Graphical Abstract

To create your abstract, type over the instructions in the template box below.
Fonts or abstract dimensions should not be changed or altered.

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

Akira Shigenaga,* Naoto Naruse and Akira Otaka

Leave this area blank for abstract info.

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

Akira Shigenaga,* Naoto Naruse and Akira Otaka

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

ARTICLE INFO

ABSTRACT

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.

2009 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 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-

∗ Corresponding author. Tel.: +81-88-633-9534; fax: +81-88-633-9505; e-mail: shigenaga.akira@tokushima-u.ac.jp

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.
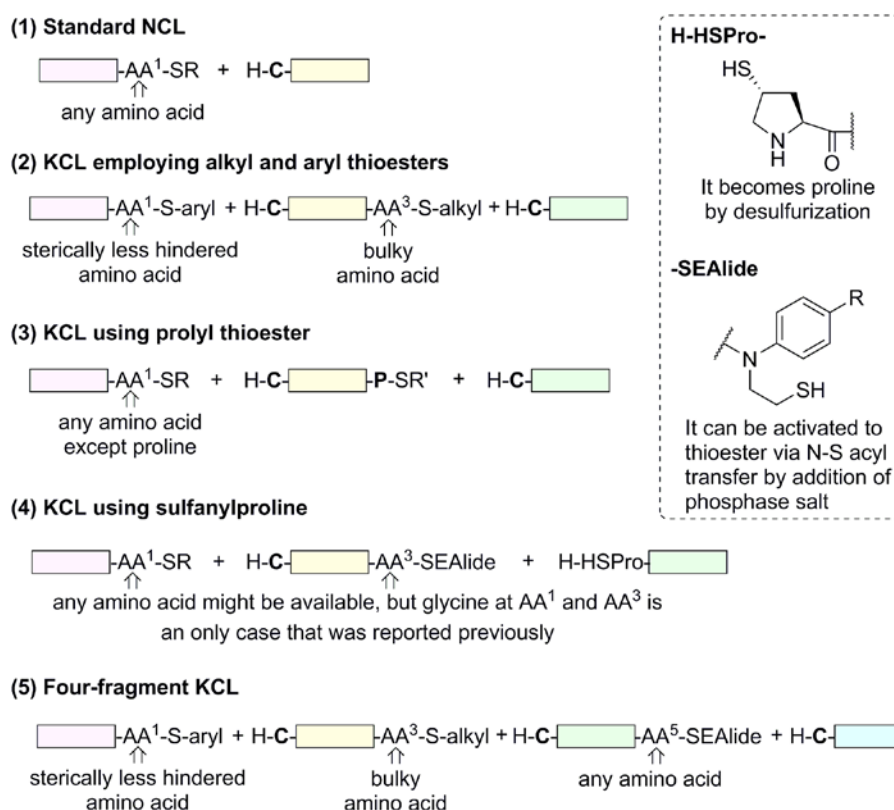
2. Results and discussion

