A Study of Genomics and Metabolic Reconstructions with Pseudomonas from Biotechnology

Dr. Sudhir Kumar Mishra

Assistant Professor, S.N.S College, Tekari, Gaya

Abstract

The best-studied pollutant-degrading bacterium, Pseudomonas putida, is used in industrial biotechnology to manufacture fine compounds. Some in silico evaluations of P. putida KT2440's metabolic and biotechnological capacities have been reported since the genome was published. However, a comprehensive knowledge of P. putida KT2440's capabilities necessitates the development of a metabolic model that allows the integration of traditional experimental data with genomic and high-throughput data. In silico genome-scale metabolic reconstructions have been successfully built and analysed using the constraint-based reconstruction and analysis (COBRA) method. The employment of microorganisms in the efficient manufacture of chemicals and the eradication of toxic waste is a cornerstone of biotechnology. Due to its metabolic plasticity, stress resilience, amenability to genetic alterations, and great potential for environmental and industrial applications, Pseudomonas putida is an exemplar of such bacteria. We built and describe here a genome-scale constraint-based model of P. putida KT2440 metabolism to address both the elucidation of the metabolic wiring in P. putida and its applications in biocatalysis, particularly for the synthesis of non-growth related biochemicals. Network reconstruction and flux balance analysis (FBA) allowed the topology of the metabolic network to be defined, knowledge gaps to be identified, and critical metabolic functions to be pinpointed, allowing for the refining of gene annotations.

Keywords: *Genomes; Biochemical; Biological; Glucose.*

I. INTRODUCTION

Pseudomonas putida is one of the well-studied species in the metabolically diverse and widespread Pseudomonads genus. It has a broad biotechnological potential as a species, with a variety of strains (some of which are solvent-tolerant) capable of producing a wide range of bulk and fine compounds. These characteristics, together with their well-known stress resilience, genetic manipulation ability, and appropriateness as a host for heterologous expression, make Pseudomonas putida a very appealing biocatalyst. P. putida strains have been used to make phenol, cinnamic acid, cis-cis-muconate, p-hydroxybenzoate, p-cuomarate, and myxochromide, among other things. P. putida enzymes have also been used in biocatalytic processes such as the resolution of D/L-phenylglycinamide into D-phenylglycinamide and L-phenylglycine, the production of non-proteinogenic L-amino acids, and the biochemical oxidation of methylated heteroaromatic compounds to form heteroaromatic monocarboxylic acids. Most Pseudomonas-based applications, on the other hand, are still in their infancy, owing to a lack of understanding of the genotype-phenotype correlations in these bacteria under industrial and environmental settings.

Safe (GRAS approved), it allowed researchers to examine the metabolic capabilities of the P. putida species and pave the way for novel biotechnological applications. A wealth of genetic determinants that play a role in biocatalysis, such as those for hyper-production of polymers (such as polyhydroxyalkanoates) and industrially relevant enzymes, the production of epoxides, substituted catechols, enantiopure alcohols, and heterocyclic compounds, were discovered using whole genome analysis. Despite the significant progress made in our understanding of P. putida as a result of this sequencing effort, the relationship between genotype and phenotype cannot be predicted simply by cataloguing and assigning gene functions to the genes found in the genome, and much

more work is required before the genome can be translated into a fully functional metabolic model useful for predicting cell phenotypes.

At the moment, constraint-based modelling is the only way to simulate an organism's metabolic and transport network at genome size. A genome-wide constraint-based model includes a stoichiometric reconstruction of all known reactions that function in the organism's metabolism, as well as a set of constraints on the fluxes of each reaction. The model does not need knowledge of the reaction kinetics, which is a key benefit of this technique. These models give a foundation for navigating through the cell's metabolic wiring by defining the organism's global metabolic space, network structural features, and flux distribution possibilities.

Constraint-based models can assist predict cellular phenotypes given certain environmental variables using a variety of analytic methodologies. Flux balance analysis (FBA) is one such approach that depends on optimising for an objective flux while ensuring mass balance in all modelled reactions to provide a set of fluxes that are compatible with the objective function's maximum output. When a biomass sink is set as the goal in FBA, the output may be linked to growth, and the model fluxes can be used to forecast growth phenotypes. FBA and other constraint-based analytic approaches have proved useful in identifying metabolic properties in a range of species, and they have been applied in a few situations for real biotechnology undertakings.

