Identifying proteins and metabolic pathways associated to the neuroprotective response mediated by tibolone in astrocytes under an induced inflammatory model.

Daniel Camilo Osorio Hurtado

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de Ingeniería de Sistemas e Industrial
Bogotá, Colombia
2016

Identifying proteins and metabolic pathways associated to the neuroprotective response mediated by tibolone in astrocytes under an induced inflammatory model.

Daniel Camilo Osorio Hurtado

Tesis presentada como requisito parcial para optar al título de:
Magister en Bioinformática

Directora:
Ph.D. Janneth Gonzalez Santos

Co-Director:
Ph.D. Andrés Mauricio Pinzón

Línea de Investigación:
Nombrar la línea de investigación en la que enmarca la tesis o trabajo de investigación
Grupo de Investigación:
Bioinformática y Biología de Sistemas Computacional

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de Ingeniería de Sistemas e Industrial
Bogotá, Colombia
2016

Resumen

El resumen es una presentación abreviada y precisa (la NTC 1486 de 2008 recomienda revisar la norma ISO 214 de 1976). Se debe usar una extensión máxima de 12 renglones. Se recomienda que este resumen sea analítico, es decir, que sea completo, con información cuantitativa y cualitativa, generalmente incluyendo los siguientes aspectos: objetivos, diseño, lugar y circunstancias, pacientes (u objetivo del estudio), intervención, mediciones y principales resultados, y conclusiones. Al final del resumen se deben usar palabras claves tomadas del texto (mínimo 3 y máximo 7 palabras), las cuales permiten la recuperación de la información.

Palabras clave: (máximo 10 palabras, preferiblemente seleccionadas de las listas internacionales que permitan el indizado cruzado).

A continuación se presentan algunos ejemplos de tesauros que se pueden consultar para asignar las palabras clave, según el área temática:

Artes: AAT: Art y Architecture Thesaurus.
Ciencias agropecuarias: 1) Agrovoc: Multilingual Agricultural Thesaurus - F.A.O. y 2)GEMET: General Multilingual Environmental Thesaurus.
Ciencias sociales y humanas: 1) Tesauro de la UNESCO y 2) Population Multilingual Thesaurus.
Ciencia y tecnología: 1) Astronomy Thesaurus Index. 2) Life Sciences Thesaurus, 3) Subject Vocabulary, Chemical Abstracts Service y 4) InterWATER: Tesauro de IRC - Centro Internacional de Agua Potable y Saneamiento.
Tecnologías y ciencias médicas: 1) MeSH: Medical Subject Headings (National Library of Medicine's USA) y 2) DECS: Descriptores en ciencias de la Salud (Biblioteca Regional de Medicina BIREME-OPS).
Multidisciplinarias: 1) LEMB - Listas de Encabezamientos de Materia y 2) LCSH- Library of Congress Subject Headings.

También se pueden encontrar listas de temas y palabras claves, consultando las distintas bases de datos disponibles a través del Portal del Sistema Nacional de Bibliotecas[1], en la sección Recursos bibliográficos.ºpción "Bases de datos".

Abstract

Es el mismo resumen pero traducido al inglés. Se debe usar una extensión máxima de 12 renglones. Al final del Abstract se deben traducir las anteriores palabras claves tomadas del

[1] ver: www.sinab.unal.edu.co

texto (mínimo 3 y máximo 7 palabras), llamadas keywords. Es posible incluir el resumen en otro idioma diferente al español o al inglés, si se considera como importante dentro del tema tratado en la investigación, por ejemplo: un trabajo dedicado a problemas lingüísticos del mandarín seguramente estaría mejor con un resumen en mandarín.