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 ($fr \times min$) and less than or equal to ($fr \times max$), 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 (3) and (4) are replaced by that shown in the dotted square on the right side of the figure. Briefly, the protein is split at the N-terminal side AA^1-AA^2 and the obtained C-terminal fragment is passed to the following processing, similar to that of (3) and (4) on the left side of the figure. Additionally, in this case, the N-terminal side AA^1-AA^2 sequence of the parent protein is replaced with an X-AA^2 (the X is a dummy residue that cannot be assigned to a ligation site), and the obtained protein is added to the protein list to check the other ligation sites [(4') and (4") in Fig. 1]. Steps (2) to (4) are repeated for the obtained C-terminal fragment while **n** [number of steps (2) examined] is less than ($fr - 1$). If **n** reaches ($fr - 1$) and the length of the remaining fragment satisfies a certain criterion [$min \leq$ the length $\leq max$, (5) in Fig. 1], the checked protein could be a synthetic target. ProteoFind applies these processes to all proteins in the list, and finally outputs a catalog of the potential synthetic targets after removal of
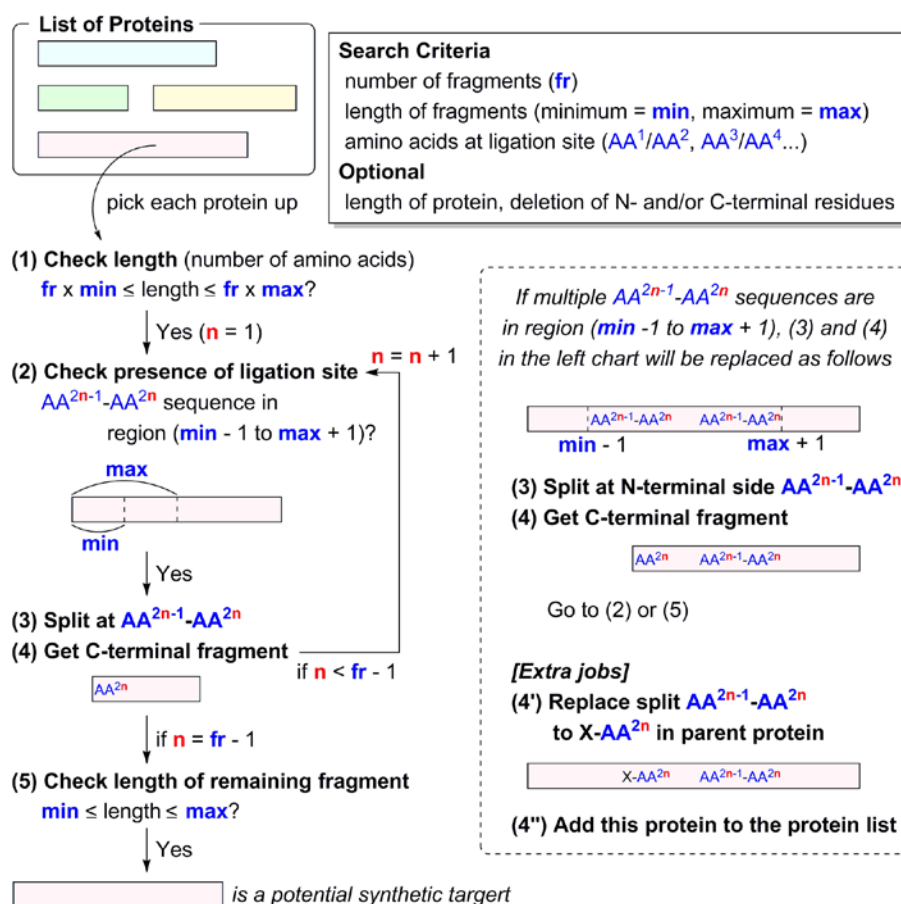
Fig. 1. A schematic diagram of workflow of ProteoFind. Colored squares are proteins or peptides. AA means an amino acid.

(1) Standard NCL

-AA¹-SR + H-C-
any amino acid

(2) KCL employing alkyl and aryl thioesters

-AA¹-S-aryl + H-C- -AA³-S-alkyl + H-C-
sterically less hindered amino acid / bulky amino acid

(3) KCL using prolyl thioester

-AA¹-SR + H-C- -P-SR' + H-C-
any amino acid except proline

(4) KCL using sulfanylproline

-AA¹-SR + H-C- -AA³-SEAlide + H-HSPro-
any amino acid might be available, but glycine at AA¹ and AA³ is an only case that was reported previously

(5) Four-fragment KCL

-AA¹-S-aryl + H-C- -AA³-S-alkyl + H-C- -AA⁵-SEAlide + H-C-
sterically less hindered amino acid / bulky amino acid / any amino acid

H-HSPro-
It becomes proline by desulfurization

-SEAlide
It can be activated to thioester via N-S acyl transfer by addition of phosphase salt

Scheme 1. Ligation reactions examined in the case studies. Colored squares represent peptide fragments.

duplications followed by replacement of the dummy residue X with the original amino acid.

Additionally, two optional functions were implemented in ProteoFind. One is to exclude large proteins by setting an upper limit of the length of the proteins. The other enables deletion of amino acids at the N- and/or C-terminus of the proteins. Many proteins with an additional sequence such as a signal sequence at their terminus are deposited in the protein list; this N- and/or C-terminal deletion mode was thus incorporated to virtually search proteins after post-translational processing.