All earlier research that employed a constraint-based method to design the production of a biochemical, on the other hand, only looked at the synthesis of chemicals that can be directly connected to the objective function used in the underlying FBA issue. The main reason for this is that FBA-based approaches anticipate a zero-valued flux for every response that does not contribute directly to the desired outcome. Because most high-added value and bulk chemical synthesis pathways run concurrently with growth-related metabolism, using FBA to these biocatalytic processes isn't a reliable predictor of output. Other constraint-based analysis methods, such as Extreme Pathways and Elementary Modes analysis, can analyse non-growth related metabolic pathways, but they haven't been able to predict fluxes or phenotypes at the genome scale for guiding biocatalysis efforts due to the combinatorial explosion inherent in numerical resolution.

To address both the elucidation of P. putida metabolic wiring and the use of P. putida for the production of non-growth related biochemicals, we developed and present a genome-scale reconstruction of the metabolic network of Pseudomonas putida KT2440, the subsequent analysis of its network properties through constraint-based modelling, and a thorough assessment of the model's potential and limits. The reconstruction is based on current bacterial genomic, biochemical, and physiological understanding. The model uses a constraint-based modelling framework to account for the operation of 877 reactions that connect 886 metabolites. Only 6% of the reactions in the network are not linked to a gene. Several genes' annotations were refined as a result of the rebuilding process. Continuous culture studies, substrate usage tests (BIOLOG), internal flow measurements, and a collection of mutant strains were used to verify the model. We looked at how biomass composition and maintenance values affected the results of flux balance analysis (FBA) simulations, and we used metabolic reconstruction to predict internal reaction fluxes, identify different mass-routing options, and figure out which genes and reaction sets were required for growth on minimal medium. Finally, we used the model to develop hypotheses for prospective improvements in the synthesis of polyhydroxy alkanoates by P. putida, a class of chemicals whose production consumes resources that might otherwise be utilised for growth, using a modified OptKnock technique. This reconstruction therefore provides a modelling framework for investigating P. putida's metabolic capacities, which will aid in comprehending the complicated genotype-phenotype correlations influencing its metabolism and expand the usability of P. putida strains for bioremediation and biotechnology.

II. MATERIALS AND METHODS

- **Constraint-Based Models**

A constraint-based (CB) technique was used to create the P. putida model presented here. A constraint-based model includes a genome-wide stoichiometric reconstruction of metabolism as well as a set of limitations on the system's reaction fluxes. Because most biological processes are catalysed by enzymes, the reconstruction depicts the stoichiometry of the set of all reactions known to act in the organism's metabolism, which may be estimated in large

part from genomic data. As a result, the model does not require any knowledge of reaction kinetics, and the required thermodynamic information is confined to reaction directionality.

In addition to the processes, the model contains a set of genes that are linked to the reactions that their protein products catalyse using Boolean logic, allowing for precise differentiation of the effects of genetic perturbations like knockouts. The metabolic reconstruction's gene-protein-reaction relationships (GPRs) are made up of these Boolean rules.

The constraints, which make up the second element of the CB-model, are a collection of criteria that restrict down the range in which a reaction's flux must fall. These limitations are based on physico-biological understanding. One of these has previously been mentioned: the information on response directionality. The Pseudo-Steady-State Assumption (PSSA), which says that a chemical compound's concentration remains constant during the simulated time frame, is another extensively used restriction in biological systems. Internal compounds are the reactants to which this restriction is applied, and they correspond to the chemical components found inside the cell or its compartments in biological models. External compounds, or remaining substances, are species that may be taken up or secreted and hence exchanged with the environment. Other restrictions include top and bottom limits, which relate to the enzymes' catalytic capacities. In the, section "Constraint based models— mathematical explanation," you may find a more extensive discussion of the constraint based modelling technique.

