AI accelerated discovery of self-assembling peptides

Yejiao Shi[1,], Honggang Hu[2, *]*

Self-assembly is a process capitalized by nature for converting chemically simple building blocks into hierarchically ordered structures and materials to function in living systems cooperatively.[1] Among all the building blocks, proteins are the most commonly utilized and form diverse structures ranging from oligomers and nanospheres to tubes and hierarchical architectures.[2] These structures with unique physical properties play crucial roles in vast biological functions, for instance, structural support, cargo transport, and microbial defence.[3] The formation of these structures is predominantly driven and governed by non-covalent interactions, conferring the dynamism and flexibility of these structures. By studying the fundamental principles that control the self-assembly processes of native proteins, functional materials can be replicated in a bottom-up approach, whereby synthetic building blocks are designed for assembling into desired architectures with specific properties.[4] In this context, biomimetic peptides with comparatively shorter sequences than proteins emerge as ideal building blocks.[5] They provide a versatile approach for mimicking complex structures and materials with structural and functional diversity in a more tractable but less expensive approach.

The self-assembly behaviours of peptides depend strongly on their amino acid sequences.[6] To date, the design of self-assembling peptides still relies on the examination of natural amino acid sequences, professional expertise in the peptide field, or laboratory discoveries by serendipity. However, these inefficient approaches fail to meet the growing demands for functional peptide materials since the design range of self-assembling peptides can be extremely broad. The amount of possible amino acid combinations for a peptide can reach up to 20^n, where 20 is the number of commonly available amino acids and n is the amino acid number in the peptide.[7] For instance, tripeptides ($n = 3$) containing 8000 combinations are intractable for any rigorous experimental study. Even though a brute-force computational search based on coarse-gained molecular dynamic (MD) simulations has been proposed by Ulijn's team to overcome the search bias and has successfully identified several self-assembling tripeptides.[8] Due to the high computational costs, it is impossible to extend it to longer sequence lengths ($n > 3$). Accurate prediction of the assembling processes and discovery of the assembling peptides remain challenging.

Artificial intelligence (AI) is an emerging inter-discipline that integrates computer science, mathematics, and psychology among others to emulate human cognitive functions for tasks requiring intelligence, such as language understanding and decision making.[9] It holds great potential to navigate through the vast search space among amino acid combinations and present a subset displaying the most promising possibilities. Recently, writing in Nature Chemistry, Sankaranarayanan's team introduced an "AI expert" that combined the Monte Carlo tree search (MCTS) algorithm with coarse-gained MD simulations for identifying self-assembling peptides that exhibit high aggregation propensities in aqueous solution.[10]

Performing in an autonomous way, the AI expert first utilized MCTS to generate peptide sequences before utilising MD simulations to estimate the aggregation propensity of the generated sequences and provide feedback to improve the quality of MCTS and guide future searches (**Figure 1**). Notably, an innovative concept of uniqueness function within the MCTS objective function, together with a random forest based surrogate model were included to boost the performance of the MCST algorithm. Compared to MCTS without the random forest model, random search, and brute-force search, the AI expert enabled by MCTS and random forest scheme was demonstrated to be the most efficient in identifying tripeptide with the highest-scoring.

1 Institute of Translational Medicine, Shanghai University, Shanghai, China;
2 School of Medicine, Shanghai University, Shanghai, China
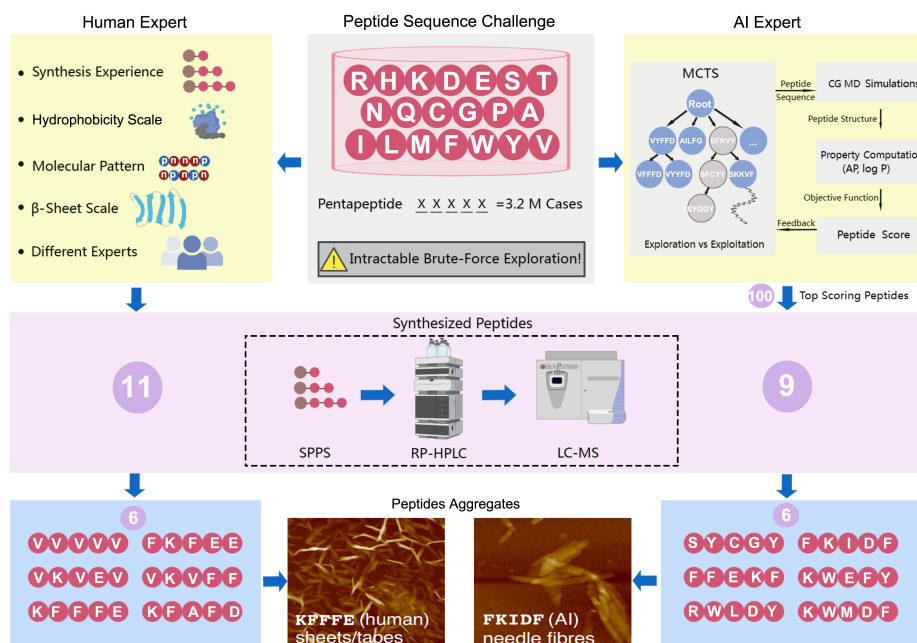
*Corresponding authors:
Yejiao Shi,
sh10yj@shu.edu.cn;
Honggang Hu,
hhu66@shu.edu.cn.