Keywords: palabras clave en inglés(máximo 10 palabras, preferiblemente seleccionadas de las listas internacionales que permitan el indizado cruzado)

Content

1 Introduction

The research presented in this thesis is organized in four manuscripts for publication (chapters 2 to 5). Chapter 2 shows:.Chapters 3 to 5 shows:

2 Neuroprotective effects of tiboline during astrocytic metabolic inflammation

Written by: *Daniel Osorio, Janneth Gonzalez and Andrés Pinzón-Velasco.*

3 minval: An R package for MINimal VALidation of stoichiometric reactions

Written by: *Daniel Osorio, Janneth Gonzalez and Andrés Pinzón-Velasco.*

3.1. Abstract

The genome-scale metabolic reconstructions, a compilation of all stoichiometric reactions that can describe the entire cellular metabolism of an organism, have become an indispensable tool for our understanding of biological phenomena, covering fields that range from systems biology to bioengineering. Evaluation of metabolic reconstructions are generally carried through Flux Balance Analysis, an optimization method where the biological sense of optimal solution is sensitive to thermodynamic unbalance, caused by the presence of stoichiometric reactions whose compounds are not produced or consumed in any other reaction (orphan metabolites) and by mass unbalanced stoichiometric reactions. The **minval** package was designed as a tool to identify orphan metabolites and the mass unbalanced reactions in a set of stoichometry reactions, it also permits to extract all reactants, products, metabolite names and compartments from a metabolic reconstruction, moreover specific functions to map compound names associated to the Chemical Entities of Biological Interest (ChEBI) database are also included.

3.2. Introduction

A chemical reaction is a process where a set of chemical compounds called *reactants* are transformed into another compounds called *products* [**?**]. The accepted way to represent a chemical reaction is called a *stoichiometric reaction*, where reactants are placed on the left and the products on the right separated by an arrow which indicates the direction of the reaction as is showed in the equation 3-1 [**?**]. In biochemistry a set of chemical reactions that transform a substrate into a product, after several chemical transformations is called a metabolic pathway [**?**]. The compilation of all stoichiometric reactions included in all metabolic pathways that can describe the entire cellular metabolism encoded in the genome of a particular organism is known as a *genome-scale metabolic reconstruction* [**?**] and has

become an indispensable tool for studying metabolism of biological entities at the systems level [**?**].

$$\overbrace{\underbrace{1}_{coefficient}\ \underbrace{cis-aconitic\ acid}_{metabolite\ name}\ \underbrace{[c_a]}_{compartment}\ +1\ water[c_a]}^{reactants}\ \underbrace{\Rightarrow}_{directionallity}\ \overbrace{1\ isocitric\ acid[c_a]}^{products} \quad (3\text{-}1)$$

Reconstruction of genome-scale metabolic models starts with a compilation of all known stoichiometric reactions for a given organism, as evidenced by the presence of enzyme coding genes in its genome. Thus the reactions in which these enzymes are known to participate in, are usually downloaded from specialized databases as KEGG [**?**], BioCyc [**?**], Reactome [**?**], BRENDA [**?**] and SMPDB [**?**], however the downloaded stoichiometric reactions are not always mass-charge balanced and don't represent complete pathways to construct a high-quality metabolic reconstruction [**?**, **?**]. The identification and curation of these type of reactions is a time consuming process which the researcher have to complete manually using available literature or experimental data [**?**].

Genome-scale metabolic reconstructions are usually interrogated through FBA (*Flux Balance Analysis*), an optimization method that allows us to understand the metabolic status of the cell, improve the production capability of a desired product or make a rapid evaluation of cellular physiology at genome-scale [**?**, **?**]. Nevertheless FBA is sensitive to thermodynamic unbalance, so in order to asses the validity of a biological extrapolation (i.e. an optimal solution) from a FBA analysis it is mandatory to avoid this type of unbalancing in mass conservation through all model reactions [**?**]. Another drawback when determining the validity of a metabolic reconstruction is the presence of reactions with compounds that are not produced or consumed in any other reaction (dead ends), generally known as orphan metabolites [**?**, **?**]. The presence of this type of metabolites can be problematic, since they lead to an artificial cellular accumulation of metabolism products which therefore bias our biological conclusions. Tracking these metabolites is also a time consuming process, which most of the time has to be performed manually or partially automatized by in-house scripting. Given that typical genome-scale metabolic reconstructions account for hundreds or thousands of biochemical reactions, the manually curation of these models is a task that can lead to both, the introduction of new errors and to overlook some others.

