Computers in Biology and Medicine 43 (2013) 471–480

Contents lists available at SciVerse ScienceDirect

Computers in Biology and Medicine

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

An algorithm for rapid computational construction of metabolic networks: A cholesterol biosynthesis example

Aleš Belič [a,*], Denis Pompon [b], Katalin Monostory [c], Diane Kelly [d], Steven Kelly [d], Damjana Rozman [e]

[a] Faculty of Electrical Engineering, University of Ljubljana, SI-1000 Ljubljana, Slovenia
[b] Centre de Genetique Moleculaire, Centre National de la Recherche Scientifique, F-91190 Gif-sur-Yvette, France
[c] Chemical Research Center, Hungarian Academy of Sciences, H-1025 Budapest, Hungary
[d] Institute of Life Science, School of Medicine, University of Swansea, SA2 8PP Swansea, United Kingdom
[e] Center for Functional Genomics and Bio-Chips, Institute of Biochemistry, Faculty of Medicine, University of Ljubljana, SI-1000 Ljubljana, Slovenia

ARTICLE INFO

Article history:
Received 21 November 2011
Accepted 16 February 2013

Keywords:
Systems biology
Binary coding
Modelling
Simulation
Metabolic networks

ABSTRACT

Alternative pathways of metabolic networks represent the escape routes that can reduce drug efficacy and can cause severe adverse effects. In this paper we introduce a mathematical algorithm and a coding system for rapid computational construction of metabolic networks. The initial data for the algorithm are the source substrate code and the enzyme/metabolite interaction tables. The major strength of the algorithm is the adaptive coding system of the enzyme–substrate interactions. A reverse application of the algorithm is also possible, when optimisation algorithm is used to compute the enzyme/metabolite rules from the reference network structure. The coding system is user-defined and must be adapted to the studied problem. The algorithm is most effective for computation of networks that consist of metabolites with similar molecular structures. The computation of the cholesterol biosynthesis metabolic network suggests that 89 intermediates can theoretically be formed between lanosterol and cholesterol, only 20 are presently considered as cholesterol intermediates. Alternative metabolites may represent links with other metabolic networks both as precursors and metabolites of cholesterol. A possible cholesterol-by-pass pathway to bile acids metabolism through cholestanol is suggested.

1. Introduction

The complexity of metabolic networks represents a serious problem for their analysis and identification. The most problematic is discovery of alternative metabolic pathways, these usually convey a negligible portion of the whole metabolic flow; however, in some cases they represent escape routes and cause the adverse effects of drugs as described in [1].

Several methods for metabolic network identification have been developed to date, some have been designed to identify networks from experimental data [2–6], and some to computationally construct networks on the basis of molecular interactions and thermodynamics [7–15].

Network identification from experimental data is problematic due to the large variety and number of experiments required to sufficiently excite the entire network structure. As a network is not completely known in advance this requires cycles of experiments and analyses. Furthermore, chemical reactions in metabolic networks are mostly enzyme catalysed, introducing feedback mechanisms and further complicating network identification.

On the other hand computational reconstruction relies on the physics of the molecular interactions that govern chemical reactions, these may not be entirely known, or may be too complex to compute within a reasonable time frame. Therefore, enzyme–substrate relations are described by various molecular topology features that define substrate suitability for a specific enzyme, rather than by first principle physical laws. The active enzyme domain may fit to the whole substrate molecule, or only to a specific structure within the substrate molecule. As a consequence, some enzymes can accept a single, or a few different molecules as possible substrates, while others can accept a wide variety of different molecular structures as substrates and can be involved in many reactions. Depending on the specificity of the defined features, algorithms can either be used to order a set of pre-defined reactions into metabolic networks [7,8], or to compute a metabolic network by trying to find the suitable domains in several substrate candidates [10–13].

In this paper we present an algorithm for rapid computation of metabolic pathways. The algorithm was developed for studying

* Corresponding author. Tel.: +386 1 4768 762; fax: +386 1 4264 631.
E-mail address: ales.belic@fe.uni-lj.si (A. Belič).

0010-4825/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiomed.2013.02.017

the late part of cholesterol biosynthesis where high structural similarity between substrates causes poor enzyme specificity. The algorithm is useful for studying metabolic networks where one or a few pathways have already been identified and many alternative pathways are expected.

1.1. Cholesterol biosynthesis

The pathway of cholesterol biosynthesis [16–18] begins with acetyl coenzyme A. After many consecutive steps lanosterol is formed. Lanosterol is the first cyclic metabolite in the pathway (see Fig. 1 for the skeletal structure).

Cholesterol is formed from lanosterol in nine consecutive steps (Fig. 2).

Most commonly, the primary and the secondary pathways with several bridges in between are described in KEGG [16]. The primary pathway begins with the step catalysed by CYP51, while the rest of the enzymes are sorted as follows: TM7SF2, SC4MOL, NSDHL, HSD17B7, EBP, SC5DL, DHCR7, and DHCR24 (for the description of the enzyme symbols see Table 1). The secondary pathway begins with DHCR24 and the rest of the enzymes are ordered as in the primary pathway. DHCR24 represents a bridge between the primary and the secondary pathways as it can take any of the metabolites of the primary pathway as a substrate.

