Modelling Gene-Protein-Reaction Associations on an FPGA

Macauley Coggins

May 2019

Abstract

Genome-Scale metabolic models have proven to be incredibly useful.Allowing researchers to model cellular functionality based upon gene expression. However as the number of genes and reactions increases it can become computationally demanding. The first step in genome-scale metabolic modelling is to model the relationship between genes and reactions in the form of Gene-Protein-Reaction Associations (GPRA). In this research we have developed a way to model GPRAs on an Altera Cyclone II FPGA using Quartus II programmable logic device design software and the VHDL hardware description language. The model consisting of 7 genes and 7 reactions was implemented using 7 combinational functions and 14 I/O pins. This model will be the first step towards creating a full genome scale metabolic model on FPGA devices which we will be fully investigating in future studies.

1 Keywords

FPGA, Genome-scale metabolic models, GPRA, Genes, Reactions, FBA

2 Introduction

Genome scale metabolic models have gone a long way since the first genome-scale metabolic model which was developed in 1995 for Haemophilus influenzae a single celled organism [1]. This model was based on assembly of unselected pieces of DNA from the whole chromosome had been applied to obtain the complete nucleotide sequence. However the preferred approach today is Flux Balance Analysis (FBA). FBA can be traced back to Papoutsakis et al who in 1984 demonstrated the first use of flux balance equations from metabolic maps [6]. This approach has an advantage over more traditional approaches as it requires very little information about enzyme kinetics or concentration of metabolites in the organism to be described. Instead it assumes that the system will be steady state ie that fluxes are balanced and that the system will always be optimal. The steady state assumption means that the system can be reduced to a set of linear equations which can then be solved using a linear solver such as Gurobi [5], [3]. The first step in constructing a genome scale metabolic model is to describe the relationship between genes and the reactions they encode for. These are called Gene-Protein-Reaction Associations (GPRAs). These are implemented as boolean rules and depending on the reaction these rules can become quite complex. In their simplest form however they model isozymes and enzyme complexes where an isoenyme can be described as

$$g_1 \vee g_2 \tag{1}$$

where g_1 and g_2 are the genes coding for the reaction enzyme.

Enzyme complexes can be described as

$$g_1 \wedge g_2 \tag{2}$$

One real world example of a GPRA would be (HGNC:6541 $\vee$ HGNC:6535 $\vee$ HGNC:19708) which would describe the reaction Lactate Dehydrogenase. All in-silico models so far have been software based however dedicated hardware such as FPGAs may be a promising alternative. FPGAs as a platform would provide several advantages as they would provide faster computational speeds. Given that GPRAs are essentially boolean rules these could be modelled naturally on an FPGA using in built logic.

3 Materials and Methods

3.1 Selection of GPRAs to be modelled

For this research pre-existing GPRAs had to be found from pre-existing literature. GPRA's were found from Virtual Metabolic Human (VIHM) a site that curates genes,reactions, metabolites and diseases as well as provide information on how all these components interact including GPRAs [4]. For this research GPRAs and reactions that were associated with Lactate Dehydrogenase were selected and are outlined in Table 1. Furthermore Lactate Dehydrogenase is an enzyme that catalyzes the conversion of pyruvate to lactate. The main reason for selecting these are that many genes involved in Lactate Dehydrogenase are also involved in a number of other reactions. Furthermore Lactate Dehydrogenase is a notable enzyme that has been studied for its potential role as a biomarker in tumours [2].

