

*Short Review*

## **An overview of pathway prediction tools for synthetic design of microbial chemical factories**

**Bashir Sajo Mienda \* and Mohd Shahir Shamsir**

Bioinformatics Research Group (BIRG), Department of Biosciences and Health Sciences, Faculty of Biosciences and Medical Engineering, Universiti Teknologi Malaysia, 81310 Skudai Johor. Malaysia

\* **Correspondence:** Email: bsmienda@gmail.com; Tel: +60149136476;  
Fax: +607-5531279

**Abstract:** The increasing need for the bio-based industrial production of compounds via microbial cell factories leads to a demand for computational pathway prediction tools. A variety of algorithms have been developed that can be used to identify possible metabolic pathways and their corresponding enzymatic parts. These prediction tools play a central role in metabolic pathway design and microbial chassis selection for industrial chemical production. Here, we briefly discuss how the development of some key computational tools, which are currently available for pathway construction, could facilitate the synthetic redesign of microbial chassis. Special emphasis is given to the characteristics and drawback(s) of some of the computational tools used in pathway prediction, and a generalized workflow for the design of microbial chemical factories is provided. Perspectives, challenges and future trends are briefly highlighted.

**Keywords:** metabolic pathways; computational tools; synthetic design; microbial chemical factories

---

### **1. Introduction**

There is a great potential for the use of microbial cell factories in the synthesis of bio-based industrial compounds. Systems metabolic engineering for strain redesign is a promising approach to improve the prospects of using microbial hosts for production of high value biofuels or natural products. Currently, a number of computational tools have been published to aid in the pathway reconstruction of a host chassis [1–8], but a number of challenges exist that must be resolved before the potential of this field is fully realized. These challenges include making the available pathway prediction tools more users friendly by the addition of graphical user interfaces or web servers which

are linked to well-curated databases of experimentally characterized enzymatic parts in an integrated framework.

Biosynthetic routes that exist naturally have been the main target of engineering in the past few years; however, little success has been seen with some of the computational tools/algorithms, therefore, this is an interesting area to explore. This review focused on a few computational tools for pathway predictions that are applicable to microbial production systems, and were freely available as web services upon request. The computational tools include the Biochemical Network Integrated Computational Explorer (BNICE) [2,3], DESHARKY [4], from Metabolite to Metabolite (FMM) [1], RetroPath [5] and additional methods developed by Cho [7] and Furusawa [8]. Although their approaches to pathway predictions vary considerably, they can still predict non-native novel biochemical routes that can be reconstructed in a designated host. It is important to note that, for the following approaches to pathway design and microbial strain improvement to deliver on their key promises, progress must be made in computational tools and algorithms to predict novel metabolic pathway routes. These predictions are not limited to native biochemical pathways; however, it remains to be seen whether understanding of the computational predictive power gained will expedite forward pathway design, and whether or not intelligent modifications based upon the understanding of microbial systems and their metabolic engineering targets can be applied to increase the production of compounds that can be synthesized microbially.

Few examples of pathway prediction tools exist with limited successes. We reason that pathway prediction tools can be synergistically combined with experimental synthetic pathway engineering for improvement of microbial chemical production. We intend to do the following in this article: (i) briefly pinpoint the advancement and significance of pathway prediction tools in microbial strain redesign for chemical synthesis (ii) highlight the characteristics of computational tools that facilitate the specification of pathway design for strain alteration to synthesize target compound of interest (see Table 1 and Figure 1) and (iii) briefly offer perspectives and challenges that must be resolved before the potential of this field can be fully realized.

## 2. Pathway prediction Tools

### 2.1. From Metabolite to Metabolite (FMM)

This is a freely available web server <http://fmm.mbc.nctu.edu.tw/> and is considered a user-friendly system for pathway identification. The server is characterized by search options that allow the user to identify possible pathways between known input and output compounds [7]. It has comparative features and combines the KEGG maps and KEGG ligands to form an integrated pathway map. The corresponding gene(s) and the organisms involved can be identified, and the system generates an output in which different pathways can be compared. FMM has the following components:

#### 2.1.1. Data collection and integration

Reaction definitions, species-specific reactions, reaction maps and enzymes list can be obtained from KEGG/L/LIGAND and KEGG/PATHWAY databases recent releases. Information such as gene names, enzyme commission numbers, and species-specific enzymes can be retrieved from

UniprotKB/Swiss-prot and NCBI taxonomy databases. Additionally, the data in FMM is usually updated on a regular basis.