- **Metabolic Network Reconstruction**

Biological databases were the primary sources of information on the makeup of Pseudomonas putida KT2440's metabolic network. The majority of the data comes from the Pseudomonas Genome Database (PGD) and the Kyoto Encyclopedia of Genes and Genomes (KEGG). The information on P. putida in these two databases is primarily based on the bacterium's published genome annotation, therefore there is a lot of overlap. In addition, data from the BRENDA database, which catalogues reaction and enzyme information, was used extensively. All of this was supplemented with information gleaned from primary research articles. The reconstruction was done in stages, with each round of model testing including the addition or removal of responses from the model. First, data on P. putida reactions was gathered from KEGG and PGD. The model included reactions that were supported by adequate data and had explicit enough functional annotations. The reversibility of each approved reaction was determined using KEGG pathway designations and information from the BRENDA database. The reversibility of reactions with conflicting assignments was determined based on the reaction's analysis as well as its reversibility in other species. An early version of the metabolic model was built in this way ($iJP815^{pre1}$).

The next stage was to see if the rebuilt metabolic network could generate energy from glucose. FBA was used with ATP generation as the goal function to achieve this. After that, the model's capacity to develop in silico on glucose was examined. When in silico growth is successful, it means that every chemical molecule in the biomass equation can be created from current sources utilising the model's processes. The composition of E. coli biomass was chosen as an estimate because the actual cellular composition of P. putida is unknown. The test was carried out by running FBA with the goal of producing each biomass ingredient. If a chemical could not be synthesised, the gaps in the route leading to it were manually identified, and a search for reactions to fill the gaps was conducted. If this method failed, gaps were filled using reactions from an E. coli model. This resulted in the reconstruction's second version ($iJP815^{pre2}$).

There were two sub-steps in the third cycle of rebuilding. First, the chemicals for which transport proteins exist were discovered, followed by the addition of relevant reactions. Second, in silico simulations of growth on those compounds were compared to the findings of BIOLOG carbon-source consumption studies. The capacity to develop in silico on a certain molecule as the only carbon source was considered to be similar to in vivo usage. A literature search was conducted to determine probable routes of usage for those chemicals that did not display in silico growth. The model was updated with the findings of the search, which came in the form of reactions and GPRs. The model's final version was the result (iJP815).

III. RESULTS

- **Highlights of the Model Reconstruction Process**

We used a technique shown in Figure 1 to recreate P. putida metabolism at the genomescale. (1) An initial data collection stage leading to a first pass reconstruction; (2) a model building stage in which simulations with iJP815pre1 were run and reactions were added until the model could grow in silico on glucose minimal medium; and (3) a model completion stage in which BIOLOG substrate utilisation data was used to guide model expansion and in silico viability on various substrates.

The final reconstruction, dubbed iJP815 after a well-known convention, is made up of 824 intracellular and 62 external metabolites linked by 877 reactions. According to the geneprotein-reaction (GPR) connections, 821 reactions (94%) have at least one assigned gene. GPR connections are made up of Boolean logic statements that use AND and OR operators to link genes to protein complexes and protein complexes to reactions. An 'AND' operator denotes a redundant function that may be catalysed by any of many genes (as in the case of multi-protein complexes), whereas a 'OR' operator denotes a redundant function that can be catalysed by any of several genes (as in the case of multi-protein complexes) (as in the case of isozymes). Only 56 reactions have no related genes, nine of which are non-enzymatic. The remaining 47 nongene-associated enzymatic processes were included to fill in metabolic network gaps discovered during the rebuilding process's various stages.

