Proceedings of the 36th European Peptide Symposium
Michal Lebl (Editor)
European Peptide Society, 2022 *https://doi.org/10.17952/36EPS.2022.262*

The Protein Chemical Synthesis Database (pcs-db.fr)

Vangelis Agouridas[1,2], Ouafâa El Mahdi[3], and Oleg Melnyk[1]

[1]Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 9017 - CIIL - Center for Infection and Immunity of Lille, F-59000 Lille, France; [2]Centrale Lille, F-59000, Lille, France
[3]Faculté Polydisciplinaire de Taza, University Sidi Mohamed Ben Abdellah, BP 1223 Taza Gare, Morocco

Introduction

Since 1994, chemoselective amide-bond forming reactions, of which the Native Chemical Ligation (NCL [1]), the Serine Threonine Ligation (STL [2]) or the KetoAcid HydroxylAmine ligation (KAHA [3]) are the more popular, have revolutionized chemical protein synthesis (CPS) by enabling the concatenation of unprotected, polyfunctionalized peptide segments in water under mild conditions (Figure 1). Such reactions have provided the means for successful applications in chemical biology, medicinal chemistry [4], materials science and nanotechnology research.

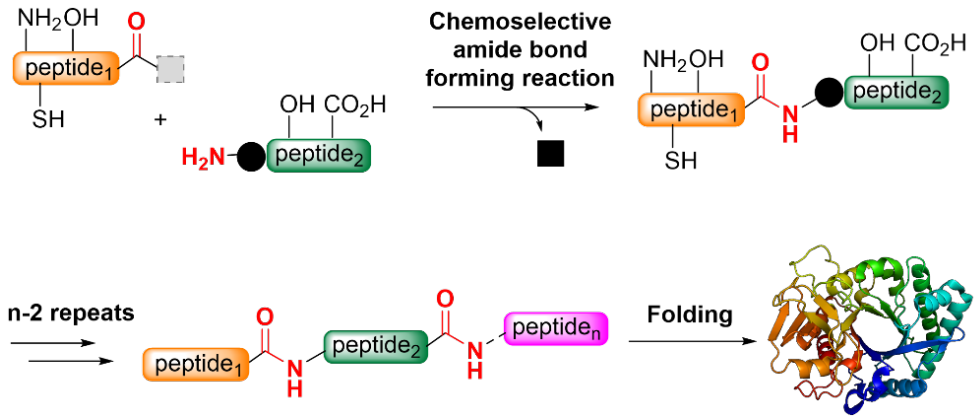

Fig. 1. General principle of amide-bond forming native ligation reactions.

These advances, which are described in a prolific literature, are presented in a freely available database: the Protein Chemical Synthesis database (pcs-db.fr) [5]. The present proceeding article describes the logical and conceptual design of the latter and how it can be exploited as a bridge between the research and teaching communities to promote chemical protein synthesis.

Material and methods

The PCS-db was built around three levels of information, mined from over 700 articles in the literature. In this context, synthetic targets of biological significance were retained whereas methodological studies using model peptides only were not considered. The different levels of information deal with:

- the identity of synthesized proteins (name, year of publication, length, bibliographical reference),
- the synthetic design of protein targets (type of ligation chemistry applied, number of ligations, nature of the junction residues, use of amino acid surrogates, thiol auxiliaries and/or the application of post-ligation treatments),
- the presence of modifications (mutations, post-translational modifications, tags or presence of non canonical amino acids).

The various items were reported in a table file, which was processed with a cloud-based self-service to provide users with an intuitive, clear and comprehensive interface (Figure 2). In total, the PCS-db counts more than 1300 entries with around 15 descriptors each and can be interrogated through four independent search modules, all connected to the main database.

