dashed lines indicating average values of the group. **(B)** Binding-site volumes for the same categories, showing a higher average value for apo than inhibitor-bound structures. **(C)** Resolution of all 690 structures (687 X-ray and 3 neutron crystallography), with the dashed line denoting mean X-ray resolution (1.75 Å).

Affinity propagation [Frey & Dueck 2007] over the spectral distance matrix returned 16 exemplars which, under the convention that a cluster is a state only where it has two or more members, resolve into eight populated conformational states and eight singletons, which we report explicitly as noise points (Figure 7).

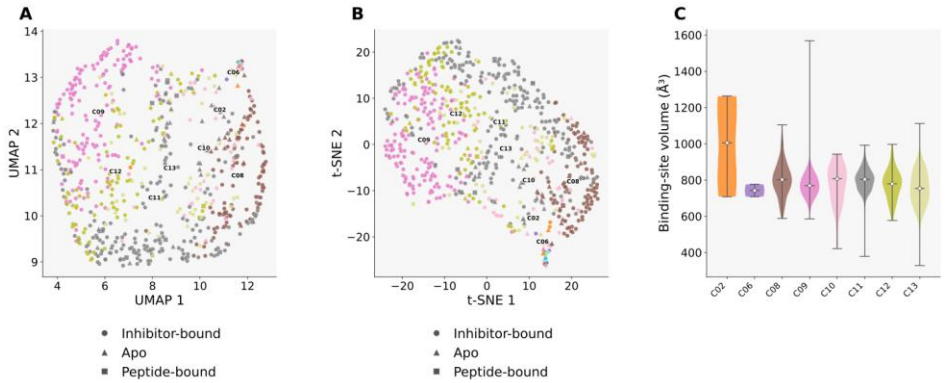

Figure 7. HIV-1 PR conformational landscape.
UMAP **(A)** and t-SNE **(B)** embeddings of per-structure spectral fingerprints, colored by conformational state, with stars denoting representative exemplars and labels shown for clusters of ≥8 members. **(C)** Per-cluster distributions of binding-site volume (Å³).

Four dominant states account for 583 of the 684 structures (85%): C11 (3VFB), the canonical darunavir-class closed-flap geometry; C09, a closed-flap state of low variability; C08, the most internally variable; and C12, the most homogeneous. The minor state C13 concentrates the drug-

resistant variants (17.3% of members against 5.5% dataset-wide), while the singleton C07 (1ZTZ) is the maximally open, metallacarborane-bound outlier with the largest cavity (2,236 Å³).

Across all 684 structures, twelve residues are exposed within the binding cavity in at least 80% of structures (Leu23, Gly27, Ala28, Asp29, Asp30, Val32, Ile47, Gly48, Ile50, Pro81, Val82, and Ile84), and together they assemble the catalytic cleft (Asp29/Asp30), the S1/S1′ pocket (Leu23, Ala28, Val32), the flap tips (Gly48, Ile50), and the S2/S2′ wall (Pro81, Val82, Ile84). Among the binders, Asp29 carries the highest hydrogen-bond frequency (58.6% and 51.1% for the two chains), and this core coincides with the established direct-contact resistance-mutation sites of HIV-1 PR. An inhibitor engaging most of the core therefore remains active against mutant strains, since any single active-site mutation weakens only one contact while the inhibitor holds at the remaining eleven. This is the substrate-envelope anchoring that darunavir exploits at the backbone atoms of Asp29/Asp30 [Ghosh et al. 2016]. Cavity size couples to ligand size (inhibitors above ~600 Da populate the largest closed-flap state C11, fragments below 400 Da the tighter C09 and C12), yet the chemical space of the 452 inhibitors, described by Morgan fingerprints computed with RDKit [Landrum 2023], is effectively continuous (mean pairwise Tanimoto distance 0.73–0.76 within each dominant state against 0.74 dataset-wide), so no state filters for a particular scaffold, and the active site is therefore amenable to *de novo* design without a new conformation having to emerge.

