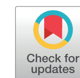
Tetrahedron 74 (2018) 2291–2297

Contents lists available at ScienceDirect

Tetrahedron

journal homepage: www.elsevier.com/locate/tet

ProteoFind: A script for finding proteins that are suitable for chemical synthesis



Akira Shigenaga*, Naoto Naruse, Akira Otaka

Institute of Biomedical Sciences and Graduate School of Pharmaceutical Sciences, Tokushima University, Shomachi, Tokushima, 770-8505, Japan

A R T I C L E I N F O

Article history:
Received 21 February 2018
Accepted 12 March 2018
Available online 15 March 2018

Keywords:
Chemical protein synthesis
Fragment ligation
Native chemical ligation
Python
Script

A B S T R A C T

ProteoFind is a computational script for finding proteins that are suitable for fragment ligation-based chemical synthesis. This paper describes the development and case studies of ProteoFind, which searches protein lists obtained from the UniProt website. Its application to visualize areas covered by several one-pot three-fragment ligation methods is also discussed. The results demonstrate that ProteoFind not only saves time when searching for synthetic target proteins, but also proposes many candidate proteins from among which biomedically interesting proteins could be found. It also enables clarification of the features of ligation methods by comparing the areas to which each ligation reaction is accessible.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Modern chemical synthesis enables the preparation of proteins from multiple peptide fragments via a fragment ligation method.[1] The most widespread ligation method is native chemical ligation (NCL) developed by Kent and coworkers.[2] NCL allows chemoselective ligation of a peptide fragment with a C-terminal thioester and another fragment with an N-terminal cysteine to generate a native protein. Inspired by NCL, intensive efforts have been made to expand the scope of NCL and develop new ligation methods.[1a,b] Four topics that have been of major focus in this research field and that are closely related to the present study are as follows: (1) Cysteine-free NCL-like reaction: One of the limitations of the original NCL is the requirement for cysteine at the ligation site. Representative methods to overcome this limitation utilize a C^{β}-mercaptoamino acid derivative[1c,3] or an amino acid with a thiol-containing auxiliary[1c,4] instead of the cysteine. (2) Preparation of peptide thioesters by 9-fluorenylmethoxycarbonyl (Fmoc)-based solid-phase peptide synthesis (Fmoc SPPS): Because of the high electrophilicity of a thioester, standard Fmoc SPPS that routinely utilizes nucleophilic piperidine for removal of an Fmoc group is not compatible with preparation of the peptide thioesters. Many auxiliaries, which enable conversion of an Fmoc SPPS-tolerant C-terminal amide or ester to the thioester after completion of peptide elongation, have therefore been developed.[5] (3) One-pot multi-fragment ligation: Preparation of large proteins requires ligation of more than two peptide fragments. Purification by high-performance liquid chromatography (HPLC) after each ligation step could be a reason for lowering the total yield of the proteins and wasting time. Therefore, one-pot multi-fragment ligation methods have been developed to reduce the number of the HPLC purification steps.[6,7] (4) New ligation methods: There are fragment ligation methods other than NCL and NCL-like reactions.[1a,b] These methods, involving an old thioester method,[8] or recently developed Ser/Thr ligation[7g] or KAHA ligation,[9] can utilize amino acids that differ from those of the original NCL at the ligation site.

If the new methods were developed in the fields mentioned above, they could often be applied to the synthesis of proteins to demonstrate their utility. The synthetic target protein should therefore satisfy the requirements of the new method. Examples of the requirements corresponding to categories (1) to (4) described above are as follows: (1) Cysteine-free NCL-like reaction: At the ligation site, the target protein should possess a certain amino acid instead of cysteine. (2) Preparation of peptide thioesters by Fmoc SPPS: The auxiliaries usually utilize N- or O-S acyl transfer reaction.[5] Therefore, the amino acid on the auxiliaries should be carefully chosen to prevent side reactions such as epimerization/racemization. (3) One-pot multi-fragment ligation: In the case of one-pot multi-fragment ligation utilizing the difference of reactivity of amino acids at the ligation site, there is sometimes a limitation of the amino acids (details are described below). (4) New ligation methods: Amino acids at the ligation site are sometimes

* Corresponding author.
E-mail address: shigenaga.akira@tokushima-u.ac.jp (A. Shigenaga).

https://doi.org/10.1016/j.tet.2018.03.030
0040-4020/© 2018 Elsevier Ltd. All rights reserved.

fixed (e.g. Ser/Thr ligation[7i]) or restricted to achieve a high ligation yield. Furthermore, in all cases, the number (e.g., 2 for standard ligation, 3 for one-pot three-fragment ligation) and length of the peptide fragments constituting the synthetic target protein should be appropriate.

