



HAL
open science

Using the Interactive Tool of the Protein Chemical Synthesis Database

Vangelis Agouridas, Oleg Melnyk

► **To cite this version:**

Vangelis Agouridas, Oleg Melnyk. Using the Interactive Tool of the Protein Chemical Synthesis Database. Peptide and Protein Engineering, pp.29-36, 2020, <10.1007/978-1-0716-0720-6_3>. <hal-02995409>

HAL Id: hal-02995409

<https://hal.science/hal-02995409v1>

Submitted on 12 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Using the interactive tool of The Protein Chemical Synthesis Database

Vangelis Agouridas*, Oleg Melnyk

Université de Lille, Institut Pasteur de Lille, UMR CNRS 8204 1 rue du Pr Calmette, 59021
Lille, France.

vangelis.agouridas@ibl.cnrs.fr

Summary

Over the last 25 years, chemoselective amide-bond forming reactions have established themselves as an essential tool for the total chemical synthesis of peptides and proteins. This spectacular development is echoed in an abundant literature that we have compiled in a database: the Protein Chemical Synthesis DataBase (<http://pcs-db.fr>). The PCS website provides an interactive tool with a user-friendly interface to get introduced to the most routinely used ligation methods including their scope. It can also be used for simply getting an overview or a track of the most recent advances made in the field of peptide and protein synthesis by means of chemoselective ligation reactions. The aim of this protocol article is to present the content of the database and showcase a typical query with the interactive web interface.

Keywords: chemical protein synthesis, chemoselective amide-bond forming reaction, database

1. Introduction

During the last 25 years, the scope of peptides and proteins amenable to chemical synthesis has been considerably extended with the advent of chemoselective amide-bond forming reactions, recently pushing the size of fully synthetic and functional proteins up to 350 residues.¹ Basically, these reactions consist in the chemoselective formation of a native peptide bond between two unprotected peptide segments under mild conditions (Figure 1).² The Native Chemical Ligation (NCL³), a ligation method based on the reaction of a peptidyl thioester with a cysteinyl peptide, has been extensively used since its discovery in 1994 (Figure 1a). Since then, some variations have been introduced to expand its scope by modifying the nature of the acyl donor (i.e. hydrazides,⁴ benzimidazolinones⁵ or *O,S*-, *N,S*- or *N,Se*-acyl shift systems⁶), of the thiol component (thiol or selenol amino acid surrogates,⁷ auxiliaries⁸) or both (diselenide selenoester ligation, DSL,⁹ Figure 1c). Besides, other mechanistically unrelated ligation methodologies were developed such as the α -ketoacid-hydroxylamine ligation (KAHA, Figure 1d),¹⁰ the serine-threonine ligation (STL, Figure 1e)¹¹ and the hydroxamate ligation (Figure 1f).¹² Finally, the protein chemist synthetic toolbox has also been enriched by a repertoire of post-ligation reactions that allow for further modifications on assembled proteins (e.g. desulfurization¹³).

