# UNIT-3

## What is metabolic control analysis?

- Metabolic Control is thus defined as the power to change the state of metabolism in response to an external signal, and it is measurable in terms of the strength of the metabolic response to external factors, without any assumption about the function/purpose/mechanism of the response.
- Metabolic control analysis is a method for analysing how the control of fluxes and intermediate concentrations in a metabolic pathway is distributed among the different enzymes that constitute the pathway.
- MCA quantifies how variables, such as fluxes and species concentrations, depend on network parameters.

Flux:is a term used in metabolic analysis to indicate the rate of a multi-component system (metabolic pathway), while "rate" is reserved for individual components (enzyme).

- In particular it is able to describe how network dependent properties, called control coefficients, depend on local properties called elasticities

What is a control coefficient?

- A control coefficient is the system property of an enzyme that expresses how some systemic variable, usually a flux or a metabolite concentration, depends on the activity of the enzyme.
- If some perturbation of an enzyme activity increases the rate of the isolated reaction by 5%, whereas the same perturbation of the same enzyme when it is embedded in a metabolic system increases the flux by 2%, the enzyme is said to have a flux control coefficient of 2/5, or 0.4.
- If the same perturbation causes the concentration of the substrate of the enzyme to decrease by 10%, the enzyme has a concentration control coefficient for that metabolite of -10/5, or -2.
- Notice that there is no mention of the concentration of the enzyme when the control coefficients are defined in this way.
- In the older literature control coefficients were known as sensitivities or control strengths.
- What is an elasticity?
- An elasticity is a local property of an isolated enzyme that expresses how its rate varies with the concentration of any metabolite that affects it: this can be its substrate, product, or any other metabolite.

- An elasticity of, say, 0.5 with respect to a substrate means that a 2% increase in substrate concentration would increase the rate of the reaction catalysed by the enzyme by 1%, i.e. by 0.5 times 2%.
- It follows that substrate elasticities are positive (except under conditions of substrate inhibition), product elasticities are negative (except in the very rare case of product activation), activator elasticities are positive, and inhibitor elasticities are negative.
- Metabolic control analysis was formerly (and is sometimes still) known as metabolic control theory, and is closely related to the engineering discipline known as sensitivity analysis.
- Alternative approaches to studying the kinetic behaviour of multi-enzyme systems are flux-oriented theory and biochemical systems theory.

What is flux-oriented theory

- Flux-oriented theory is an approach developed by E. A. Newsholme, B. Crabtree and their associates that can regarded as an alternative to metabolic control analysis for formalizing control relationships in metabolism.
- Although it has some similarities with metabolic control analysis it differs in important respects.
- In particular, it incorporates the concept of partially external regulators, whose concentrations are partially variable and partially constant.

What is biochemical systems theory

- Biochemical systems theory is an approach developed by M.A.Savageau and his associates, who regard it as a general theory of metabolic control that includes metabolic control analysis and flux-oriented theory as special cases.

- It places much more emphasis on predicting how systems will behave when the conditions are changed than on understanding in physical terms how they are controlled.

In general, the process to build a reconstruction is as follows:

1. Draft a reconstruction
2. Refine the model
3. Convert model into a mathematical/computational representation
4. Evaluate and debug model through experimentation

Drafting a reconstruction:

An initial fast reconstruction can be developed automatically using resources like PathoLogic or ERGO in combination with encyclopedias like MetaCyc, and then manually updated by using resources like PathwayTools.

Databases

Kyoto Encyclopedia of Genes and Genomes (KEGG): a bioinformatics database containing information on genes, proteins, reactions, and pathways. The 'KEGG Organisms' section, which is divided into eukaryotes and prokaryotes, encompasses many organisms for which gene and DNA information can be searched by typing in the enzyme of choice.

BioCyc, EcoCyc, and MetaCyc: BioCyc Is a collection of 3,000 pathway/genome databases (as of Oct 2013), with each database dedicated to one organism. For example,

EcoCyc is a highly detailed bioinformatics database on the genome and metabolic reconstruction of *Escherichia coli*, including thorough descriptions of *E. coli* signaling pathways and regulatory network. The EcoCyc database can serve as a paradigm and model for any reconstruction.

MetaCyc, an encyclopedia of experimentally defined metabolic pathways and enzymes, contains 2,100 metabolic pathways and 11,400 metabolic reactions .

ENZYME: An enzyme nomenclature database (part of the ExPASy proteonomics server of the Swiss Institute of Bioinformatics). After searching for a particular enzyme on the database, this resource gives you the reaction that is catalyzed. ENZYME has direct links to other gene/enzyme/literature databases such as KEGG, BRENDA, and PUBMED.

