Insights of techniques applied into protein engineering and proteomics

Author: Marc Fernandez; Tutor: Josep Vendrell; Treball de Fi de Grau

Universitat Autònoma de Barcelona

Introduction

The field of protein engineering has been around there for a long time, starting when the capacity of altering the structure of proteins was achieved by the evolution of the recombinant DNA technologies. In all this time there has been a lot of changes in this research field. In this work I will try to show you some nowadays techniques that a lot of people that work with proteins have use, and my goal in doing this is that by the end of each part you might have noticed a feature you did not know about it.
For each technique exposed in this work, I will briefly introduce it, with a little relevant history so far, and its role in protein studies; then, I will present some recent improvements or achievements in it, thus resulting in an extended and updated introduction of each one.

Bioinformatics

Introduction

For some time now, bioinformatics have emerged as a very useful tool in every protein engineering laboratory. For example, nowadays all protein engineers use bioinformatics support when designing and studying proteins. From predicting the function of a desired protein to predicting the structure of a well-known one, there are several improvements and issues of both approaches

Computational design of proteins

For this use of bioinformatics, two major facts can be observed:

- Today's best function protein prediction algorithms surpass those used previously, with significant gains in practically all aspects
- Although those algorithms and other widely used methods perform its function well enough, there is a need to improve these tools

Nevertheless, improvements in computing energethics are making possible advances on this field a few years ago were utopian. One example of this is the emerging de novo protein design, a promising tool for generating new biocatalysts. This generation consists in several parts:

- The selection of a target reaction and the definition of the catalytic mechanism
- Modeling of an idealized active site
- Docking and optimization of the active site into protein scaffolds
- The best resulting designs are synthesized, produced and characterized

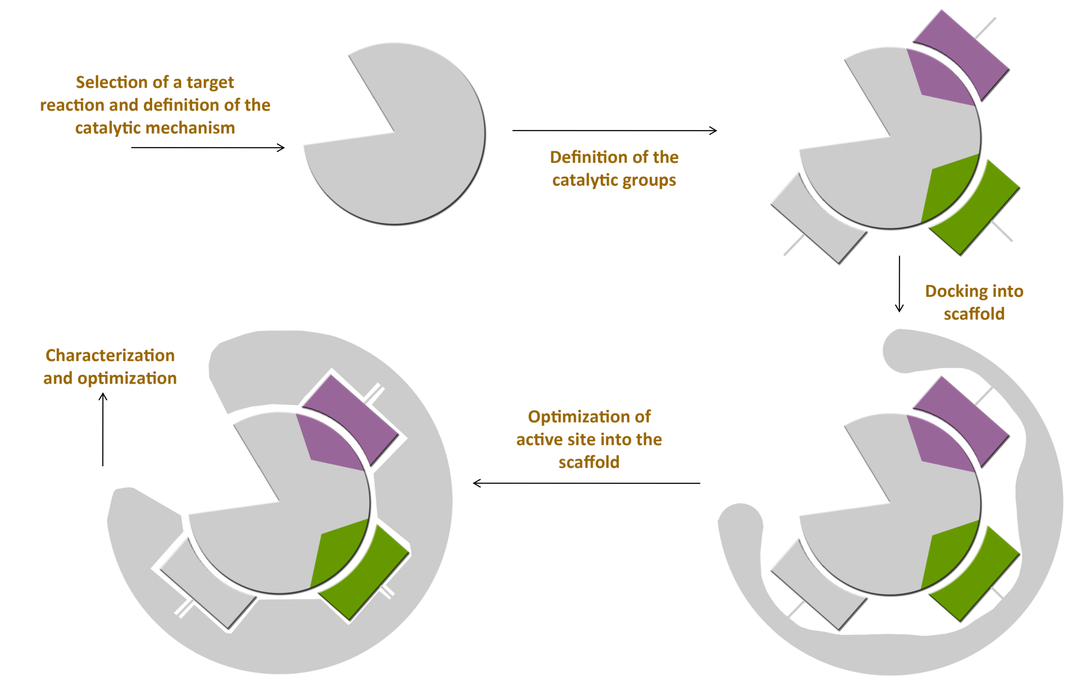

Fig. 1 – Scheme of the steps of *de novo* biocatalysts generation; each step is mentioned above.

References

[1] H. Kries, R. Blomberg, & D. Hilvert , *De novo enzymes by computational design* (2013) Current opinion in Chemical Biology, 17:221-228.

[2] K. Gevaert, M. Goethals, L. Martens, J. Van Damme, A. Staes, G. R. Thomas & J. Vandekerckhove (2003) Nature Biotechnology, 21:566-569.

[3] P.Kraft, A. Bergamaschi, Ch. Broennimann, R. Dinapoli, E. F. Einkenberry, B. Henrich, I. Johnson, A. Mozzanica, C. M. Schlepütz, P. R. Willmott & B. Schmitt, Performance of single-photon-counting PILATUS detector modules (2009) Journal of Synchrotron Radiation, 16:368-375.

