Computers in Biology and Medicine 61 (2015) 19–35

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Evolutionary growth of certain metabolic pathways involved in the functioning of GAD and INS genes in Type 1 *Diabetes Mellitus*: Their architecture and stability

Somnath Tagore [a,*], Rajat K. De [b]

[a] Department of Biotechnology and Bioinformatics, Padmashree Dr. D.Y. Patil University, Navi Mumbai 400614, India
[b] Machine Intelligence Unit, Indian Statistical Institute, Kolkata 700108, India

ARTICLE INFO

Article history:
Received 1 August 2014
Accepted 13 March 2015

Keywords:
Community structure
Evolution
Hamming distance
Hasse diagram
Payoff matrix
Tanimoto coefficient

ABSTRACT

Background: Studying biochemical pathway evolution for diseases is a flourishing area of Systems Biology. Here, we study Type 1 *Diabetes Mellitus* (T1D), focusing on growth of *glutamate*, *β-alanine*, *taurine and hypotaurine*, and *butanoate* metabolisms involved in onset of *GAD* and *INS* genes in *Homo sapiens* with comparative analysis in non-obese diabetic *Mus musculus*, biobreeding Diabetes-prone *Rattus norvegicus*, *Pan troglodytes*, *Oryctolagus cuniculus*, *Danio rerio* and *Drosophila melanogaster* respectively.

Methods: We propose an algorithm for growth analysis for four metabolic pathways involved in T1D. It has three modules, *pattern* finding, *interaction* identification and *growth* detection. The first module identifies patterns using *Community structures* using *Hamming distances* and the *Tanimoto coefficient*. We have performed functional analysis by representing patterns using ODEs, and identified *Stoichiometric*, *Gradient* and *Jacobian* matrices. The second module identifies interactions among patterns using cut-sets and network-partitioning by 'Divide-and-conquer'. The third module identifies functions of patterns using interactions, thereby highlighting their nature of growth.

Results: We observed that metabolites that are genetically robust and resist alterations against stable state during evolution, account for emergence of a scale-free network.

Discussion: New modules get acquired to the fundamental cluster in a preferential manner, an instance of *micro-evolution theory*. For instance, *(S)-3-hydroxy butanoyl-CoA, acetoacetyl-CoA, acetoacetate, acetyl-CoA, (S)-3-hydroxy-3-methyl glutaryl-CoA* acts as a fundamental cluster in *butanoate* metabolism. Moreover, the interactions among metabolites are divergent in nature.

1. Introduction

The behavior and evolutionary properties of metabolic pathways have been studied for a long time using architectural and network-based methods [1]. Among various research attempts undertaken in network-based studies, an essential one is based on identifying meaningful patterns within metabolic pathways. It aims at identifying essential regions within metabolic pathways and is useful for studying evolution. Furthermore, it also identifies functionally important metabolites that help in determining alternate paths in metabolic pathways [2,3]. This also help in identifying analogous pathways that exist in different species and provide a key step in discerning essential properties of evolution in terms of interactions among metabolic pathways [4]. To analyze the importance of interactions among metabolic pathways in evolution, we have selected certain metabolic pathways involved in the prevalence of Type 1 *Diabetes Mellitus* (T1D).

Diabetes Mellitus (DM), one of the most prevalent diseases in India, is mostly due to the food habit and lifestyle. We have implemented our proposed algorithm on certain metabolic pathways involved in the emergence of Type 1 *Diabetes Mellitus* (T1D). It is a disorder characterized by high blood sugar, resulting in improper production of insulin. It results in failure and dysfunction of multiple organs. Moreover, till date around 250 genes have been studied for analyzing their role with T1D, one of which is *anti-Glutaric acid decarboxylase* (GAD) that leads to autoimmune processes for *β*-cell destruction [5]. Similarly, the *insulin* gene (INS) is the second well established susceptibility locus in *Diabetes Mellitus*, contributing to about 10% towards T1D susceptibility [6]. We have considered four metabolic pathways which have shown role in the expression of *GAD* and *INS* genes, namely, *glutamate*, *β-alanine*, *taurine and hypotaurine*

* Corresponding author.
E-mail addresses: somnathtagore@yahoo.co.in (S. Tagore), rajat@isical.ac.in (R.K. De).

http://dx.doi.org/10.1016/j.compbiomed.2015.03.012
0010-4825/

and *butanoate* metabolisms, for the purpose of studying their development from an evolutionary point-of-view. This study is very critical, as analyzing the growth history and lineage of these metabolic pathways can lead to understanding the circumstances and role which might be responsible for a diseased state.

The major hypotheses proposed till date regarding metabolic pathway evolution are based on gene duplication events. One of the earliest attempts in this regard has been the 'Retrograde' hypothesis, which suggested that the biosynthesis of metabolites requires sequential activities and sub-step conversions of precursors [7]. Alternatively, the 'Granick' hypothesis suggested that biosynthesis of metabolites are results of forward evolution from simple precursor compounds [8]. Likewise, the 'Patchwork' hypothesis suggested that metabolic pathways have evolved through the reactions of certain primitive enzymes with a variable number of chemically related metabolites, followed by duplication events, resulting in metabolite diversification [9]. Similarly, the 'micro-evolution' theory states that certain divergent features are significant among growing entities in metabolic pathways [10]. Lastly, the 'Semienzymatic' origin of metabolic pathways suggested that certain prebiotic stable metabolites, mediated by specific and non-specific enzymes resulted in their evolution.