Reaction	GPRA
LDHLM	6541 or 6535 or 19708
LDHL	(6541 and 6535) or (28335) or (6544) or (6535) or (30866) or (6541) or (21481)
2HBO	(21481) or (6544) or (30866) or (6541 and 6535) or (28335) or (6541) or (6544)
GLX01	(6535) or (30866) or (6541 and 6535) or (19708) or (21481) or (6541) or (28335) or (6544)
HPYRR2x	(6544) or (19708) or (6541) or (28335) or (6541 and 6535) or (21481) or (6535)
MCLOR	(21481) or (6541 and 6535) or (30866) or (6544) or (6541) or (28335) or (6535)
R0173	(6541) or (6535) or (6544)

Table 1: Modelled Reactions and their GPRAs. Note the gene IDs are HGNC

3.2 Design and Synsthesis on Quartus II

Quartus II 64-bit is software for analysis and synthesis of HDL designs. Using Quartus II GPRAs were implemented using VHDL a hardware description language used in the synthesis of HDL designs. Genes were implemented as

Std logic input signals using while reactions were implemented as Std logic output signals.

GPRAs were then mapped to the reactions using Boolean combinations of genes. Due to the boolean nature of the model gene expression of any gene is simplified to an On/Off state. If a gene signal is true then this corresponds to that particular gene being knocked in. The same can be said for reactions where if the reaction output is true then the reaction is turned on. Next the GPRAs were compiled and a vector waveform file created.A Block Diagram was also created as seen in Figure 1. The vector waveform file provides a way to analyze the functionality of the compiled design whose output is a timing diagram allowing the user to see the inputs and outputs as seen in Figure 1. In the waveform file the genes were grouped together and counted from 0000000 to 1111111 in grey code in order to go through each combination of gene knock-in/knock-out

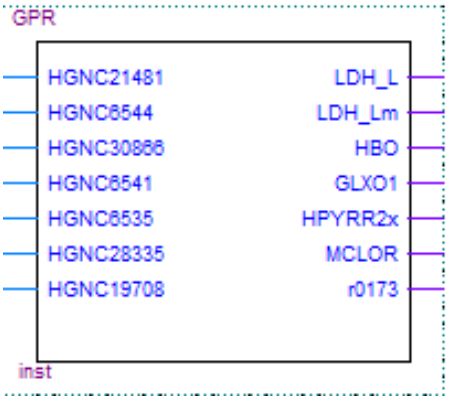

Figure 1: Block Diagram of GPRA Model. Note the genes as input signals to the left and reactions as outputs to the right

4 Results

The results shown in Figure 2 show all combinations of gene knock-in and knock-out were simulated in 989ns with each knock-in/knock-out taking 10ns.The waveform shows that reactions were modelled correctly with no logical hazards. However the risk of hazards is increased as the time between input changes is decreased. Analysis and Synthesis summary shown in Figure 3 shows that the model is very compact as it uses only 7 combinational functions and 14 pins.