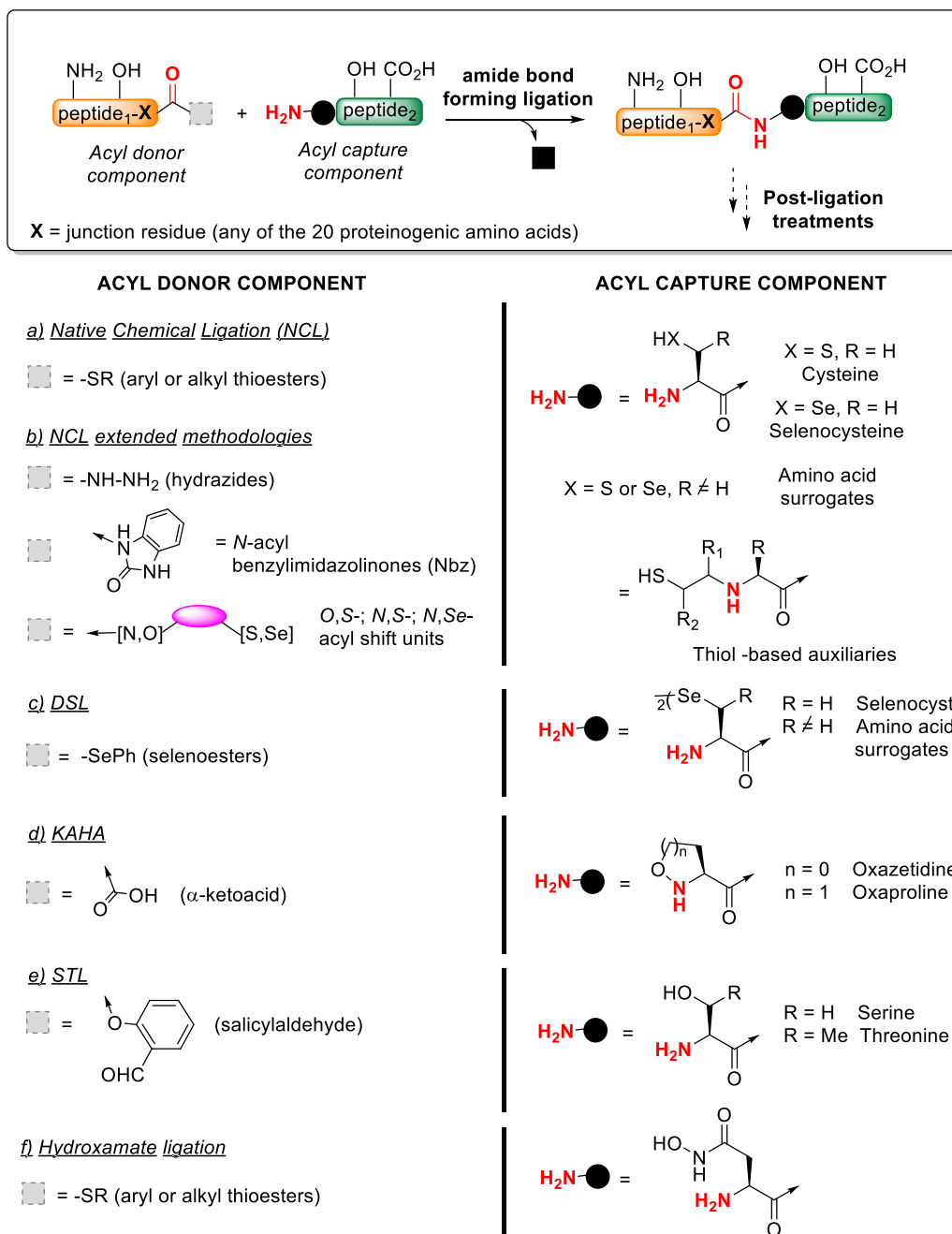


Figure 1. General principle of most commonly used chemoselective amide-bond forming reactions.

Since 1994, all of these methods have been used alone or combined in complex reaction schemes to prepare hundreds of proteins. The “Protein Chemical Synthesis Database”¹⁴ (pcsd.fr) is a comprehensive database that was created in order to facilitate the collection or the retrieval of information about the synthetic design of proteins. Moreover, the PCS proves a

particularly powerful tool when it comes to put the domain in perspective or compare where a specific method stands compared to the other ones for a specific application.

2. Presentation of the PCS database: Conceptual and logical design

The PCS database (PCS-db) was built by collecting different types of descriptors from more than 600 articles reporting the chemical synthesis of proteins by means of chemoselective amide-bond forming reactions since 1994. Only targets of biological significance were retained whereas model peptides used for methodological studies, polymers or hybrid materials were systematically discarded. First level of information available for registered peptides and proteins concerns their inherent characteristics: name, year of publication and length. Second level of information is relative to their synthetic design. It provides some details about the type of ligation chemistry used, the number of ligations achieved to assemble the target proteins but also the nature of the junction residues, the use of amino acid surrogates, thiol auxiliaries and/or the application of post-ligation treatments.

All the above-mentioned data were compiled in a table file which was processed with a cloud-based self-service usually used for data management in business intelligence. A copy of the table in MS excel format remains available in the download section of the PCS database website, though with a limited number of functionalities. The web interface of the PCS-db provides a user-friendly tool where queries are simply made by mouse-clicking buttons which activate / deactivate various filters and display refined subsets of the collected data in a table.

Additionally, a “graphical overview” module (PCS-GO) complements the database. The PCS-GO module is composed of interactive charts based on the PCS-db dataset that visitors can manipulate at their convenience to quickly get a synoptic view of the domain.

