F

Faceted Browsing

▶ Search Engines with Faceted Search

Faceted Navigation

▶ Search Engines with Faceted Search

Fact Extraction

▶ Information Extraction

Factors Modulating Nucleosome Structure

▶ Nucleosome Acting Factors

False Discovery Rate (FDR)

Sigrid Rouam
Université Paris-Sud, Villejuif, France

Definition

The false discovery rate (FDR) is a statistical approach used in multiple hypothesis testing to correct for multiple comparisons. It is typically used in high-throughput experiments in order to correct for random events that falsely appear significant. When testing a null hypothesis to determine whether an observed score is statistically significant, a measure of confidence, the p-value, is calculated and compared to a confidence threshold α. When k hypotheses are tested simultaneously with a confidence level α, the chances of occurrence of false positives (i.e., rejecting the null hypothesis when in fact it is true) is equal to $1 - (1 - \alpha)^k$, which can lead to a high error rate in the experiment. Therefore, a multiple testing correction, such as the FDR, is needed to adjust our statistical confidence measures based on the number of tests performed.

The FDR is defined as the expected proportion of false discoveries, i.e., incorrectly rejected null hypothesis, among all discoveries (Benjamini and Hochberg 1995). Consider the problem of testing simultaneously m null hypothesis H_0 versus the alternative hypothesis H_1, of which m_0 are true. We define the following random variables (see Table 1): TN is the number of true negatives; FP is the number of false positives (or type I errors); FN is the number of false negatives; TP is the number of true positives; R is the number of hypotheses rejected. R is an observable random variable, while TN, FN, TP, and FP are unobservable random variables.

The FDR is given by

$$\mathrm{FDR} = E\left[\frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TP}}\right] = E\left[\frac{\mathrm{FP}}{\mathrm{R}}\right] \text{ if } R > 0; 0 \text{ otherwise.} \quad (1)$$

W. Dubitzky et al. (eds.), *Encyclopedia of Systems Biology*, DOI 10.1007/978-1-4419-9863-7,

False Discovery Rate (FDR), Table 1 Number of errors committed when testing m null hypothesis H_0 versus H_1

	Accept H_0	Reject H_0	Total
H_0 true	TN	FP	m_0
H_1 true	FN	TP	$m - m_0$
Total	$m - R$	R	m

The FDR is the proportion of the rejected null hypotheses which are erroneously rejected. In practice, the level of acceptable FDR is fixed by the investigator, depending on the desired stringency.

An example of the practical use of the FDR is in genomic association studies, for which a large number of statistical tests are performed simultaneously. In such studies, the objective is to identify genomic factors worthy of further analysis. If the multiplicity of the data is not taken into account, the probability that a false identification (type I error) is committed can increase sharply when the number of tested genes becomes large. The FDR controlling approach is widely used to avoid such errors.

References

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc Ser B 57(1):449–518

False Positive Rate

Haiying Wang and Huiru Zheng
School of Computing and Mathematics, Computer Science Research Institute, University of Ulster, Jordanstown, UK

Synonyms

Type I error rate

Definition

In statistical analysis, the false positive rate of a test is defined as the probability of rejecting the null hypothesis H_0 when it is true, which can be denoted as:

False Positive Rate, Table 1 Sample confusion matrix

		Predicted	
		A	Non-A
Actual	A	TP	FN
	Non-A	FP	TN

$$false\ positive\ rate(\alpha) = \{reject\ H_0 | H_0\ true\}$$

In machine learning (▶ Model Validation, Machine Learning), the false positive rate is closely related to the notion of specificity, one of statistical measures widely used to assess the performance of prediction models.

Let TP be true positives (samples correctly classified as class A), FN be false negatives (samples incorrectly classified as not belonging to class A), FP be false positives (samples incorrectly classified as class A), and TN be true negatives (samples correctly classified as not belonging to class A). The relationship between these prediction outcomes can then be summarized using a confusion matrix (Kohavi and Provost 1998) as illustrated Table 1.

The false positive rate is the proportion of samples not belonging to class A that were incorrectly classified as class A, i.e.,

$$\begin{aligned} false\ positive\ rate\ (\alpha) &= {}^{FP}/_{(FP + TN)} \\ &= 1 - specificity \end{aligned}$$

As an example, suppose we want to build a ▶ classification model using gene expression microarray data to predict whether a subject has a disease or not. In a total of 100 subjects known to be free of a disease, the model actually predicts 10 subjects having the disease. In this scenario, FP = 10 and TN = 90. Thus, the false positive rate is 10%.

Cross-References

▶ Model Validation, Machine Learning

References

Kohavi R, Provost F (1998) Glossary of terms. Mach Learn 30:271–274

FBA Analysis, Plant-Pathogen Interactions

Andrés Mauricio Pinzón Velasco[1], Silvia Restrepo[2] and Andrés Fernando González Barrios[3]
[1]Department of Systems Engineering, Universidad Jorge Tadeo Lozano, Bogotá, Colombia
[2]Department of Biological Sciences, Universidad de los Andes, Bogotá, Colombia
[3]Department of Chemical Engineering, Universidad de los Andes, Bogotá, Colombia

Synonyms

Flux balance analysis; Pathosystem

Definition

Flux Balance Analysis (FBA) linear programming based optimization approach that departs from an objective cell or system function restricted by mass and energy conservation, which has been widely used for the analysis of biochemical fluxes in metabolic networks. This approach allows researchers to determine changes in metabolic phenotypes given a set of biochemical constraints. In the case of plant-pathogen interactions, or any host-pathogen model, this methodology is particularly promising since it offers the opportunity to analyze the structure, dynamics, and complex behavior of metabolic networks during a particular infective process.

Characteristics

Mathematical modeling of plant-pathogen interactions has been carried out since the 1950s (Pinzón et al. 2009). Typically, plant pathologists have approached the quantitative and computational modeling of this type of interactions focusing on processes at high levels of resolution, such as the description of temporal dynamics of crops diseases and spatial patterns of dispersion. From its very beginning these models belonged to the family of logistic equations and its application to a broad spectrum of plant diseases has been a constant since then (Pinzón et al. 2009).

Nevertheless, this type of modeling lacks the capacity to represent the underlying biological processes that take place during an infection, for instance, those that modify infected cell behavior at the molecular level. Recently, due to the availability of high-throughput biological data as well as the development of sophisticated computational tools, there is a growing interest in the modeling of these underlying molecular processes.