Our proposed algorithm has three modules, *first*, 'pattern finding' identifying patterns in metabolic pathways based on structural and functional features. We have used *Hamming distance*, the *Tanimoto coefficient* and *Community structures* for detecting meaningful patterns [11] (discussed in 'Methodology' section). The *second* module, i.e., 'interaction identification' identifies interactions based on the patterns [12], wherein interaction is 'the effect that is conferred upon one biological moiety to another'. Fig. 1 illustrates interactions occurring within carbohydrate metabolism in *H. sapiens* at levels of representations [13]. We use cut-sets and a 'Divide-and-conquer' strategy for studying interactions [14]. For example, in Fig. 2, *butanoate* metabolism is partitioned into S_1 and S_2 by removing a link between *acetoacetyl CoA* and *(S)-3-hydroxybutanoyl CoA*. The *third* module, i.e., 'growth detection' identifies growth using the patterns, their interactions and certain perturbations, for which we have associated weights to metabolites and reaction links (*butanoate* metabolism in Fig. 3). But, analyzing metabolic pathways on the basis of *architectural* features only may not facilitate functional significance. For this purpose, we have identified *cycles* and *Hasse* [15] (*qualitative*), ODEs for identifying *Stoichiometric* [16], *Gradient* [17], *Jacobian* [18] and *payoff* matrices [19] (*quantitative*); and 'bi-directional best hit criterion' and nutrient availability (*functional*) features respectively. For evolutionary studies, we have studied scale-free properties [20] (discussed in 'Methodology' section) (Fig. 4), wherein the evolutionary growth resides on novel link attachment in existing metabolic network, as per the 'micro-evolution' theory. We term this evolutionary process as 'divergent root-to-leaf' growth model (discussed in detail in 'Methodology' section).

2. Methodology

In this section we describe the proposed methodology. We have integrated the pathway related information from 36 public repositories. We initiate with this integration followed by describing the proposed methodology.

2.1. Integration and construction of glutamate, β-alanine, taurine and hypotaurine and butanoate metabolic pathways

We have performed this integration step, since none of the repositories have complete information about the metabolic pathways and suffers from redundancy too. We observed that in case of *glutamate* metabolic pathway, the highest number of links (12) is predicted by KEGG, followed by Gold.db (11), HumanCyc (11), HMDB (11), INOH (11), Kappa-View4 (11), KOBAS (11), MetaCyc (11), MANET (11), PathCase (11) and Pathguide (11). Likewise, the repositories predicting lowest number of links (1) are Ambion Pathway Atlas,

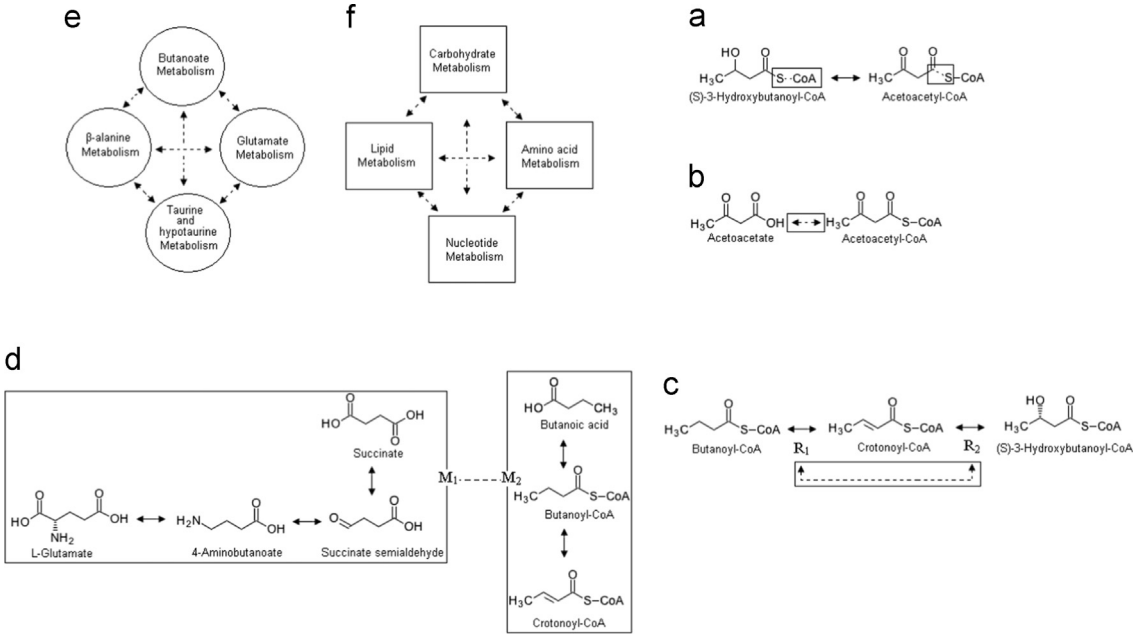

Fig. 1. Types of interactions at (a) atomic level interactions occur among atoms present in metabolites within a metabolic reaction in butanoate metabolism,viz., S-CoA, S-C, (b) metabolite level interactions exist among different metabolites within reactions in butanoate metabolism, viz., among acetoacetyl CoA and acetoacetate, (c) reaction level interactions constitute interactions among the various reactions within a network (denoted by R_i's) in butanoate metabolism, viz., between R1 and R2, (d) module level interactions are existent among different reaction blocks (groups) within pathways (denoted by M_i's) as shown between M_1 and M_2, created on the basis of some common occurring features, (e) pathway level interactions constitute interactions among various pathways present within a metabolic network, as in between butanoate metabolism and other metabolic pathways, (f) network level interactions constitute interactions within various metabolic networks in a cell as illustrated interactions within T1D metabolic network and other metabolic networks within a single cell; dotted lines denote interactions.