Finally, the PCS website also features a full page dedicated to instructions which provides detailed information on the meaning of each filter and a bibliographical page presenting a

selection of landmark books, reviews and papers to get introduced to the main concepts of chemoselective peptide ligation reactions.

3. Materials

The PCS website (pcs-db.fr) proposes a dynamic, interactive and intuitive environment that requires no particular computer skills. It is accessible from any personal computer, tablet or mobile phone connected to the internet with an up-to-date web browser. The very basic example below details the procedure for making queries on the PCS database (update of March 25th 2019: 931 entries which correspond to the total number of proteins available in this version of the database). Of course, all search criteria can be combined at will in more complex scenarios.

4. Methods

Scenario 1 (use of the PCS-db module): In scenario 1, one would like to design a synthetic approach for a protein of 180 residues with the following constraints: i) the protein will be produced chemically, without resorting to recombinant technologies; ii) the exclusive use of NCL or NCL-derived methods is required; iii) a Q-C (Gln-Cys) junction will be assembled by ligation. How can the PCS-db help him/her retrieve works from the literature responding to this query?

1. Go to the “pcs-db.fr” website. On the homepage, select the “PCS-DB” menu to display the PCS-db control panel (Figure 2).
2. [OPTIONAL] On the bottom right corner of the interactive table, click the double-headed arrow to toggle full screen mode.
3. To have a relevant answer set regarding the size of your target, search for proteins whose length is comprised between 160 and 200 residues using the numeric range slicer (Figure 2,

step 1). Application of this first search criterion results in a significant shrinking of the answer set (from 931 to 41 answers).

4. Exclude the recombinant proteins by selecting “No” in the “EPL filter” box (Figure 2, step 2). 34 entries remain available in the database.

5. The answer set can still be refined according to the initial search criteria. In the “type of ligation” menu, select all NCL and NCL-extended methods still available (i.e. NCL, hydrazides, *N,S*-shift, *N,Se*-shift, *O,S*-shift, Nbz) (Figure 2, step 3). Removal of the KAHA ligation which is not a thiol-thioester exchange based reaction discards two additional references.

6. In the “C-terminal Residue Control Panel”, click Q to select synthetic works which describe the assembly of at least one Q-C junction (Figure 2, step 4).

7. Check the result in the bottom table. You are now redirected to 7 entries describing either the total chemical synthesis of the NK1 domain of the human hepatocyte growth factor or the synthesis of various EPO glycoforms. The assembly of these proteins is described in five different publications whose abstracts can be directly accessed by clicking the link right next to the reference whenever available (Figure 2, step 5).

8. If needed, the answer set can still further be refined by mouse clicking the unexploited descriptors. As an example, consider that the total synthesis of the targeted protein is finally to be conducted on a solid support, clicking on the solid phase button will point to only one publication discussing the production of biotinylated NK1 (Figure 2, step 6).¹⁵

