Biochimica et Biophysica Acta 1801 (2010) 646–654

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

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

Systematic construction of a conceptual minimal model of plasma cholesterol levels based on knockout mouse phenotypes

Niek C.A. van de Pas [a,b,c,*], Ans E.M.F. Soffers [b], Andreas P. Freidig [a,1], Ben van Ommen [a], Ruud A. Woutersen [a,b,c], Ivonne M.C.M. Rietjens [b,c], Albert A. de Graaf [a]

[a] The Netherlands Organization for Applied Scientific Research (TNO) Quality of Life, Utrechtseweg 48, PO Box 360, 3700 AJ Zeist, The Netherlands
[b] Division of Toxicology, Wageningen University, Tuinlaan 5, 6703 HE Wageningen, The Netherlands
[c] TNO/WUR Centre for Innovative Toxicology, Tuinlaan 5, 6703 HE Wageningen, The Netherlands

ARTICLE INFO

Article history:
Received 1 December 2009
Accepted 12 February 2010
Available online 20 February 2010

Keywords:
HDL
LDL
Mouse Genome Database
Key genes
Plasma levels

ABSTRACT

Elevated plasma cholesterol, a well-known risk factor for cardiovascular diseases, is the result of the activity of many genes and their encoded proteins in a complex physiological network. We aim to develop a minimal kinetic computational model for predicting plasma cholesterol levels. To define the scope of this model, it is essential to discriminate between important and less important processes influencing plasma cholesterol levels. To this end, we performed a systematic review of mouse knockout strains and used the resulting dataset, named KOMDIP, for the identification of key genes that determine plasma cholesterol levels. Based on the described phenotype of mouse knockout models, 36 of the 120 evaluated genes were marked as key genes that have a pronounced effect on the plasma cholesterol concentration. The key genes include well-known genes, e.g., *Apoe* and *Ldlr*, as well as genes hardly linked to cholesterol metabolism so far, e.g., *Plagl2* and *Slc37a4*. Based on the catalytic function of the genes, a minimal conceptual model was defined. A comparison with nine conceptual models from literature revealed that each of the individual published models is less complete than our model. Concluding, we have developed a conceptual model that can be used to develop a physiologically based kinetic model to quantitatively predict plasma cholesterol levels.

1. Introduction

Cholesterol is an important molecule in fat metabolism, a precursor for vitamin D, and is involved in maintaining cellular integrity [1]. Elevated plasma concentrations of LDL-cholesterol (LDLc) have been correlated with the risk of atherosclerosis and cardiovascular diseases [2], which are leading causes of death in Western societies [3].

A large part of cholesterol research uses the mouse as a model organism [4] to benefit from many practical advantages such as short generation times and reduced genetic variation (inbred strains). Additionally, several modern biological techniques are applicable in mice, but not in humans. A very powerful one is the gene knockout technology [5], which allows direct study of the effect of individual genes *in vivo*. Using knockout mouse models, the involvement of many genes in cholesterol metabolism and plasma cholesterol levels has been studied. From these studies, it has become clear that a complex system of biochemical processes is involved in the enzymatic conversions and transport of cholesterol in the body. Unfortunately, the quantitative interplay of these processes *in vivo* is not fully understood. As a consequence, it is generally difficult to predict the effect of interventions on plasma cholesterol.

Physiologically based kinetic (PBK) models can be effective tools to assess this issue. PBK models allow to predict the effect of system perturbations (e.g., genetic defects, therapies), help to gain quantitative insight, and can play a role in translational research [6–8]. These computational models are generally built from conceptual models that present knowledge in an integrated fashion [9].

Several conceptual models of whole-body cholesterol metabolism have been published [4,10–12]. These models are graphical representations of the set of organs, metabolite pools, and fluxes that together comprise the most important enzymatic conversions in cholesterol metabolism and pathways of cholesterol transport between different organs. However, the conceptual models on cholesterol metabolism defined so far [4,10,12–18] do not explicitly focus on plasma cholesterol.

Therefore, the aim of the present study was to construct a comprehensive minimal conceptual model of cholesterol metabolism

Abbreviations: Ko, knockout; Wt, wild type; KOMDIP, Knockout Mouse Data Inventory of cholesterol Phenotype; LDLc, low-density lipoprotein cholesterol; HDLc, high-density lipoprotein cholesterol; CE, cholesteryl ester; C, cholesterol; TF, transcription factor; G6P, glucose-6-phosphate; CM, chylomicron

* Corresponding author. The Netherlands Organization for Applied Scientific Research (TNO) Quality of Life, Utrechtseweg 48, PO Box 360, 3700 AJ Zeist, The Netherlands. Tel.: +31 30 694 48 81; fax: +31 30 694 49 86.
E-mail address: niek.vandepas@tno.nl (N.C.A. van de Pas).
[1] Present address: Amsterdam Molecular Therapeutics, Meibergdreef 61, 1105 BA Amsterdam, The Netherlands.

1388-1981/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbalip.2010.02.009

in the mouse, explicitly focusing on processes that will affect plasma cholesterol. This will subsequently allow the development of a quantitative PBK model for prediction of plasma cholesterol concentrations and of the effect of cholesterol-lowering therapies in preclinical research. In order to avoid overparameterization, this model should contain as few processes and parameters as possible (i.e., it should represent a minimal conceptual model).