The most popular FBA implementations as COBRA and RAVEN includes similar functions (`checkMassChargeBalance` and `getElementalBalance` respectively) implemented under the commercial MATLAB® environment. These functions identify orphan metabolites and mass unbalanced reaction based in the chemical formula or the IUPAC International Chemical Identifier (InChI) supplied manually by the user for each metabolite included in the genome-scale metabolic reconstruction. With the aim to automatize the identification of orphan metabolites as well as the unbalanced stoichiometric reactions in a genome-scale metabolic reconstruction, we have developed the **minval** package. It includes thirteen functions to

evaluate mass balance and extract all reactants, products, orphan metabolites, metabolite names and compartments for a set of stoichiometric reactions, moreover specific functions to map compound names associated to the Chemical Entities of Biological Interest (ChEBI) database are also included.
For this version we use the included "`glugln`" dataset [**?**], 128 non-exchange/sink stoichiometric reactions from the reconstruction of the glutamate/glutamine cycle constructed in-house using the KEGG database , as an example for each function included in the **minval** package with the aim to show their potential use.

3.3. Installation and functions

`minval` includes thirteen functions and is available for download and installation from CRAN, the Comprehensive R Archive Network. To install and load it, just type:

```
> install.packages("minval")
> library(minval)
```

The `minval` package requires R version 2.10 or higher. Development releases of the package are available on the GitHub repository `http://github.com/gibbslab/minval`.

3.3.1. Inputs and syntaxis

The functions included in `minval` package take as input a string list with stoichiometric reactions. The data loading from traditional human-readable spreadsheets can be carried through other CRAN-available packages as `gdata`, `readxl` or `xlsx`. Each reaction string must contain metabolites, with an optional compartment label between square brackets. The metabolites should be separated by a plus symbol (`+`) between two blank spaces and may have just one stoichiometric number before the name. The reactants should be separated of products by an arrow using the following symbol `=>` for irreversible reactions and `<=>` for reversible reactions.

3.3.2. Syntax Validation

Flux Balance Analysis method is implemented in a variety of software under different programming languages. Some of the most popular implementations are `COBRA` [**?**] and `RAVEN` [**?**] under matlab language as well as `sybil` and `abcdeFBA` under R language. The `is.validsyntax` function validate the well accepted compartmentalized stoichiometric syntax (Equation 3-1) for several FBA implementations and returns a boolean value TRUE if syntax is correct. In this example we show the stoichiometric syntax for the inter-convertion of malate to fumaric acid and water in astrocytes cytoplasm.

```
> is.validsyntax("(S)-malate(2-)[c_a] <=> fumaric acid[c_a] + water[c_a]")
```

```
[1] TRUE
```

3.3.3. Reactants and Products

As defined in introduction, stoichiometric reactions represent the transformation of reactants into products in a chemical reaction. The `reactants` and `products` functions extract and return all reactants and products present in a stoichiometric reaction as a vector. In this example we show the extraction of the reactants (quinone and succinic acid) and products (hydroquinone and fumaric acid) in a reaction that occurs in astrocytes mitochondrias.

```
> reactants("Quinone[m_a] + succinic acid[m_a] => Hydroquinone[m_a] + fumaric acid[m_a]")

[1] "Quinone[m_a]"       "succinic acid[m_a]"

> products("Quinone[m_a] + succinic acid[m_a] => Hydroquinone[m_a] + fumaric acid[m_a]")

[1] "Hydroquinone[m_a]" "fumaric acid[m_a]"
```

