

REVIEW

of the official opponent on the dissertation of

Hamlet Narek Khachatryan

“Active Site Conformational Analysis Framework of Viral Proteins for Structure-Based Antiviral Drug Discovery”,

submitted for the degree of Candidate of Biological Sciences (PhD)

in the field 03.00.03. “Molecular and Cellular Biology”

The dissertation addresses an important yet barely studied problem of structure-based drug discovery and that is rarely treated as a problem at all. Both virtual screening and generative *de novo* design take a three-dimensional receptor model as their input, and for any well-studied viral enzyme the Protein Data Bank offers tens to hundreds of alternative crystallographic conformations of the same active site. In practice the choice among them is empirical, driven by resolution, R-factors and other similar global quality descriptors. However, the question if these descriptors have anything to do with the capacity of a given structure to bind to an active compounds has not been tested systematically. The candidate selects this important issue as the subject of the work and answers it thoroughly.

The dissertation follows the classical structure and consists of an Introduction, a Literature Review, Materials and Methods, Results and Discussion, a Conclusion, Inferences and a list of 174 references. Its text is accompanied with 35 figures, 7 tables and 10 equations that increase the comprehensiveness of the thesis. The main results have been published in four papers and were presented in local and international meetings.

The literature review is compact and argued rather than cataloged, which I regard as an advantage. It covers docking algorithms and the four families of scoring functions, virtual-screening metrics with the appropriate emphasis on early recognition, reinforcement-learning *de novo* design, and the end-state rescoring and ADME estimation used later in the work. It well prepared a reader to the main still-open question of how the representative ensemble should be selected, and to the observation that the available selection criteria compare backbones rather than the chemistry of the cavity. The treatment of conformational selection against induced fit, of cryptic pockets, and of the flap dynamics of HIV-1 protease is accurate and appropriately selective, and the structural rationale for each choice of enzyme is clearly stated.

The Q-Pocket framework is specified in sufficient detail to be reproduced and reused. A binding site is represented as a point cloud at 0.70 Å spacing, each point annotated with six physicochemical fields: the electrostatic potential obtained by solving the linearized Poisson-Boltzmann equation in APBS with PARSE charges, a Coulomb potential from atomic partial charges, a Fermi-Dirac-weighted lipophilic potential, Gaussian hydrogen-bond donor and acceptor densities, and an aromatic density over ring centroids. An ϵ -ball geodesic graph is built over the cloud, all-pairs shortest paths are computed, and the symmetric normalized Laplacian yields both a low-frequency eigenvalue spectrum as a shape signature and a graph Fourier power spectrum of the six fields as a chemical fingerprint; the two are combined into a single pairwise distance and exemplars are selected by affinity propagation. I note with approval that the weighting parameters were held fixed across every target analyzed, so that no reported result depends on per-target parameter optimization. Curation criteria are stated separately for each dataset, the experimental section is

described adequately for reproduction, and the ethical approval for the animal work is documented. The release of Q-Pocket as an openly available Python package under the MIT license is a substantive contribution to the respective field.

The results are presented in three interconnected parts. In the first, 454 SARS-CoV-2 Mpro holo-structures were curated to 342, from which eight representative conformations were benchmarked against a library of 297 co-crystallized active molecules and 8,910 property-matched decoys using three independently developed docking engines. Performance varied sharply and reproducibly with the receptor structure: 7DDC, 7DPU and 5RHE underperformed on enrichment, ROC-AUC and BEDROC across all three engines, and for 7DDC no program recovered a single pose within 2 Å of the crystallographic reference. The candidate then supplies the structural explanation, namely loss of the S1 subsite in 7DDC and 7DPU and an aberrant subpocket opening within S2 in all three, and shows by ensemble normal-mode analysis that the catalytic residues are among the least mobile in the ensemble, so that backbone flexibility alone cannot account for the differences. The demonstration that crystallographic quality and screening competence are independent properties of a structure is, in my judgment, the most valuable result of the dissertation.

In the second part the framework is applied to the HIV-1 protease, where the cavity forms at the homodimer interface. Of 690 retrieved structures, 684 were selected and clustered into 16 groups, of which eight are singletons reported explicitly as outliers. The identification of a twelve-residue binding-site core coincident with the cataloged direct-contact resistance positions, and the concentration of resistant variants in a single minor state, are interesting observations that link the geometric analysis to the pharmacology of the target.

The third part carries the approach into two prospective campaigns. Against Mpro, a pharmacophore query applied to approximately 200 million commercially available compounds, followed by consensus rank-summed docking across eight representative conformations and two engines, MM/PBSA rescoring and ADME-Tox filtering, yielded sixteen compounds for assay, of which seven inhibited the recombinant enzyme below 100 µM. Against Influenza neuraminidase, a reinforcement-learning agent generated 60,291 compounds against Q-Pocket-selected representatives of Influenza A and B; nine of the filtered candidates were synthesised and four proved active. The two leads inhibit both the IAV and IBV enzymes and, importantly, retain activity against the oseltamivir-resistant H275Y neuraminidase with no measurable cytotoxicity, and confer dose-dependent protection in lethally challenged BALB/c mice with significant reduction of lung viral titers. That these compounds are non-sialic-acid scaffolds with a mean Tanimoto similarity of 0.13 to oseltamivir is a meaningful result in a target class dominated by substrate mimetics.