ProteoFind and related scripts were written in Python because it allows easy customization by users. Protein lists were obtained from the free and publicly accessible UniProt website in FASTA (canonical) format. To accelerate the search process, the obtained lists were converted to a one-line/one-protein format by our script ProteoFind_list_converter before the use of ProteoFind. To save time or to focus on certain species, protein lists of not only all species, but also some individual species are enrolled in the default setting of the ProteoFind [the available lists: all proteins (reviewed) containing 555,100 proteins, human (reviewed) containing 20,244 proteins, human (not reviewed) containing 141,277 proteins, mouse (reviewed) containing 16,909 proteins, rat (reviewed) containing 7,993 proteins, *A. thaliana* (reviewed) containing 15,423 proteins, and *S. cerevisiae* (reviewed) containing 6,721 proteins]. Users can therefore choose an appropriate list suitable for their purpose. Furthermore, an option to search the user's original list is also provided.

To simplify the output file, similar results are omitted when the two amino acids at the ligation site are identical and the region (**min** – 1 to **max** + 1) contains a long repeat of the same amino acid. If multiple amino acids are assigned to one side of the ligation site at once (e.g., AA¹ = alanine or glycine or…), some redundant results, in which the proteins are the same but

the ligation site is slightly different, are sometimes removed for clarity of the output file. In both cases, the number of hits is reduced to prevent redundancy; however, the number of proteins contained in the hits is not affected (examples are shown in the following case studies).

2.2. Case studies

ProteoFind was employed to find the potential synthetic target proteins of the following five ligation reactions (Scheme 1): (1) standard NCL;[2] (2) kinetically controlled ligation (KCL) employing alkyl and aryl thioester fragments developed by Kent's group;[13] (3) KCL using a prolyl thioester developed by us;[14] (4) KCL using sulfanylproline instead of cysteine developed by us;[15,16] and (5) four-fragment KCL based on Kent's KCL and our *N*-sulfanylethylanilide (SEAlide) technology.[10,17] In these case studies, a list of "all proteins (reviewed)" including 555,100 proteins was employed, and minimum and maximum lengths of fragments were fixed to 10 and 50, respectively, because they could be routinely prepared by SPPS. Details are described below and the results are summarized in Table 1.

2.2.1. Case study 1: Standard NCL

First, target proteins that could be synthesized by one-step standard NCL were searched. NCL allows the use of any amino acid including proline[14] at the AA¹ position in (1) of Scheme 1, but its counterpart should be fixed to cysteine. After inputting the amino acids of the ligation site followed by the number of fragments (**fr** = 2) and other search criteria described above, ProteoFind found 32,803 hits within a few minutes by using an acceleration mode (details of the acceleration mode are described in the Experimental section). Part of the obtained output file is depicted in Fig. 2. Following the number of hits and the search criteria, information on potential synthetic targets including a brief explanation that aids finding the protein in the UniProt website, the full sequence, amino acids at the ligation site, and

the fragments for synthesis are listed. The hits can contain the same protein with different ligation sites; therefore, the Python script ProteoFind_analyzer was prepared to count the number of proteins in the hits. Analysis of the output file by the script revealed that the number of proteins potentially accessible by the one-step standard NCL is 13,973.

2.2.2. Case study 2: KCL employing alkyl and aryl thioesters

A search was performed for potential synthetic targets of KCL utilizing the difference of reactivity of alkyl and aryl thioesters[13] [(2) in Scheme 1]. It is reported that a sterically less hindered amino acid and a bulky amino acid at AA[1] and AA[3] positions, respectively, are preferred to achieve high selectivity.[18] Therefore, glycine for AA[1] and isoleucine, leucine, threonine, or valine for

AA[3] were considered. Here, N-terminal amino acids of the second and third fragments were fixed as cysteine. First, the four amino acids at the AA[3] position were inputted at once as "ILTV," and the number of hits was 893 (Table 1). When each amino acid was inputted individually, the sum of the hits was 900. This number is slightly larger than that in the case of inputting "ILTV" at once; however, the number of proteins in the hits is the same as mentioned in section 2.1 (744 proteins in both cases). If detailed information of the available ligation sites is preferred, inputting amino acids one by one might be recommended.