Application to antiviral discovery and experimental confirmation

Having characterised the Mpro landscape, we screened for inhibitors with a workflow combining a ligand-based pharmacophore with structure-based advanced virtual screening, the docking stage evaluating candidates against multiple representative Mpro conformations rather than one. Known Mpro inhibitors (IC50 < 20 μM) from ChEMBL and BindingDB [Gaulton et al. 2012; Gilson et al. 2016] yielded 155 reliable actives, which clustered into 94 groups. We built two pharmacophores in LigandScout [Wolber & Langer 2005] and carried forward the nine-feature model, which was stronger on selectivity and specificity. We screened this model through the Pharmit server [Sunseri & Koes 2016] against roughly 200 million compounds across six libraries (MolPort returned the most matches, 385), giving 168 pharmacophore hits. We docked these into eight representative structures with ICM-Pro and AutoDock Vina [Abagyan et al. 1994; Trott & Olson 2010], retained 43 by consensus ranking, and re-ranked them by MM/PBSA and ADME-Tox to a final 16. In an enzymatic assay against recombinant Mpro (nirmatrelvir control IC50 = 0.013 μM), seven of the sixteen showed IC50 below 100 μM, and the two most potent were ID11 (54.39 μM) and ID35 (59.39 μM). At a hit rate above 40%, well above the roughly 10% typical of standard virtual screening, this is a substantive outcome. The compounds are early-stage scaffolds for optimization rather than drug candidates [Pinzi et al. 2021; Dos Santos et al. 2022].

The methodology was then carried to influenza neuraminidase, which provided the *in vitro* and *in vivo* confirmation of the framework. Neuraminidase is a harder case on two counts: the target is a pair, the neuraminidases of influenza A and B, that share only moderate sequence identity (≈ 35% over the aligned catalytic region), and the confirmation sought is not merely enzymatic but antiviral in cells and protective in animals. A Q-Pocket query returned 209 neuraminidase structures. After biological filtering, 32 influenza A HxN1 and 23 influenza B structures remained, resolving into eight clusters (Table 2), and we selected the representative wild-type influenza A structure 6HP0 and the influenza B structure 4CPY for the design campaign. Against these representatives, a deep reinforcement-learning platform [Matevosyan et al. 2022] generated 60,291 unique compounds from a benzene seed, 86.5% scoring better than oseltamivir, which successive filters (synthetic accessibility [Ertl & Schuffenhauer 2009], QED [Bickerton et al. 2012], structural alerts, vendor availability, MM/PBSA) reduced to 31 chemically diverse leads of minimal resemblance to oseltamivir (mean Tanimoto 0.13 [Bajusz et al. 2015]). These are non-sialic-acid scaffolds, distinct from the substrate-mimetic inhibitors that dominate the class (Figure 8).

Table 2. Descriptions of clusters and representative structure corresponding to influenza A HxN1 subtypes.

PDB ID	Cluster ID	Is representative	Mutant/Wild Type	Subtype
2HTY	0	FALSE	Wild	H5N1
2HU4	0	FALSE	Wild	H5N1
3CKZ	0	FALSE	H274Y	H5N1
3CL0	0	TRUE	H274Y	H5N1
3NSS	1	FALSE	Wild	H1N1
3TI3	1	FALSE	Wild	H1N1
3TI4	1	FALSE	Wild	H1N1
3TI5	1	FALSE	Wild	H1N1
3TI6	1	FALSE	Wild	H1N1
6HP0	1	TRUE	Wild	H1N1
9O4N	1	FALSE	Wild	H1N1
4B7J	2	FALSE	I223R	H1N1
4B7R	2	TRUE	I223R	H1N1
5NWE	2	FALSE	H274Y	H1N1
5NZ4	2	FALSE	I223V	H1N1
5NZE	2	FALSE	S247N	H1N1
5NZF	2	FALSE	H274Y/ I223V	H1N1
5NZN	2	FALSE	H274Y/ S247N	H1N1
6G01	2	FALSE	Wild	H1N1
6G02	2	FALSE	Wild	H1N1
9NJ2	2	FALSE	Wild	H5N1
5HUG	3	TRUE	Wild	H5N1
9EJF	4	TRUE	Wild	H1N1
9O9V	4	FALSE	Wild	H1N1
4B7M	5	FALSE	I223R	H1N1
9GQQ	5	TRUE	Unidentified	xN1
9GQT	5	FALSE	Unidentified	xN1
9HLG	5	FALSE	Wild	xN1
9GQX	6	TRUE	Unidentified	xN1
8YVL	7	FALSE	Wild	H1N1
9HLH	7	FALSE	Wild	xN1
9HLI	7	TRUE	Wild	xN1

