Transactions of the Korean Nuclear Society Spring Meeting
Jeju, Korea, May 7-8, 2026

Design and Operational Verification of an Agentic AI System for SSTR2-Binding Peptide Candidate Screening

Dongju Kim [a], Soyeon Kim [a, b], Yonggyun Yu [a, b], Min-Kyu Kim[c], Ki Bum Ahn[c], Ho Seong Seo[b,c*], Yujong Kim [a, b*]
[a]*Applied Artificial Intelligence Section, KAERI, Daejeon 34057, Republic of Korea*
[b]*University of Science & Technology, Daejeon 34113, Republic of Korea*
[c]*Cyclotron Application Research Section, KAERI, Jeongeup 56212, Republic of Korea*
[*]*Corresponding author: hoseongseo@kaeri.re.kr, yjkim@kaeri.re.kr*

***Keywords :** Agentic AI, Multi-agent, AI Scientist, Peptide Screening, SSTR2, SST14, Workflow Automation

1. Introduction

Peptide candidate screening is not a single computational task but rather an iterative exploration workflow comprising candidate generation, structure-based evaluation, quality control (QC), result compilation, and subsequent planning. As the number of candidates grows, the manual burden increases not only in computational execution but also in result interpretation and next-round experimental planning. This burden becomes a bottleneck particularly when rapid triage of numerous candidates is required at the pre-wet-lab stage. Furthermore, when computation, review, documentation, and planning are distributed across separate tools and procedures, the traceability and reproducibility of the iterative search degrade, leading to inefficient linkage with downstream experimental design.

Existing docking and simulation tools are useful for executing individual computational steps, but their capacity to integrate candidate generation–evaluation–selection–reporting–next-iteration planning into a single iterative loop remains limited. In problems such as SSTR2-binding candidate screening, where prioritization is critical, a systematic architecture that reliably executes and logs iterative loops has a more direct impact on productivity than the generation of a single score. However, domain-specific implementations of such automated workflow designs remain scarce.

In this work, we design and implement an Agentic AI-based multi-agent system using Qwen3:8b[1] that combines LLM-driven planning with rule-based quality control to automate the iterative workflow of SSTR2-binding peptide candidate screening. The proposed system consists of five agents (Planner, QCRanker, DiversityManager, Critic, and Reporter) and executes a mutate–dock–QC–critique–report loop where each agent uses natural language processing for planning, simulation, and feedback. The objective of this paper is not to improve a specific binding-affinity prediction accuracy but to present the design, implementation, and operational verification of a system that supports candidate prioritization and subsequent experimental candidate derivation at the pre-wet-lab stage.

The system's operational validity is examined through a directional-consistency check against a small literature-based reference set and a prospective candidate-proposal demonstration using the LLM Planner.

2. Methods and Results

This section describes the architecture and operational results of the proposed system. The system structure and iterative workflow are presented first, followed by the candidate generation–simulation–QC pipeline. A brief check of prospective candidate-proposal example for directional-consistency is also presented.

2.1. System Overview and Iterative Workflow

The proposed system is designed as an iterative automation workflow for SSTR2-binding peptide candidate screening. As illustrated in Fig. 1, the system comprises five agents (Planner, QCRanker, DiversityManager, Critic, and Reporter) connected through a mutate–dock–QC–critique–report loop. In each iteration, the Planner formulates a mutation strategy informed by prior-iteration results and Critic feedback; the generated candidates undergo PyRosetta[2] based simulation and rule-based QC for evaluation and ranking; the DiversityManager suppresses duplicate or biased candidates; the Critic identifies failure patterns and improvement points; and the Reporter compiles structured logs including top candidates, QC outcomes, failure reasons, and follow-up recommendations. The Reporter's output is fed back to the Planner as input for the next iteration, thereby closing the loop.

The purpose of this architecture is not merely to automate individual docking calculations but to connect candidate generation–evaluation–selection–logging–planning into a single traceable workflow. By explicitly linking inputs and outputs at each iteration, the system reduces inter-step discontinuities and maintains decision rationale in structured logs, where the information tends to become scattered in manual search processes. The role separation among agents also ensures that each step's responsibility boundary is clearly defined, facilitating extensibility and auditability.