2.2.3. Case study 3: KCL using prolyl thioester

Although it is known that prolyl thioester hardly contributes to NCL under standard NCL conditions,[19] we clarified that it can

Table 1. Summary of the case studies[a]

Reaction[b]	fr[c]	ligation site	number of hits	number of proteins in hits[d]
Standard NCL	2	any/C	32,803	13,973
KCL employing alkyl and aryl thioesters	3	G/C I, L, T or V[e]/C	893	744
KCL using prolyl thioester	3	other than P[f]/C P/C	2,591	1,053
KCL using sulfanylproline	3	G/C G/P	299	253
four-fragment KCL	4	G/C I, L, T or V[e]/C any/C	1,838	631
one-pot ligations listed above	-	-	-	14,953

[a] The protein list "all proteins (reviewed)" consisting of 555,100 proteins was searched. Minimum and maximum lengths of the fragments were fixed as 10 and 50, respectively.

[b] The reaction names correspond to those in Scheme 1.

[c] Number of fragments.

[d] The hits can contain the same protein with different ligation sites; therefore, the number of proteins after the removal of duplications was estimated by using ProteoFind_analyzer.

[e] "ILTV" was inputted at the AA[3] position.

[f] "ACDEFGHIKLMNQRSTVWY" was inputted at the AA[1] position.

```
32803 hit(s)/555100 proteins!

================ Search Criteria =================
data base: all proteins (reviewed)
number of fragments: 2
minimum length of fragments: 10
maximum length of fragments: 50
list of ligation site: ['any', 'C']
N-terminal deletion: 0 residue(s)
C-terminal deletion: 0 residue(s)
==============================================
>sp|Q9W7K1|3S32_PSETE Short neurotoxin 2 OS=Pseudonaja textilis PE=1 SV=1
MKTLLLTLVMVTIMCLDLGYTLTCYKGYHDTVVCKPHETICYRYLIPATHGNAIPARGCGTSCPGGNHPVCCSTDLCNK (1-79)
[Ligation site]
IC
[Fragments]
MKTLLLTLVMVTIMCLDLGYTLTCYKGYHDTVVCKPHETI (1-40)
CYRYLIPATHGNAIPARGCGTSCPGGNHPVCCSTDLCNK (41-79)
==============================================
>sp|Q9W7K1|3S32_PSETE Short neurotoxin 2 OS=Pseudonaja textilis PE=1 SV=1
MKTLLLTLVMVTIMCLDLGYTLTCYKGYHDTVVCKPHETICYRYLIPATHGNAIPARGCGTSCPGGNHPVCCSTDLCNK (1-79)
[Ligation site]
VC
[Fragments]
MKTLLLTLVMVTIMCLDLGYTLTCYKGYHDTVV (1-33)
CKPHETICYRYLIPATHGNAIPARGCGTSCPGGNHPVCCSTDLCNK (34-79)
==============================================
......
```

Fig. 2. An example of an output file of ProteoFind.

participate in NCL by raising the reaction temperature to 50 °C.[14] NCL of prolyl thioester was then applied to one-pot three-fragment ligation [(3) in Scheme 1]. In this reaction, the first and second fragments should contain a C-terminal thioester of an amino acid other than proline and N-terminal cysteine, respectively, to carry out the ligation under standard NCL conditions. The second ligation site is located between proline and cysteine. Therefore, the first NCL is conducted under standard NCL conditions, and then the third fragment is added and the reaction mixture is heated to 50 °C to start the second NCL. The ProteoFind search employing "ACDEFGHIKLMNQRSTVWY, C, P, C" as amino acids at the ligation site suggested 2,591 hits (Table 1). The search employing not "ACDEFGHIKLMNQRSTVWY" but the individual amino acids was then examined, and the sum of the hits was 2,636. In both cases, the number of proteins in the hits was 1,053 whether or not amino acids at the AA^1 position were inputted at once, similar to that described in the above section.

2.2.4. Case study 4: KCL using sulfanylproline

We previously reported one-pot three-fragment ligation utilizing sulfanylproline, as shown in Scheme 1 (4).[15] This reaction employs two technologies. One is sulfanylproline that can participate in an NCL-like reaction as a cysteine surrogate and then be converted to proline by desulfurization.[15,16] The other is SEAlide,[17] which can be activated from an amide to a thioester by the addition of phosphate salt. In this one-pot ligation, any amino acid might be available at the AA^1 and AA^3 positions; however, glycine at these positions is the only case that has been reported so far.[15] Therefore, amino acids at the ligation sites were fixed as glycine/cysteine and glycine/proline for the ProteoFind search, and 253 proteins in 299 hits were found as potential synthetic targets (Table 1).