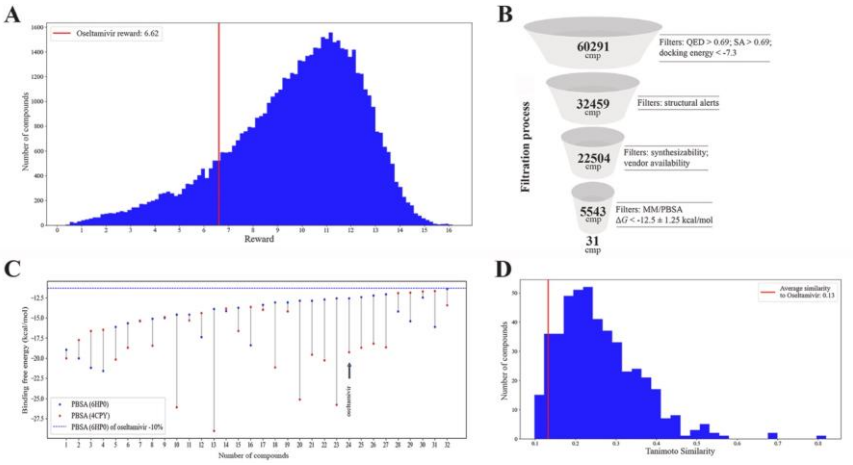

Figure 8. Summary of properties of RL-generated compounds and generation workflow.
(A) Reward-score distribution, with the red line marking oseltamivir's score. **(B)** Filtration workflow, showing the applied filters and threshold values. **(C)** MM/PBSA scores of compounds and oseltamivir against the 6HP0 and 4CPY **(D)** Average pairwise similarity among the 31 compounds, with the red line indicating average similarity to oseltamivir.

We synthesized nine leads, four of which inhibited the neuraminidase of influenza A (A/WSN/33), a hit rate above 40%. The two most potent, DS-22-inf-009 and DS-22-inf-021, inhibited both neuraminidases at low micromolar concentrations (IAV IC50 24.8 and 11.2 μM; IBV IC50 21.2 and 9.1 μM) and, critically, retained activity against the oseltamivir-resistant H275Y mutant (IC50 18.2 and 14.7 μM; Figure 9).

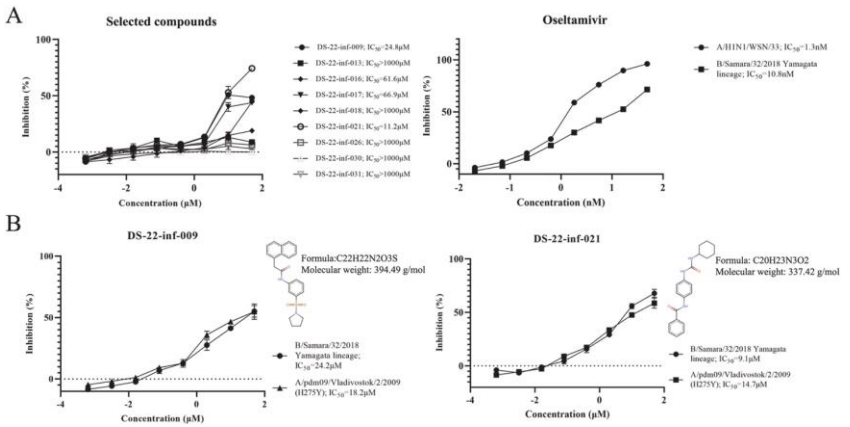

Figure 9. Neuraminidase inhibition (IC_{50}) by the tested compounds against IAV, IBV, and the H275Y mutant.
(A) Inhibitory effect of the NCEs and oseltamivir on NA enzymatic activity. **(B)** Inhibition of H275Y-mutant IAV and native IBV Yamagata-lineage NA activity by DS-22-inf-009 and DS-22-inf-021. Data are mean ± SD.

Molecular dynamics in GROMACS [Abraham et al. 2015] against the Q-Pocket-selected structures 6HP0 and 4CPY, validated on oseltamivir (whose pose was reproduced to within 0.26 and 0.65 Å), showed stable binding (backbone RMSD < 0.2 nm) with oseltamivir-like contacts to conserved arginine and aspartate residues. Because the leads are not sialic-acid mimics, they do not depend on the contacts the H275Y substitution disrupts [Moscona 2005a; Gubareva 2004].