BRENDA: A comprehensive enzyme database that allows for an enzyme to be searched by name, EC number, or organism.

BiGG: A knowledge base of biochemically, genetically, and genomically structured genome-scale metabolic network reconstructions.

metaTIGER: Is a collection of metabolic profiles and phylogenomic information on a taxonomically diverse range of eukaryotes which provides novel facilities for viewing and comparing the metabolic profiles between organisms.

Tools for Metabolic Modeling:

Pathway Tools:

- A bioinformatics software package that assists in the construction of pathway/genome databases such as EcoCyc.Developed by Peter Karp and associates at the SRI International Bioinformatics Research Group, Pathway Tools has several components.
- Its PathoLogic module takes an annotated genome for an organism and infers probable metabolic reactions and pathways to produce a new pathway/genome database.
- Its MetaFlux component can generate a quantitative metabolic model from that pathway/genome database using flux-balance analysis.
- Its Navigator component provides extensive query and visualization tools, such as visualization of metabolites, pathways, and the complete metabolic network.
- ERGO: A subscription-based service developed by Integrated Genomics. It integrates data from every level including genomic, biochemical data, literature, and high-throughput analysis into a comprehensive user friendly network of metabolic and nonmetabolic pathways.
- KEGGtranslator:an easy-to-use stand-alone application that can visualize and convert KEGG files (KGML formatted XML-files) into multiple output formats.
- Model SEED: An online resource for the analysis, comparison, reconstruction, and curation of genome-scale metabolic models. Users can submit genome sequences to the RAST annotation system, and the resulting annotation can be automatically piped into the Model SEED to produce a draft metabolic model. The Model SEED automatically constructs a network of metabolic reactions, gene-protein-reaction associations for each reaction, and a biomass composition reaction for each genome to produce a model of microbial metabolism that can be simulated using Flux Balance Analysis.
- MetaMerge: algorithm for semi-automatically reconciling a pair of existing metabolic network reconstructions into a single metabolic network model.

Tools for Literature:

PUBMED: This is an online library developed by the National Center for Biotechnology Information, which contains a massive collection of medical journals. Using the link provided by ENZYME, the search can be directed towards the organism of interest, thus recovering literature on the enzyme and its use inside of the organism.

Methodology to draft a reconstruction:

- A reconstruction is built by compiling data from the resources.
- Database tools such as KEGG and BioCyc can be used in conjunction with each other to find all the metabolic genes in the organism of interest.

- These genes will be compared to closely related organisms that have already developed reconstructions to find homologous genes and reactions.
- These homologous genes and reactions are carried over from the known reconstructions to form the draft reconstruction of the organism of interest.
- Tools such as ERGO, Pathway Tools and Model SEED can compile data into pathways to form a network of metabolic and non-metabolic pathways.
- These networks are then verified and refined before being made into a mathematical simulation

Model Refinement:

- An initial metabolic reconstruction of a genome is typically far from perfect due to the high variability and diversity of microorganisms. Often, metabolic pathway databases such as KEGG and MetaCyc will have "holes", meaning that there is a conversion from a substrate to a product (i.e., an enzymatic activity) for which there is no known protein in the genome that encodes the enzyme that facilitates the catalysis.

- What can also happen in semi-automatically drafted reconstructions is that some pathways are falsely predicted and don't actually occur in the predicted manner.

- Because of this, a systematic verification is made in order to make sure no inconsistencies are present and that all the entries listed are correct and accurate.

- Furthermore, previous literature can be researched in order to support any information obtained from one of the many metabolic reaction and genome databases. This provides an added level of assurance for the reconstruction that the enzyme and the reaction it catalyzes do actually occur in the organism.

A number of algorithms and bioinformatics resources have been developed for refinement of sequence homology-based assignments of protein functions:

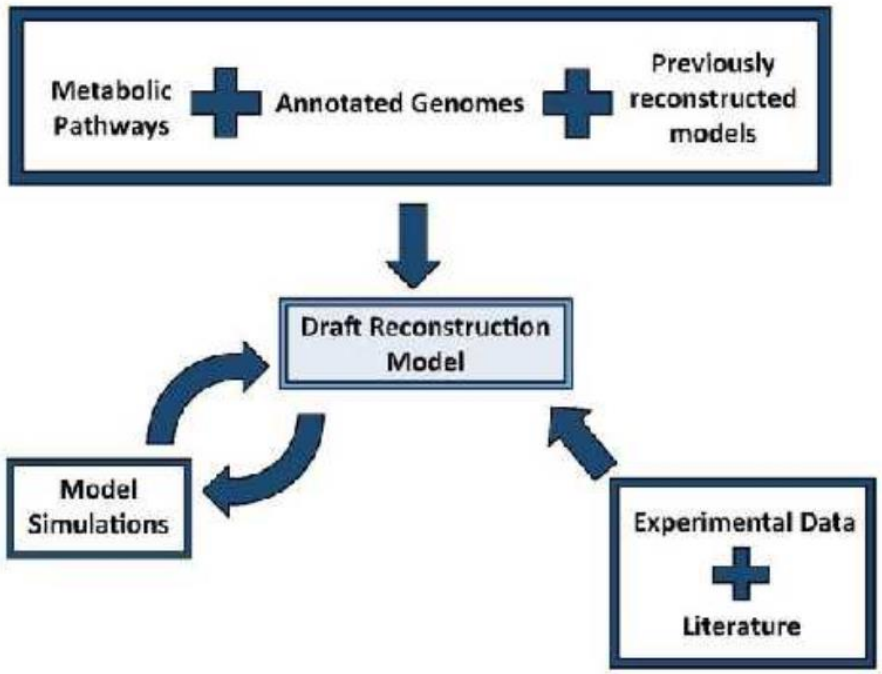
1. InParanoid: Identifies eukaryotic orthologs by looking only at in-paralogs.
2. CDD: Resource for the annotation of functional units in proteins. Its collection of domain models utilizes 3D structure to provide insights into sequence/structure/function relationships.
3. InterPro: Provides functional analysis of proteins by classifying them into families and predicting domains and important sites.
4. STRING: Database of known and predicted protein interactions.

Applications of a reconstruction:

1. Several inconsistencies exist between gene, enzyme, and reaction databases and published literature sources regarding the metabolic information of an organism. A

reconstruction is a systematic verification and compilation of data from various sources that takes into account all of the discrepancies.

2. The combination of relevant metabolic and genomic information of an organism.
3. Metabolic comparisons can be performed between various organisms of the same species as well as between different organisms.
4. Analysis of synthetic lethality
5. Predict adaptive evolution outcomes
6. Use in metabolic engineering for high value outputs




**Genome-scale reconstruction and *in silico* analysis of the *Clostridium acetobutylicum* ATCC 824 metabolic network**

- *Clostridium acetobutylicum* is a Gram-positive anaerobic bacterium that produces several solvents including acetone, butanol, and ethanol. Due to demand for solvents in the chemical industry and during the two World Wars in the twentieth century, acetone–butanol–ethanol (ABE) fermentation using *Clostridium* spp. had been well developed and used on a large scale.

- However, biobutanol production by *C. acetobutylicum* stopped for the most part in the middle of twentieth century (with a few exceptions) due to the rapid growth of the petrochemical industry.As the oil price sharply increased together with our environmental concerns, biobutanol production has come into focus as an alternative to gasoline. Therefore, interest in solvent-producing clostridia has come back to the forefront.

- Since native solventogenic clostridia produce by-products such as acetone (or isopropanol), acetate, and butyrate (Jones and Woods 1986), it needs to be metabolically engineered to improve butanol yield.

- Recent advances in genomics made it possible for us to reconstruct "genome-scale" metabolic network of an organism.

- A useful and popular tool to investigate such reconstructions is flux balance analysis (FBA). FBA is based on linear programming and is used to evaluate flux distribution in a metabolic network under governing constraints.