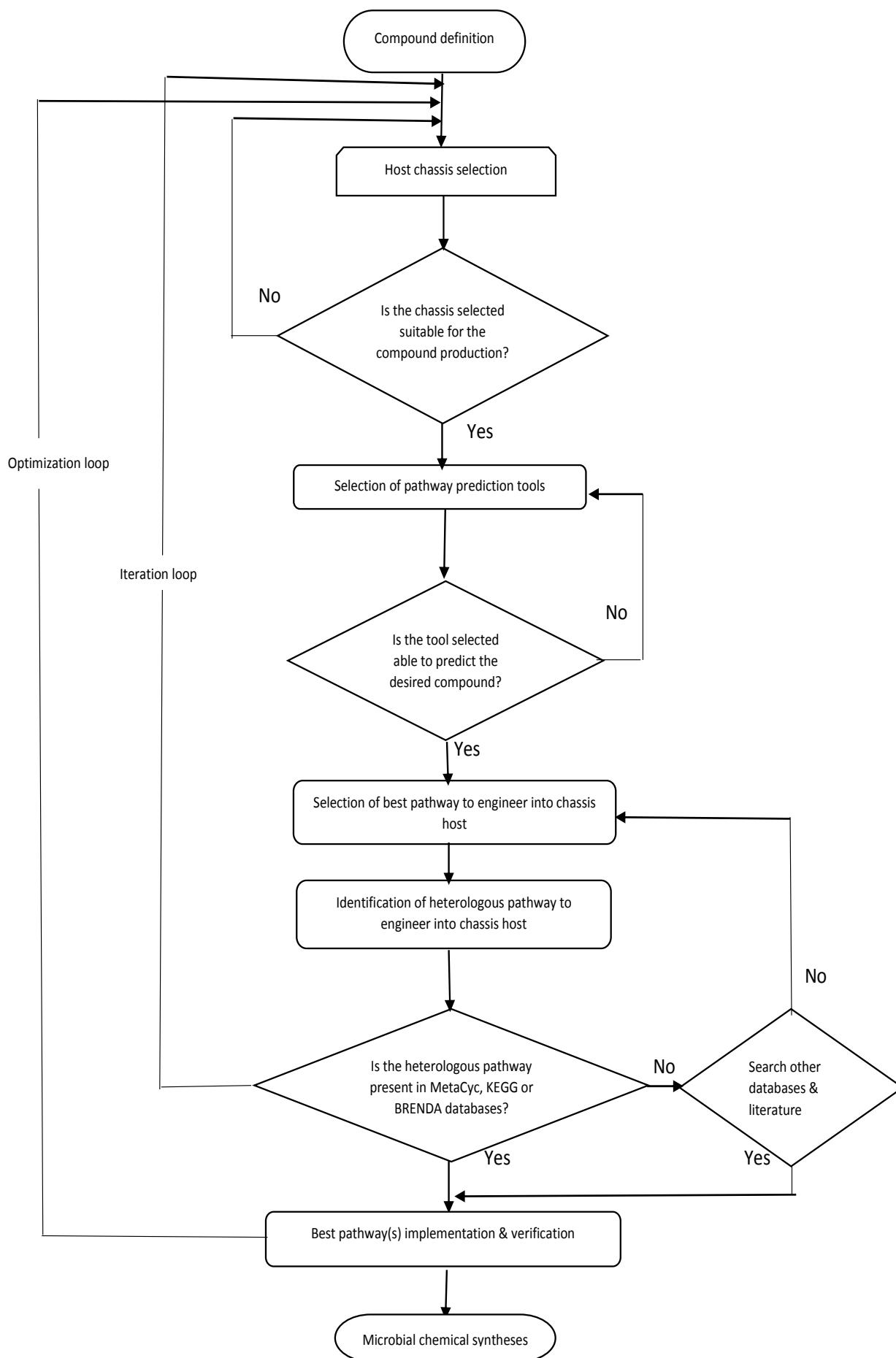
### 2.1.2. Construction of reaction matrices

Information on reactions and enzymes can be obtained from KEGG maps and the equation of each reaction can be determined. Therefore reaction matrices can be constructed based on maps, reactions, and enzyme data [1].