In a cells-based assay with MDCK cells, the two compounds inhibited IAV and IBV dose-dependently (EC50 0.29–2.31 μM) with no measurable cytotoxicity (CC50 > 50 μM) and greater than 95% reduction in virus yield at 10–50 μM. The cell-based EC50 values fell consistently below the enzymatic IC50 values, consistent with genuine target engagement. In the animal model eight-week-old BALB/c mice (n = 6 per group), infected with IAV (H1N1 pdm09) or IBV and dosed intraperitoneally for five days, showed dose-dependent survival (Figure 10).

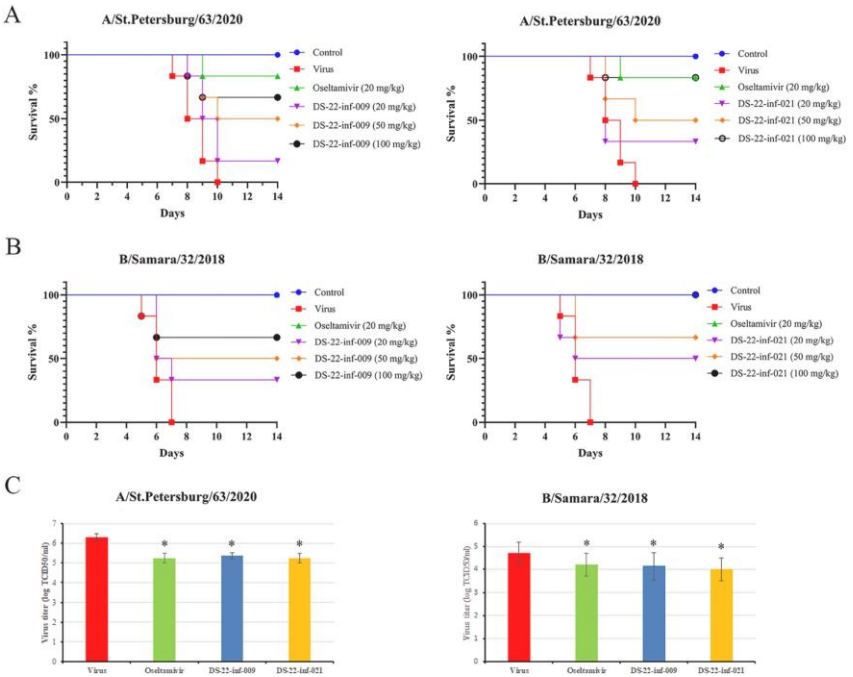

Figure 10. *In vivo* therapeutic efficacy of DS-22-inf-009 and DS-22-inf-021 against Influenza virus infection.
(A–B) Survival curves of mice infected with **(A)** IAV and **(B)** IBV following a 5-day, once-daily regimen of compounds or oseltamivir at varying doses. Mortality was monitored for 14 days post-infection (Mantel-Cox analysis; significant difference defined as $P < 0.01$). **(C)** Pulmonary viral titers quantified in lung tissue at day 5 post-infection. Data are presented as mean ±SD. Statistical significance relative to the infected control group is indicated by $* P < 0.05$.

For DS-22-inf-009 against IAV, survival was 15%, 50%, and 65% at 20, 50, and 100 mg/kg, and DS-22-inf-021 reached 85% at the top dose, while against IBV the two compounds gave 65% and 100% at 100 mg/kg. Lung viral titres fell significantly ($p < 0.05$), and we saw no weight loss or adverse clinical signs at the efficacious dose. These directly measured *in vivo* outcomes are the

strongest evidence in the work that the framework can deliver compounds with confirmed antiviral activity in an animal model, covering the whole computational pipeline.

Limitations. Q-Pocket operates on the experimentally evidenced conformational ensemble and can only resolve states that structure determination has sampled. For a sparsely characterised target the recovered ensemble will be incomplete, and cryptic or transient states seen only by NMR or molecular dynamics would not appear. This defines a gap and highlights the need for integration of a simulation-based sampling stage. The benchmarking metrics depend on the composition of the active and decoy sets, so absolute performance figures are library-specific even though the relative ranking of structures was reproducible across different docking engines. The experimental attribution of activity specifically to representative-conformation selection is only partly established, since screening outcome is also influenced by engine, library, and generation choices. Finally, generalization has been shown on three enzymes (two proteases and a sialidase), which is a meaningful but bounded range.