3.3.4. Metabolites

Two of the more popular packages that implement FBA analysis such as `COBRA` [**?**] and `RAVEN` [**?**] require the complete list of metabolites included in the metabolic reconstruction, in a particular section of the human-readable input file. The `metabolites` function automatically identifies and lists all metabolites (with and without compartments) for a specific or a set of stoichiometric reactions. In this example we show how to extract all metabolites (reactants and products) with and without compartment for the Ubiquinol and FAD production reaction in astrocytes mitochondrias.

```
> metabolites("FADH2[m_a] + ubiquinone-0[m_a] => FAD[m_a] + Ubiquinol[m_a]")

[1] "FADH2[m_a]"        "ubiquinone-0[m_a]" "FAD[m_a]"
[4] "Ubiquinol[m_a]"
```

As was mentioned before, the report option without compartment was added:

```
> metabolites("FADH2[m_a] + ubiquinone-0[m_a] => FAD[m_a] + Ubiquinol[m_a]",
+             woCompartment = TRUE)

[1] "FADH2"        "ubiquinone-0" "FAD"          "Ubiquinol"
```

3.3.5. Orphan Metabolites

Orphan metabolites, compounds that are not produced or consumed in any other reaction are one of the main causes of mass unbalance in metabolic reconstructions. The `orphan.reactants` function, identifies compounds that are not produced internally by any other reaction and should be added to the reconstruction as an exchange reaction following the protocol proposed by [?]. In this examples we show how to extract all orphan compounds for all reactions included in the glutamate/glutamine cycle.

```
> data("glugln")
> orphan.reactants(glugln)

 [1] "alpha-D-Glucose 6-phosphate[r_n]"    "water[r_n]"
 [3] "2,3-bisphospho-D-glyceric acid[r_n]" "GTP[c_n]"
 [5] "oxaloacetic acid[m_n]"               "citric acid[c_n]"
 [7] "coenzyme A[c_n]"                     "Quinone[m_n]"
 [9] "D-Glutamine[m_n]"                    "L-Glutamine[m_n]"
[11] "FADH2[m_n]"                          "oxygen atom[m_n]"
[13] "Ferrocytochrome c2[m_n]"             "diphosphate(4-)[m_n]"
[15] "alpha-D-Glucose 6-phosphate[r_a]"    "water[r_a]"
[17] "2,3-bisphospho-D-glyceric acid[r_a]" "GTP[c_a]"
[19] "hydrogencarbonate[m_a]"              "citric acid[c_a]"
[21] "coenzyme A[c_a]"                     "Quinone[m_a]"
[23] "L-glutamic acid[c_a]"                "Ammonia[c_a]"
[25] "FADH2[m_a]"                          "oxygen atom[m_a]"
[27] "Ferrocytochrome c2[m_a]"             "diphosphate(4-)[m_a]"
```

The `orphan.products` function, identifies compounds that are not consumed internally by any other reaction and should be added to the reconstruction as an sink reaction following the protocol proposed by [?]. In this example we show the option added to `orphan.*` functions, that permits to report the orphan metabolites as a list grouped by compartment:

```
> orphan.products(glugln, byCompartment = TRUE)

$r_n
[1] "alpha-D-Glucose[r_n]"           "phosphate(3-)[r_n]"
[3] "2-phospho-D-glyceric acid[r_n]"

$c_n
[1] "GDP[c_n]"         "(S)-Lactate[c_n]" "acetyl-CoA[c_n]"

$m_n
```

```
[1] "Hydroquinone[m_n]"       "D-glutamic acid[m_n]"
[3] "FAD[m_n]"                "Ferricytochrome c2[m_n]"

$r_a
[1] "alpha-D-Glucose[r_a]"           "phosphate(3-)[r_a]"
[3] "2-phospho-D-glyceric acid[r_a]"

$c_a
[1] "GDP[c_a]"         "(S)-Lactate[c_a]" "acetyl-CoA[c_a]"  "L-Glutamine[c_a]"

$m_a
[1] "Hydroquinone[m_a]"       "FAD[m_a]"
[3] "Ferricytochrome c2[m_a]"
```