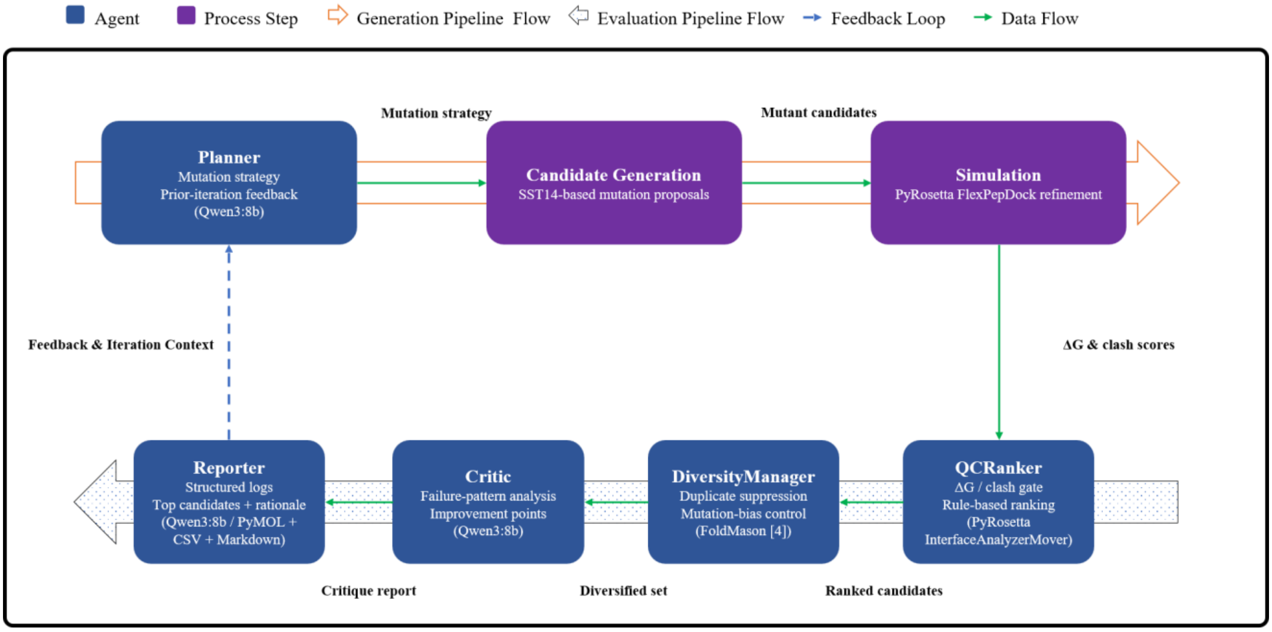

Fig. 1. Iterative search workflow of the proposed system (data flow among Planner–Candidate Generation-Simulation–QCRanker–DiversityManager-Critic–Reporter and the mutate–dock–QC–critique–report loop).

2.2. Candidate Generation, Simulation, and QC Pipeline

In the candidate generation step, SST14-based sequences serve as starting points, and mutation proposals are made centered on predefined variable positions. In initial iterations, mutation strategies inspired by hotspot residues which are protein-ligand interaction sites with distinctive structural or functional characteristics, reported in the literature or analogue-inspired variations may be employed; in subsequent iterations, the Planner adjusts mutation positions and search intensity based on prior results. The DiversityManager provides constraints that suppress repeated generation of identical or similar candidates and mitigate mutation-position bias.

Generated candidates undergo PyRosetta FlexPepDock refinement and InterfaceAnalyzerMover scoring [2,3] using an SSTR2–ligand complex structure (PDB anchor) [5,6] as the reference. For each candidate, 10 independent docking trials are performed, and the top-3 mean ΔG is used as the primary ranking metric to mitigate the inherent stochasticity of Monte Carlo-based refinement. Trials yielding $\Delta G > 0$ are excluded as refinement failures. Simulation outputs including ΔG, interface scores, and clash counts are used as internal criteria for relative comparison among candidates under identical computational conditions, and they are not interpreted as direct predictions of wet-lab binding affinity.