CONCLUSION

This work develops and validates Q-Pocket, a framework that represents a protein binding site not by its backbone coordinates but by the chemistry and geometry of the cavity itself, and on that basis reduces the target's experimentally evidenced conformational ensemble to a minimal set of representative conformations. The Q-Pocket uses sequence-agnostic and alignment-free spectral distance logic to compare cavities, and its implementation has been systematically demonstrated on three viral use cases (the SARS-CoV-2 main protease, the HIV-1 protease, and the influenza neuraminidase) and validated through experimentally confirmed antiviral compounds obtained from the resulting representative ensembles.

On the SARS-CoV-2 Mpro we established that crystallographic quality and virtual-screening competence are independent properties of a structure. Certain conformations underperformed on enrichment, ROC-AUC, and pose prediction reproducibly across three independent docking engines, so the deficiency is a property of the structure itself rather than an artefact of any single scoring function. The cause of that was also characterised by structural analysis, which means that the failure of a conformation has a defined, identifiable basis and that performance benchmarking must be treated as a mandatory, separate stage of structure selection rather than being inferred from crystallographic quality metrics.

Q-Pocket was applied also to the HIV-1 protease, which has 684 structures available and a binding pocket located between two homodimer chains. After clustering, 16 clusters (8 of which were singletons) were identified, four of which accounted for 85% of all structures, so the use of 4 representatives of those clusters can cover most of evidenced conformational states during SBDD campaigns. The ensemble is therefore finite and small even for a pocket within a homodimer interface.

Selection of representative conformations then supported experimental discovery against two targets. Against Mpro, an advanced virtual screen of approximately 200 million compounds returned confirmed micromolar-range inhibitors in an enzymatic assay. Against influenza neuraminidase, a *de novo* reinforcement-learning campaign produced novel, non-sialic-acid compounds that inhibited both the influenza A and B neuraminidases, retained activity against the oseltamivir-resistant H275Y variant, and demonstrated protection in an animal model.

Taken together, the research done for this dissertation converges on a single conclusion: the choice of receptor conformation is a decisive and benchmarkable input to structure-based antiviral drug discovery, a target's conformational ensemble is finite and can be represented with a few physicochemically distinct representative states, and SBDD campaigns against that representative ensemble yield experimentally validated compounds, up to protection *in vivo*. The framework is

limited by the structural record available. Within these bounds, the work reframes conformation selection from a crystallographic-quality-based step into a defined part of the discovery workflow, and the representative receptor ensemble constructed during these studies is transferable and should be recorded as a conformational atlas, which we plan to develop.

INFERENCES

1. Crystallographic quality and screening competence are independent properties: performance benchmarking must be treated as a mandatory, separate selection stage, not inferred from quality metrics.
2. Underperformance for studied structures is reproducible across three docking engines, so it is a property of the structure itself, not an artefact of one scoring function.
3. A target's holo ensemble is finite and reduces to a small set of representative conformations, making the ensemble, not a single structure, the correct input for structure-based design.
4. Failure of certain structures in structure-based design has a defined mechanistic basis, which can be analyzed and reported as computable inclusion/exclusion criteria for newly deposited structures.
5. Screening on framework-selected structures delivers large, experimentally confirmed gains in prospective success rate (> 40%) and recovers chemically diverse scaffolds that single-structure screening would miss
6. Generative (RL)-based drug design campaign with selected representative-structures resulted in in vitro and in vivo efficacy (success rate > 40%), serving as stronger validation of the framework.
7. The framework generalizes beyond high-conservation single targets to moderate-homology, multi-subtype targets.
8. Based on selected representative structures RL platform yields novel, non-sialic-acid scaffolds (mean Tanimoto 0.13 to oseltamivir), expanding chemical space beyond substrate-mimetic inhibitors.

LIST OF PUBLICATIONS

1. **Khachatryan H.** Systematic Analysis of the HIV-1 Protease Active-Site Conformational Space Across 690 Crystal Structures. Reports NAS RA. 2026, 126(2), 2. doi:10.54503/0321-1339-2026.126.2-2.