The choice of the processes to be included in the model was made on the basis of key genes. To discriminate key genes in determining plasma cholesterol concentrations from less important ones, we systematically reviewed the cholesterol phenotype of available knockout mouse models. These key genes were defined in two different ways: (1) genes were marked key genes type A if the cholesterol plasma level of the knockout mouse of the gene is highly affected compared to the wild-type level and (2) genes that lead to lethality early in embryogenesis when knocked out (i.e., no plasma cholesterol data available for the knock out) were labeled key genes type B, if the primary function of the gene could be assumed to directly affect plasma cholesterol concentrations.

The outcomes were used to construct a minimal conceptual model for plasma cholesterol levels that was subsequently compared with conceptual models described in literature [4,10,12–18].

2. Methods

2.1. Data set construction

To rank the impact of genes on cholesterol plasma levels and to define key genes type A (highly affecting plasma cholesterol levels in knockout strains compared to the wild type), an inventory of all relevant knockout mouse models was made using the Mouse Genome Database (available via http://www.informatics.jax.org) [19]. The database was searched for alleles that correspond to the phenotypic categories (1) 'abnormal cholesterol homeostasis' and (2) 'abnormal bile salt homeostasis'. The second search term was included since bile salts are a degradation product of cholesterol and bile salt metabolism is closely related to cholesterol metabolism [4]. These two categories are the most comprehensive phenotypic categories having a direct link with cholesterol. They comprise the following 16 different daughter categories: (1) 'decreased cholesterol level', (2) 'increased cholesterol level', (3) 'abnormal circulating cholesterol level', (4) 'decreased circulating cholesterol level', (5) 'increased circulating cholesterol level', (6) 'abnormal circulating HDL cholesterol level', (7) 'decreased circulating HDL cholesterol level', (8) 'increased circulating HDL cholesterol level', (9) 'abnormal circulating LDL cholesterol level', (10) 'decreased circulating LDL cholesterol level', (11) 'increased circulating LDL cholesterol level', (12) 'abnormal circulating VLDL cholesterol level', (13) 'decreased circulating VLDL cholesterol level', (14) 'increased circulating VLDL cholesterol level', (15) 'abnormal bile salt level', and (16) 'increased bile salt level'. The final database search was performed on August 28, 2008. To increase the homogeneity of the data set, knockin and transgenic mouse strains were not included and this also holds for mouse strains that carried alleles with non-targeted mutations. The latter included Quantitative Trait Loci, chemically induced and spontaneous mutations.

For all alleles, data on the wild type strain and the reference to the literature source were obtained from the Mouse Genome Database. This information was linked to data on diet, gender, total plasma cholesterol for both wild-type (wt) and knockout (ko), plasma LDLc levels (wt + ko), and plasma HDLc levels (wt + ko) extracted from the original publications. This resulted into a manually checked KnockOut Mouse Data Inventory of cholesterol Phenotype of models (KOMDIP) containing plasma cholesterol phenotype data together with data on the experimental design.

2.2. Key genes type A identified from plasma data in KOMDIP

To distinguish the most important genes affecting plasma cholesterol levels, a gene effect (E) was calculated for every experiment with a given knockout strain present in KOMDIP.

$$E = {}^{2}\log\frac{[TC]_{ko}}{[TC]_{wt}} \quad (1)$$

In Eq. (1), $[TC]_{ko}$ stands for the reported average total plasma cholesterol concentration of the knockout mouse and $[TC]_{wt}$ for the average total plasma cholesterol concentration of the wild-type counterpart. E is based on a 2-log ratio as commonly used in gene expression analysis, because it produces a continuous spectrum of values for increased and decreased cholesterol concentrations as compared to wild type [20]. Genes corresponding to absolute values of E larger than a cutoff value are henceforth referred to as key genes type A. Since plasma cholesterol levels typically vary approximately a factor of two between background strains [21], we chose the cutoff value to be 1 or −1. An E value larger than +1 means that the knockout mouse has more than twofold higher cholesterol levels than the wild type and an E value smaller than −1 means that the knockout mouse has a more than twofold lower cholesterol concentration than the wild type.

2.3. Impact of gender and diet

To judge the importance of gender and dietary effects, wild-type total plasma cholesterol data (knockout data were not considered since these would focus on the genetic factors) were distributed in different groups according to the reported experimental design. Differences in average total plasma concentrations between the groups were tested for significance using the Wilcoxon rank sum test as implemented in the Statistics toolbox of MATLAB (version 7.5 R2007b). Tests for normality were performed using the Lilliefors test as implemented in the same toolbox. Differences with $p<0.05$ were considered significant.

2.4. Key genes type B identified from non viable knockout mice

An obvious reason why genes might be missed in the previous analysis is that the corresponding knockout mouse is not viable. To find the genes that lead to a non-viable mouse strain and are tightly linked to plasma cholesterol levels, the Mouse Genome Database was searched for targeted alleles that were grouped in the categories 'embryonic lethality at implantation', 'embryonic lethality before implantation', 'embryonic lethality before turning of embryo', 'embryonic lethality before somite formation', and 'embryonic lethality during organogenesis'. References to the corresponding literature sources were obtained from the Mouse Genome Database. To verify whether these genes can be considered a key gene type B, Pubmed was used to search the titles and abstracts of these publications for the term 'cholesterol'. If the term *cholesterol* is present and the gene was not integrated in the KOMDIP data set and the hypothetical function of the gene is directly related to (plasma) cholesterol, the gene was marked as a key player type B.

2.5. Conceptual model

The key genes of both types, which resulted from the aforementioned analyses, were used to construct a conceptual model including the processes that influence plasma cholesterol. The choices on which processes are relevant enough to include in the model were made based on the (suggested) functions of the key genes. Since we aim to construct a minimal conceptual model, we chose to include only those genes that can be directly appointed to a specific metabolic process in