The QCRanker applies rule-based QC gates to the simulation results to eliminate clearly unsuitable candidates and performs relative ranking on those that pass QC. Where necessary, re-sorting reflecting diversity constraints is performed to prioritize the candidate pool for subsequent review. This pipeline targets "consistent application of evaluation procedures and selection criteria" rather than "accuracy optimization". The actual thresholds and metric combinations are designed to be adjustable to operational objectives.

From the pipeline-operation perspective, the candidate pool is evaluated through the same simulation and QC procedures, ensuring inter-iteration comparability and criterion consistency. The automated filtering and ranking procedure also provides a structure that reduces the manual burden of result compilation and prioritization.

2.3. Checking of Directional-Consistency and Proposal of Prospective Candidate

To examine whether the pipeline produces structurally plausible rankings, a small reference set was evaluated through the same simulation–QC–ranking pipeline, as summarized in Table I. The reference set consists of three categories: a baseline, positive controls, and negative controls. The baseline is LIT-01, the wild-type SST14 sequence, which serves as the reference point for all comparisons. Two positive controls were derived from clinically established SST14-based analogues: LIT-02, an octreotide-mapped 14-mer [5] representing a clinically validated SSTR2-selective compound, and LIT-03 (CST-14) [7], a naturally occurring SSTR2 agonist with confirmed nanomolar binding affinity. Two negative controls were also included, each targeting a known hotspot residue: SAN-01 (W8A) and SAN-02 (K9A). According to structural studies [5,6], Trp8 and Lys9 are positioned at the bottom of the SSTR2 binding pocket and form critical contacts with receptor residues; alanine substitution at these positions is therefore expected to be structurally unfavorable. In addition, two prospective candidates (NOV-01 and NOV-02),

Table I: Directional-Consistency Check and Prospective Candidate Examples

ID	Class	Key Features/ Mutation	Expected Direction	System Judgement	Mean ΔG(kcal/mol)	Stdev (σ)	Delta vs WT (d ΔG)	Match
LIT-01	WT SST14	Endogenous SSTR2 ligand	Favorable	Upper Tier	**-43.78**	**15.01**	-	--
LIT-02	Octreotide	SST14-derived 8-mer; SSTR2-selective clinical drug	Favorable	Upper Tier	-42.11	8.10	+1.67	Y
LIT-03	CST-14	Natural SSTR2 agonist (14-mer)	Favorable	Upper Tier	-37.30	7.57	+6.47	N*
SAN-01	Virtual mutant	W8A-disrupts key Trp8-pocket contact	Unfavorable	Lower Tier	-38.22	9.64	+5.56	Y
SAN-02	Virtual mutant	K9A-disrupts Lys9-D122/Q126/Y302 interactions	Unfavorable	Lower Tier	-39.53	13.38	+4.25	Y
NOV-01	**SST14 Mutant**	**Pipeline-designed**	-	**Upper Tier**	**-43.92**	**13.62**	**-0.15**	-
NOV-02	**SST14 Mutant**	**Pipeline-designed**	-	**Upper Tier**	**-41.47**	**3.53**	**+2.31**	-

* *Proline-induced backbone rigidity penalized by FlexPepDock*

proposed by the LLM Planner across two successive iterations, were included to demonstrate the system's automated candidate-generation capability.

The negative controls behaved as expected: both SAN-01 (dΔG = +5.56 kcal/mol) and SAN-02 (dΔG = +4.25 kcal/mol) exhibited consistently weaker binding relative to WT, confirming the pipeline's ability to discriminate structurally disruptive mutations at known hotspot positions. The positive control LIT-02 (dΔG = +1.67 kcal/mol) was ranked slightly below WT but clearly above the negative controls, maintaining directional consistency.

LIT-03 (CST-14) presented a notable discrepancy: despite being a confirmed nanomolar SSTR2 agonist in vitro [7], it scored dΔG = +6.47 kcal/mol, comparable to the negative controls. This discrepancy is attributable to the G2→P substitution in LIT-03, which introduces proline-induced backbone rigidity that is penalized under the flexible-backbone refinement protocol of FlexPepDock. This case illustrates a known simulator limitation rather than a failure in the ranking logic of the system, and it highlights the importance of interpreting computational rankings in conjunction with structural context.