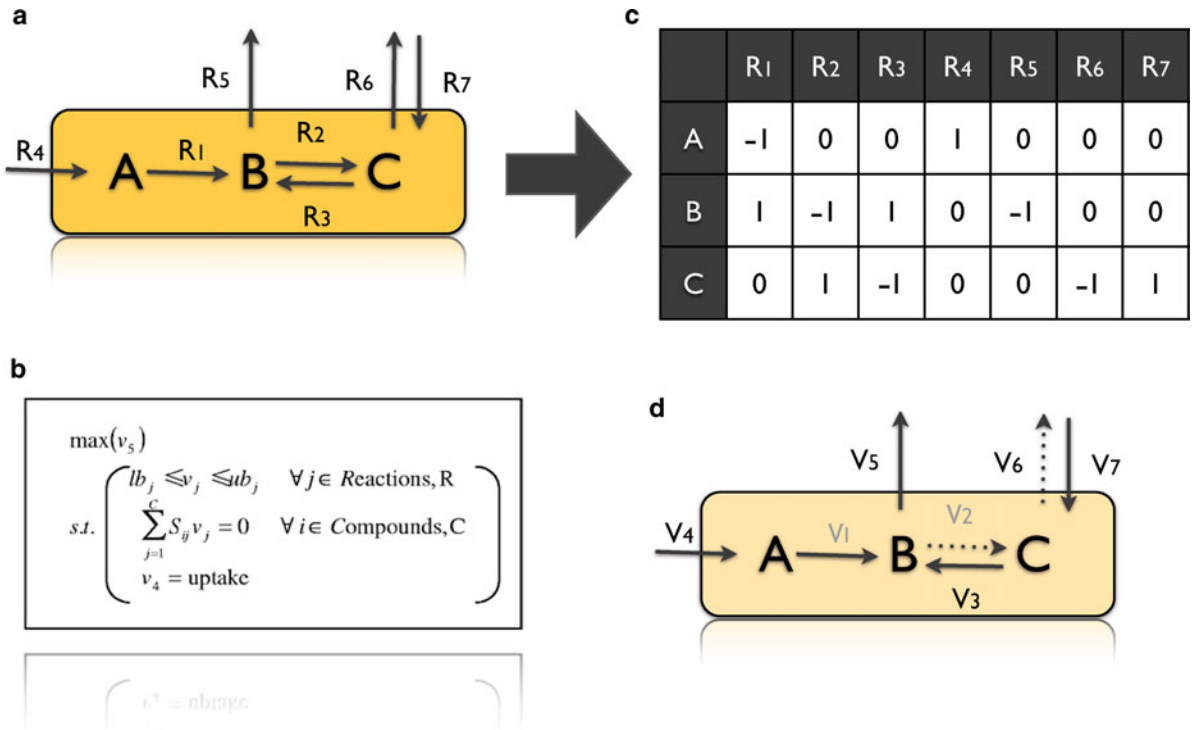
Computational Modeling of Metabolism

Although the way plants resist a particular pathogen attack vary according to plant species and specific pathogen characteristics, in general terms, plants face a pathogen attack by shifting their defense mechanisms. These mechanisms consists of a complex and highly interconnected set of networks, in which host defense genes interact with each other as well as with pathogen proteins present in the cell. Metabolic and regulatory networks are of particular interest when studying this kind of biological processes. It is at this level that one expects that pathogen manipulation lead to phenotypic and behavioral differences among plants under attack. In this context, a particular approach for the in silico representation and analysis of the set of biochemical reactions that take place in a given organism, known as metabolic reconstructions (MRs), is of particular interest.

Typically MRs are based on genomics information and therefore known as GEMRs or Genome Scale Metabolic Reconstructions (Thiele and Palsson 2010). This type of reconstructions are based on the gene content of a complete genome, from where the set of genes that code for enzymes is selected and the reactions they belong to are gathered.

Another approach for metabolic reconstruction is known as TMRs or Targeted Metabolic Reconstruction (Pinzón et al. 2010), which are not based on genome content but in transcriptomics information. Due to its main characteristic TMRs do not cover the complete set of potential genes present in an organism genome, but they do provide a way to analyze *active metabolic networks*, as the set of enzymes expressed at a given time point and under particular conditions.

Overall, the main aim of metabolic reconstructions is the identification of suitable targets for metabolic engineering, the improvement of production yields and nutritional value of crops, as well as the understanding of certain processes such as resistance or defense from pathogens, for instance, through the study of time



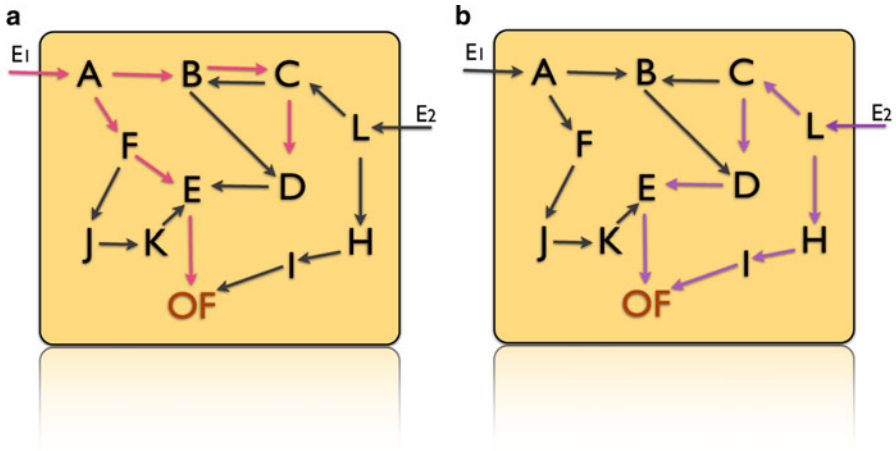
FBA Analysis, Plant-Pathogen Interactions, Fig. 1 FBA is a constraint-based optimization method that facilitates the computational prediction of systemic phenotypes in the form of fluxes of reactions. For instance, given a set of available nutrients for an organism, FBA allows for the prediction of the set of fluxes of metabolic reactions that optimize the growth for that organism. FBA requires the conversion of the metabolic system (A) into a stoichiometric matrix "S" (C). In this matrix rows represent metabolites and columns reactions. In R1, metabolite A is consumed (−1) and metabolite B is produced (1), metabolite C does not participate in the reaction (0). The FBA model usually optimizes for a particular characteristic in the organism, which is commonly called the *Objective Function (OF)*. Usually growth is used as OF, and it is represented in the form of biomass. However, under particular conditions for some metabolic reconstructions in plants, other OFs can be described. For example, in *S. tuberosum* and other tuber-like plants it is feasible to use starch production and storage as OF instead of a typical biomass set of reactions, given that starch storage could be seen as an indicator of growth. Part B shows a standard form of a FBA problem where flux through the OF is maximized subject to some constraints such as reactions stoichiometric and uptake. In this way reaction fluxes (V_j) are between lower (*lb*) and upper (*ub*) bounds. Since FBA assumes a steady state for all reactions, metabolite concentrations are fixed ($S.v = 0$)

evolution of a host network under pathogen attack. One of the most important characteristics of MRs is that they can be interrogated by means of computational modeling and therefore derive hypothesis based on model predictions.