9. To reset all filters at once and make a new query, click the “clear filters” image (Figure 2, step 7).

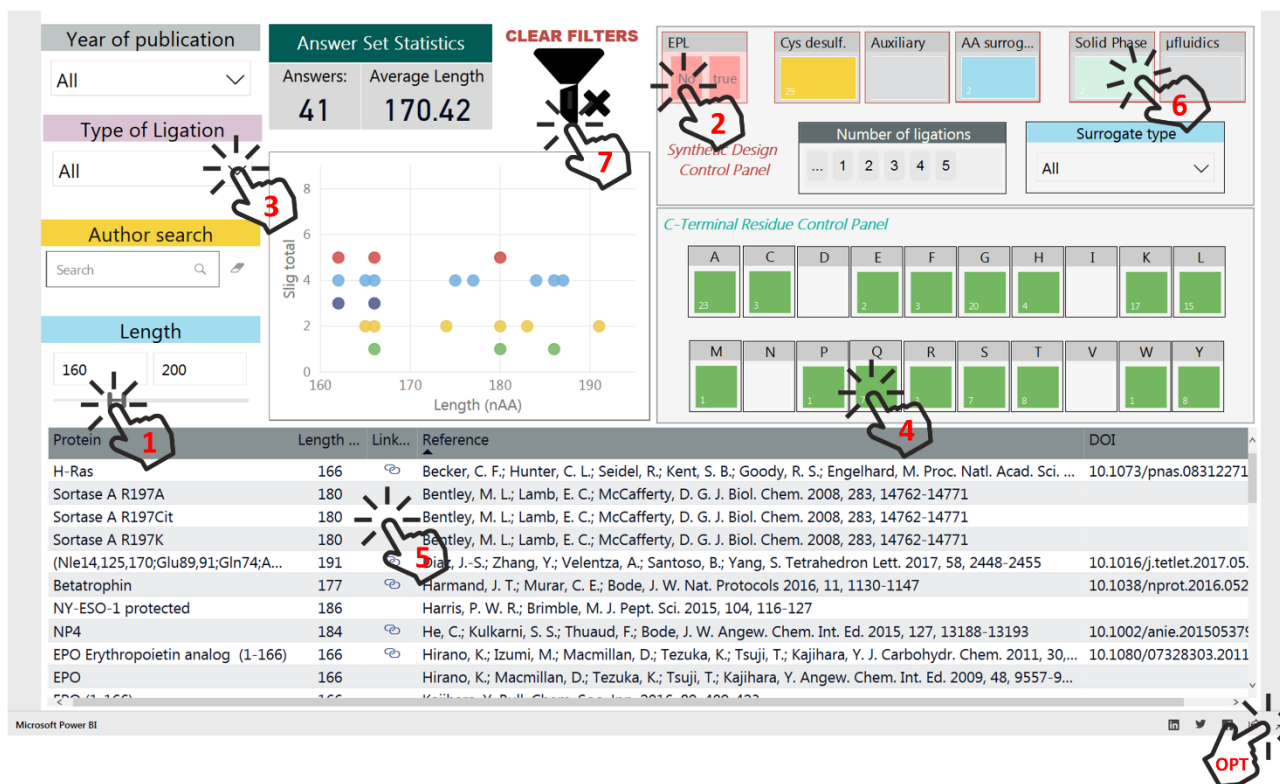


Figure 2. General control panel of the Protein Chemical Synthesis DataBase (PCS-db).

Scenario 2 (use of the PCS-GO module): The PCS-GO module is a particularly powerful tool to generate statistical overviews of the domain either for enriching a course material or for illustrating an oral presentation. The following example shows how to use it. The objective here is to evaluate the occurrence of synthetic designs involving 3 or more ligation reactions to assemble a protein.

1. Go to the “pcs-db.fr” website. On the homepage, select the “PCS-GO” menu to display the PCS-GO control panel (Figure 3). The PCS-GO module is composed of various graphical elements which represent the number of peptides and proteins produced each year by means of chemoselective ligation reactions (Figure 3A), provide quantitative information about the refined datasets (Figure 3B) or are informative of the number or the type of ligation reactions

used to assemble proteins (Figure 3C,D). The use of amino acid surrogates, auxiliaries or desulfurization approaches is reported in Figure 3E.

2. [OPTIONAL] On the bottom right corner of the interactive table, click the double-headed arrow to toggle the full screen mode.

3. In rectangle C, left-click on the area corresponding to 3 ligation reactions. While holding down the Ctrl key, left-click the areas corresponding to 4, 5, 6 and 8 ligations (Figure 3, step 1).

4. Other graphical representation instantly refresh to deliver adjusted statistics that allow to appreciate the importance of multisegment design over the recent years.

