Scoring the Synthesizability of Any Novel Non-natural Amino Acid

Gökçe Geylan[ab*], Mikhail Kabeshov[a], Samuel Genheden[a], Christos Kannas[a], Thierry Kogej[a], Ola Engkvist[ac], Leonardo De Maria[d]

a. Molecular AI, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden.
b. Division of Systems and Synthetic Biology, Department of Life Sciences, Chalmers University of Technology, Gothenburg, Sweden.
c. Department of Computer Science and Engineering, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden
d. Medicinal Chemistry, Research and Early Development, Respiratory & Immunology, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

* Email: gokce.geylan@astrazeneca.com

https://doi.org/10.17952/37EPS.2024.P2276

Abstract

The introduction of solid-phase peptide synthesis (SPPS) has transformed peptide design and synthesis. While nowadays dozens of non-natural amino acids (NNAAs) are available from vendors to be used in peptide synthesizers, the NNAA chemical space is vast and still unexplored. Enabling any in-silico screened NNAAs for solid-phase synthesis requires to: i) introduce the right protecting groups and ii) synthesize them in adequate amounts. In this work, we introduce tools for cheminformatically protect any novel amino acid to be used in solid-phase synthesis and to score the synthesizability of the so obtained molecule. Our tool not only offers a synthesizability retrosynthesis score of the NNAAs but also can be used for ranking of the potential protection strategies, thus increasing the synthetic feasibility.

1. Introduction

While dozens of NNAAs are currently available for peptide synthesizers, there is still a vast unexplored chemical space awaiting discovery. We have recently shown how this space can be significantly expanded, enumerating ~400000 novel alpha NNAAs and using 10,000 representatives to predict improved peptide binding affinities and solubility (Fig 1).[1,2] In this work we show how to add synthesizability considerations in our computational workflow.

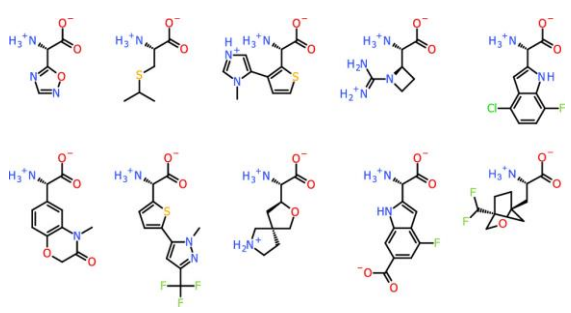

Figure 1. Examples of novel non-natural amino acids from our virtual library[2]

Computational protocols work on un-protected amino-acids and incorporating a novel NNAA on a peptide requires the synthesis of the SPPS specifically protected compound. The protection strategy requires careful consideration of the protecting groups for sidechains and backbones. Here, we introduce a tool for evaluating the synthetic feasibility of any protected non-natural amino acid, after finding the optimal protection strategy. The proposed synthesizability score can be used as additional in-silico parameter for compound selection.

2. Methods

I. Generate Protected Compounds. A cheminformatics tool was built to obtain multiple protected versions of each NNAA, suitable for SPPS.[3]

II. Retrosynthesis scoring. We used our single-step retrosynthesis model (AiZynthFinder[4]) trained on USPTO data and the stocks for eMolecules[5]. The generated routes were scored by accumulating scores for the reactions in the route and were filtered with the feasibility predicted by a forward Chemformer model[6]. Finally, the routes were scored with a rank score for reaction class based on historical data, referred as the Synthetic Feasibility (SF) score.

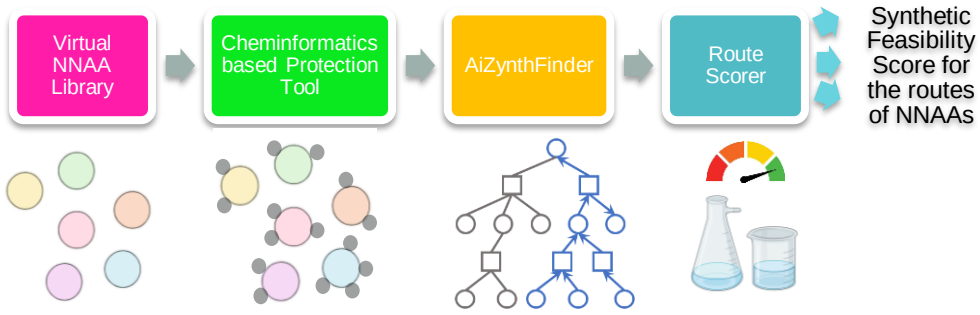

Figure 2. Schematic representation of the scoring workflow

3. Results

A. Retrosynthesis Planning for a novel NNAA

Problem: A peptide designed *in-silico* included a novel NNAA. Next, the NNAA must be synthesized and protected to be used in SPPS for experimental testing the predicted improved peptide.

Solution: Our tool enables ranking and choosing the most feasible protection strategy for a novel NNAA upon retrosynthesis planning.

