Using the interactive tool of The Protein Chemical Synthesis Database

Vangelis Agouridas and Oleg Melnyk

Abstract

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, interactive 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 (*Erreur ! Source du renvoi introuvable.*).[2] The Native Chemical Ligation (NCL[3]), a ligation method based on the reaction of a peptidyl thioester with a cysteinyl peptide, has been extensively used since its discovery in 1994 (*Erreur ! Source du renvoi introuvable.**a***). Since then, some variations have been introduced to expand its scope by modifying the nature of the acyl donor (i.e. hydrazides,[4] benzimidazolinones[5] or *O,S*-, *N,S*- or *N,Se*-acyl shift systems[6]), of the thiol component (thiol or selenol amino acid surrogates,[7] auxiliaries[8]) or both (diselenide selenoester ligation, DSL,[9] *Erreur ! Source du renvoi introuvable.**c***). Besides, other mechanistically unrelated ligation methodologies were developed such as the α-ketoacid-hydroxylamine ligation (KAHA, *Erreur ! Source du renvoi introuvable.**d***),[10] the serine-threonine ligation (STL, *Erreur ! Source du renvoi introuvable.**e***)[11] and the hydroxamate ligation (*Erreur ! Source du renvoi introuvable.**f***).[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. desulfurization[13]).

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] (pcs-db.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-db proves a particularly powerful tool when it comes to put the domain in perspective or compare where a specific method stands in relation to the other ones for a specific application.

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

The 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. The first level of information available for registered peptides and proteins concerns their inherent characteristics: name, year of publication and length. The 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-db 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-db (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 (*Erreur ! Source du renvoi introuvable.*).
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 (*Erreur ! Source du renvoi introuvable.*, 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 (*Erreur ! Source du renvoi introuvable.*, 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) (*Erreur ! Source du renvoi introuvable.*, 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 (*Erreur ! Source du renvoi introuvable.*, 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 (*Erreur ! Source du renvoi introuvable.*, 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 (*Erreur ! Source du renvoi introuvable.*, step 6).[15]

9. To reset all filters at once and make a new query, click the “clear filters” image (*Erreur ! Source du renvoi introuvable.*, step 7).

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 (*Erreur ! Source du renvoi introuvable.*). 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 (*Erreur ! Source du renvoi introuvable.**A***), provide quantitative information about the refined datasets (*Erreur ! Source du renvoi introuvable.**B***) or are informative of the number or the type of ligation reactions used to assemble proteins (*Erreur ! Source du renvoi introuvable.**C,D***). The use of amino acid surrogates, auxiliaries or desulfurization approaches is reported in *Erreur ! Source du renvoi introuvable.**E***.

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 (*Erreur ! Source du renvoi introuvable.*, step 1).

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

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: the Proteofind script[16] and the Aligator software[17], respectively.

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). [18]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 TF (2017) Total chemical synthesis of a thermostable enzyme capable of polymerase chain reaction. Cell Discov 3:17008.

[2] Agouridas V, El Mahdi O, Diemer V, Cargoët M, Monbaliu JCM, Melnyk O (2019) Native Chemical Ligation and Extended Methods. Chem Rev 119:7328-7443.

[3] Dawson PE, Muir TW, Clark-Lewis I, Kent SBH (1994) Synthesis of proteins by native chemical ligation. Science 266:776-779.

[4] Fang GM, Li YM, Shen F, Huang YC, Li JB, Lin Y, Cui HK, Liu L (2011) Protein chemical synthesis by ligation of peptide hydrazides. Angew Chem Int Ed 50:7645–7649.

[5] Blanco-Canosa JB, Dawson PE (2008) An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation. Angew Chem Int Ed 47:6851–6855.

[6] Melnyk O, Agouridas V (2014) 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 22:137-145.

[7] Wong CTT, Tung CL, Li X (2013) Synthetic cysteine surrogates used in native chemical ligation. Mol BioSyst 9:826-833.

[8] Burke HM, McSweeney L, Scanlan EM (2017) Exploring chemoselective S-to-N acyl transfer reactions in synthesis and chemical biology. Nat Commun 8:15655.

[9] Mitchell NJ, Malins LR, Liu X, Thompson RE, Chan B, Radom L, Payne RJ (2015) Rapid Additive-Free Selenocystine-Selenoester Peptide Ligation. J Am Chem Soc 137:14011-14014.

[10] Bode JW, Fox RM, Baucom KD (2006) Chemoselective Amide Ligations by Decarboxylative Condensations of N-Alkylhydroxylamines and a-Ketoacids. Angew Chem Int Ed. 118:1270-1274.

