Biotechnology Advances 33 (2015) 566–604

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

Bioinformatic analysis of a PLP-dependent enzyme superfamily suitable for biocatalytic applications

Fabian Steffen-Munsberg [a,c], Clare Vickers [a], Hannes Kohls [a,b], Henrik Land [c], Hendrik Mallin [a], Alberto Nobili [a], Lilly Skalden [a], Tom van den Bergh [d], Henk-Jan Joosten [d], Per Berglund [c], Matthias Höhne [b,*], Uwe T. Bornscheuer [a,**]

[a] Dept. of Biotechnology & Enzyme Catalysis, Institute of Biochemistry, Greifswald University, Felix-Hausdorff-Str. 4, 17487 Greifswald, Germany
[b] Protein Biochemistry, Institute of Biochemistry, Greifswald University, Felix-Hausdorff-Str. 4, 17487 Greifswald, Germany
[c] KTH Royal Institute of Technology, School of Biotechnology, Division of Industrial Biotechnology, AlbaNova University Center, SE-106 91 Stockholm, Sweden
[d] Bio-Prodict, Nieuwe Marktstraat 54E, 6511 AA Nijmegen, The Netherlands

ARTICLE INFO

Available online 7 January 2015

Keywords:
Protein function
Annotation
PLP-dependent enzymes
Bioinformatics
Biocatalysis
Enzyme discovery
Transaminase

ABSTRACT

In this review we analyse structure/sequence–function relationships for the superfamily of PLP-dependent enzymes with special emphasis on class III transaminases. Amine transaminases are highly important for applications in biocatalysis in the synthesis of chiral amines. In addition, other enzyme activities such as racemases or decarboxylases are also discussed. The substrate scope and the ability to accept chemically different types of substrates are shown to be reflected in conserved patterns of amino acids around the active site. These findings are condensed in a sequence–function matrix, which facilitates annotation and identification of biocatalytically relevant enzymes and protein engineering thereof.

Contents

Abbreviations: AA, amino acid; AAA, amino acid amide; 3AcOc, 3-acetyloctanal; AcOrn, *N*-acetylornithine; DAIB, D-aminoisobutyrate; ATA, amine transaminase; CoAβAA, coenzyme A β-amino acid thioester; DABA, α,γ-diaminobutyrate; DAPA, 7,8-diaminopelargonic acid; DGD, 2,2-dialkylglycine decarboxylase; DTS, dethiobiotin synthase; $fumB_1$, fumonisin B_1; GABA, γ-aminobutyrate; glyox, glyoxylate; GSAM, glutamate-1-semialdehyde aminomutase; Lysε, lysine ε-amino group; $HfumB_1$, hydrolysed fumonisin B_1; Orn, ornithine; KAPA, 7-keto-8-aminopelargonic acid; αKG, α-ketoglutarate; OrnTL DB, ornithine transaminase-like database; βAla, β-alanine; βPhe, β-phenylalanine; PLP, pyridoxal 5'-phosphate; PUT, putrescine; pyr, pyruvate; SAM, *S*-adenosylmethionine; SuOrn, *N*-succinylornithine; TA, transaminase; tau, taurine

* Corresponding author. Tel: +49 3834 8622832; fax: +49 3834 86 794391.
** Corresponding author. Tel.: +49 3834 864367; fax: +49 3834 86 794367.
E-mail addresses: matthias.hoehne@uni-greifswald.de (M. Höhne), uwe.bornscheuer@uni-greifswald.de (U.T. Bornscheuer).

http://dx.doi.org/10.1016/j.biotechadv.2014.12.012
0734-9750/

1. Introduction

1.1. Motivation and learning objectives

What's the function of a certain gene or protein? Answering this question precisely is still a challenging, but very important task. An overwhelming number of potentially interesting enzymes for biocatalysis are available in public protein databases. However, this resource is only partially useful, because often the function and properties of enzymes cannot be predicted reliably. This review exemplifies how structural knowledge of enzymes and bioinformatics tools can be integrated to increase the precision of function prediction. As an example, we analysed enzymes of the PLP fold type I superfamily with special focus on class III transaminases.

With the help of the review, the reader should be able to:

- understand the fascinating mechanisms and features that govern reaction and substrate specificity of PLP-fold type I enzymes,
- understand how bioinformatics tools and structural knowledge can be combined to study structure–function relationships,
- understand how the enzymes' activities are reflected in small amino acid sequence fingerprints,
- take a class III transaminase amino acid sequence and easily assign the most probable function (out of 28 different known functions),
- apply this knowledge to guide experiments for the discovery of novel enzymes,
- apply the guidelines and tools covered in this review to analyse other enzyme superfamilies

1.2. How the review is structured and where do I find what?

Some basic introduction about the diversity of PLP chemistry, PLP-dependent enzyme classification and the biotechnological relevance of transaminases is given in the introductory sections 1.3 and 1.4. The section 1.5 introduces the active site fingerprint concept, which forms the basis of our structure–function relationship analysis. The most important terms and concepts of sections 1.3, 1.4 and 1.5, which are used throughout the review, are summarised in Boxes 1 and 2. Section 2 condenses all information from literature and our bioinformatic analysis: first, in section 2.1 we provide a brief description of the algorithms behind 3DM, the bioinformatics platform used for our analyses. General structure and sequence features of the class III transaminase family and specificity determining residues are analysed in sections 2.2 and 2.3. Section 2.4 presents the sequence–activity matrix, the central part of our analysis. It shows a correlation of the function of different proteins with amino acid patterns of a few active site residues (fingerprint). The most important structural details behind these analyses are presented in section 3. In this section we aim to illustrate the artful mechanisms and active site adaptations that facilitated the development of 28 different enzyme activities. On the one hand, specificity is created by providing a binding pocket that is complementary to the substrate in shape and polarity and provides electrostatic interactions. On the other hand, different mechanisms render the active site very flexible and allow two or more chemically different substrates to bind in the same pocket (so called dual substrate recognition). An overview of section 3 is given by Table 3, which contains structures of substrates and