**Table 1. Characteristics of key computational tools for pathway prediction.**

Tools/System	Characteristics/ Descriptions	Advantage(s)	Drawback(s)	References
FMM	This web server uses two metabolites present within KEGG database to find their biosynthetic routes	It produces an output that compares different pathways.  It quickly generates a clear overview of different possible metabolic routes to a particular target compound.	It is limited to the KEGG framework.  It does not provide a practical insight into thermodynamic feasibility of the pathway in question.	[1]
BNICE	The reaction rules and enzymes commission classification system are used as a basis for predictions.  It considers the starting compound and/or products.  Thermodynamic feasibility was later introduced in to the framework as a prioritization approach by (Henry et al., 2010).  Thermodynamic feasibility, pathway length, maximum achievable yield and	Unknown pathways that are potentially chemically feasible could be predicted.  Unlike <b>FMM</b> , it is not only limited to the KEGG database	The BNICE framework is restricted to a first step, so much more analysis of the results is needed to obtain a useful outcome.	[2,3]

	maximum activity have been incorporated into the framework			
DESHARKY	<p>It is based on enzymatic reactions to predict novel pathways.</p> <p>The first step is the design and selection of the target compound and host chassis.</p>	<p>Thermodynamic favorability and energy loss in the transcription and translation are usually calculated in this framework</p>	<p>Pathway is required in this prediction system. Therefore one must search among myriad possibilities for the pathway that will efficiently work in the selected host chassis</p>	[4]
RETROPATH	<p>A web based server that hosts a unified framework for retrosynthetic pathway design.</p> <p>The first step is compound design and selection of the host chassis.</p> <p>It has a ranking function that integrates data from different sources such as host compatibility for inserted genes, estimation of steady-state fluxes from the genome wide reconstructions of the organism's metabolism. It also features compound toxicity prediction in selected hosts such as <i>E. coli</i></p> <p>[19]</p>	<p>This method uniquely enables candidate pathways to be ranked to determine which heterologous pathways are best to engineer.</p> <p>Only the top pathways are enumerated which reduces the problem of selecting the best pathway to engineer</p>	<p>It is limited to few host chassis e. g., <i>Escherichia coli</i>, <i>Bacillus subtilis</i> and <i>Saccharomyces cerevisiae</i></p>	[5,6,19]



**Figure 1. Generalized workflow for the design of microbial chemical factories, from initial idea to final product. First, the target compound is defined and the host chassis is selected. Then, the pathway prediction tool is applied based on chemical reaction rules and/or metabolic maps. Subsequently, the best pathway prediction tool is selected based on a number of criteria, such as ranking functionalities. The best predicted heterologous pathway to engineer will be subsequently selected and identified from the databases, such as KEGG. Finally, pathway implementation and verification will immediately follow for the host chassis of interest. Sometimes information obtained in a “later” stage suggests revision of an “earlier” decision.**

### 2.1.3. Reconstructions of metabolic pathways from various KEGG pathway maps

All of the possible reaction paths can be identified together with calculation of the pathway maps. Found paths usually occurred not only in a single pathway map, but also in a complicated fashion in several maps. Pathway maps that contain the most paths are usually selected and pathway maps that have only one reaction are avoided. A matrix of maps versus reactions can be implemented to reconstruct a metabolic pathway from different KEGG maps as fully described in the original documentation [1].

This framework is similar to RetroPath in some aspects; as they are both user-friendly and provide web services that are either freely available, in the case of FMM, or available on request as in the case of RetroPath. The steps in the flow diagram in Figure 1 can be applied using FMM and RetroPath to effectively design a microbial chassis to produce a designated compound of interest. This approach relies solely on characterized pathways limited to only the KEGG database [9] which is often considered incomplete and does not provide significant insight into the practical or thermodynamic feasibility of the pathway[1,10]. Therefore, FMM only provides an overview of different possible metabolic routes to the target product of interest and, hence, can only serve as a starting point for many preliminary investigations. This method requires additional work for further validation and if necessary apply the use of a proof of concept approach to demonstrate the application of the tool in metabolic pathway engineering for the production of desired compound of interest.

### 2.2. *Biochemical Network Integrated Computational Explorer (BNICE)*

This framework predicts novel pathways on the basis of broader rules of the enzyme commission classification system. Unlike FMM, this framework can also predict pathways that are completely unknown but potentially chemically feasible while at the same time taking into account their thermodynamic properties [2]. The pathways that are usually generated using this framework are not limited to only one database. It incorporates compounds that exist in different biological and chemical databases, as well as novel compounds, by suggesting novel biochemical routes for compound generation while also suggesting the existence of biochemical compounds that are not yet discovered or synthesized via enzymes and/or pathway engineering [2].

Furthermore, this framework searches for pathways by considering the starting compound and/or products, the requested length of the pathway, and the range of reactions to search over [3,10].