Lanosterol is known to be metabolised by CYP51 or DHCR24, however, suitable domains within this structure also could be predicted to interact with enzymes SC4MOL, EBP, and SC5DL. Similarly for several intermediates of the late cholesterol pathway, structural features that may interact with several enzymes of the pathway can also be found, suggesting the possibility of a higher complexity of the network than at first envisaged. This hypothesis was confirmed in a yeast model where a large number of alternate pathways have been described in strains engineered in the late part of ergosterol synthesis [19,20]. It was clearly evidenced that the yeast sterol Δ 7,8-isomerase can biotransform sterols in cholesterol and ergosterol pathways with different side chains, as well as with 14-methyl sterols and it is inhibited by structurally related compounds e.g. pregnenolone. Poor specificity is a general feature for side chain modification and cycle A and B targeted reactions where almost all combinations are possible. For example, the yeast sterol Δ 7,8-isomerase, sterol Δ 7-reductase, sterol Δ 5-desaturase work with the intermediates from cholesterol and ergosterol biosynthesis containing either a double, or a single bond at positions 22(23), 24(25) or 24(28). The yeast DHCR7 works both on cholesta- or ergosta-5,7- and 7-enols. The pre-zymosterol yeast enzymes tolerate the retention of a 14-methyl group in the substrate, while the 4- and 4′-methyl group in a substrate can be metabolised by the enzymes in the late part of the pathway [21,22]. Cholesterol biosynthesis is one of the central metabolic processes in the body, its deregulation is observed in several metabolic diseases (hypercholesteremia, diabetes, non-alcoholic fatty liver disease, etc.). With high prevalence in western populations, these are associated with high treatment costs [23,24]. The networks generated by the algorithm should provide better insight into the cholesterol biosynthesis and thus possibly suggest alternative treatments of the above mentioned diseases with less side effects than observed in the current treatment strategies, which would also reduce the additional treatment costs arising from side effects treatment.

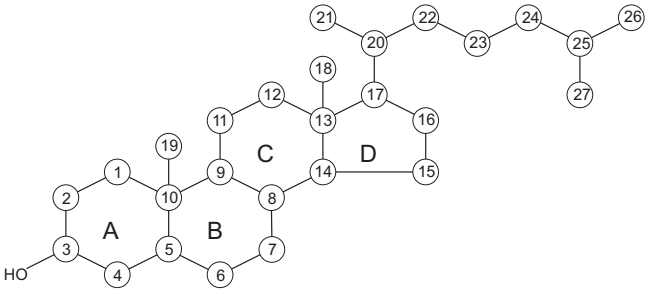

Fig. 1. Skeleton of lanosterol, cholesterol, and intermediates with numbered C-atoms and marked rings.

2. Materials and methods

2.1. The binary coding system and the algorithm

In order to computationally predict the possible intermediates in the formation of cholesterol we developed a procedure that enables modelling of pathways by encoding the metabolites and reactions with a simple binary code and by applying binary operations to the code. The coding system incorporates the information of the metabolites and the enzymes in a compatible

Table 1
Description of enzyme symbols.

Enzyme symbol	Enzyme description
CYP51	Lanosterol 14-α-demethylase
DHCR24	24-Dehydrocholesterol reductase
DHCR7	7-Dehydrocholesterol reductase
EBP	Emopamil binding protein (sterol C7,8-isomerase)
HSD17B7	17-β-Hydroxysteroid dehydrogenase type 7
NSDHL	NAD(P) dependent steroid dehydrogenase-like
SC4MOL	Sterol-C4-methyl oxidase-like
SC5DL	Sterol C5 desaturase
TM7SF2	14-*Delta*-reductase (transmembrane 7 superfamily 2)

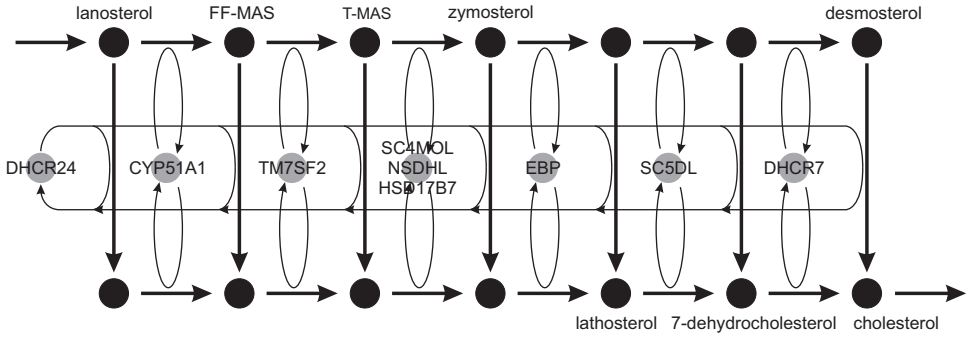

Fig. 2. The late part of the cholesterol biosynthesis network. Black circles, metabolites; gray circles, enzymes; thick black arrows, metabolic flow; thin black arrows, enzyme involvement. Some well known metabolites of the network are named.