Flux Balance Analysis (FBA) in the Context of Plant-Pathogen Interactions

There are several methodologies than can be used to interrogate a metabolic reconstruction (Oberhardt et al. 2009). A common constraint-based method, known as Flux Balance Analysis or FBA (Becker et al. 2007), has been widely used for this purpose (Fig. 1).

Using FBA it is possible to determine how microorganisms utilize their metabolism, and predict their changes under environmental perturbations. For instance, it is possible to assess the effect that a single gene deletion can have over the flux of all reactions in a metabolic network (Fig. 2). These fluxes of reactions correspond to intracellular biochemical networks that, due to the lack of kinetics information, are assumed to operate under pseudo steady-state conditions, which seems to be in agreement with experimental data. For instance, this approach showed an 85% consistency of gene essentiality for the genome scale reconstruction of *Escherichia coli* and 70% consistency for gene



FBA Analysis, Plant-Pathogen Interactions, Fig. 2 (**a**) metabolic profile before virtual knockout. (**b**) Metabolic profile after virtual knockout of reactions between B and C and between F and E. OF: Objective function. In part (**a**) of the figure it is clear how a different flux of reactions is present for OF optimization. Based on biological information, such as microarray, SAGE or protein-protein interaction data, it is possible to silence some reactions and let FBA to predict flux change under these new conditions. In the case of plant-pathogen interactions, it is possible to use data from known host R-proteins and pathogen effectors interaction, as well as genes with differential expression obtained by essays of plants challenged with the pathogen

essentiality for the reconstruction of *Pseudomonas aeruginosa* (Reviewed in Pinzón et al. 2009).

This characteristic of FBA is of particular interest in the study of plant-pathogen interactions, but before trying to clarify how this approach can be used in this context, some background on the pathogenicity and host defense response is necessary.

Plant-pathogen interactions can be divided in two main mechanisms. During the first one, known also as *basal defense*, the plant recognizes general features in pathogens that are universally associated with microbes. These features are known as MAMPs (Microbe-Associated Molecular Patterns) which are recognized by membrane receptors in plants. The recognition of these MAMPs triggers the first line of defense, which for most plants it is enough to stop pathogen attack. However, some plant pathogens are specific to a host and have developed special strategies for the manipulation of host defenses. This is the case for the pathogen *Phytophthora infestans* in *Solanum tuberosum*. During the pathogenicity process *P. infestans* releases an arsenal of proteins known as *effectors*, which are injected into the host plant through specialized secretion systems, suppressing the first line of defense. *S. tuberosum* in turn, have evolved to recognize these effectors using Resistance proteins (R-proteins), which directly or indirectly interact with them. Not always this interaction between host R-proteins and pathogen effectors is effective and then the pathogen is able to suppress host recognition and the plant defense response derived from it, leading to a disease known as *Late blight of potato*.

This pattern of interaction is also typical for many other pathosystems such as *Arabidopsis thaliana* and *Pseudomonas syringae* or *A. thaliana* and *Hyaloperonospora parasitica*, among many others.

Although for many of these pathosystems some molecular details are known, this is not the reality for most of them. In the case of *P. infestans* and *S. tuberosum*, some phenotypical responses to plant attack are recognized to date, such as a decrease in plant's photosynthetic capacity several hours after infection, but the precise biochemical mechanisms that lead to that phenotype are not known. Although typical molecular approaches (proteomics, transcriptomics, etc.) have shown to be effective in revealing how some of these mechanisms work, the methodologies are not tractable when trying to understand a complex network of biochemical interactions. Therefore, FBA over a metabolic reconstruction can be an ideal alternative for the comprehension of this type of networks.

In general, what a FBA analysis describes is the set of fluxes of reactions that take place in a metabolic

network under a particular environment. Therefore, if we take into consideration different environment conditions, it is possible to evaluate changes over the same network given those new conditions. If we know, for instance, that a particular effector protein acts as repressor for the expression of a host gene involved in metabolism, we can represent this situation by a virtual knockout of this gene, for which different software tools are available (Hoppe et al. 2011; Rocha et al. 2010; Becker et al. 2007). By the study of the network after and before this knockout it is possible to obtain a different profile of the reactions implied in both stages and derive important biological conclusions from their analysis (Fig. 2).

Although metabolism is crucial for most cellular activities, it is also important to take into account that metabolism response is highly integrated to signaling that comes from different routes. In the case of plant-pathogen interactions, it is then possible to integrate into the FBA information regarding typical signal defense routes, such as ethylene, jasmonic, and salicylic acids pathways. The information related to this signaling process can be integrated into FBA as a set of restrictions, similar to the virtual knockouts described before.

References

Becker SA, Feist AM, Mo ML, Hannum G, Palsson BØ, Herrgard MJ (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox. Nat Protoc 2(3):727–738. doi:10.1038/nprot.2007.99

Hoppe A, Hoffmann S, Gerasch A, Gille C, Holzhutter H-G (2011) FASIMU: flexible software for flux-balance computation series in large metabolic networks. BMC Bioinformatics 12(1):28. doi:10.1186/1471-2105-12-28

Oberhardt MA, Palsson BØ, Papin JA (2009) Applications of genome-scale metabolic reconstructions. Mol Syst Biol 5:320. doi:10.1038/msb.2009.77

Pinzón A, Barreto E, Bernal A, Achenie L, Barrios AFG, Isea R et al (2009) Computational models in plant-pathogen interactions: the case of *Phytophthora infestans*. Theor Biol Med Model 6:24. doi:10.1186/1742-4682-6-24

Rocha I, Maia P, Evangelista P, Vilaca P, Soares S, Pinto JP, Nielsen J et al (2010) OptFlux: an open-source software platform for in silico metabolic engineering. BMC Syst Biol 4(1):45. doi:10.1186/1752-0509-4-45