To find synthetic target proteins that satisfy all of the criteria of a reaction that they have developed, many protein chemists survey easily accessible information such as papers in their reference list and catalogs of suppliers. However, this is time-consuming and sometimes results in no appropriate protein being found (in this case, a biologically unnecessary mutation is introduced to satisfy the reaction's criteria). Furthermore, proteins presented in catalogs are already commercially available, so there is no need to synthesize them by new methods. We actually prepared an α-conotoxin ImI derivative with a biologically unnecessary mutation and commercially available human atrial natriuretic peptide for the demonstration of one-pot four- and three-fragment ligation, respectively; unfortunately, the obtained proteins could not contribute to further biological and medicinal studies.[11] Taking this background in account, we thought that a computational method of searching for synthetic target proteins from among a large amount of protein data would be indispensable for protein chemists because it would not only save time in the search, but also propose many candidate proteins from among which biomedically interesting proteins might be found. Furthermore, it could reveal areas for which chemical protein synthesis employing the new method would be applicable.[11]

This paper first describes the development of ProteoFind, which is a computational script for finding synthetic target proteins from within a publicly accessible protein list.[12] It then discloses the results of case studies of ProteoFind. Finally, it discusses visualization of the difference of the areas covered by three one-pot three-fragment ligation methods.

2. Results and discussion

2.1. Design of ProteoFind

An overview of the workflow of ProteoFind is shown in Fig. 1. Users are first required to input search criteria: number of fragments of which the synthetic target proteins consist (**fr**), maximum and minimum lengths of the fragments (**max** and **min**), and amino acids at the ligation sites (AA^1/AA^2, AA^3/AA^4 …). Then ProteoFind checks each protein in a protein list. The length of the protein is initially reviewed [(1) in Fig. 1]. If the length is more than or equal to (**fr** × **min**) and less than or equal to (**fr** × **max**), the number of AA^1-AA^2 sequences in a region (**min** − 1 to **max** + 1) is counted [(2) in Fig. 1]. If this number is 0, the protein checked is not appropriate as a synthetic target. If number is 1, the protein is split between AA^1 and AA^2, and the remaining C-terminal fragment containing AA^2 is used for further processing [(3) and (4) on the left side of Fig. 1]. When the number of AA^1-AA^2 sequences is more than 1, processes

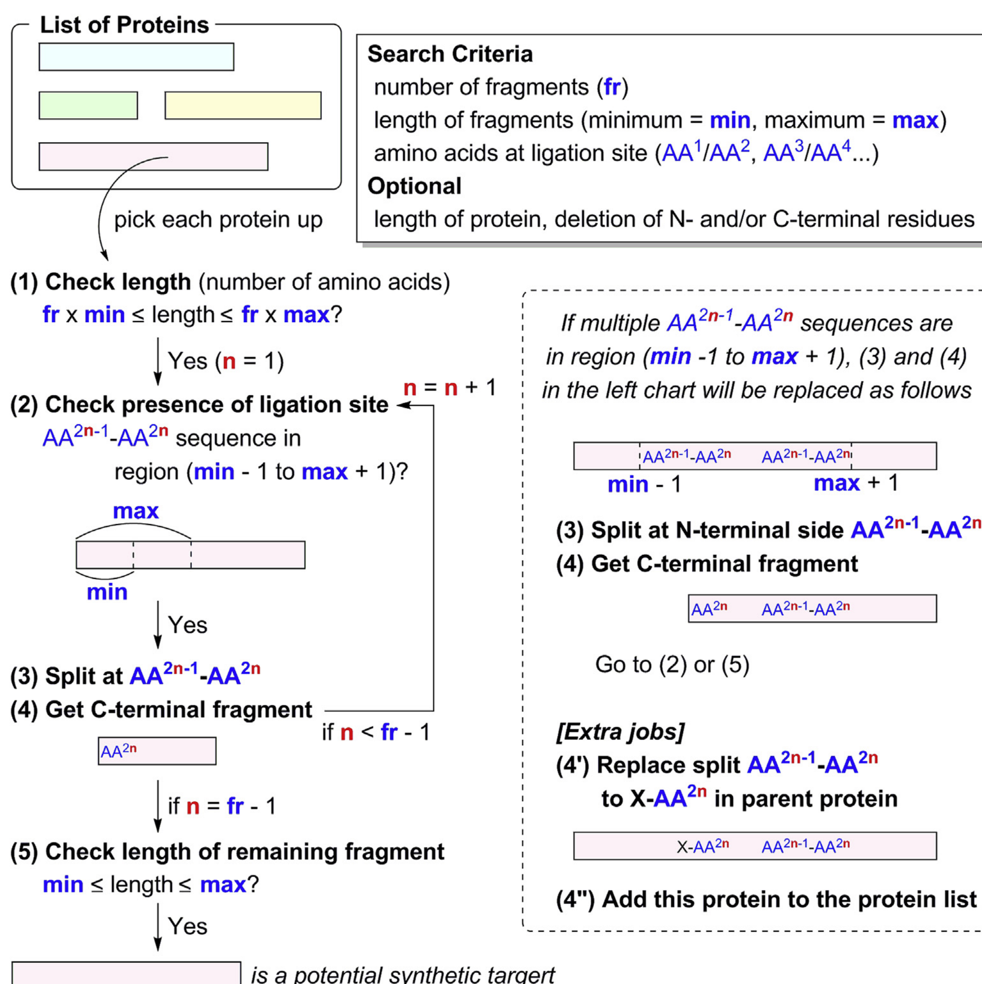


Fig. 1. A schematic diagram of workflow of ProteoFind. Colored squares are proteins or peptides. AA means an amino acid.

Download English Version:

https://daneshyari.com/en/article/7827219

Download Persian Version:

https://daneshyari.com/article/7827219

Daneshyari.com