A user can choose to search for a myriad of possibilities, such as searching for a pathway using enzyme reactions from known pathways, a combination of multiple pathways, or the whole metabolic network [2,3]. This framework can be considered as a first step in finding possible pathways, but a useful result is less likely to be obtained without a thorough subsequent analysis. Limitations exist in using this framework as it predicts more than 10,000 different pathways for the biosynthesis and degradation of the compound of interest, due to the fact that the system relies on few criteria. It provides a myriad of possible pathways that become difficult to select; as such a ranking of the generated pathways would add significant utility. However, Henry and colleagues [3] have pioneered a prioritization approach in the BNICE framework, in which generated pathways are ranked according four criteria: pathway length, thermodynamic feasibility, maximum achievable yield and maximum achievable activity (see Table 1). Despite the introduced prioritization approach, BNICE is not as efficient as RetroPath, because it has unique ranking functionalities where only the top pathways are enumerated.

The BNICE framework requires a graph-theoretic matrix representation of biochemical compound and enzyme reaction rules; and molecules are represented using the bond-electron matrix (BEM) where each atom in a molecule is represented by a row and column. The BEM is characterized by diagonal elements, which denote the non-bonded valence electrons and non-diagonal elements, which give the connectivity via bonding between different atoms as well as the bond order between atoms [2]. Similar notation can be used to represent enzyme catalyzed reactions where the reactive site for each enzyme class is pre-defined as a two dimensional (2D) molecule fragment and coded in BNICE. A set of molecules is given as an input and every molecule is evaluated to determine if it has the appropriate functionality to undergo reactions corresponding to the specified reaction classes [2]. Subsequently, the reaction can be implemented through matrix addition, generating negative and positive reaction matrix numbers. The former reaction matrix numbers correspond to the cleavage of bonds while the latter correspond to the formation of bonds [2]. The matrix representing the enzyme catalyzed reaction can be added to the BEM for the substrate, and the BEM formed specifies the product of the reaction [2].

The framework developed can be applied to huge numbers of different systems of biotechnological significance. One of the notable examples is isoprenoid and polyketide synthesis pathways, which have been shown to produce an enormous number of structurally different compounds, and their diversity has been expanded via metabolic engineering [11]. The application of this framework to the biochemistry of central carbon pathways will facilitate the identification of potential novel routes to essential metabolic and biosynthetic compounds or the synthesis of new compounds, based on renewable resources [2]. The application of this type of tool in host strain design can have significant implications for the development of bioprocesses of industrial chemicals in sustainable technology. We previously reported [12] the need for integration of various disciplines and use of computational tools for DNA synthesis in microbial strain improvement.

### 2.3. *DESHARKY*

This pathway prediction system is based on enzymatic reactions, but it approaches the search in a way that is unique among the aforementioned systems. Unlike BNICE, this tool uses a heuristics algorithm based on a Monte Carlo method to find a possible route connecting the specified target metabolism with the host metabolism, instead of using pathway selection by enumeration of possible

metabolic routes [4]. The first step in this system is compound design followed by selection of the chassis. DESHARKY outputs a biochemical route leading to the host metabolism, together with a novel approach to cellular processes that uses mathematical models of the cellular resources and metabolism [4]. The algorithm is implemented in C/C++, it is easily compiled and runs in UNIX environment (e.g. in Linux or in Windows using Cygwin). The algorithm calculates thermodynamic favorability and energy loss in transcription and translation. The capabilities of the algorithm can be enlarged by accounting for reversible reactions. Additionally, one can choose to introduce reactions which are not found in KEGG. The input of the algorithm is usually the target compound while its output is the designed metabolic pathway together with quantification of the transcriptional, translational, and metabolic load [4]. This framework also provides the sequence of amino acids of the enzyme involved in the pathway. The amino acid sequences provided are usually the closest phylogenetically to *Escherichia coli* according to KEGG classification of organisms [4]. The tool becomes extremely useful if the chosen chassis has already been determined and the user must search for the pathway that will most efficiently generate the desired compound of interest [4,10].

In addition, DESHARKY was applied to design several metabolic pathways including the biodegradation of toluene or phenol and biological production of sorbitol and glucaric acid [4]. The significance of the microbial production of glucaric acid has been shown to have therapeutic applications that include cholesterol reduction and cancer chemotherapy, in addition to the synthesis of new nylons and hyper branched polyesters. DESHARKY finds a proper pathway and computes its associated genetic burden within the shortest possible time; at the same time, it can be used in distributed computing to sample most of the solutions space [4].