Thiele I, Palsson BØ (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc 5(1):93–121. doi:10.1038/nprot.2009.203

Feasibility

Frank Allgöwer, Jan Hasenauer and Steffen Waldherr
Institute for Systems Theory and Automatic Control, University of Stuttgart, Stuttgart, Germany

Definition

A feasibility problem denotes the mathematical problem of showing that there exist values for an unknown variable which satisfy specific constraints. In an n-dimensional vector space over the real numbers, the problem is to find $x \in \mathbb{R}^n$ such that $f_i(x) \leq 0$, $i = 1,\ldots, m$, for a given list of m scalar-valued functions f_i. This problem is mathematically written as

$$\begin{aligned} &\text{find} && x \in \mathbb{R}^n \\ &\text{subject to} && f_i(x) \leq 0, \quad i = 1, \ \ldots, \ m. \end{aligned}$$

Feasibility problems are tightly linked to optimization. Any feasibility problem can be formulated as an optimization problem with a constantly zero objective function:

$$\begin{aligned} &\underset{x \in \mathbb{R}^n}{\text{minimize}} && 0 \\ &\text{subject to} && f_i(x) \leq 0, \quad i = 1, \ \ldots, \ m. \end{aligned}$$

If the constraints cannot be satisfied, the optimal value is ∞ by definition. On the other hand, if an x exists that satisfies the constraints, the optimal value will be 0.

The formulation of a feasibility problem as optimization problem is particularly attractive for constraint classes where a solver exists which is guaranteed to find the optimal value, e.g., ▶ semidefinite programs. In this case, the optimization algorithm can directly be used to decide the feasibility problem.

Cross-References

▶ Model Falsification, Semidefinite Programming
▶ Semidefinite Program

References

Boyd S, Vandenberghe L (2004) Convex optimization. Cambridge University Press, Cambridge

Feasible Parameter Space

► Dynamic Metabolic Networks, k-Cone

Feature Reduction

► Feature Selection

Feature Selection

Chenglei Sun
Institute of Systems Biology, Shanghai University, Shanghai, Shanghai, China

Synonyms

Attribute selection; Feature reduction; Variable selection; Variable subset selection

Definition

Feature selection is a must in data mining. The main idea of feature selection is to choose a subset of input variables by eliminating the noisy or redundant features and keeping the quality patterns. Feature selection can significantly improve the effectiveness and robustness of the resulting classifier or regression models. More importantly, feature selection can help to discover the features that are really important in the classification.

There are two types of feature selections: filter and wrapper. Filter methods evaluate the goodness of the feature subset by using the intrinsic characteristic of the data. They are relatively computationally cheap, since they do not involve the induction algorithm. However, they also take the risk of selecting subsets of features which may not match the chosen induction algorithm. Wrapper methods, on the contrary, directly use the induction algorithm to evaluate the feature subsets. They generally outperform filter methods in terms of prediction accuracy, but are generally computationally more intensive. In summary, filter and wrapper methods can complement each other.

References

Dash M, Liu H (1997) Feature selection for classification, intelligent data analysis. An Int'l J 1(3):131–156
Kim YS, Street WN, Menczer F (2003) Feature selection in data mining. In: Wang J (ed) Data mining: opportunities and challenges. Idea Group, Hershey, pp 80–105

Feed Forward Loop

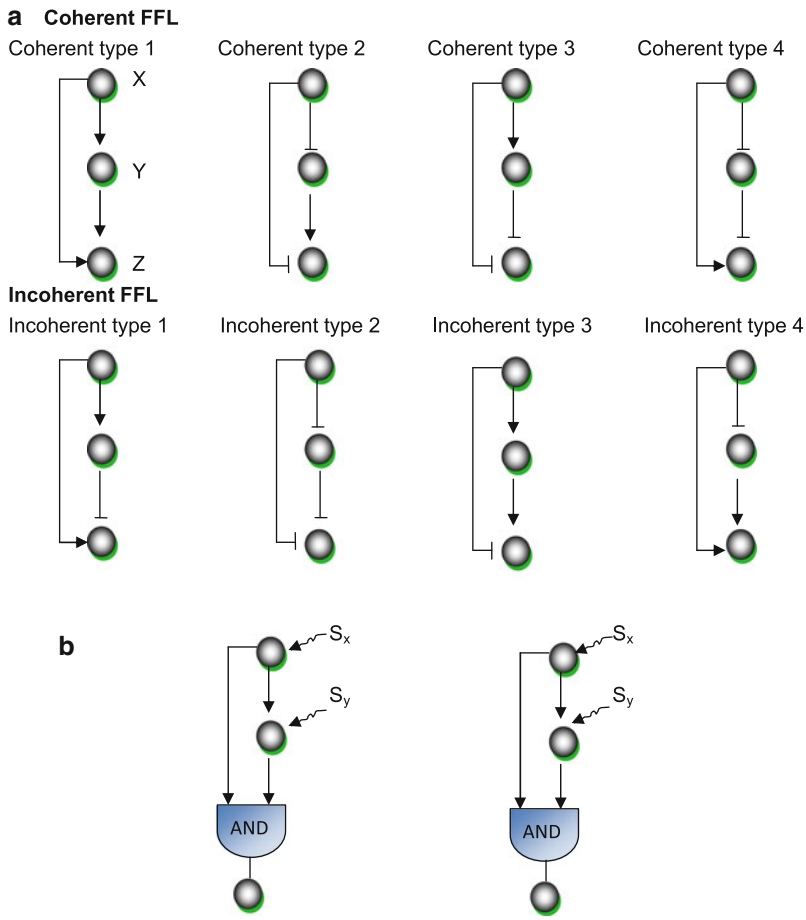
Guangxu Jin
Systems Medicine and Bioengineering, Bioengineering and Bioinformatics Program, The Methodist Hospital Research Institute, Weill Medical College, Cornell University, Houston, TX, USA

Synonyms

Feed forward loop motif, FFL

Definition

Feed forward loop (FFL) motif is one of the most significant one in both *E. coli* and yeast. The FFL is composed of a transcription factor X, which regulates a second transcription factor Y. X and Y both bind the regulatory region of target gene Z and jointly modulate its transcription rate. The FFL has three transcription interactions. Each of these can be either positive (activation) or negative (repression). There are therefore eight possible structural configurations of activator and repressor interactions. Four of these configurations are termed "coherent": the sign of the direct regulation path (from X to Z) is the same as the overall sign of the indirect regulation path (from X through Y to Z). The other four structures are termed "incoherent": the signs of the direct and indirect regulation paths are opposite. Mathematical modeling indicates that FFLs can serve as a novel mechanism for accelerating the expression of the target genes. Both coherent and incoherent FFL behaviors are sign sensitive: they accelerate or delay responses to stimulus steps (Fig. 1).