2.2.5. Case study 5: Four-fragment KCL

The combination of our SEAlide technology[17] and Kent's KCL[13] enables one-pot four-fragment ligation, as shown in Scheme 1 (5). Some amino acids at the ligation sites are restricted as mentioned above (Table 1). Whereas we previously could not find appropriate synthetic target proteins and therefore synthesized a protein with a biologically unnecessary mutation, ProteoFind gave a list of potential synthetic targets of the one-pot four-fragment ligation involving 631 proteins within a few minutes. Finally, the number of synthetically accessible proteins by the one-pot ligations discussed in the case studies was estimated as 14,953 by using ProteoFind_analyzer.

2.3. Analysis of areas covered by three one-pot three-fragment ligation reactions

Analyses of the output files of ProteoFind by using ProteoFind_analyzer can elucidate differences of areas covered by the ligation reactions. In this paper, the difference of three one-pot three-fragment ligations shown in Scheme 1 was estimated (A: KCL employing alkyl and aryl thioesters; B: KCL using prolyl thioester; C: KCL using sulfanylproline). By using ProteoFind_analyzer, the number of proteins was counted using a Venn diagram mode and is summarized as follows (area, number of proteins there): **A**, 744; **B**, 1053; **C**, 253; $A \cap \overline{B} \cap \overline{C}$, 592; $\overline{A} \cap B \cap \overline{C}$, 930; $\overline{A} \cap \overline{B} \cap C$, 176; $A \cap B \cap \overline{C}$, 95; $A \cap \overline{B} \cap C$, 49; $\overline{A} \cap B \cap C$, 20; and $A \cap B \cap C$, 8. The Venn diagram mode can output a script for Python with matplotlib and matplotlib_venn to visualize the result. The Venn diagram obtained using the script with slight modification is shown in Fig. 3. The diagram clearly shows that the number of proteins that can be accessed by only one of the three reactions is large compared with that in the overlapped areas. This result demonstrates that the combination of ProteoFind and ProteoFind_analyzer could be of benefit to compare features of the new ligation method with those of the existing ones.

3. Conclusion

This paper describes the development and case studies of ProteoFind, which is a computational script for finding proteins that are suitable for chemical synthesis from protein lists provided at the UniProt website. It then presents a discussion of the application of ProteoFind to visualize areas covered by several one-pot three-fragment ligation methods. The results demonstrated that ProteoFind not only saves time in the search for synthetic target proteins, but also proposes many candidate
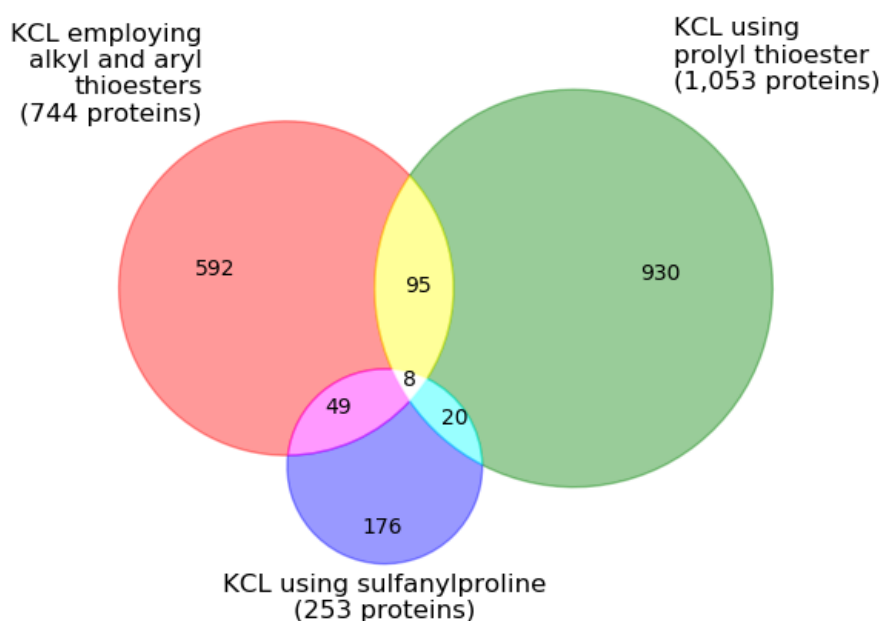

Fig. 3. Venn diagram for visualization of areas covered by the three one-pot three-fragment ligations.

proteins from among which biomedically interesting proteins could be found. Furthermore, comparison of features of the ligation methods is enabled by the use of ProteoFind and ProteoFind_analyzer by elucidating areas for which the ligation methods are accessible. We believe that ProteoFind and its related scripts will become indispensable tools for synthetic protein chemists in the near future.