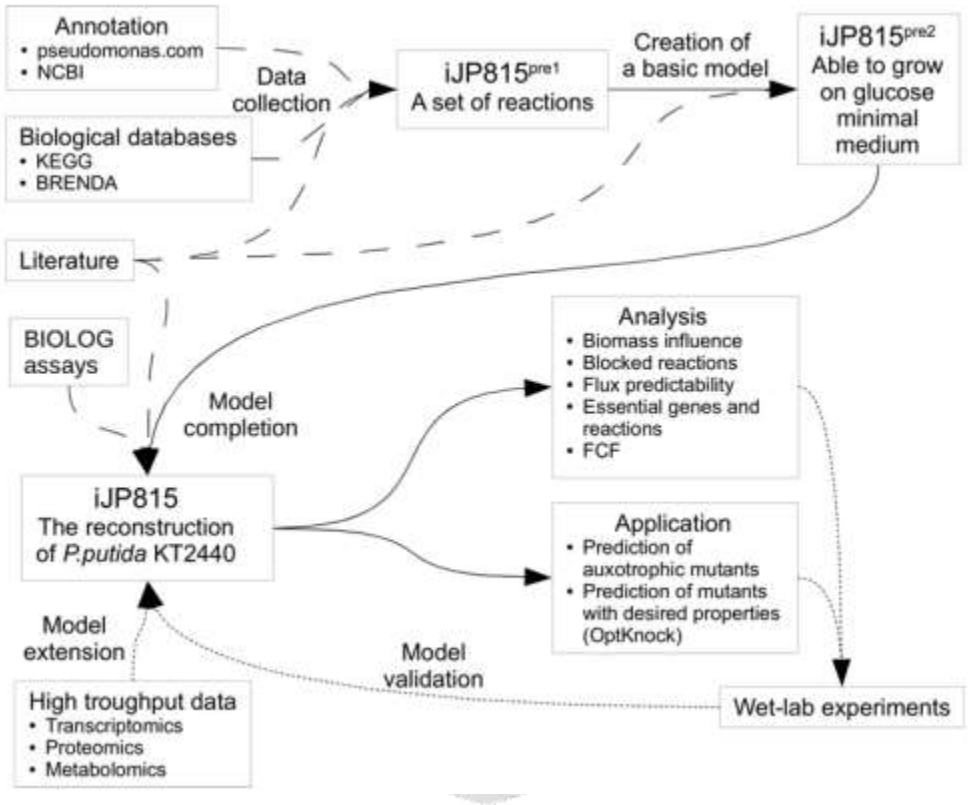

Figure 1: Schematic diagram of the metabolic reconstruction and analysis processes.

During the second phase of reconstruction, most network gaps were detected and repaired by comprehensive literature mining, allowing iJP815 to grow in silico on glucose on minimum media. The remaining gaps in the model completion phase (Figure 1) were primarily single missing stages in the pathway for which experimental proof of functioning was available (e.g., a compound is consumed but not produced, and no alternative pathways exist). It should be noted that there are other combinations of responses that might be used to fill various gaps. When more than one gap closing approach was available, similarity queries to related bacteria were used to determine which to apply.

While imposing the pseudo steady-state assumption, the iJP815 model includes 289 reactions for which non-zero flux values cannot be produced under any environmental scenario (PSSA). These reactions are said to as "unconditionally blocked," which means they are unable to operate since not all connections could be created with the information supplied. "Unbalanced metabolites" are metabolites that are solely engaged in certain reactions and are categorised as such. The "weakly annotated" collection of model reactions is another interesting subset, as all of the genes associated to these 57 reactions are now annotated as coding for "putative" or "family" proteins. The connections between all of the subgroups. The final reconstruction explains the function of 815 genes, which accounts for 15% of all genes in the P. putida genome and 65% (1253) of those currently classified as 'Metabolism' (K01100) and 'Membrane Transport' (K01310) in the Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology classification system. These results are in line with other prokaryotes' metabolic reconstructions that have recently been published.

- **Model Assessment and Extension through High Throughput Phenotyping Assays**

To verify and extend the model, P. putida was subjected to a high-throughput BIOLOG phenotypic experiment. P. putida's capacity to oxidise 95 carbon substrates in minimum media was examined in this experiment. P. putida oxidised 95 of the 95 substrates. Despite a negative BIOLOG result, we added two more carbon sources to the positive-oxidation group (Lphenylalanine and L-threonine) because these substrates had previously been shown to be growth substrates and we confirmed these results experimentally (data not shown), giving a total of forty-seven compounds used in vivo. iJP815^{pre2} accounted for 47 of the 95 carbon sources investigated, allowing a comparison of these BIOLOG results with FBA simulations of iJP815 grown on in silico minimum media with the corresponding molecule as the only carbon source.