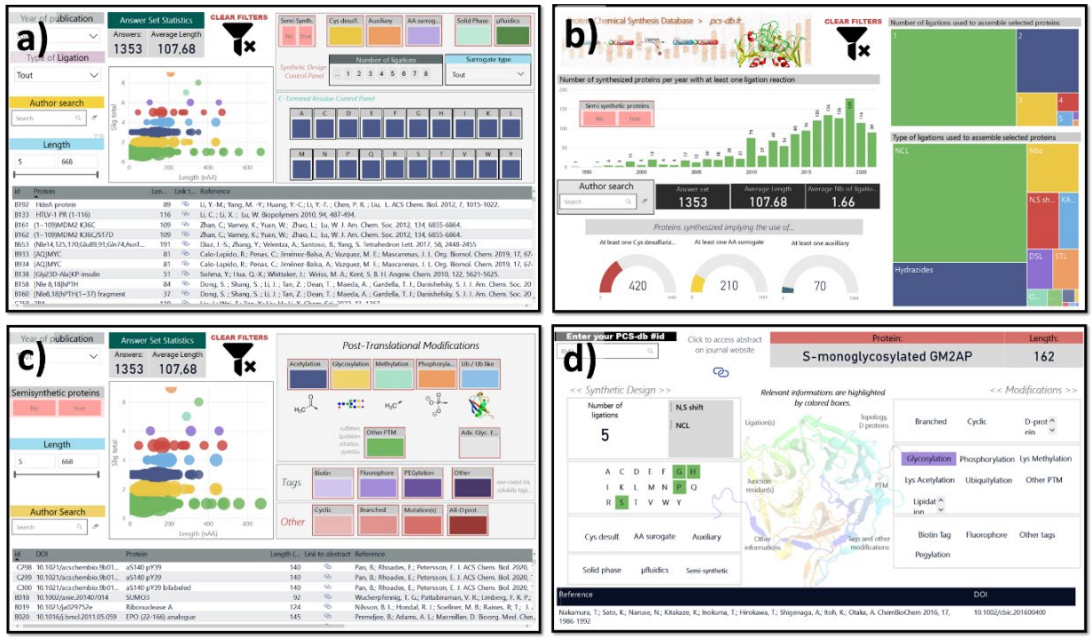

Fig. 2. The various modules of the PCS-db; a) the PCS-db module is the main module and is devoted to synthetic design; b) PCS-GO is a graphical companion providing synoptic views; c) PCS-PTM is a module devoted to modified synthetic proteins; d) PCS-id is a reverse look-up directory that allows to retrieve an entry from a #PCS-id.

The website also contains instructions, introductory bibliography and much more to discover online.

Results and discussion

The PCS-db offers researchers the possibility to achieve targeted bibliographical research ideally complementing well-established resources. By providing straightforwardly accessible information on the synthetic design of protein targets, the database can be perceived as a support decision tool for synthetic design selection. The database also reinforces the Research-Teaching Nexus by providing a structured learning environment.

Apart from practical aspects that allow instructors to find up-to-date articles or refresh their slides, the database can prove a useful support to teaching. Thanks to its comprehensive nature, data can be exploited to put the topic in perspective which is a concern hardly overcome by students when they are first introduced to new concepts or techniques. In the specific case of the PCS-db, we mean for example how the domain has quantitatively evolved over the years in terms of successfully synthesized targets and where a specific methodology stands compared to the other ones for a given period of time.

Also, the use of a database might orient teaching towards an inductive and more student-centered instruction. As databases are most often designed according to an entity-relationship model, deciphering the conceptual and logical design levels of the PCS-DB may help students acquiring a deeper understanding of the topic. For example, by formulating relevant queries in the context of driven homeworks, they should be able, alone or in a group, to identify, to evaluate and to analyze the main factors that govern the reactivity in amide-bond forming ligation reactions. Eventually, the results

of their research can be discussed in the classroom with the instructor and serve as a basis to develop the main concepts of ligation chemistry more in detail. Conversely, teachers can easily browse the PCS-DB and quickly build whole sets of relevant research articles to design a suitable learning environment for case-based studies.

The PCS-DB can also be further exploited for assessments. In this context, the instructor will let the students define their own search criteria, pick up, summarize and present a series of relevant articles to address a specific issue or limitation of amide-bond forming ligation reactions.

Conclusion

In its current state, the PCS database offers an efficient, interactive and free web-based tool on chemoselective amide-bond forming ligation reactions. Researchers, instructors or students can use it regardless of their knowledge of the field. The content of the PCS-DB is regularly updated to include the most recent contributions in the field and new features are to be implemented to offer a more accurate view of the topic.

References

1. Dawson, P.E., et al. *Science* **266**, 776-779 (1994), http://dx.doi.org/10.1126/science.7973629
2. Zhang, Y., et al. *Proc. Natl. Acad. Sci.* **110**, 6657-6662 (2013), http://dx.doi.org/ 10.1073/pnas.1221012110
3. Bode, J.W., et al. *Angew. Chem. Int. Ed.* **45**, 1248-1252 (2006), http://dx.doi.org/10.1002/anie.200503991
4. Agouridas, V., et al. *J. Med. Chem.* **63**, 15140-15152 (2020), http://dx.doi.org/10.1021/acs.jmedchem.0c01082
5. Agouridas, V., et al. *Bioorg. Med. Chem*. **25**, 4938-4945 (2017), http://dx.doi.org/10.1016/j.bmc.2017.05.050