[11] Zhang Y, Xu C, Lam HY, Lee CL, Li X (2013) Protein chemical synthesis by serine and threonine ligation. Proc Natl Acad Sci USA 110:6657-6662.

[12] Dunkelmann DL, Hirata Y, Torato KA, Cohen DT, Zhang C, Gates ZP, Pentelute BL (2018) Amide-forming chemical ligation via O-acyl hydroxamic acids. Proc Natl Acad Sci USA 115:201718356.

[13] Wan Q, Danishefsky SJ (2007) Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. Angew Chem Int Ed 46:9248-9252.

[14] Agouridas V, El Mahdi O, Cargoët M, Melnyk O (2017) A statistical view of protein chemical synthesis using NCL and extended methodologies. Bioorg Med Chem 25:4938-4945.

[15] Ollivier N, Desmet R, Drobecq H, Blanpain A, Boll E, Leclercq B, Mougel A, Vicogne J, Melnyk O (2017) A simple and traceless solid phase method simplifies the assembly of large peptides and the access to challenging proteins. Chem Sci 8:5362-5370.

[16] Shigenaga A, Naruse N, Otaka A (2018) Proteofind: A script for finding proteins that are suitable for chemical synthesis. Tetrahedron 74:2291-2297.

[17] Jacobsen MT, Erickson PW, Kay MS (2017) Aligator: A computational tool for optimizing total chemical synthesis of large proteins. Bioorg Med Chem 25:4946-4952.

[18] Agouridas V, El Mahdi O, Melnyk O. Protein Chemical Synthesis in Medicinal Chemistry. Manuscript in preparation.