4. Experimental section

4.1. General

All scripts are written for Python 2.7. The scripts, the protein lists in the one-line/one-protein format, batch files, results of case studies, and a simple manual are provided as Supplemental Files. The procedures shown below are written for Windows users (operating system: Windows 7), but Linux and Mac users can also run the scripts through a terminal after slight modification of the source codes. Searching the codes using the keyword "Linux and Mac users" will show the parts in which the modification is required.

4.2. Set-up

Python 2.7.6 was obtained from Python.org (https://www.python.org/downloads/) and installed on a computer before running ProteoFind. A folder "proteofind" provided in the Supplemental Files was then copied in a Python27 folder (C:\Python27).

4.3. Preparation of the protein list

The protein lists in one-line/one-protein format are incorporated in the Supplemental Files. To update the lists, protein lists in FASTA (canonical) format are downloaded from the UniProt website (http://www.uniprot.org/uniprot/) and the extension ".txt" is added if necessary. After placing the downloaded files in the ProteoFind folder (C:\Python27\proteofind), double clicking ProteoFind_list_converter.bat and following the guidance shown in the command prompt generate the new list files in the one-line/one-protein format.

4.4. Activation of acceleration mode of ProteoFind

Depending on the spec of the computer, the acceleration mode sometimes significantly reduces the time required for the search. Windows users can activate the acceleration mode by searching the source code of ProteoFind using the keyword "Windows users" and following the guidance described there.

4.5. Searching for synthetic target proteins using ProteoFind

After double clicking ProteoFind.bat in the ProteoFind folder to run the script, the following search criteria were inputted according to guidance shown in the command prompt: a protein list, number of peptide fragments of which the synthetic target protein consists, minimum and maximum lengths of the fragments, amino acids at ligation sites, maximum size of the proteins (optional), and number of N- and/or C-terminal amino acids that are deleted before the search (optional). The result was outputted as a text file. The results can contain the same proteins with different ligation sites; therefore, ProteoFind_analyzer was provided to count the number of proteins in the output file.

4.6. Case studies

Search criteria and results of the case studies shown in Table 1 are as follows. In all cases, "all proteins (reviewed)" containing 555,100 proteins was employed as a protein list, and fragment length was limited to 10–50 residues. The output files of the case studies are found in the Supplemental Files.

Standard NCL: number of fragments = 2, amino acids at a ligation site = "any/C", 32,190 hits.

KCL employing alkyl and aryl thioesters: number of fragments = 3, amino acids at ligation sites = "G/C" and "AA³/C" from an N-terminal side. AA³ and number of the hits: ILTV, 893; I, 122; L, 142; T, 222; V, 414.

KCL using prolyl thioester: number of fragments = 3, amino acids at ligation sites = "AA¹/C" and "P/C" from an N-terminal side. AA¹ and number of the hits: ACDEFGHIKLMNQRSTVWY, 2,545; A, 142; C, 287; D, 161; E, 109; F, 91; G, 143; H, 64; I, 112; K, 180; L, 139; M, 17; N, 147; Q, 126; R, 221; S, 217; T, 178; V, 213; W, 18; Y, 71.

KCL using sulfanylproline: number of fragments = 3, amino acids at ligation sites = "G/C" and "G/P" from an N-terminal side, 279 hits.

Four-fragment ligation: number of fragments = 4, amino acids at ligation sites = "G/C", "ILTV/C" and "any/C" from an N-terminal side, 1,830 hits.

Number of proteins in the above mentioned results after removal of duplication: The number of proteins in Table 1 was estimated using ProteoFind_analyzer (standard NCL, 13,973 proteins; KCL employing alkyl and aryl thioesters, 744 proteins whether amino acids at the AA³ position were inputted at once or individually; KCL using prolyl thioester, 1,053 proteins whether amino acids at the AA¹ were inputted at once or individually; KCL using sulfanylproline, 253 proteins; four-fragment ligation, 631 proteins).