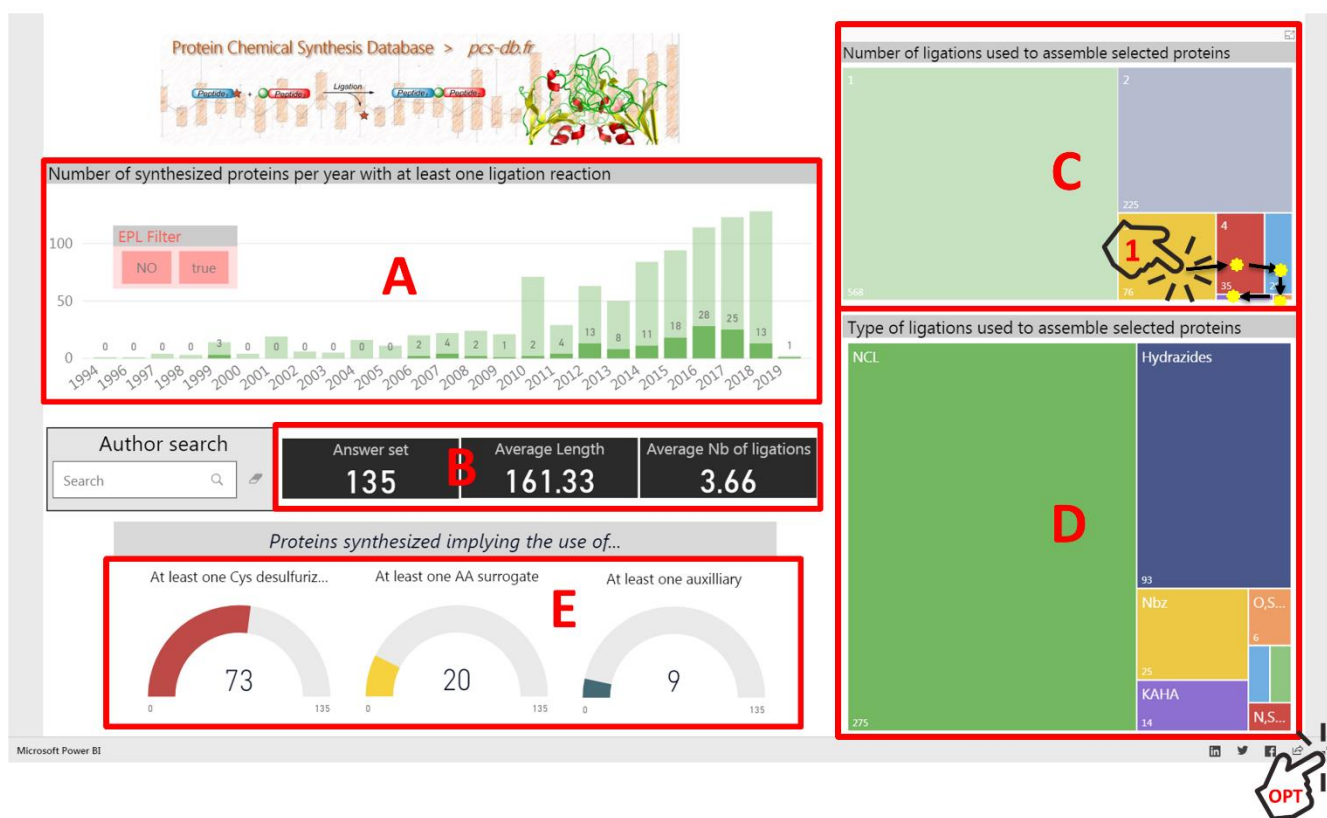


Figure 3. Interface of PCS-GO module

5. Notes

1. The PCS database can be cited with mention of our 2017 publication in *Bioorganic and Medicinal Chemistry* (see ref. 14).
2. More comprehensively, the PCS-db interactive table can filter entries according to year of publication, type or number of ligations, author names or length of proteins. It can also display results according to whether or not synthetic targets have been produced using recombinant technologies, solid phase synthesis, microfluidics, use of amino acid surrogates, auxiliaries or post-ligation treatments such as desulfurization. It is also possible to quickly visualize all peptides or proteins whose synthesis has involved a ligation step at a particular junction residue. Finally, an interactive plot-chart located at the center of the interface can be clicked to select one particular entry in the database. Each point represents one or more proteins as a function of its length and the number of ligation steps required to assemble it.
3. The language of the PCS-DB module is set to be the default language of the web browser used to display the database. Language can be changed by modifying the language settings of your web browser.
4. Https traffic, which is the kind of protocols used for the PCS-db might be blocked by firewalls. In such a case, an error message is displayed and the database will not load. The problem can be solved by changing the site permission parameters of your firewall or asking your computer and network related service to allow access to the PCS-db website at your institution.
5. Other tools have been recently developed in order to facilitate the identification of targets of interest or to assist the protein chemist in the synthetic design of peptides and proteins: respectively the Proteofind script¹⁶ and the Aligator software.¹⁷

6. In upcoming PCS updates, new options will be added that will enable to filter results according to whether or not synthesized peptides or proteins are cyclic, possess post-translational modifications or feature tags (peptidic or non-peptidic such as fluorophores). Besides, the content of the database will be regularly updated.

Acknowledgements

This work was supported by the Centre National de la Recherche Scientifique (CNRS), the University of Lille and Institut Pasteur de Lille.