Proteomics-based substrate identification

Introduction

A key point in the study of enzymes is to know the substrate they catalyze. In proteases, this identification is difficult, because the products must be identified among the entire cellular proteins, and the isolation of these products provide little information regarding the protease that catalyzed the reaction. However, the link between the product and the protease can be defined by several techniques, some of them being proteomics-based ones.

One of these techniques is based on the analysis of proteolytically digested protein mixtures by mass spectrometry (MS), before applying a peptide isolation based on diagonal electrophoresis and diagonal chromatography.

COFRADIC

COFRADIC stands for combined fractional diagonal chromatography, and this technique is capable of reduce the complexity of the MS protein sample by isolating each peptide and therefore improves and/or allows the identification of each processed peptide.

Recently, a COFRADIC-based approach has been used to build two peptide libraries that has been used to determine the substrate preferences of several peptidases. In this assay, the peptide libraries are generated by treating whole proteomes with library-generating proteases (such as chymotrypsin). Then, the peptide library is separated by HPLC. The primary fractions are then re-separated on the same column using identical conditions. The shifted peptidase products are then collected and can therefore be distinguished from their C-terminal unmodified counterparts following LC-MS/MS analyses and database searching.

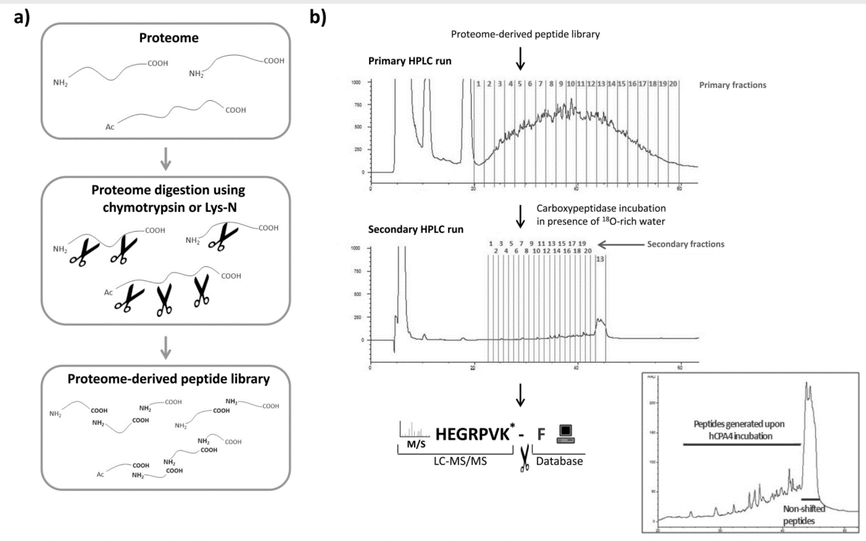

Fig .2 – Proteome-derived peptide library generation and peptidase substrate screen. **a)** Peptide library generation workflow. **b)** Chymotryptic proteome-derived peptide library-based peptidase substrate screen. Image obtained from *S. Tanco, J. Lorenzo, J. García-Pardo, S. Degroeve, L. Martens, FX. Avilés, K. Gevaert & P. Van Damme (2013) Mol. Cell Proteomics, Manuscript.*

X-ray Crystallography

Introduction

The final goal of protein engineering and proteomics is to fully understand all the characteristics of the studied protein. To achieve that, one must solve its structure, because is a key point of all the protein knowledge, and probably the best way to obtain that is to determine the 3D structure by X-ray crystallography.

The development and history of this technique has been so important that has resulted in some very prestigious prices, most of them Nobel Prices.

Improvements

The most notable change between first structures obtained by X-ray crystallography and recent ones is the resolution enhancement.

In the last years, this enhancement has been largely possible thanks to the improvements made to the equipment of X-ray sources, most common being synchrotrons. The improvements include, among others, the inclusion of novel detectors, such as single-photon counting (SPC) X-ray detectors, capable of acquiring data five times faster than the since then standards, the CCD detectors. We have near us one of these SPC detectors, mounted in the macromolecular crystallography beamline (named XALOC) at synchrotron ALBA.

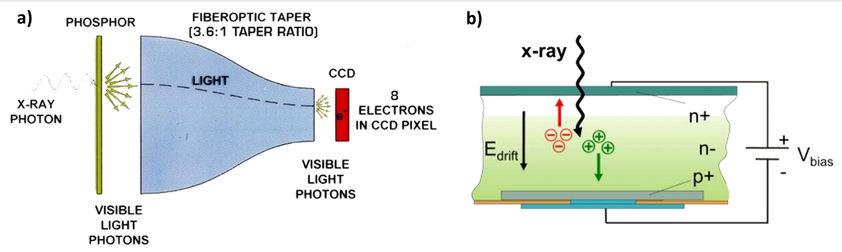

Fig. 3 – Schemes of: **a)** CCD (Charge Coupled Device) detector **b)** SPC (Single-photon counting) detector. Images obtained from *proteincrystallography.org* and *dectris.com*, respectively.