#### 2.4. RetroPath

The wide adoption of retrosynthesis in the manufacturing pipeline has been hampered by the complexity of enumerating all feasible biosynthetic pathways for a target compound of interest. This process is much easier today, with the development of a RetroPath webserver <http://www.issb.genopole.fr/~faulon/retroPath.php> [5]. This server applies a retrosynthetic approach, a concept originally proposed for synthetic chemistry, which uses reverse chemical transformations (reverse enzyme-catalyzed reactions in the metabolic space) starting from the desired target compound to identify the reactants (precursors) that are indigenous to the selected host [5].

This method of metabolic pathway design is unique because it addresses the complexity problem by coding substrates, products and reactions into molecular signatures. The approach used by RetroPath is characterized by metabolic maps, which are represented in hypergraphs. The complexity involved in the reactions is controlled by varying the specificity of the molecular signature [5]. Each signature has different "heights",  $h$ , that correspond to levels of structural detail. The height can be varied, which reduces the number of reactions that can be generated [5]. This number varies from the large number of reactions found using BNICE to the small number of original reactions that are present in the KEGG database [9].

The proliferation of several metabolic databases with rich information is considered to be a significant breakthrough. KEGG [9] is a database resource that integrates chemical and systematic functional information and genomics. This database is linked to RetroPath, where information on the reactions predicted using this framework can be found in KEGG. BRENDA [13] is another database that contains one of the largest collections functional enzyme data [6]. Incomplete knowledge or gaps



still exist in many cases, especially when looking for novel ways to synthesize a target compound of interest [6]. To this end, computational approaches such as RetroPath, can provide promising new alternatives by predicting putative heterologous pathways that produce the desired compound. To successfully achieve a heterologous pathway design, the process need to be rationalized by following the principles of synthetic biology: modelling of the biological system of interest, modular design through standardization, goal-oriented optimization and experimental validation [6]. An approach using retrosynthetic design represents a promising alternative that provides a streamlined methodology for addressing the general problem of obtaining successful high yield production of target compounds in microbial cell factories [6].

Finally, a basic methodology as reported by Pablo and colleagues [6] for implementing the retrosynthetic design of heterologous pathways will consist of the following steps: (1) host chassis selection, (2) *in silico* model selection for the chassis from BiGG [14] or BiModels[15], (3) definition of the metabolic space, (4) pathway enumeration, (5) gene selection, (6) estimation of yields by metabolic analysis software, e.g., COBRA, OptFlux [16] and COPASI [17,18], (7) toxicity prediction of pathway metabolite [19], (8) definition of an objective function to select the best pathway to engineer, and (9) pathway implementation and validation. This method was recently applied computationally and validated by Faulon and colleagues [20].

For specific case studies and a more detailed explanation on the inner workings of each step in retrosynthetic design of heterologous pathways, we refer the reader to a range of excellent reviews published recently [6,21,22].

### 2.5. The Cho System framework

This system framework was developed to suggest promising enzyme candidates to synthesize desired chemicals in selected microbial chassis based on combined information on chemical structural changes, enzyme characteristics and reaction mechanisms present in systems databases [7]. The developed system framework identifies structurally qualified enzymes for the synthesis of predetermined target compounds and subsequently ranks the enzymes through a novel method known as the prioritization scoring algorithm. This algorithm is applicable for an enzymatic reaction that has the same reaction rule with a novel reaction in a desired pathway. This helps to clarify which enzymatic reactions will be more promising in the microbial production of desired chemicals.

The prioritization method used in this system framework composed of the following five (5) factors: (i) binding site covalence, (ii) chemical similarity, (iii) thermodynamic favorability, (iv) pathway distance and (v) organism specificity. On the basis of these factors, a final priority score can be established which is classified into three (3) groups: structural similarity of reaction step in a route, thermodynamic benefits among the intermediates, and co-expression probability of enzymes [7]. This framework allows a user to establish a final priority equation taking into account the following parameters as described in the original documentation [7].