References

-
- ¹ Xu, W.; Jiang, W.; Wang, J.; Yu, L.; Chen, J.; Liu, X.; Liu, L.; Zhu, T. F. Total chemical synthesis of a thermostable enzyme capable of polymerase chain reaction. *Cell Discov.* **2017**, *3*, 17008.
- ² Agouridas, V.; El Mahdi, O.; Diemer, V.; Cargoët, M.; Monbaliu, J.-C. M.; Melnyk, O. Native Chemical Ligation and Extended Methods. *Chem. Rev.* **2019**, *119*, 7328-7443.
- ³ Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. Synthesis of proteins by native chemical ligation. *Science* **1994**, *266* (5186), 776-779.
- ⁴ Fang, G. M.; Li, Y.M.; Shen, F.; Huang, Y. C.; Li, J. B.; Lin, Y.; Cui, H. K.; Liu, L. Protein chemical synthesis by ligation of peptide hydrazides. *Angew. Chem. Int. Ed.* **2011**, *50*, 7645–7649.
- ⁵ Blanco-Canosa, J.B.; Dawson, P.E. An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation. *Angew. Chem. Int. Ed.* **2008**, *47*, 6851–6855.

-
- ⁶ Melnyk, O.; Agouridas, V. From protein total synthesis to peptide transamidation and metathesis: playing with the reversibility of *N,S*-acyl or *N,Se*-acyl migration reactions. *Curr. Opin. Chem. Biol.* **2014**, *22*, 137-145.
- ⁷ Wong, C. T. T.; Tung, C. L.; Li, X. Synthetic cysteine surrogates used in native chemical ligation. *Mol. BioSyst.* **2013**, *9*, 826-833.
- ⁸ Burke, H. M.; McSweeney, L.; Scanlan, E. M. Exploring chemoselective S-to-N acyl transfer reactions in synthesis and chemical biology. *Nat. Commun.* **2017**, *8*, 15655.
- ⁹ Mitchell, N. J.; Malins, L. R.; Liu, X.; Thompson, R. E.; Chan, B.; Radom, L.; Payne, R. J. Rapid Additive-Free Selenocystine-Selenoester Peptide Ligation. *J. Am. Chem. Soc.* **2015**, *137*, 14011-14014.
- ¹⁰ Bode, J. W.; Fox, R. M.; Baucom, K. D. Chemoselective Amide Ligations by Decarboxylative Condensations of *N*-Alkylhydroxylamines and α -Ketoacids. *Angew. Chem. Int. Ed.* **2006**, *118*, 1270-1274.
- ¹¹ Zhang, Y.; Xu, C.; Lam, H. Y.; Lee, C. L.; Li, X. Protein chemical synthesis by serine and threonine ligation. *Proc. Natl. Acad. Sci. USA.* **2013**, *110* (17), 6657-6662.
- ¹² Dunkelmann, D. L.; Hirata, Y.; Torato, K. A.; Cohen, D. T.; Zhang, C.; Gates, Z. P.; Pentelute, B. L. Amide-forming chemical ligation via *O*-acyl hydroxamic acids. *Proc. Natl. Acad. Sci.* **2018**, *115*, 201718356.
- ¹³ Wan, Q.; Danishefsky, S. J. Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angew. Chem. Int. Ed.* **2007**, *46*, 9248-9252.
- ¹⁴ Agouridas, V.; El Mahdi, O.; Cargoët, M.; Melnyk, O. A statistical view of protein chemical synthesis using NCL and extended methodologies. *Bioorg. Med. Chem.* **2017**, *25*, 4938-4945.

-
- ¹⁵ Ollivier, N.; Desmet, R.; Drobecq, H.; Blanpain, A.; Boll, E.; Leclercq, B.; Mougel, A.; Vicogne, J.; Melnyk, O. A simple and traceless solid phase method simplifies the assembly of large peptides and the access to challenging proteins. *Chem. Sci.* **2017**, *8*, 5362-5370.
- ¹⁶ Shigenaga, A.; Naruse, N.; Otaka, A. Proteofind: A script for finding proteins that are suitable for chemical synthesis. *Tetrahedron* **2018**, *74*, 2291-2297.
- ¹⁷ Jacobsen, M. T.; Erickson, P. W.; Kay, M. S. Aligator: A computational tool for optimizing total chemical synthesis of large proteins. *Bioorg. Med. Chem.* **2017**, *25*, 4946-4952.