2. **Khachatryan H**, Matevosyan M, Harutyunyan V, Gevorgyan S, Shavina A, Tirosyan I, Gabrielyan Y, Ayvazyan M, Bozdaganyan M, Fakhar Z, Gharaghani S, Zakaryan H. Computational evaluation and benchmark study of 342 crystallographic holo-structures of SARS-CoV-2 Mpro enzyme. Sci Rep. 2024 Jun 20;14(1):14255. doi: 10.1038/s41598-024-65228-5.

3. Gevorgyan S, **Khachatryan H**, Shavina A, Gharaghani S, Zakaryan H. Targeting SARS-CoV-2 main protease: a comprehensive approach using advanced virtual screening, molecular dynamics, and *in vitro* validation. Virol J. 2024 Dec 21;21(1):330. doi: 10.1186/s12985-024-02607-4.

4. Izmailyan R, Matevosyan M, **Khachatryan H**, Shavina A, Gevorgyan S, Ghazaryan A, Tirosyan I, Gabrielyan Y, Ayvazyan M, Martirosyan B, Harutyunyan V, Zakaryan H. Discovery of new antiviral agents through artificial intelligence: *in vitro* and *in vivo* results. Antiviral Res. 2024 Feb;222:105818. doi: 10.1016/j.antiviral.2024.105818.

ԽԱՉԱՏՐՅԱՆ ՀԱՄԼԵՏ ՆԱՐԵԿԻ

Վիրուսային սպիտակուցների ակտիվ կենտրոնների կոնֆորմացիոն վերլուծության հարթակ հակավիրուսային միացությունների կառուցվածքահեն հայտնաբերման համար

ԱՄՓՈՓԱԳԻՐ

Բանալի բառեր՝ դեղերի կառուցվածքահեն հայտնաբերում, վիրտուալ սքրինինգ, սպիտակուցի կոնֆորմացիա, ներկայացուցչական կառուցվածք, մոլեկուլային դոկինգ, հակավիրուսային միջոցներ, նեյրամինիդազ, SARS-CoV-2 գլխավոր պրոտեազ, ՄԻԱՎ-1 պրոտեազ:

Դեղերի կառուցվածքահեն նախագծումը և հայտնաբերումը ժամանակակից դեղագիտական հետազոտությունների առանցքային մոտեցումներից են: Դրանք հիմնվում են թիրախ-սպիտակուցի եռաչափ կառուցվածքի վրա՝ պոտենցիալ արգելակիչների կամ ակտիվատորների համակարգչային մոդելավորման նպատակով: Մեթոդի արդյունավետությունը գլխավորապես կախված է թիրախի կոնֆորմացիայի ճիշտ ընտրությունից: Քանի որ սպիտակուցների կապման տեղամասերը կոնֆորմացիոն առումով հետերոգեն են, միևնույն թիրախի համար Սպիտակուցների տվյալների բանկում (PDB) հասանելի են բազմաթիվ փորձարարական կառուցվածքներ՝ տարբեր կոնֆորմացիոն վիճակներում:

Կառուցվածքների ընտրությունն ընդունված է հիմնավորել բյուրեղագրական կամ այլ փորձարարական մեթոդների որակի ստանդարտ ցուցանիշներով (Resolution, R-գործակիցներ, Ramachandran-ի վիճակագրություն, Clashscore): Սակայն այս ցուցանիշները բնութագրում են կառուցվածքի փորձարարական որակը, այլ ոչ թե դրա պիտանիությունը վիրտուալ սքրինինգի համար, ինչն էլ ստեղծում է մեթոդաբանական էական բաց: Հետևաբար, սույն աշխատանքի նպատակն է ստեղծել և փորձարարական եղանակով վավերացնել համակարգված, ֆիզիկաքիմիական առումով հիմնավորված հարթակ՝ դեղերի կառուցվածքահեն դիզայնի և հայտնաբերման գործընթացում սպիտակուցների ներկայացուցչական կոնֆորմացիաների ընտրության համար:

Աշխատանքում ցույց է տրվել, որ բյուրեղագրական որակի ստանդարտ ցուցանիշները բավարար չեն վիրտուալ սքրինինգի համար կառուցվածքների ընտրության հարցում: SARS-CoV-2-ի գլխավոր պրոտեազի (Mpro) 342 հոլո-կառուցվածքների կլաստերավորմամբ առանձնացվել է 8 ներկայացուցչական կոնֆորմացիա: Որակի բոլոր չափանիշներին բավարարող երեք կառուցվածքներ (5RHE, 7DDC, 7DPU) երեք անկախ դոկինգային ծրագրերում (Vina, rDock, Glide) հետևողականորեն ցուցաբերել են ամենացածր EF, ROC-AUC և BEDROC ցուցանիշները, ինչպես նաև կապման դիրքի կանխատեսման ցածր ճշգրտություն: Բացահայտվել է, որ սքրինինգի անարդյունավետության կառուցվածքային պատճառը S1 ենթակենտրոնի փակ լինելն է (Met49-ի գոշ կոնֆորմացիա, Leu141/Asn142 կողմնային շղթաների վերակողմնորոշում) և S2 ենթագրպանի առկայությունը:

Նախագծված հարթակի արդյունավետությունն ու կիրառելիությունը ցույց է տրվել նաև ՄԻԱՎ-1 պրոտեազի օրինակով: 690 կառուցվածքների ակտիվ կենտրոնների

կոնֆորմացիոն բազմազանությունը հարթակի միջոցով ներկայացվել է 16 դիսկրետ կոնֆորմացիաների տեսքով, որոնցից չորսն ընդգրկում են կառուցվածքների մոտ 85%-ը: Կոնֆորմացիոն վերլուծությամբ ընտրված Mpro կառուցվածքների վրա իրականացված վիրտուալ սքրինինգով ուսումնասիրվել է շուրջ 200 միլիոն միացությունից բաղկացած գրադարան: Ֆիլտրված և փորձարարական եղանակով ստուգված 16 միացություններից 2-ը (≈12,5% հաջողության մակարդակ, ինչը շուրջ 100 անգամ գերազանցում է ստանդարտ բարձր թողունակությամբ սքրինինգի արդյունքները) դրսևորել են *in vitro* ակտիվություն:

Մեթոդաբանության կիրառելիությունը չափավոր հոմոլոգիա ունեցող թիրախների համար ցուցադրվել է գրիպի A (IAV) և B (IBV) վիրուսների նեյրամինիդազի (NA) օրինակով: Մշակված հարթակի միջոցով ընտրվել են ներկայացուցչական հոլո-կառուցվածքները (IAV՝ 6HP0, IBV՝ 4CPY), և RL մոդելով գեներացվել է 60,291 *de novo* միացություն, որոնց 86,5%-ը ֆիզիկաքիմիական հատկություններով գերազանցել է օսելտամիվիր (oseltamivir) դեղամիջոցին: Ֆիլտրումից հետո սինթեզվել է 9 ոչ-սիալաթթվային միացություն (օսելտամիվիրի հետ միջին Տանիմոտոյի նմանությունը՝ 0,13), որոնցից չորսն ակտիվ են եղել: Երկու առաջատար միացությունները (DS-22-inf-009 և DS-22-inf-021) ցածր միկրոմոլային կոնցենտրացիաներում (IC50 ≈ 9–25 մկՄ) արգելակել են ինչպես IAV-ի, այնպես էլ IBV-ի NA-ն՝ ներառյալ օսելտամիվիրի նկատմամբ կայուն H275Y մուտանտը (18,2 / 14,7 մկՄ): *In vivo* փորձերում (մկներ, 100 մգ/կգ) DS-22-inf-021-ն ապահովել է 85% (IAV H1N1pdm09) և 100% (IBV Yamagata) ապրելիություն՝ թոքային հյուսվածքում վիրուսային տիտրի վիճակագրորեն նշանակալի նվազմամբ:

Ամփոփելով՝ աշխատանքը հաստատում է, որ ներկայացուցչական կառուցվածքի ընտրությունը գիտականորեն էական, հաշվարկելի բնութագրիչ է, որը պահանջում է հատուկ վերլուծություն: Մշակված հարթակը, որպես մեթոդից անկախ ընտրության շերտ, կիրառելի է ինչպես վիրտուալ սքրինինգի, այնպես էլ մեքենայական ուսուցման վրա հիմնված *de novo* նախագծման աշխատանքներում՝ ապահովելով փորձարարական ճանապարհով վավերացված արդյունքներ:

ХАЧАТРЯН ГАМЛЕТ НАРЕКОВИЧ

«Платформа конформационного анализа активного центра вирусных белков для структурно-ориентированного дизайна противовирусных препаратов»

РЕЗЮМЕ

Ключевые слова: структурно-ориентированный поиск лекарств, виртуальный скрининг, конформация белка, репрезентативная структура, молекулярный докинг, противовирусные препараты, нейраминидаза, главная протеаза SARS-CoV-2, протеаза ВИЧ-1.