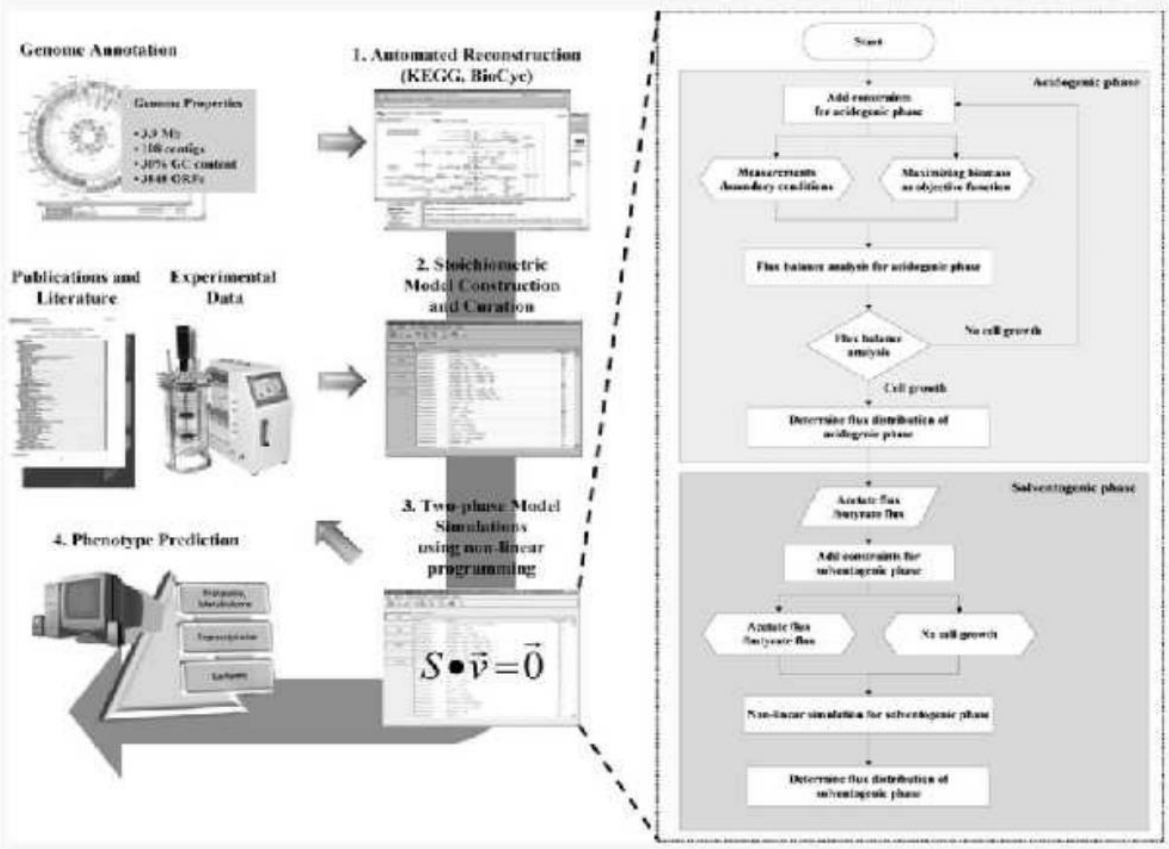
- Using this formalism, gene deletion simulations can provide qualitative and quantitative predictions of network robustness and the production rates of specific metabolites. Such predictions have proven useful in strain improvement for metabolite production.

Metabolic network reconstruction:

A metabolic network model of *C. acetobutylicum* ATCC 824 was constructed using the combined information from many different sources such as public databases, literature, and experiments.

The primary genome annotation of *C. acetobutylicum* ATCC 824 was obtained from NCBI (http://www.ncbi.nlm.nih.gov/), and it was used as a guideline to build gene–protein–reaction (GPR) relationships.

In addition, for efficient network reconstruction, we collected the annotation data from various databases, including BioSilico (Hou et al. 2004), KEGG (Kanehisa et al. 2006), TIGR (http://www.tigr.org), and MetaCyc (Caspi et al. 2006). Information on most biochemical reactions was obtained from the BioSilico and KEGG databases and was used to develop the draft reconstruction.




The process utilized to reconstruct the metabolic network of *C. acetobutylicum*.

*1* .The *C. acetobutylicum* ATCC 824 annotation was used along with several databases to generate an automated reconstruction.

*2* Next, based on the automated reconstruction, a draft model was developed and curated using various literature and experimental data sources. Gaps in the automated reconstruction were filled during this step.

*3* Results from two-phase analysis were compared with experimental data to validate the content and modeling approach.

*4* After validation, the curated model, *Cac*MBEL502, was utilized for phenotype predictions of *in silico* knockout strains and genome-wide metabolic engineering applications of *C. acetobutylicum*.

*Right panel* of the figure shows flow chart diagram for two-phase flux analysis using nonlinear programming.

First, typical FBA using linear programming was carried out for acidogenic phase. In this case, maximization of cellular growth was the objective function. Acetate and

butyrate production rates in the acidogenic simulation were used as additional constraints for nonlinear solventogenic constraints

## metaTIGER

- Metabolic networks are a subject that has received much attention, but existing web resources do not include extensive phylogenetic information.
- Phylogenomic approaches (phylogenetics on a genomic scale) have been shown to be effective in the study of evolution and processes like horizontal gene transfer (HGT).
- To address the lack of phylogenomic information relating to eukaryotic metabolism, metaTIGER (www.bioinformatics.leeds.ac.uk/metatiger) has been created, using genomic information from eukaryotes and prokaryotes and sensitive sequence search techniques to predict the presence of metabolic enzymes
- metaTIGER is a collection of metabolic profiles and phylogenomic information on a taxonomically diverse range of eukaryotes. Phylogenomic information is provided by 2,257 large phylogenetic trees which can be interactively explored.
- High-throughput tree analysis can also be carried out to identify trees of interest, e.g. trees containing horizontal gene transfers. metaTIGER also provides novel facilities for viewing and comparing the metabolic profiles.