Feed Forward Loop, Fig. 1 The structures for coherent and incoherent FFL

References

Mangan S, Alon U (2003) Structure and function of the feed-forward loop network motif. Proc Natl Acad Sci USA 100:11980–11985

Shen-Orr SS, Milo R, Mangan S, Alon U (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. Nat Genet 31:64–68

Feed Forward Loop Motif, FFL

► Feed Forward Loop

Feedback Loops

► MicroRNA Regulation, Feedback Loop

Feedback Regulation

Yufei Huang
Picower Institute for Learning and Memory, Massachusetts Institute of Technology, Cambridge, MA, USA
Greehey Children's Cancer Research Institute, University of Texas at Texas Health Science Center at San Antonio, San Antonio, TX, USA
Department of Epidemiology and Biostatistics, University of Texas at Texas Health Science Center at San Antonio, San Antonio, TX, USA

Definition

Feedback regulation describes the particular type of gene regulation, where the current status of gene

expression will influence the future status of the same set of genes. In a cell, feedback regulation is an essential mechanism to ensure the stability of the cellular system under changing environmental conditions.

Cross-References

► Gene Regulation

References

Campbell N, Mitchell L et al (2003) Biology: concepts and connections. Benjamin Cummings, New York
Reik W (2007) Stability and flexibility of epigenetic gene regulation in mammalian development. Nature 447(7143):425–432

Feed-Forward Loops

► MicroRNA Regulation, Feed-Forward Loops

FFL

► Canonical Network Motifs

Fibroblasts

Mary Helen Barcellos-Hoff
Department of Radiation Oncology and Cell Biology, New York University School of Medicine, New York, NY, USA

Synonyms

Fibrocytes; Interstitial cells; Mesenchymal cells

Definition

Fibroblasts are cells that reside in connective tissues and produce extracellular matrix (ECM). Quiescent fibroblasts, known as fibrocytes, are abundant in the interstitial tissues of all organs.

Characteristics

Fibroblasts reside in the stroma, are derived from the mesoderm, and are members of the cell type known as mesenchymal. Dormant fibroblasts become fibrocytes that can be easily identified along the margins of many types of connective tissue.

Fibroblasts are important to the structural integrity of tissues and organisms, in part because they generate type I collagen, the most abundant protein in animals. Collagen I is a triple-helical ECM protein that provides mechanical strength. Fibroblasts are also a major source of fibronectin, which serves as a scaffold for cell adhesion, and generate many minor components of the ► extracellular matrix. Fibronectin production by fibroblasts is highly responsive to stress and is alternatively spliced in response to specific signals resulting in a highly variable protein with a multiplicity of functions.

Fibroblasts are described by their common characteristics that include spindle cell morphology, vimentin cytoskeleton, and (relatively) proliferative quiescence; yet fibroblasts are highly heterogeneous. The distribution of fibroblasts varies widely between organs, from single cells tucked between epithelium and endothelium to densely packed. Another feature of fibroblasts is that they are readily cultured because they attach to tissue-culture plastic and proliferate in response to serum used in most culture media. Because of this common trait, many experimental studies have used fibroblasts as "garden-variety" cells. Yet genome-wide patterns of gene expression in cultured fetal and adult human fibroblasts derived from skin at different anatomical sites revealed that fibroblasts from each site displayed distinct and characteristic transcriptional patterns (Chang et al. 2002).

A broad classification discriminate fibroblasts derived from visceral organs and cutaneous origin. It was suggested that the site specificity of the molecular signals and extracellular proteins expressed by fibroblasts provide "home addresses" that are monitored by epithelial cells to restrict their migration, survival, and proliferation (Chang et al. 2002). Fibroblasts isolated from different anatomical sites and from the same anatomical site but of different diseases display topographic differentiation and positional memory. Although this suggests fibroblasts at different locations in the body could be considered distinct differentiated cell types, there has been little success if classification schemes

based on exclusive markers. Few genes are uniquely expressed in fibroblasts from any particular site; rather, combinatorial patterns of large groups of genes define fibroblasts from different sites.

Fibroblasts play important roles in tissue repair and wound healing, while aberrant fibroblasts contribute to disease processes and cancer (Kalluri and Zeisberg 2006). In response to wounding, dermal fibrocytes undergo a phenotypic change to myofibroblasts that serve to actively contract during wound closure. However, ▶ wound healing differs in adults and fetuses; the latter do not undergo myofibroblasts differentiation and also do not form scars. In the adult wound, fibroblast synthesis of collagen is delayed while fibroblasts proliferate, but fetal fibroblasts simultaneously proliferate and synthesize collagen (Buchanan et al. 2009). Fibroblasts also induce specific transition from acute inflammation to acquired immunity, while inappropriate production of chemokines and matrix components by fibroblasts leads to the establishment of chronic inflammation (Buckley 2011).

Fibroblasts rarely undergo spontaneous apoptosis or in response to DNA damage. A response of primary fibroblasts to stress and aging is a permanent arrest called replicative senescence. The senescence-associated secretory phenotype (SASP) is distinct from that of replication competent fibroblasts that is rich in cytokines and factors that may mediate aging processes. Fibroblasts are also culprits in tissue-compromising fibrosis that can starve, restrict, and ultimately replace functional parenchyma. In this case, exuberant collagen production is a major characteristic, often mediated by response to elevated TGFβ, a cytokine both produced by fibroblasts and to which fibroblasts are exquisitely sensitive. An novel source of fibroblasts in fibrotic conditions is the epithelium via epithelial to mesenchymal transition (EMT) (Kalluri and Neilson 2003). The parenchymal response to injury, as occurs in ▶ wound healing, causes EMT, but in the disease state, cells fail to revert to type due to an imbalance in cytokine signaling, leading to accumulation in the mesenchymal compartment (the significance and occurrence of EMT is the subject of a controversy [Tarin 2011]). Similarly, cancer cells can recruit and induce a highly activated fibroblast phenotype called cancer-associated fibroblasts (CAF). Together, these examples underscore the remarkably plasticity and diversity of fibroblasts, whose nature should be explicitly evaluated when considering tissue as systems.