How to cite this article:
Shi, Y.; Hu, H. AI accelerated discovery of self-assembling peptides. *Biomater Transl.* **2023**, *4*(4), 291-293.

Figure 1. Comparison of the workflow using the input of self-assembling pentapeptides from AI expert and human expert. The search range for peptides increases dramatically due to the multiple combinations of 20 amino acids, and 3.2 million pentapeptides are unmanageable to be calculated with the brute-force method. Human expert adopted rationally design methods such as hydrophobic scales, molecular patterns (npnpn), and personal experience. Six of the eleven pentapeptides proposed by six human experts were synthesised and observed to be clustered together. In contrast, AI expert developed a combination of MCTS. Six of the nine pentapeptides proposed by the AI expert were synthesised and found to be clustered. AI expert has also proposed novel sequences involving multiple amino acids to recover some intuitive sequences, reflecting its advantages in overcoming human bias. Atomic force microscope images of some promising pentapeptides from both groups are shown. Created with MedPeer (www.medpeer.cn). AI: artificial intelligence; AP: aggregation propensity; CG: coarse-gained; LC-MS: liquid chromatography-mass spectrometry; MCTS: Monte Carlo tree searches; MD: molecular dynamic; n: non-polar; p: polarity; RP-HPLC: reverse phase high performance liquid chromatography; SPPS: solid phase peptide synthesis.

Having validated the efficiency of AI experts, Sankaranarayanan and colleagues went on using it to discover self-assembling pentapeptides. Among the 3.2 million (20^5) pentapeptide permutations, the AI expert first evaluated approximately 6600 cases based on MD simulations and then selected the top 100 candidates for further MD simulations using more rigorous parameters and longer timescales to improve the estimates of aggregation propensity. Nine candidates with top scores were filtered out and six of them were observed to aggregate according to experimental dynamic light scattering and atomic force microscope measurements. Comparatively, six out of eleven pentapeptides proposed by human experts were found to aggregate. By analysing the sequence similarity of identified peptides, the AI expert was found could not only recover known sequences similar to the human experts but also discover unknown sequences that deviate notably from the existing ones, indicating its potential to overcome the bias from humans and accelerate the discovery of peptides.

Integrated with coarse-grained MD, machine learning, and experiments, the AI expert was demonstrated to be an efficient "human-in-the-loop" framework for the discovery of self-assembling peptides. Future efforts were anticipated to be made to connect the developed AI expert to a robotic platform that can synthesise and characterise innovative peptides. As such experimental feedback could be digested by AI expert directly, supporting the search to progress in an iterative manner. Additional information obtained from the simulations, for instance, the aspect ratio and number of the peptides, their morphologies, and their moments of inertia were expected to be included to improve the MCTS scoring function. Similar AI strategies could also be developed for screening and discovering more functional peptides and peptide assemblies, providing great potential for further innovations in peptide-based novel therapeutics and functional materials.

Conflicts of interest statement
The authors declare no conflicts of interest.

1. Whitesides, G. M.; Grzybowski, B. Self-assembly at all scales. *Science.* **2002**, *295*, 2418-2421.

2. Luo, Q.; Hou, C.; Bai, Y.; Wang, R.; Liu, J. Protein assembly: versatile approaches to construct highly ordered nanostructures. *Chem Rev.* **2016**, *116*, 13571-13632.

3. Desai, M. S.; Lee, S. W. Protein-based functional nanomaterial design for bioengineering applications. *Wiley Interdiscip Rev Nanomed Nanobiotechnol.* **2015**, *7*, 69-97.

4. Zhang, S. Fabrication of novel biomaterials through molecular self-assembly. *Nat Biotechnol.* **2003**, *21*, 1171-1178.

5. Pearce, A. K.; Wilks, T. R.; Arno, M. C.; O'Reilly, R. K. Synthesis and applications of anisotropic nanoparticles with precisely defined dimensions. *Nat Rev Chem.* **2021**, *5*, 21-45.

6. Mendes, A. C.; Baran, E. T.; Reis, R. L.; Azevedo, H. S. Self-assembly in nature: using the principles of nature to create complex nanobiomaterials. *Wiley Interdiscip Rev Nanomed Nanobiotechnol.* **2013**, *5*, 582-612.

7. Ulijn, R. V.; Smith, A. M. Designing peptide based nanomaterials. *Chem Soc Rev.* **2008**, *37*, 664-675.

8. Frederix, P. W.; Scott, G. G.; Abul-Haija, Y. M.; Kalafatovic, D.; Pappas, C. G.; Javid, N.; Hunt, N. T.; Ulijn, R. V.; Tuttle, T. Exploring the sequence space for (tri-)peptide self-assembly to design and discover new hydrogels. *Nat Chem.* **2015**, *7*, 30-37.

9. Hamet, P.; Tremblay, J. Artificial intelligence in medicine. *Metabolism.* **2017**, *69s*, S36-S40.

10. Batra, R.; Loeffler, T. D.; Chan, H.; Srinivasan, S.; Cui, H.; Korendovych, I. V.; Nanda, V.; Palmer, L. C.; Solomon, L. A.; Fry, H. C.; Sankaranarayanan, S. Machine learning overcomes human bias in the discovery of self-assembling peptides. *Nat Chem.* **2022**, *14*, 1427-1435.