- The prediction of enzymes and in turn organisms metabolic profiles is based upon the program SHARKhunt.
- SHARKhunt is a high-throughput genome annotation program which takes genomic DNA sequence and searches it with enzyme proifles.
- The enzymes profiles are based upon alignment of the amino acids sequence of conserved regions of genes of known function.

- These are used to search the genomes using a combination of two sensitive bioinformatics techniques, PSI-BLAST and Hidden Markov Models, which means distant homologs can be detected in highly diverged organisms.

metaTIGER provides a variety of ways of exploring and comparing these enzyme predictions, including:

- search functions,
- organism specific pathway images,
- comparative pathway images and the potential to compare 10's of organisms at once through table based comparison.
- To accompany each of the enzyme profiles in metaTIGER is a maximum-likelihood phylogenetic tree that was produced from an alignment of the amino acid sequences.
- The phylogenetic trees can be viewed interactively on the site or they can be search with complex user designed tree queries.
- These enzyme sequences were used to create a comprehensive database of 2257 maximum-likelihood phylogenetic trees, some containing over 500 organisms.
- The trees can be viewed using iTOL, an advanced interactive tree viewer, enabling straightforward interpretation of large trees.
- Complex high-throughput tree analysis is also available through user-defined queries, allowing the rapid identification of trees of interest, e.g. containing putative HGT events.
- metaTIGER also provides novel and easy-to-use facilities for viewing and comparing the metabolic networks in different organisms via highlighted pathway images and tables.
- metaTIGER is demonstrated through evolutionary analysis of *Plasmodium*, including identification of genes horizontally transferred from chlamydia.
- Searching metaTIGER:
- A specific eukaryote or groups of eukaryotes can be searched for a particular enzyme or compound.
- For all searches the user can determine the search type, the e-value cut-off, the group of organism or the organism to search through, and the text to be searched for.
- Depending on how the search facility is used the search program will construct its queries in a way that will aim to give the fastest and most user friendly output.

<u>Enzyme searches in a</u>

single organisms using

- a ec number - returns a link to the enzyme info
- a enzyme name - returns a list of enzyme that match the name imputed each with a link to its enzyme info page
- blank input - returns all the enzymes that are predicted to be in that organism

group of organisms using

- a ec number - returns a list of eukaryotes that have that ec number each with a link to the enzyme info page. A fasta format file of all the

profile hits that meet search criteria can be downloaded from the results page

- a enzyme name - returns a list of enzymes that match the input. A particular enzyme can then be selected to search for within the group of eukaryotes that you are interested in

- blank input - returns a list of all the enzymes. A particular enzyme can then be selected to search for within the group of eukaryotes that you are interested in.

- Enzyme Information Pages:
- The enzyme info page contains several sections:-
- Title - which is made up of the enzymes name, ec number, e-value
- BLAST - there are two different BLAST options; the first runs a search against the UniPro database, the second option is a recomputed search of the NCBI's non-redundant protein database
- Pathway Links - a list links that take you to the pathways in which the enzyme occurs, where all the reactions that are predicted to occur within the organism are highlighted
- Data - a link to a fasta file of the AA sequences of the hits
- External Links - a list of links that take you to the information pages for the enzyme in other databases
- Profile Trees - the phylogenetic trees of the profiles. The organisms of relevance is highlighted by a black square on the outside of the tree.

Viewing metaTIGER Pathways

There are two ways of exploring the pathway images contained within metaTIGER.

1. The first is to us the all pathways option that allows users to select any organism and pathway that is contained in metaTIGER.
2. The second option is to use the filtered pathways menu; this allows users explore the pathways that are most relevant their organism of choice.

Images of KEGG pathways specific to each of the eukaryotes can be viewed.

Enzymes are colored with the following scheme:-

Red = hits <= 1.0e-30

Orange = hits >1.0e-30 and <= 1.0e-20

Yellow = hits > 1.0e-20 and <= 1.0e-10

Silver = enzymes which do not have a profile and therefore are undetectable

Each of the enzymes acts a link to the enzyme info page and each of the compounds act as links to the corresponding KEGG compound information page. Pathways can be navigated through by using the pathway name links on the images or by selecting a different organism/pathway at the top of the page and clicking view.

Compounds searches can be carried out in four different ways depending upon the involvement of the compound in the reaction that you are interested in.

The different compound involvement types are:-

Metabolites

Cofactors
Effectors
Inhibitors