The parameters are as follows:

- (a) priority score of an enzyme route candidate
- (b) parameter for binding site covalence
- (c) parameter for chemical similarity
- (d) parameter for thermodynamic favorability

- (e) parameter for pathway distance
- (f) parameter for organism specificity

The novel strategy using the aforementioned parameters was applied in the system framework for effective identification of a desired synthetic pathway. As a proof of concept the authors have shown clearly that the framework can be used to predict synthetic pathway for the production of higher alcohols such as 1- propanol, 1- butanol, 2- methyl-1-butanol, 3- methyl-1-butanol, isobutanol, and 2-phenylethanol in *E. coli* as experimentally demonstrated elsewhere [23]. Prediction of synthetic pathway for the production of other compound such as 3-hydroxypropionic acid (3HP) was also demonstrated in a similar manner. This suggests that the system framework would find application in strain improvement for microbial production of biofuels and other industrial compound of interest.

## 2.6. The Furuuwa platform

This is an *in silico* platform that uses a developed algorithm for finding feasible heterologous pathways by which non-native target metabolites are produced by microorganisms, using *Escherichia coli*, *Corynebacterium glutamicum* and *Saccharomyces cerevisiae* as templates [8]. The implementation of this platform for heterologous pathway design entails four (4) steps:

### 2.6.1. Construction of an in-house database of metabolic reactions

In this approach, in-house database can be constructed taking in to account all known metabolic reactions from KEGG ligand section database and BRENDA [8]. These metabolic reactions are considered as candidate heterologous reactions that could be added to the host metabolic networks [8]. All metabolic reaction information regarding genes, enzymes, pathways, and organism in the KEGG database [9] can be collected into the database. The user can collect the information in a constructed database using PostgreSQL 9.0 that was developed by the postgresSQL Global Development Group. The enzymatic information employed can be retrieved from BRENDA and python script can be used to access the constructed in-house database [8]. This system might be disadvantaged, as it requires high level programming language to use Python script in order to achieve the desired objectives.

### 2.6.2. Genome scale metabolic models of host microorganisms

Three microorganisms that are widely used in industry were adopted as chassis templates to demonstrate the viability of this *in silico* platform. These microorganisms include *Escherichia coli*, *C. glutamicum* and *S. cerevisiae* which were selected based on a number of criteria such as having high growth activity under various conditions, ease of genetic manipulation and hence are considered as ideal hosts for bioengineered products [8]. The genome scale models of *E. coli* [24], *C. glutamicum* [25] and *S. cerevisiae* [26] were used as proof of concept examples to validate the developed algorithm based on earlier metabolic reconstructions with slight modifications. The KEGG database is considered as the focal point of this platform.

### 2.6.3. Heterologous pathway identification for target production

The developed algorithm has been applied to identify heterologous reactions producing a target metabolite within a host microorganism. The in-house databases constructed in the aforementioned step served as the source of the heterologous metabolic reactions. The algorithm expands the host metabolic network by sequentially adding heterologous metabolic reactions from the in-house constructed database [8].

The general step involved in heterologous pathway identification procedure is as follows:

- (i) Sets of native metabolites and reactions present in the host genome metabolic model are designated as  $M_0$  and  $R_0$  respectively
- (ii) From the in-house database constructed, a set of heterologous reactions and metabolites that do not exist in  $M_0$  and  $R_0$  are defined as  $R_1$  and  $M_1$ .
- (iii) Following the same trend as above,  $R_i$  is designated as the set of reaction that are not present in  $\{R_0, R_1, R_{i-1}\}$  which can produce metabolites not existing in  $\{M_0, M_1, \dots, M_{i-1}\}$  from metabolites included in those sets [8].

The expansion procedure applied in this framework is iterated until no further reaction can be connected to the expanded metabolic network [8].

In summary the strategy introduced above was a development of a pathway search algorithm that identifies the shortest pathway between a host metabolic network and target metabolites as heterologous reactions are added [8]. The developed algorithm can be used to screen all producible target metabolites listed in the database by adding heterologous reactions to host microorganisms [8]. For all producible target metabolites the user can estimate the production yields using FBA, assuming steady-state conditions and the maximum biomass production rate [8]. The entire list of producible target metabolites in different hosts can be analyzed and a set of rational heterologous pathways and hosts can be selected that will likely produce the desired targets.

#### 2.6.4. Flux balance analysis (FBA)

FBA is based on a genome scale metabolic model and optimization of a specific objective flux by linear programming. One can use FBA to estimate the metabolic flux profile of metabolic networks expanded with heterologous reactions [8]. All FBA simulations in this framework can be performed under the MATLAB interface as fully described in the original documentation [8]. Other software platforms that can be used to perform FBA include the COBRA [27,28] toolbox which is run under the MATLAB interface and the OptFlux [16] open source software platform. In addition, Elementary Flux Mode analysis has been reported to be used for metabolic pathway analysis with the genome scale metabolic model, and it is run under the MATLAB interface, the details of which have been reviewed elsewhere [29].