Figures Captions

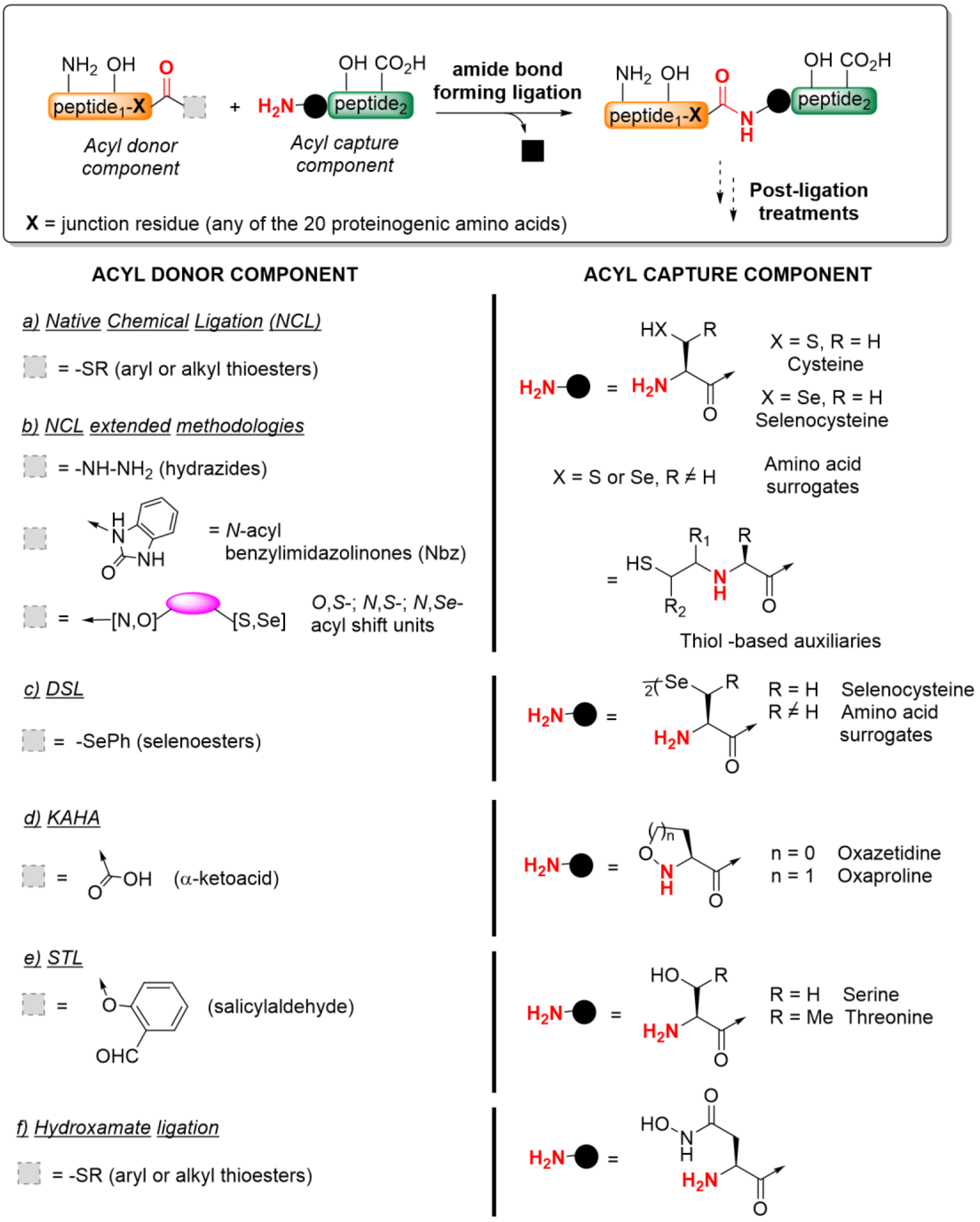
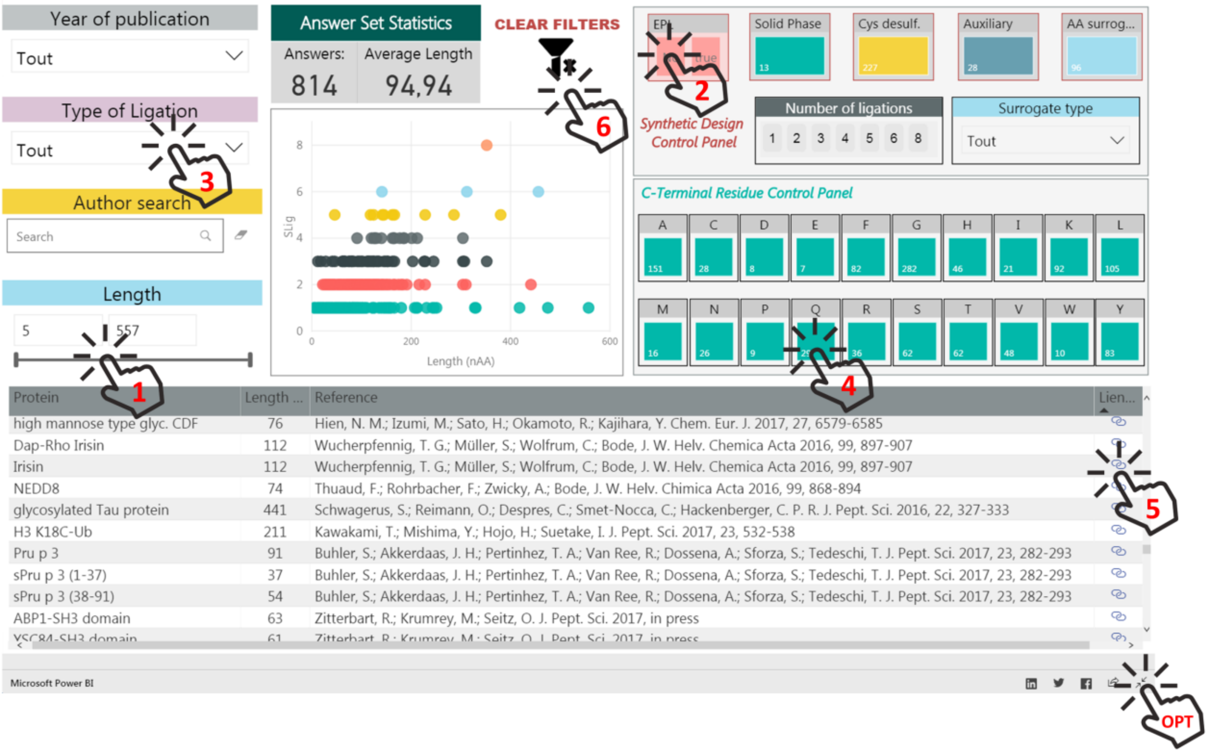


Figure Erreur ! Document principal seulement.**.** General principle of most commonly used chemoselective amide-bond forming reactions.