The novelty of the work lies in treating receptor-conformation selection as a definable and benchmarkable stage of the structure-based discovery workflow rather than as an informal preliminary, and in supplying an alignment-free, physicochemically grounded procedure for performing it. The demonstration that global crystallographic quality does not predict screening competence, together with a computable structural criterion for inclusion or exclusion, is directly actionable. The practical significance is supported by extensive validation, including *in silico* benchmarking, enzymatic and cell-based assays, and protection in an animal model. The framework itself, the curated ensembles and the representative receptor sets are reusable, and the candidate's stated intention to develop them into a conformational atlas seems to me a sound direction.

The results are reliable. The central computational claim rests on three independent docking engines and three complementary metrics, the structural datasets are large and their curation is documented, the framework parameters were not tuned per target, and the biological conclusions are supported by orthogonal experimental readouts, with the molecular-dynamics protocol validated against the crystallographic pose of oseltamivir before being applied to the designed compounds. I also note that the candidate sets out the limitations of the work with a candor that is not always encountered, including the dependence on the experimentally evidenced ensemble, the deposition bias toward inhibitor-bound states, and the partial nature of the causal link between conformation selection and the experimental outcomes. This is the correct scientific posture and it strengthens rather than weakens the dissertation.

I would offer the following remarks, which concern the presentation of the material rather than its substance and none of which affects the conclusions of the work.

1. The Introduction (p. 12) announces that the main results are published in four papers, but the work contains no separate statement of the personal contribution of the author. Given that the candidate is first or sole author on two of the four publications and a co-author on the two experimental papers it would be useful to the Council.
2. A number of internal inconsistencies should be corrected before the text is deposited. Section 3.2.1 refers the reader to Section 3.1.1 for the Mpro conformational landscape, which is presented in Section 3.1.2, and the YAML configuration file is Figure 13 but is cited twice as Figure 9 (pp. 47, 49). The six point-cloud descriptors are listed in Section 3.1 as including a “lone-pair potential”, whereas Section 2.2.1 and the list of abbreviations define the sixth field as the Coulomb potential. In Section 3.1.6 the catalytic cleft of HIV-1 protease is attributed to Asp29/Asp30, whereas Section 1.3.2 correctly assigns catalysis to the Asp25/Asp25' dyad, and only ten of the twelve core residues are named. The caption to Figure 19 reports apo cavity volumes as higher than those of inhibitor-bound structures while the text describes them as comparable. Table 6 lists clusters 0 to 7, all of them Influenza A, although the text states that the largest cluster consisted exclusively of Influenza B entries represented by 4CPY. The in vivo survival values of 15, 65 and 85 percent cannot be reconciled with the stated group size of six animals.
3. The text would benefit from a final language-editing pass. Several typographical and grammatical slips remain, among them “farmwork” for framework (p. 11), “cauterizes” (p. 88), “genitalization” (p. 16), “craterisation” (p. 49) and “684 hundred structures” (p. 88); the abbreviation “IAB” is used for Influenza B in several places, including the heading of Section 3.2.2, while “IBV” is used elsewhere; Matevosyan et al. is cited as 2022 in the text and listed as 2023 in the references; and the size of the screened library is given variously as 198 million, approximately 200 million and “some two hundred million”.

I would also ask the candidate to address two points, raised for discussion rather than as objections.

1. The eight representative Mpro conformations used in the benchmark were obtained, according to Sections 2.3 and 3.1.2, by hierarchical clustering of the C α coordinates of the binding-site residues in Bio3D, whereas Section 3.2.1 describes the same eight structures as selected by Q-Pocket. Section 2.3 further announces a comparison of this backbone grouping against the chemistry-based Q-Pocket

clustering of the same set, which is not reported. Does Q-Pocket, applied to the full ensemble of 342 structures, recover the same partition and the same exemplars? I would add that 187 of the 297 benchmark actives are covalent binders of Cys145 yet were docked and ranked with non-covalent scoring functions, and it would be instructive to know how the ranking of the eight structures changes if the benchmark is restricted to the 110 non-covalent ligands.

2. The prospective hit rate of seven actives out of sixteen assayed compounds is compared with a reference figure of roughly ten percent for conventional virtual screening. What is the source of that reference and on what activity threshold is it based, given that activity here is defined as an IC₅₀ below 100 μ M? Since no single-conformation arm was run in parallel, on what basis can the improvement be attributed specifically to representative-conformation selection rather than to the pharmacophore prefilter, the consensus of two docking engines or the MM/PBSA rescoring?

The remarks set out above do not diminish the novelty or the significance of the work and do not call into question the results submitted for defense. The dissertation of Hamlet Khachatryan is an original, independent, complete and scientifically significant study. The candidate has demonstrated command of computational chemistry, structural bioinformatics and antiviral pharmacology, and the ability to formulate and solve a well-defined scientific problem and to interpret the results in their biological and clinical context.

The dissertation “Active Site Conformational Analysis Framework of Viral Proteins for Structure-Based Antiviral Drug Discovery” meets the requirements applied to dissertations submitted for the degree of Candidate of Biological Sciences in the field 03.00.03. “Molecular and Cellular Biology”, and its author, Hamlet Narek Khachatryan, deserves the award of the degree of Candidate of Biological Sciences.

Official opponent,

Prof. Bing Xiong, PhD

Shanghai Institute of Materia Medica, Chinese Academy of Sciences

signature of Bing Xiong:



2026.9.19