3.3.6. Compartments

As well as in cells, where not all reactions occur in all compartments, stoichiometric reactions in a metabolic reconstruction can be labeled to be restricted for a single compartment during FBA, by the assignment of a compartment label after the stoichiometric coefficient and name of each metabolite. Some FBA implementations require the report of all compartments included in the metabolic reconstruction as an independent part of the human-readable input file. In this example we show how to extrac all compartments for all reactions included in the glutamate/glutamine cycle.

```
> compartments(glugln)

[1] "c_n" "r_n" "m_n" "c_a" "r_a" "m_a"
```

3.3.7. Association with ChEBI

The Chemical Entities of Biological Interest (`ChEBI`) database is a freely available dictionary of molecular entities focused on 'small' chemical compounds involved in biochemical reactions [**?**]. Amongst other characteristics, the release 136 of ChEBI database contains a set of standardized metabolite names, synonyms and molecular formula for at least 52521 chemical compounds. The use of standardized metabolite names facilitate the sharing process and inter-convertion to another metabolite names or identifiers [**?**, **?**]. The `minval` package contains five functions to check and extract values from a local copy of the ChEBI database release 136. The `is.chebi` function takes a compound name as input, compares it against all the compounds names in ChEBI and returns a logical value TRUE if a match is found. In this next four examples we show the potential use of the functions using as input the acetyl-CoA compound.

```
> is.chebi("acetyl-CoA")

[1] TRUE
```

The `chebi.id` function takes a compound name as input, compares it against all the compounds names in ChEBI and returns the compound identifier if a match is found.

```
> chebi.id("acetyl-CoA")

[1] "15351"
```

The `chebi.formula` function takes a compound name as input, compares it against all the compounds names in ChEBI and returns the molecular formula if a match is found.

```
> chebi.formula("acetyl-CoA")

[1] "C23H38N7O17P3S"
```

The `chebi.candidates` function takes a compound name as input, compares it against all the compounds synonyms in ChEBI and returns possible compound names if a match is found.

```
> candidates<-chebi.candidates("acetyl-CoA")
> head(candidates)

[1] "acetoacetyl-CoA"                 "acetyl-CoA"
[3] "(1-hydroxycyclohexyl)acetyl-CoA" "cinnamoyl-CoA"
[5] "2-methylacetoacetyl-CoA"         "phenylacetyl-CoA"
```

The `to.ChEBI` function translates the compounds names of a stoichiometric reaction into their corresponding identifier or molecular formula in the ChEBI database. In this example we show how to use the `to.ChEBI` function for the Ubiquinol and FAD production reaction in astrocytes mitochondrias.

```
> toChEBI("FADH2[m_a] + ubiquinone-0[m_a] => FAD[m_a] + Ubiquinol[m_a]")

[1] "1 17877 + 1 27906 => 1 16238 + 1 17976"

> toChEBI("FADH2[m_a] + ubiquinone-0[m_a] => FAD[m_a] + Ubiquinol[m_a]",formula = TR

[1] "1 C27H35N9O15P2 + 1 C9H10O4 => 1 C27H33N9O15P2 + 1 C9H12O4(C5H8)n"
```

3.3.8. Mass Balance Validation

Thermodynamic unbalance of genome-scale metabolic reconstructions can also be promoted by stoichiometric mistakes. In a well balanced stoichiometric reaction according to the Lomonósov-Lavoisier law, the mass comprising the reactants should be the same mass present in the products. The `unbalanced` function converts the metabolites identifiers to molecular formulas, multiplies the atom numbers by their respective stoichiometric coefficient, and establishes if the atomic composition of reactants and products are the same, it returns a logical value TRUE if mass is unbalanced. In this example we show the mass balance evaluation for the first twenty reactions of the glutamate/glutamine cycle.

```
> unbalanced(glugln[1:20])
```

```
 [1] FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[13] FALSE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE
```