Preparation of a Venn diagram: ProteoFind_analyzer was employed to count the number of proteins shown in Fig. 3, and the script for drawing the Venn diagram of the result was generated automatically. The obtained script was slightly modified to improve the appearance of the diagram, and it was then passed to python.exe, in which matplotlib and matplotlib_venn had been installed, in a drag-and-drop manner to give the Venn diagram. Matplotlib and matplotlib_venn can be obtained by typing "C:\Python27\Scripts\pip.exe install matplotlib" and "C:\Python27\Scripts\pip.exe install matplotlib_venn" in a command prompt, respectively.

References

1. (a) Liu, L., Ed. *Protein Ligation and Total Synthesis I*, Springer International Publishing AG, Cham, Switzerland 2015; (b) Liu, L., Ed. *Protein Ligation and Total Synthesis II*, Springer International Publishing AG, Cham, Switzerland 2015; (c) Burke, H. M.; McSweeney, L.; Scanlan, E. M. *Nat. Commun.* **2017**, *8*, 15655.
2. (a) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. *Science* **1994**, *266*, 776-779; Dawson, P. E.; Kent, S. B. H. *Annu. Rev. Biochem.* **2000**, *69*, 923-960; (c) Kent, S. B. H. *Chem. Soc. Rev.* **2009**, *38*, 338-351.
3. (a) Rohde, H.; Seitz, O. *Biopolymers (Pept. Sci.)* **2010**, *94*, 551-559; (b) Dawson, P. E. *Israel J. Chem.* **2011**, *51*, 862-867; (c) Malins, L. R.; Payne, R. J. *Aust. J. Chem.* **2015**, *68*, 521-537.
4. (a) Offer, J. *Biopolymers (Pept. Sci.)* **2010**, *94*, 530-541; (b) Payne, R. J.; Wong, C.-H. *Chem. Commun.* **2010**, *46*, 21-43.

5. (a) Kang, J.; Macmillan, D. *Org. Biomol. Chem.* **2010**, *8*, 1993-2002; (b) Zheng, J.-S.; Tang, S.; Huang, Y.-C.; Liu, L. *Acc. Chem. Res.* **2013**, *46*, 2475-2484; (c) Macmillan, D.; Adams, A.; Premdjee, B. *Israel J. Chem.* **2011**, *51*, 885-899; (d) Tailhades, J.; Patil, N. A.; Hossain, M. A.; Wade, J. D. *J. Pept. Sci.* **2015**, *21*, 139-147; (e) Macmillan, D. *Synlett* **2017**, *28*, 1517-1529.