Cross-References

▶ Bone Marrow-derived Cells
▶ Extracellular Matrix
▶ Wound Healing

References

Buchanan EP, Longaker MT, Lorenz HP (2009) Chapter 6 Fetal skin wound healing. In: Gregory SM (ed) Advances in clinical chemistry, vol 48. Elsevier, Amsterdam, pp 137–161

Buckley CD (2011) Why does chronic inflammation persist: an unexpected role for fibroblasts. Immunol Lett 138(1):12–14

Chang HY, Chi J-T, Dudoit S, Bondre C, van de Rijn M, Botstein D, Brown PO (2002) Diversity, topographic differentiation, and positional memory in human fibroblasts. Proc Natl Acad Sci 99(20):12877–12882. doi:10.1073/pnas.162488599

Kalluri R, Neilson EG (2003) Epithelial-mesenchymal transition and its implications for fibrosis. J Clin Invest 112(12):1776–1784

Kalluri R, Zeisberg M (2006) Fibroblasts in cancer. Nat Rev Cancer 6(5):392–401

Tarin D (2011) Cell and tissue interactions in carcinogenesis and metastasis and their clinical significance. Sem Cancer Biol 21:72–82

Fibrocytes

▶ Fibroblasts

Fick's Second Law

▶ Parabolic Differential Equations, Diffusion Equation

Figurate Networks

▶ Stem Cell Networks

File Standard

▶ Proteomics Data Formats

Final Cause

▶ Teleology

Fisher's Linear Discriminant

▶ Linear Discriminant Analysis

Fisher's Test

Kejia Xu
Institute of Systems Biology, Shanghai University, Shanghai, China

Definition

Fisher's exact test is a statistical test used to examine the significance of the association between two categorical variables. The test is useful for categorical data that result from classifying objects in two different ways, and it is used to examine the significance of the association between the two kinds of classification.

For an example application of the 2 × 2 test, let X be a journal, either a mathematics magazine or a science magazine, and let Y be the number of articles on the topics of mathematics and biology.

	Mathematics magazine	Science magazine	Total
Mathematics	a	b	a + b
Biology	c	d	c + d
Total	a + c	b + d	a + b + c + d

Fisher calculates the conditional probability of getting the actual matrix given the particular row and column sums:

$$P = \binom{a+b}{a}\binom{c+d}{c}\Big/\binom{a+b+c+d}{a+c} = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!(a+b+c+d)!}, \quad (1)$$

Cross-References

▶ Exact Test For Independence

References

Fisher RA. On the interpretation of χ2 from contingency tables, and the calculation of P. J Roy Stat Soc. 1922;85(1):87–94.

Fitness

Philippe Huneman
Institut d'Histoire et de Philosophie (IHPST), des Sciences et des Techniques Université Paris 1 Panthéon-Sorbonne, Paris, France

Definition

The word comes from the phrase "survival of the fittest," which Darwin borrowed from Spencer in the last edition of the *Origin of species* due to his own hand. Fitness has two components, survival and reproduction. If an organism is very well adjusted to its milieu, but does not reproduce, it has no evolutionary impact; hence, reproduction is often taken as crucial, and survival considered as a proxy for reproduction (the longer X survives, the higher the chances it has offspring). However, even if often correct, those approximations prove to be controversial, for example, when some organisms grow rather than reproduce. In some contexts, one has to model the two components of fitness separately, for instance, when the issue is to understand how individual resources are partitioned between reproduction and survival.

More formally, fitness can be defined as the probability distribution of the representation of a gene, a trait or an individual at the next generation; yet often equations can consider only the expectancy (fitness meaning *expected* offspring number of an individual). Because it concerns expected rather than actual offspring, fitness is often metaphysically considered as a propensity (sometimes called "expected fitness") rather than as a categorical property (then called "realized fitness").

Sometimes, several generations have to be taken into account to understand the evolutionary dynamics

(e.g., when explaining the constancy of sex ratio, which involves considering the effect on grandchildren) (Sober 2002). The case of altruism compels biologists to consider selection at the level of genes. According to ▶ Hamilton's rule, the ▶ relatedness between actor and beneficiary may account for the selection of the altruistic act because the degree of relatedness mitigates the cost: if the act increases the number of altruistic alleles at the next generation (as compared to the selfish alleles), be they directly alleles of the offspring of the actor, or alleles of the offspring of the beneficiaries of its altruistic acts, then altruism evolves. One can therefore reason by including within fitness all those alleles due to the altruist activities of the focal individual. *Inclusive fitness* is therefore the number of genes directly passed on to the next generation by a focal individual, plus the ones that are passed on by its kin. What is therefore increased by selection is rather inclusive than individual fitness, even though calculating inclusive fitness may be difficult in practice (Grafen 2009).

Even if mathematically speaking the construal of fitness is clear and how to construe it in a given problem is often straightforward, the issue of the bearer of fitness is difficult. Originally, only organisms had fitness, which was often computed as the number of offspring; with Modern Synthesis and the formulation of evolution in terms of gene frequencies in the context of population genetics, fitness is also ascribed to genes and genotypes. These values are interdependent of each other because the fitness of a gene can be seen as the contribution it makes to the fitness of the organism, but only in the case of asexual organisms the number of offspring equals the number of copies of a given gene. Moreover, fitness is often seen as lifetime fitness, that is, computed along the whole life of the organism; yet in behavioral ecology, one mainly considers individually each act and ascribes fitness to it (e.g., costs and benefits of various strategies are measured in terms of fitness).

Often, absolute fitness cannot be measured but relative fitness can, and only the latter is evolutionarily important (individuals with identical fitnesses do not undergo natural selection). Sometimes, one measures a posteriori the fitness of types of organisms by counting the number of offspring. Otherwise, one can consider fitness as strictly correlated to the way individuals face environmental demands, and then it can be computed a priori by estimating the performances of various trait types (race speed, rate of metabolism, visual acuity, etc.), provided that one has an idea about the relative importance of all factors for survival and reproduction.