The `unbalanced` function also include an option to show the molecular formula of mass unbalanced formulas through the option `show.formulas`.

```
> unbalanced(glugln[1:20], show.formulas = TRUE)
```

```
     [,1]
[1,] "alpha-D-Glucose 6-phosphate[r_n] + water[r_n] => alpha-D-Glucose[r_n] + phos ..."
[2,] "beta-D-fructofuranose 1,6-bisphosphate[c_n] + water[c_n] => beta-D-fructofur ..."
[3,] "D-Glyceraldehyde 3-phosphate[c_n] + phosphate(3-)[c_n] + NAD(+)[c_n] <=> 3-p ..."
[4,] "ATP[c_n] + 3-phosphoglyceric acid[c_n] <=> ADP[c_n] + 3-phosphonato-D-glycer ..."
[5,] "3-phosphonato-D-glyceroyl phosphate(4-)[c_n] => 2,3-bisphospho-D-glyceric ac ..."
[6,] "2,3-bisphospho-D-glyceric acid[c_n] + water[c_n] => 3-phosphoglyceric acid[c ..."
     [,2]
[1,] "1 C6H13O9P + 1 H2O => 1 C6H12O6 + 1 O4P"
[2,] "1 C6H14O12P2 + 1 H2O => 1 C6H13O9P + 1 O4P"
[3,] "1 C3H7O6P + 1 O4P + 1 C21H28N7O14P2 <=> 3 C3H4O10P2 + 1 C21H29N7O14P2 + 1 H"
[4,] "1 C10H16N5O13P3 + 3 C3H7O7P <=> 1 C10H15N5O10P2 + 3 C3H4O10P2"
[5,] "3 C3H4O10P2 => 2 C3H8O10P2"
[6,] "2 C3H8O10P2 + 1 H2O => 3 C3H7O7P + 1 O4P"
```

3.4. Summary

We introduced the `minval` package to evaluate mass balancing correctness of metabolic reconstructions and to extract all reactants, products, orphan metabolites, metabolite names and compartments for a set of stoichiometric reactions. We show step by step the minimal evaluation process of mass balance using the 128 non-exchange reactions included in the glutamate/glutamine cycle included in the "`glugln`" dataset. Also some examples of metabolites names - ChEBI database association was showed.

3.5. Bibliography

4 g2f: An R package to Find and Fill Gaps for genome-scale metabolic networks

Written by: *Kelly Botero, Daniel Osorio, Janneth Gonzalez and Andrés Pinzón-Velasco.*
Tissue-specific metabolic reconstructions are subsets of reactions highly associated to the metabolism of a specific tissue. They are constructed from measured expression data and permit characterize or predict the metabolic behavior for each tissue under any physiological condition. Due to only the reactions associated to an enzyme or gene can be mapped from the expression data, the spontaneous reactions and non facilitated diffusion are missing in first stages of a tissue-specific reconstruction. Those missing reactions create gaps that block the biomass flux inside the metabolic pathways. Gaps should be filled before model evaluation through Flux Balance Analysis (FBA) to obtain a biological response. The g2f package was designed as a tool to find the gaps (metabolites not produced or not consumed in any other reaction), and fill it from the stoichiometric reactions of a reference metabolic reconstruction using a weighting function. Also the option to download all the set of gene-associated stoichiometric reactions for a specific organism from the KEGG database is available.

4.1. Introduction

4.2. Installation and functions

4.3. Summary

4.4. Bibliography

5 exp2flux: An R package to convert expression data to FBA fluxes

Written by: *Daniel Osorio, Kelly Botero, Janneth Gonzalez and Andrés Pinzón-Velasco.*

5.1. Introduction

5.2. Installation and functions

5.3. Summary

5.4. Bibliography

6 Conclusion

6.1. Conclusion