### 3. Perspectives

In general, the ability to use computational approaches for the prediction of heterologous pathways for microbial strain improvement depends largely on headway made in software development and data systemization. The limited successes and breakthroughs in the pathway prediction tools/systems mentioned earlier (see Table 1) are of utmost significance in the field of systems and synthetic biology.

Challenges remain because there are other subsequent analyses after the pathway prediction that may require the use of other computational approaches. Once a pathway has been predicted and selected for introduction in to specific host bacterium, the repercussions of these genetic perturbations must be predicted in the new metabolic context [10]. One notable example is OptStrain [30], a system which aims to monitor the effects of a new pathway on the host. This framework uses flux analysis to provide advice on how production could be optimized by altering the gene expression of the selected chassis [10,30].

Furthermore, strain redesign using computational pathway prediction tools may require the host gene knockout to be optimized in microbial production systems. Algorithms such as Optknock [31], COBRA [27,28] and OptFlux [16] are readily available to supply the researchers with suggestions regarding which pathway(s) to knockout on the basis of metabolic flux simulations. We recently reported the use of the OptFlux software platform for metabolic engineering interventions using an *Escherichia coli* genome scale model. The metabolic interventions primarily targeted different gene knockouts to increase the production of ethanol on glucose and gluconate [32], xylose and glycerol in *Escherichia coli* [33]. In a similar study, enhanced production of D-lactate in *E. coli* from glycerol was predicted using the OptFlux software platform [34].

The currently available pathway tools will need to be made user friendly, with GUIs or web servers as in the case of RetroPath, and they should be linked to well-curated databases of experimentally characterized enzymatic parts in an integrated framework [6,10]. Enzymes have been extensively characterized that are involved in rational redesign of microbial production systems over the past decades, but little systematic data archiving has been performed to date. This lack of archiving has hampered the efficient application of these approaches, while databases such as KEGG [9] provide limited information, being largely focused on primary metabolism [9,10].

#### 4. Conclusion

The successful development of microbial chemical factories is essential to add value to the production of compounds ranging from biofuels and pharmaceuticals (see Figure 1). These successes pose a number of challenges including: a lack of well characterized enzymes, poor activity of selected pathway enzymes, low product titers, poor yield and selectivity, metabolic burden and unfavorable cofactor balancing [35]. Some of these challenges could be addressed by advances in computational tools for pathway engineering (metabolic engineering), biochemistry, protein engineering, and synthetic and molecular biology. A number of experimental and *in silico* tools have been produced to address some of these challenges (see Table 1). It is necessary to recognize that experimental and computational breakthroughs remain the key aspects of any progress made in synthetic microbiology for the development of robust microbial chemical factories. It is now clear that experimental and computational breakthroughs go hand-in-hand, and when fully integrated, synthetic pathway engineering could play a significant role in the synthetic redesign of microbial production systems and may serve as a major driver of applied synthetic biology.

#### Conflict of interests

Authors declare that they have no competing interests.

## Acknowledgements

We would like to thank Pablo Carbonell for his invaluable input during the preparation of the manuscript.

## References

1. Chou CH, Chang WC, Chiu CM, et al. (2009) FMM: a web server for metabolic pathway reconstruction and comparative analysis. *Nucleic Acids Res* 37: W129–34.
2. Hatzimanikatis V, Li C, Ionita JA, et al. (2005) Exploring the diversity of complex metabolic networks. *Bioinformatics* 21: 1603–1609.
3. Henry CS, Broadbelt LJ, Hatzimanikatis V (2010) Discovery and analysis of novel metabolic pathways for the biosynthesis of industrial chemicals: 3-hydroxypropanoate. *Biotechnol Bioeng* 106: 462–473.
4. Rodrigo G, Carrera J, Prather KJ, et al. (2008) DESHARKY: Automatic design of metabolic pathways for optimal cell growth. *Bioinformatics* 24: 2554–2556.
5. Carbonell P, Planson AG, Fichera D, et al. (2011) A retrosynthetic biology approach to metabolic pathway design for therapeutic production. *BMC Syst Biol* 5: 122.
6. Carbonell P, Planson AG, Faulon JL (2013) Retrosynthetic design of heterologous pathways, in *Methods Mol Biol*. Springer Science+Business Media, LLC. 149–173.
7. Cho A, Yun H, Park JH, et al. (2010) Prediction of novel synthetic pathways for the production of desired chemicals. *BMC Syst Biol* 4: 35.
8. Chaturachai S, Furusawa C, Shimizu H (2012) An in silico platform for the design of heterologous pathways in nonnative metabolite production. *BMC Bioinformatics* 13: 93.
9. Kanehisa M, Goto S, Sato Y, et al. (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40: D109–114.
10. Medema MH, van Raaphorst R, Takano E, et al. (2012) Computational tools for the synthetic design of biochemical pathways. *Nat Rev Microbiol* 10: 191–202.
11. Khosla C, Keasling JD (2003) Metabolic Engineering for drug discovery and development. *Nat Rev Drug Discov* 2: 1019–1025.
12. Mienda BS, Shamsir MS (2013) Thermotolerant micro-organisms in Consolidated Bioprocessing for ethanol production: A review. *Res Biotechnol* 4: 1–6.
13. Schomburg I, Chang A, Placzek S, et al. (2013) BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res* 41: D764–772.
14. Schellenberger J, Park JO, Conrad TM, et al. (2010) BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics* 11: 213.
15. Le Novère N, Bornstein B, Broicher A, et al. (2006) BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res* 34: D689–691.
16. Rocha I, Maia P, Evangelista P, et al. (2010) OptFlux: an open-source software platform for in silico metabolic engineering. *BMC Syst Biol* 4: 45.