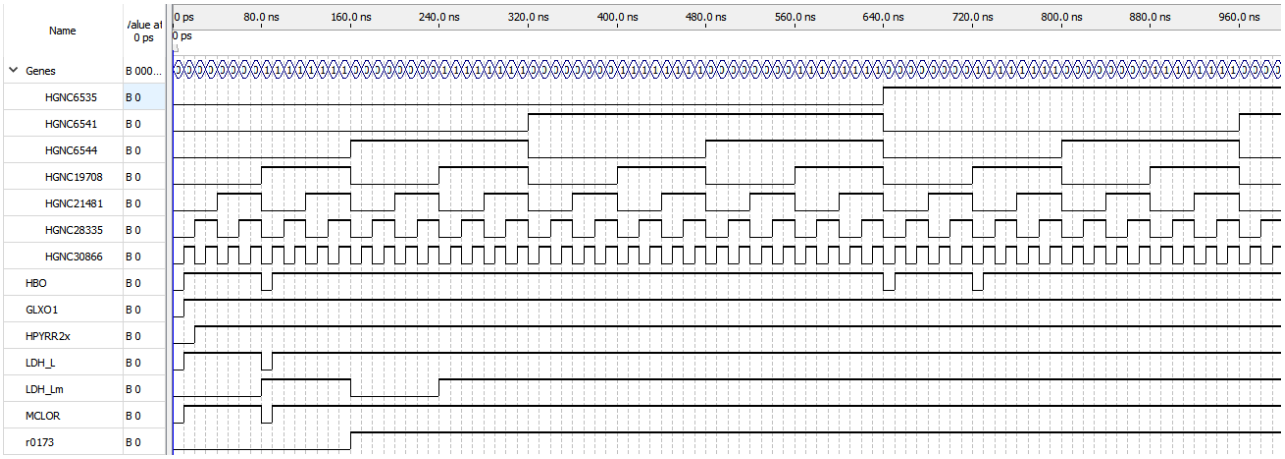

Figure 2: Timing Diagram in Quartus II

5 Discussion

From the results it can be seen that modelling GPRAs on an FPGA could be a very promising first step in creating genome scale metabolic models on hardware. This is given that modelling GPRAs is very natural on an FPGA as they are boolean rules that can be implemented well using in-built logic in FPGA devices. Furthermore it can be seen that FPGAs provide an advantage in terms of speed compared to software implemented models given that the hardware in the FPGA will be fully dedicated to the model. From Figure 2 it is seen that all combinations of gene knock-in and knock-out were completed in 989 ns. However for a fully fledged genome scale model to be implemented on FPGAs more research will have to be done to implement

Flux Balance Analysis on these devices which we will investigate in future studies.

Analysis & Synthesis Summary	
Analysis & Synthesis Status	Successful - Sun Apr 28 14:31:09 2019
Quartus II 64-Bit Version	13.0.1 Build 232 06/12/2013 SP 1 SJ Web Edition
Revision Name	GPR
Top-level Entity Name	GPR
Family	Cyclone II
Total logic elements	7
Total combinational functions	7
Dedicated logic registers	0
Total registers	0
Total pins	14
Total virtual pins	0
Total memory bits	0
Embedded Multiplier 9-bit elements	0
Total PLLs	0

Figure 3: Analysis and Synthesis Summary

References

[1] R.D. Fleischmann, M.D. Adams, and O. White. Whole-genome random sequencing and assembly of haemophilus influenzae rd. *Science*, 269(5223), 7 1995.

[2] Vladimir Jurisic, Sandra Radenkovic, and Gordana Konjevic. The actual role of ldh as tumor marker, biochemical and clinical aspects, 2015.

[3] Jan Kronqvist, David E. Bernal, Andreas Lundell, and Ignacio E. Grossmann. A review and comparison of solvers for convex minlp. *Optimization and Engineering*, 20(2):397–455, Jun 2019.

[4] Alberto Noronha, Jennifer Modamio, Yohan Jarosz, Elisabeth Guerard, Nicolas Sompairac, German Preciat, Anna Dröfn Daníelsdóttir, Max Krecke, Diane Merten, Hulda S Haraldsdóttir, Almut Heinken, Laurent Heirendt, Stefanía Magnúsdóttir, Dmitry A Ravcheev, Swagatika Sahoo, Piotr Gawron, Lucia Friscioni, Beatriz Garcia, Mabel Prendergast, Alberto Puente, Mariana Rodrigues, Akansha Roy, Mouss Rouquaya, Luca Wiltgen, Alise Žagare, Elisabeth John, Maren Krueger, Inna Kuperstein, Andrei Zinovyev, Reinhard Schneider, Ronan M T Fleming, and Ines Thiele. The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Research*, 47(D1):D614–D624, 10 2018.

[5] Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. What is flux balance analysis?, Mar 2010.

[6] Eleftherios Terry Papoutsakis. Equations and calculations for fermentations of butyric acid bacteria. *Biotechnology and Bioengineering*, 26(2):174–187.

6 Supplementary Materials

VHDL Code can be found in GitHub: https://github.com/mcoggins96/Modelling-Gene-Protein-Reaction-Associations-on-an-FPGA

7 Dedication

This paper is dedicated to Francesca Marcovecchio who supported me during the writing of this paper.