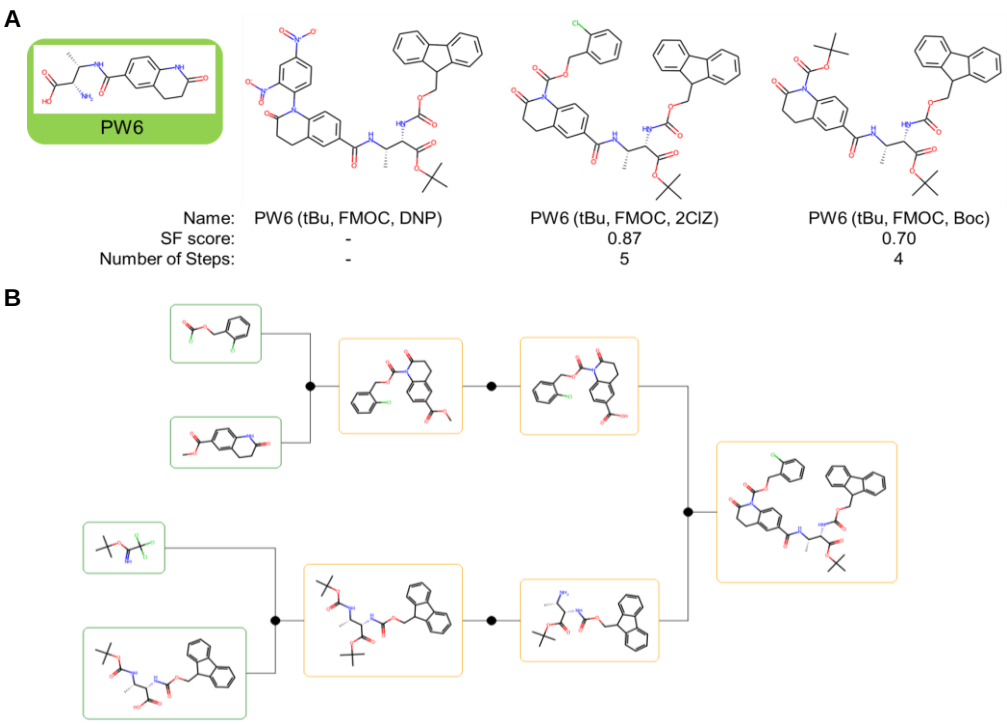

Figure 4. Route Enumeration & Scoring
A) The protected series for amino acid "PW6" and B) the most synthesizable route of the protected PW6

B. Application to Virtual Screening

Problem: Adding a synthesizability score to binding-affinity predictions for compound selection. We used as model system the N-acetylated 9-mer peptide, $^{(76)}$LDEETGEFL$^{(84)}$ and targeted the Glu78 position to find novel NNAAs leading to improve its binding affinity to KEAP1. A site-specific mutation docking protocol was used for the 10,000 representative novel a-NNAAs[2].

Outcome: Our tool facilitates synthesizability-aware compound selection.

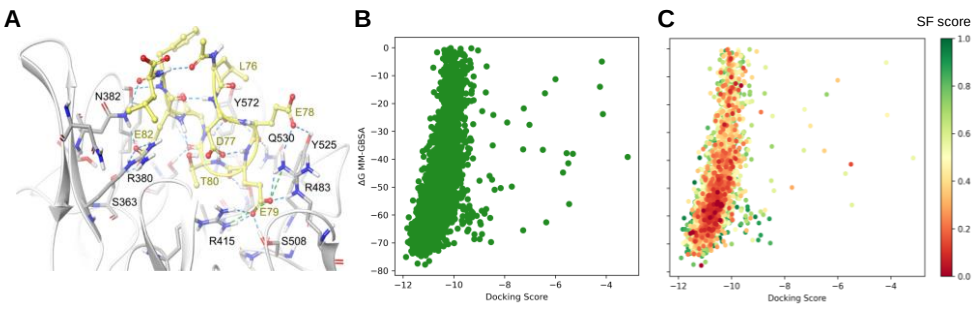

Figure 3. Virtual Screening plus Synthetic Feasibility (SF) Score
The docking of the NNAA library to position Glu78 in A) Keap1 in Keap1−Neh2 binding site structure, adapted from Amarasinghe et al[2]. The plots showing B) the docking scores for NNAAs, C) the docking scores colored by the SF score.

Conclusions & Future Directions

- Our synthesizability tool:
  - Creates multiple protected versions of a given non-natural amino acid
  - Considers SPPS specifical protection groups
  - Scores the synthesizability of each route
  - Facilitates the evaluation of retrosynthesis planning
- Our tool can be used to:
  - Rank the synthetic feasibility of all the protection strategies for a given novel NNAA
  - In addition to other peptide optimization scoring functions

We are conducting validation studies on the synthesis and characterization of some of the non-natural amino acids for the KEAP1-NEH2 model system.

References

1. Oeller et al (2023) *Sequence-based prediction of the intrinsic solubility of peptides containing non-natural amino acids* Nat. Comm. **14**, 1. DOI: 10.1038/s41467-023-42940-w
2. Amarasinghe et al (2022) *Virtual Screening Expands the Non-Natural Amino Acid Palette for Peptide Optimization* J. Chem. Inf. Mod. **62,** 2999. DOI: 10.1021/acs.jcim.2c00193
3. Wuts & Greene (2006) *Greene's Protective Groups in Organic Synthesis* DOI: 10.1002/0470053488
4. Saigiridharan et al (2024) AiZynthFinder 4.0: developments based on learnings from 3 years of industrial application. J. Cheminf. **16,** 1. DOI: 10.1186/s13321-024-00860-x
5. eMolecules (https://www.emolecules.com)
6. Westerlund et al (2024) Do Chemformers Dream of Organic Matter? Evaluating a Transformer Model for Multistep Retrosynthesis. J Chem. Inf. Model. **64**, 3021. DOI: 10.1021/acs.jcim.3c01685

Supported by AstraZeneca

Presented at the 37th European Peptide Symposium and 14th International Peptide Symposium, Florence, Italy 25-29 August 2024