Among the prospective candidates, NOV-01 achieved a top-3 mean ΔG of −43.92 kcal/mol (dΔG = −0.15), nearly matching WT despite introducing three mutations (A1Y, G2S, S13N) at non-hotspot positions. This result suggests that the Planner correctly identified mutation-tolerant sites while preserving the critical Trp8–Lys9 binding motif. NOV-02 exhibited the lowest standard deviation among all candidates (σ = 3.53 kcal/mol), indicating a highly reproducible binding pose, a property desirable for downstream experimental prioritization. These results demonstrate that the Planner can incorporate prior-iteration feedback to generate structurally plausible novel candidates and that the pipeline can evaluate them in a consistent manner.

Overall, the directional-consistency check confirmed that the relative rankings produced by the pipeline are structurally plausible for cases where the simulator's assumptions hold (i.e., flexible backbone without proline constraints). The identified limitation with proline-containing peptides provides a concrete criterion for future QC-gate refinement.

3. Conclusions

In this work, an Agentic AI-based multi-agent system was designed and implemented to automate the iterative workflow of SSTR2-binding peptide candidate screening. The proposed system integrates candidate generation–simulation–QC–critique–reporting into an iterative loop through the Planner–QCRanker–DiversityManager–Critic–Reporter architecture and provides a feedback structure in which inter-iteration results are reflected in subsequent planning.

A directional-consistency check using FlexPepDock binding-energy evaluation confirmed that the pipeline correctly ranks known hotspot-disrupting mutations (SAN-01, SAN-02) below LIT-01 and established analogues, while a known simulator limitation with proline-containing peptides LIT-03 was identified and documented. The LLM Planner-driven prospective workflow demonstrated that the system can propose novel SST14 variants that preserve the critical binding motif and achieve WT-level binding energy (NOV-01, $d\Delta G = -0.15$ kcal/mol) or exhibit highly reproducible binding poses (NOV-02, $\sigma = 3.53$ kcal/mol).

This study finds its significance not in absolute prediction-accuracy verification but in the design, implementation, and operational demonstration of an automated workflow that supports candidate prioritization and subsequent experimental candidate derivation at the pre-wet-lab stage. In future work, we plan to extend the system into a closed-loop optimization framework by incorporating wet-lab validation results into the Planner and Critic feedback loops, and to refine the QC gates to account for simulator-specific limitations such as proline backbone rigidity.

ACKNOWLEDGEMENTS

This work was supported in part by Korea Atomic Energy Research Institute R&D Program under Grant KAERI-526140-26, and in part by the ZODIAC project of the International Atomic Energy Agency (IAEA) in the Asia and the Pacific region (CRP D32039, contract number: 26513).

REFERENCES

[1] Yang, An, et al. *Qwen3 technical report.* arXiv preprint arXiv:2505.09388 (2025).
[2] S. Chaudhury, S. Lyskov, and J. J. Gray, *PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta,* Bioinformatics, Vol. 26, No. 5, pp. 689–691, 2010.
[3] B. Raveh, N. London, L. Zimmerman, and O. Schueler-Furman, *Rosetta FlexPepDock ab-initio: simultaneous folding, docking and refinement of peptides onto their receptors*, PLoS ONE, Vol. 6, No. 4, e18934, 2011.
[4] Gilchrist, Cameron L.M., Milot Mirdita, and Martin Steinegger. *Multiple protein structure alignment at scale with FoldMason.* Science 391.6784 (2026): 485-488.
[5] Y. Heo, E. Yoon, Y.-E. Jeon, J.-H. Yun, N. Ishimoto, H. Woo, S.-Y. Park, J.-J. Song, and W. Lee, *Cryo-EM structure of the human somatostatin receptor 2 complex with its agonist somatostatin delineates the ligand-binding specificity,* eLife, Vol. 11, e76823, 2022.
[6] Q. Bo, F. Yang, Y. Li, X. Meng, H. Zhang, Y. Zhou, S. Ling, D. Sun, P. Lv, L. Liu, P. Shi, and C. Tian, *Structural insights into the activation of somatostatin receptor 2 by cyclic SST analogues,* Cell Discovery, Vol. 8, 47, 2022.
[7] L. de Lecea et al., *A cortical neuropeptide with neuronal depressant and sleep-modulating properties,* Nature, Vol. 381, pp. 242–245, 1996.