6. Raibaut, L.; Ollivier, N.; Melnyk, O. *Chem. Soc. Rev.* **2012**, *41*, 7001-7015.

7. Examples of recently reported one-pot multi-fragment ligations: (a) Asahina, Y.; Kamitori, S.; Takao, T.; Nishi, N.; Hojo, H. *Angew. Chem. Int. Ed.* **2013**, *52*, 9733-9737; (b) Pan, M.; He, Y.; Wen, M.; Wu, F.; Sun, D.; Li, S.; Zhang, L.; Li, Y.; Tian, C. *Chem. Commun.* **2014**, *50*, 5837-5839; Thompson, R. E.; Liu, X.; Alonso-García, N.; Pereira, P. J. B.; Jolliffe, K. A.; Payne, R. J. *J. Am. Chem. Soc.* **2014**, *136*, 8161-8164; (c) Li, J.; Li, Y.; He, Q.; Li, Y.; Li, H.; Liu, L. Org. Biomol. Chem. 2014, 12, 5435-5441; (d) Tang, S.; Si, Y.-Y.; Wang, Z.-P.; Mei, K.-R.; Chen, X.; Cheng, J.-Y.; Zheng, J.-S.; Liu, L. *Angew. Chem. Int. Ed.* **2015**, *54*, 5713-5717; (e) Raibaut, L.; Drobecq, H.; Melnyk, O. *Org. Lett.* **2015**, *17*, 3636-3639; (f) Maity, S. K.; Jbara, M.; Laps, S.; Brik, A. *Angew. Chem. Int. Ed.* **2016**, *55*, 8108-8112; (g) Tsuda, S.; Mochizuki, M.; Nishio, H.; Yoshiya, T.; Nishiuchi, Y. *Biopolymers (Pept. Sci.)* **2016**, *106*, 503-511; (h) Thuaud, F.; Rohrbacher, F.; Zwicky, A.; Bode, J. W. *Helv. Chim. Acta* **2016**, *99*, 868-894; (i) Lee, C. L.; Liu, H.; Wong, C. T. T.; Chow, H. Y.; Li, X. *J. Am. Chem. Soc.* **2016**, *138*, 10477-10484; (j) Asahina, Y.; Kawakami, T.; Hojo, *Chem. Commun.* **2017**, *53*, 2114-2117; (k) Aihara, K.; Yamaoka, K.; Naruse, N.; Inokuma, T.; Shigenaga, A.; Otaka, A. *Org. Lett.* **2016**, *18*, 596-599; (l) Takei, T.; Andoh, T.; Takao, T.; Hojo, H. *Angew. Chem. Int. Ed.* **2017**, *56*, 15708-15711; (m) Si, Y.-Y.; Liang, L.-J.; Tang, S.; Qi, Y.-K.; Huang, Y.; Zheng, J.-S. *Tetrahedron Lett.* **2018**, *59*, 268-271; (n) Wang, Y.; Han, L.; Yuan, N.; Wang, H.; Li, H.; Liu, J.; Chen, H.; Zhang, Q.; Dong, S. *Chem. Sci.* in press (doi: 10.1039/c7sc04148a); (o) Takeda, N.; Takei, T.; Asahina, Y.; Hojo, H. Chem. Eur. J. in press (doi: 10.1002/chem.201706127).

8. (a) Hojo, H.; Aimoto, S. *Bull. Chem. Soc. Jpn.* **1991**, *64*, 111-117; (b) Aimoto, S. *Biopolymers (Pept. Sci.)* **1999**, *51*, 247-265.

9. Bode, J. W. *Acc. Chem. Res.* **2017**, *50*, 2104-2115.

10. Sato, K.; Shigenaga, A.; Tsuji, K.; Tsuda, S.; Sumikawa, Y.; Sakamoto, K.; Otaka, A. *ChemBioChem* **2011**, *12*, 1840-1844.

11. A statistical review of chemically synthesized proteins, see: Agouridas, V.; Mahdi, O. E.; Cargoët, M.; Melnyk, O. *Boorg. Med. Chem.* **2017**, *25*, 4938-4945.

12. A program not to search the synthetic target proteins but to propose a synthetic route of the target is also available. Jacobsen, M. T.; Erickson, P. W.; Kay, M. S. *Bioorg. Med. Chem.* **2017**, *25*, 4946-4952.

13. Bang, D.; Pentelute, B. L.; Kent, S. B. H. *Angew. Chem. Int. Ed.* **2006**, *45*, 3985-3988.

14. Nakamura, T.; Shigenaga, A.; Sato, K.; Tsuda, Y.; Sakamoto, K.; Otaka, A. *Chem. Commun.* **2014**, *50*, 58-60.

15. Ding, H.; Shigenaga, A.; Sato, K.; Morishita, K.; Otaka, A. *Org. Lett.* **2011**, *13*, 5588-5591.

16. The NCL-like reaction employing sulfanylproline was indipendently developped by us[15] and Danishefsky's group. Shang, S.; Tan, Z.; Dong, S.; Danishefsky, S. J. *J. Am. Chem. Soc.* **2011**, *133*, 10784-10786.

17. (a) Otaka, A.; Sato, K.; Ding, H.; Shigenaga, A. *Chem. Record* **2012**, *12*, 479-490; (b) Otaka, A.; Sato, K.; Shigenaga, A. In *Protein Ligation and Total Synthesis II*; Liu, L., Ed.; Springer International Publishing AG, Cham, Switzerland, 2015, pp 33-56.

18. Lee, J.; Kwon, Y.; Pentelute, B. L.; Band, D. *Bioconjugate Chem.* **2011**, *22*, 1645-1649.

19. Pollock, S. B.; Kent, S. B. H. *Chem. Commun.* **2011**, *47*, 2342-2344.