Protein	Length ...	Reference	Lien...
high mannose type glyc. CDF	76	Hien, N. M.; Izumi, M.; Sato, H.; Okamoto, R.; Kajihara, Y. Chem. Eur. J. 2017, 27, 6579-6585	🔗
Dap-Rho Irisin	112	Wucherpfennig, T. G.; Müller, S.; Wolfrum, C.; Bode, J. W. Helv. Chemica Acta 2016, 99, 897-907	🔗
Irisin	112	Wucherpfennig, T. G.; Müller, S.; Wolfrum, C.; Bode, J. W. Helv. Chemica Acta 2016, 99, 897-907	🔗
NEDD8	74	Thuaud, F.; Rohrbacher, F.; Zwicky, A.; Bode, J. W. Helv. Chimica Acta 2016, 99, 868-894	🔗
glycosylated Tau protein	441	Schwagerus, S.; Reimann, O.; Despres, C.; Smet-Nocca, C.; Hackenberger, C. P. R. J. Pept. Sci. 2016, 22, 327-333	🔗
H3 K18C-Ub	211	Kawakami, T.; Mishima, Y.; Hojo, H.; Suetake, I. J. Pept. Sci. 2017, 23, 532-538	🔗
Pru p 3	91	Buhler, S.; Akkerdaas, J. H.; Pertinhez, T. A.; Van Ree, R.; Dossena, A.; Sforza, S.; Tedeschi, T. J. Pept. Sci. 2017, 23, 282-293	🔗
sPru p 3 (1-37)	37	Buhler, S.; Akkerdaas, J. H.; Pertinhez, T. A.; Van Ree, R.; Dossena, A.; Sforza, S.; Tedeschi, T. J. Pept. Sci. 2017, 23, 282-293	🔗
sPru p 3 (38-91)	54	Buhler, S.; Akkerdaas, J. H.; Pertinhez, T. A.; Van Ree, R.; Dossena, A.; Sforza, S.; Tedeschi, T. J. Pept. Sci. 2017, 23, 282-293	🔗
ABP1-SH3 domain	63	Zitterbart, R.; Krumrey, M.; Seitz, O. J. Pept. Sci. 2017, in press	🔗
YSC84-SH3 domain	61	Zitterbart, R.; Krumrey, M.; Seitz, O. J. Pept. Sci. 2017, in press	🔗

Figure *Erreur ! Document principal seulement.*. General control panel of the Protein Chemical Synthesis DataBase (PCS-db).

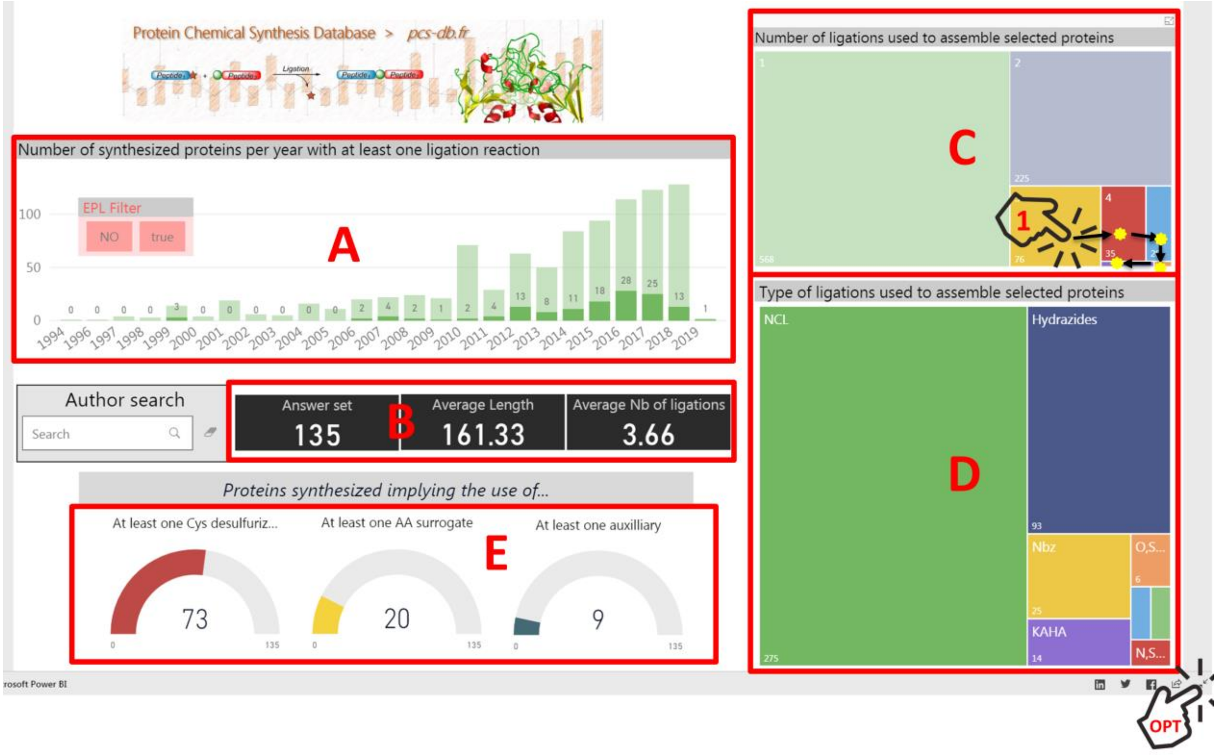


Figure *Erreur ! Document principal seulement.*. Interface of PCS-GO module.