The model's first functioning version (iJP815pre2) was able to simulate growth using 14 of the 47 chemicals tested by BIOLOG as sole carbon sources. Only a few transport processes were included in this version of the reconstruction, enabling us to identify chemicals that could not be used in silico due to the lack of a transporter. This was accomplished by enabling each chemical of interest's intracellular pool to be exchanged with the environment in silico and measuring biomass production in each scenario using FBA simulations. This method boosted the number of drugs that may be used to 34, but it also resulted in six false positives (i.e., substances that support in silico growth, but which gave a negative phenotype in the BIOLOG assay). Three metabolites involved in central metabolic pathways (D-glucose 1-phosphate, D-glucose 6-phosphate, and glycerol-3-phosphate), an intermediate in the L-histidine metabolism pathway (urocanate), an intermediate in the biosynthesis of branched amino acids (2-oxobutanoate), and the storage compound glycogen were among them. According to this research, P. putida's inability to use these chemicals in vivo is most likely owing to a lack of adequate transport machinery.

The final P. putida model (iJP815) grew on 39 of the 51 chemicals evaluated in the BIOLOG assay and was accounted for in the model at the same time. Thirty-three of these were genuine positives (compounds utilised in vivo and allowing for growth in silico). The remaining fourteen in vivo oxidised compounds (also known as false negatives) were not used in any way. The remaining 42 compounds were genuine negatives, with eight of them accounting for in the reconstruction. Ten of the chemicals used had no known transport reactions because their translocation into the cell is unknown. Despite this, the comparison of in silico growth projections with BIOLOG substrate use data suggests that P. putida's core metabolism has been correctly recreated.

IV. CONCLUSION

We offer a genome-scale reconstruction and constraint-based model of the P. putida strain KT2440, which includes 815 genes, 877 processes, and 886 metabolites. The reconstruction was based on the bacterium's most recent annotation, the content of numerous biological databases, primary research articles, and particularly developed functional genomics studies. During the rebuilding process, several genes received new or improved annotations. A range of experimental sets, including continuous culture data, BIOLOG substrate usage tests, 13C flow measurements, and a collection of specially created mutant strains, were used to verify the model. FBA and FVA were used to determine the resource distribution in KT2440, to thoroughly examine gene and reaction essentiality, and to assess the metabolic network's resilience. As a result, this study offers one of the most comprehensive sets of studies undertaken for an organism using constraint-based modelling to date, offering a robust genome-scale foundation for the research of this intriguing and adaptable bacterium's metabolism. However, because this

modelling effort is based on a variety of approximations, the analyses' limitations, potential, and application must be explicitly stated and described.

V. REFERENCES: -

1. Bosi, E., Monk, J.M., Aziz, R.K., Fondi, M., Nizet, V., and Palsson, B.O. (2016) Comparative genome-scale modelling of Staphylococcus aureus strains identifies strainspecific metabolic capabilities linked to pathogenicity. Proc Natl Acad Sci USA 113: E3801–E3809.
2. Belda, E., van Heck, R.G.A., José Lopez-Sanchez, M., Cruveiller, S., Barbe, V., Fraser, C., et al. (2016) the revisited genome of Pseudomonas putida KT2440 enlightens its value as a robust metabolic chassis. Environ Microbiol 18: 3403–3424.
3. Berger, A., Dohnt, K., Tielen, P., Jahn, D., Becker, J., and Wittmann, C. (2014) Robustness and plasticity of metabolic pathway flux among uropathogenic isolates of Pseudomonas aeruginosa. PLoS One 9: e88368.
4. Chang, R.L., Andrews, K., Kim, D., Li, Z., Godzik, A., and Palsson, B.O. (2013) Structural systems biology evaluation of metabolic thermotolerance in Escherichia coli. Science 340: 1220–1223.
5. Escapa, I.F., García, J.L., Bühler, B., Blank, L.M., and Prieto, M.A. (2012) the polyhydroxyalkanoate metabolism controls carbon and energy spillage in Pseudomonas putida. Environ Microbiol 14: 1049–1063.
6. Bugg, T.D., Ahmad, M., Hardiman, E.M., and Rahmanpour, R. (2011) Pathways for degradation of lignin in bacteria and fungi. Nat Prod Rep 28: 1883–1896.
7. Ebert, B.E., Kurth, F., Grund, M., Blank, L.M., and Schmid, A. (2011) Response of Pseudomonas putida KT2440 to increased NADH and ATP demand. Appl Environ Microbiol 77: 6597–6605.
8. Cornelis, P. (2010) Iron uptake and metabolism in pseudomonads. Appl Microbiol Biotechnol 86: 1637–1645.