Структурно-ориентированное проектирование и поиск лекарственных средств — один из ключевых подходов современных фармакологических исследований, основанный на использовании трёхмерной структуры белка-мишени для компьютерного моделирования потенциальных ингибиторов или активаторов. Эффективность данного подхода решающим образом зависит от выбора правильной конформации мишени. Сайты связывания белков конформационно гетерогенны, и для одной и той же мишени в Банке данных белков (PDB) доступно множество экспериментальных структур в различных конформационных состояниях.

Выбор структур традиционно обосновывают стандартными показателями качества кристаллографического или иного экспериментального метода (разрешение, R-факторы, статистика Рамачандрана, Clashscore). Однако эти показатели отражают лишь экспериментальное качество структуры, а не её пригодность для виртуального скрининга, что создаёт существенный методологический пробел. Цель настоящей работы — разработать и экспериментально валидировать систематизированную, физико-химически обоснованную платформу для отбора репрезентативных конформаций белков при структурно-ориентированном поиске и дизайне лекарственных соединений.

В работе показано, что стандартных показателей кристаллографического качества недостаточно для корректного отбора структур для виртуального скрининга. В результате кластеризации 342 holo-структур главной протеазы SARS-CoV-2 (Mpro) было выделено 8 репрезентативных конформаций. При этом три структуры (5RHE, 7DDC, 7DPU), удовлетворявшие всем формальным критериям качества, в трёх независимых программах докинга (Vina, rDock, Glide) последовательно демонстрировали наименьшие значения факторов обогащения (EF), ROC-AUC, BEDROC и низкую точность предсказания позиции связывания. Установлено, что структурной причиной неэффективности скрининга являются закрытое состояние субсайта S1 (гош-конформация Met49, переориентация боковых цепей Leu141/Asn142) и наличие субкармана S2.

Эффективность и универсальность разработанной платформы продемонстрированы также на примере протеазы ВИЧ-1. Конформационное разнообразие активных центров 690 структур было сведено к 16 дискретным конформациям, четыре из которых охватывают около 85% структур. С применением отобранных структур Mpro проведён виртуальный скрининг библиотеки, содержащей около 200 миллионов соединений. Из 16 отобранных и протестированных соединений 2 продемонстрировали активность *in*

vitro (уровень успеха ≈12,5%, что примерно в 100 раз превышает результативность стандартного высокопроизводительного скрининга).

Применимость методологии к мишеням с умеренной гомологией показана на примере нейраминидазы (NA) вирусов гриппа A (IAV) и B (IBV). С помощью разработанной платформы отобраны репрезентативные holo-структуры (IAV — 6HP0, IBV — 4CPY), а с применением RL-модели сгенерировано 60 291 соединение *de novo*, из которых 86,5% превосходили препарат осельтамивир по физико-химическим свойствам. После фильтрации синтезировано 9 несиаловых соединений (среднее сходство по Танимото с осельтамивиром — 0,13), четыре из которых оказались активными. Два лидирующих соединения (DS-22-inf-009 и DS-22-inf-021) ингибировали NA вирусов гриппа A и B в низких микромолярных концентрациях (IC50 ≈ 9–25 мкМ), включая активность против резистентного к осельтамивиру мутанта H275Y (18,2 / 14,7 мкМ). В экспериментах *in vivo* (мыши, 100 мг/кг) соединение DS-22-inf-021 обеспечило выживаемость 85% (IAV H1N1pdm09) и 100% (IBV Yamagata) при статистически значимом снижении вирусного титра в лёгочной ткани.

В целом работа подтверждает, что выбор репрезентативной структуры является критически важным и поддающимся строгому расчёту этапом, требующим специализированного анализа. Разработанная платформа представляет собой универсальный, метод-независимый уровень отбора, применимый как в традиционном виртуальном скрининге, так и в кампаниях *de novo* дизайна на основе машинного обучения с получением экспериментально валидированных результатов.