17. Hoops S, Sahle S, Gauges R, et al. (2006) COPASI--a COMplex PATHway Simulator. *Bioinformatics* 22: 3067–74.
18. Schaber J (2012) Easy parameter identifiability analysis with COPASI. *Biosystems* 110: 183–5.
19. Planson AG, Carbonell P, Paillard E, et al. (2012) Compound toxicity screening and structure-activity relationship modeling in *Escherichia coli*. *Biotechnol Bioeng* 109: 846–850.
20. Fehér T, Planson AG, Carbonell P, et al. (2014) Validation of RetroPath, a computer-aided design tool for metabolic pathway engineering. *Biotechnol J* 9: 1446–1457.
21. Planson AG, Carbonell P, Grigoras I, et al. (2012) A retrosynthetic biology approach to therapeutics: from conception to delivery. *Curr Opin Biotechnol* 23: 948–956.
22. Fernandez-Castane A, Feher T, Carbonell P, et al. (2014) Computer-aided design for metabolic engineering. *J Biotechnol* 192: 302–313.
23. Atsumi S, Hanai T, Liao JC (2008) Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* 451: 86–89.
24. Reed JL, Vo TD, Schilling CH, et al. (2003) An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol* 4: R54.
25. Shinfuku Y, Sorpitiporn N, Sono M, et al. (2009) Development and experimental verification of a genome-scale metabolic model for *Corynebacterium glutamicum*. *Microb Cell Fact* 8: 43.
26. Mo ML, Palsson BO, Herrgard MJ (2009) Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst Biol* 3: 37.
27. Becker SA, Feist AM, Mo ML, et al. (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc* 2: 727–738.
28. Schellenberger J, Que R, Fleming RM, et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc* 6: 1290–1307.
29. Trinh CT, Wlaschin A, Sreenc F (2009) Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Appl Microbiol Biotechnol* 81: 813–826.
30. Pharkya P, Burgard AP, Maranas CD (2004) OptStrain: A computational framework for redesign of microbial production systems. *Genome Res* 14: 2367–2376.
31. Burgard AP, Pharkya P, Maranas CD (2003) Optknoack: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng* 84: 647–657.
32. Mienda BS, Shamsir MS, Shehu I, et al. (2014) *In Silico* Metabolic Engineering Interventions of *Escherichia coli* for Enhanced Ethanol Production, Based on Gene Knockout Simulation. *IIOAB J* 5: 16–23.
33. Mienda BS, Shamsir MS (2014) *In silico* Gene knockout metabolic interventions in *Escherichia coli* for Enhanced Ethanol production on Glycerol. *Res J Pharm Biol Chem Sci* 5: 964–974.
34. Mienda BS, Shamsir MS, Salleh FM (2014) *In silico* metabolic engineering prediction of *Escherichia coli* genome model for production of D-lactic acid from glycerol using the OptFlux software platform. *Int J Computational Bioinformatics In Silico Modeling* 3: 460–465.
35. Dhamankar H, Prather KL (2011) Microbial chemical factories: recent advances in pathway engineering for synthesis of value added chemicals. *Curr Opin Struct Biol* 21: 488–494.

© 2015, Bashir Sajo Mienda, et al., licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)