Cross-References

▶ Explanation, Evolutionary

References

Bouchard F, Rosenberg A (2008) Fitness. In: Stanford on-line encyclopedia of philosophy

Grafen A (2009) Formalizing Darwinism and inclusive fitness theory. Phil Trans R Soc B 364(1):3135–3141

Sober E (2002) The two faces of fitness. In: Krimbas R (ed) Thinking about evolution: historical, philosophical, and political perspectives. Cambridge University Press, Cambridge

Fitness Function

Ke-Qin Liu
Institute of Systems Biology, Shanghai University, Shanghai, China

Synonyms

Objective function

Definition

In evolutionary algorithm, fitness function is actually an objective function that is used to determine which solution within a population is better when solving a particular problem (Holland 1975; Nelson et al. 2009). In evolution algorithm, the solution space consists of a population of chromosomes where each chromosome is one solution that can be evaluated by the fitness function. Finally, the chromosomes can be ranked by calculating the fitness function value (Holland 1975). A proper fitness function is very important for the speed of the search and the solution of optimization.

References

Holland JH. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control and artificial intelligence. (2nd edition. 1992, MIT) Ann Arbor: University of Michigan Press; 1975.

Nelson AL, Barlow GJ, Doitsidis L. Fitness functions in evolutionary robotics: a survey and analysis. Robot Auton Syst. 2009;57(4):345–370.

Fitting of Continuous and Deterministic Models

▶ Grid Computing, Parameter Estimation for Ordinary Differential Equations

Flemming Body

▶ Midbody

Floquet Multiplier

Tianshou Zhou
School of Mathematics and Computational Sciences, Sun Yet-Sen University, Guangzhou, Guangdong, China

Definition

Consider a linear differential equation of the form:

$$\dot{x} = A(t)x$$

where $A(t)$ is a continuous periodic function with period T. If $\Phi(t)$ is a fundamental matrix solution to this system, then for all $t \in R$,

$$\Phi(t+T) = \Phi(t)\Phi^{-1}(0)\Phi(T)$$

In addition, for each matrix B (possibly complex) such that $e^{TB} = \Phi^{-1}(0)\Phi(T)$, there is a periodic (period T) matrix function $t \mapsto P(t)$ such that $\Phi(t) = P(t)e^{tB}$ for all $t \in R$. Also, there is a *real* matrix S and a *real* periodic (period-$2T$) matrix function $t \mapsto Q(t)$ such that $\Phi(t) = Q(t)e^{tS}$ for all $t \in R$.

This mapping $\Phi(t) = Q(t)e^{tS}$ gives rise to a time-dependent change of coordinates ($y = Q^{-1}(t)x$), under which the original system becomes a linear system with real constant coefficients $\frac{dy}{dt} = Sy$. Since $Q(t)$ is continuous and periodic it must be bounded. Thus, the stability of the zero solution for $y(t)$ and $x(t)$ is determined by the eigenvalues of S. The representation $\Phi(t) = P(t)e^{tB}$ is called a *Floquet normal form* for the fundamental matrix $\Phi(t)$.

The eigenvalues of the matrix e^{TB} are called the Floquet multipliers of the system (some called characteristic multipliers). They are also the eigenvalues of the (linear) Poincaré maps $x(t) \to x(t+T)$. A Floquet exponent (sometimes called a characteristic exponent) is a complex μ such that $e^{\mu T}$ is a Floquet multiplier of the system. Notice that Floquet exponents are not unique since $e^{(\mu+(2\pi ik/T))T} = e^{\mu T}$ (where k is an integer), but Floquet multipliers are unique. The real parts of the Floquet exponents are called Lyapunov exponents. The zero solution is asymptotically stable if all Lyapunov exponents are negative, Lyapunov stable if the Lyapunov exponents are nonpositive, and unstable otherwise.

Flow Cytometry

Xiaojun Liu
Internal Medicine, The Second Hospital of Hebei Medical University, Shijiazhuang, Hebei, China

Definition

Flow cytometry is a technology that simultaneously measures and then analyzes multiple physical characteristics of single particles, usually cells, as they flow in a fluid stream through a beam of light. The properties measured include a particle's relative size, relative granularity or internal complexity, and relative fluorescence intensity. These characteristics are determined using an optical-to-electronic coupling system that records how the cell or particle scatters incident laser light and emits fluorescence.

Cross-References

▶ Cell Sorting

Flow Cytometry, Flow Microfluorimetry

▶ Cell Cycle Analysis, Flow Cytometry

Fluorescence

Xiaojun Liu
Internal Medicine, The Second Hospital of Hebei Medical University, Shijiazhuang, Hebei, China

Definition

Fluorescence occurs when a fluorescent compound absorbs light energy over a range of wavelengths that is characteristic for that compound. It is a transition of energy produced by the electron when it decays from higher energy level raised by the absorption light to its ground state. In Fluorescence-activated cell sorting fluorescence is often used to show the particles' characters according to the specific antibody conjugated by the fluorescent compound.

Cross-References

▶ Cell Sorting

Fluorescence Microscope

Xiaohua Wu
Department of Pediatrics, Herman B Wells Center for Pediatric Research, Indiana University School of Medicine, Indianapolis, IN, USA

Definition

Fluorescence microscope is an optical microscope used to study properties of organic or inorganic substances using the phenomena of fluorescence and phosphorescence, instead of, or in addition to, reflection and absorption.

Cross-References

▶ Fluorescence Microscopy
▶ Spectroscopy and Spectromicroscopy

Fluorescence Microscopy

Yi Zeng
Department of Pediatrics, University of Arizona, Tucson, AZ, USA

Synonyms

Epifluorescence microscope; Fluorescence microscope

Definition

Fluorescence microscopy utilizes fluorescence as a means of detecting objects through a microscope.

Characteristics

Fluorescence microscopy (FP) has been extensively used in biomedical research. It allows users to observe the structure and dynamics of the cell or tissue with great granules. FP is designed to obtain temporal and spatial information of objects that are either autofluorescent, or have been labeled with extrinsic fluorescent molecules. The combination of fluorescent specificity and the sensitivity of the latest optical instruments has enabled the detection of small amount of materials with high resolution and precision. Furthermore, the application of the FP offers not just the qualitative description, but also the quantitative attributes of the specimens.

The entire procedure of fluorescence microscopy starts with specimen preparation. The specimen is then irradiated by excitation light of specific wavelength. The much weaker emitted ▶ fluorescence is separated from the brighter excitation light and reaches human eyes or other detectors usually a digital or conventional film camera. The fluorescent areas shine brightly against a dark background with sufficient contrast to permit detection.

Fluorescence microscopy has many advantages that are not readily available in other optical microscopy techniques:

- Specificity. Fluorescence excitation and emission spectra are usually inherent characteristics of a molecule. This property is the foundation for selective analysis of complex mixtures of molecular species.
- Sensitivity. By distinguishing autofluorescence from specific fluorescence, fluorescence microscopy can reveal the presence of fluorescent material with a scale of even single fluorescent molecule.
- Environmental sensitivity. The high sensitivity of fluorescence to physical and chemical environment enables the investigation of pH, viscosity, refractive index, ionic concentrations, membrane potential, and solvent polarity in living cells and tissues.
- High temporal resolution. Fluorescence measurements can be used to track fast chemical and molecular changes in specimens.
- High spatial resolution. Fluorescence can be measured from single molecules if the molecules contain a sufficient number of ▶ fluororphores. The interactions of cellular components whose dimensions are below the diffraction-limited resolution of the microscope can be visualized using fluorescence resonance energy transfer techniques.
- Quantitation. Direct correlation between emitted ▶ fluorescence and the fluorescence quantum yield (the ratio of photon absorption to emission) allows quantitative measurements by fluorescent microscope. Quantification can be achieved at relatively low concentrations due to the greater sensitivity of emission when compared to absorption processes.

Fluorescence microscope has five basic components: excitation light sources, wavelength selection devices, objectives, detectors, and stages and specimen chambers. Powerful compact light sources are needed to generate sufficient excitation light intensity to produce detectable emission. Tungsten or halogen lamps are used in transmitted or incident illumination. The most common lamps for epifluorescence microscopes are mercury, xenon, or metal halide arc lamps. In recent years, there has been increasing use of lasers as light sources. Lasers are most widely used for confocal microscopy and total internal reflection fluorescence microscopy. Commonly used objectives can be classified into transmitted-light and reflected-light categories. Transmitted-light objectives are designed to be used with coverslips. Reflected-light objectives feature specially collated glass surfaces to avoid reflection in the optics. Detectors allow visualization of low levels of emitted ▶ fluorescence without photobleaching or photodamage to the specimen. They also allow real time recording of living cell and tissue physiology. A variety of specimen chambers are available to allow analysis of live or fixed cells and tissues.

One of the most important applications of fluorescence microscopy is in the field of ▶ immunofluorescence, which combines the sensitivity of fluorescence microscopy and the high degree of specificity exhibited by antigen-antibody binding. Two commonly used ▶ immunofluorescence techniques are direct immunofluorescence and indirect immunofluorescence. Direct immunofluorescence uses a single antibody labeled with a ▶ fluorochrome to bind the target ▶ antigen in the specimen. The reaction of the chemically attached fluorescent conjugate and antigen is demonstrated when the fluorochrome is excited. The subsequent emission intensity at various wavelengths can then be observed visually or captured by a detector system. In indirect immunofluorescence, an unlabeled primary antibody first incubates with the target antigen. A secondary fluorochrome-labeled antibody then incubates with the primary antibody because it recognizes the primary as an antigen. Subsequently, the labeled complex of antigen and antibodies is excited at the peak wavelength intensity of the fluorochrome, and any resulting emission is observed. The fact that each antigen is able to bind to multiple antibodies allows indirect immunofluorescence to produce greater fluorescence intensity compared to direct immunofluorescence. In addition, this approach reduces the necessity of keeping in stock large numbers of fluorochrome-labeled antibodies.

Other popular fluorescence microscopy applications include: fluorescence in situ hybridization, fluorescence differential interference contrast, automated fluorescence image cytometry, fluorescence recovery after photobleaching, total internal reflectance fluorescence microscopy, fluorescence resonance energy transfer microscopy, digitized fluorescence polarization microscopy, fluorescence lifetime imaging microscopy, Fourier spectroscopy/spectral dispersion microscopy, delayed luminescence microscopy.

Cross-References

- ▶ Antigen
- ▶ Fluorescence
- ▶ Fluorochrome
- ▶ Fluorophore
- ▶ Immunofluorescence

References

Herman B (1998) Fluorescence microscopy. Microscopy handbooks. Bios Scientific, Oxford. ISBN 9780387915517

Huang K, Murphy RF (2004) From quantitative microscopy to automated image understanding. J Biomed Opt 9(5):893–912, ISSN 1083-3668

MedlinePlus. Medical encyclopedia, August 2011. URL http://www.nlm.nih.gov/medlineplus/ency/encyclopedia_A.htm

Molecular expressions. Fluorescence microscopy, August 2011. URL http://micro.magnet.fsu.edu/primer/techniques/fluorescence/fluorhome.ht%ml

Rost FWD (1991) Quantitative fluorescence microscopy. Cambridge University Press, Cambridge/New York. ISBN 9780521394222

Spring KR (2003) Fluorescence microscopy. In: Driggers RG (ed) Encyclopedia of optical engineering, Dekker encyclopedias series. Marcel Dekker, New York

Wolf DE (2007) Fundamentals of fluorescence and fluorescence microscopy. In Greenfield Sluder, DE Wolf (eds) Digital microscopy. Methods in cell biology, 3rd edn, vol 81. Academic, pp 63–91

Wouters FS (2006) The physics and biology of fluorescence microscopy in the life sciences. Contemp Phys 47(5): 239–255, ISSN 0010-7514

Fluorescence Spectroscopy

Xiaohua Wu
Department of Pediatrics, Herman B Wells Center for Pediatric Research, Indiana University School of Medicine, Indianapolis, IN, USA

Definition

Fluorescence spectroscopy uses higher-energy photons to excite a sample, which will then emit lower-energy photons. This technique has become popular for its biochemical and medical applications, and can be used for confocal microscopy, fluorescence resonance energy transfer, and fluorescence lifetime imaging.

Cross-References

- ▶ Spectroscopy and Spectromicroscopy

Fluorescence-activated Cell Sorting

Steven D. Rhodes
School of Medicine, Indiana University, Indianapolis, IN, USA

Definition

Fluorescence-activated cell sorting, or FACS for short, is a specialized type of cell sorting that utilizes flow cytometry. Fluorescently tagged cells are isolated into charged droplets which are separated based on their deflection between two electrodes.

Cross-References

- ▶ Single Cell Assay, Mesenchymal Stem Cells

Fluorescent Dye

- ▶ Fluorochrome

Fluorescent Marker

- ▶ Fluorescent Markers

Fluorescent Markers

Yongzheng He and Yan Li
Department of Pediatrics, Indiana University School of Medicine, Indianapolis, IN, USA

Synonyms

Cell marker; Cluster of differentiation (CD); Fluorescent marker; Fluorophore; Multicolor flow cytometry

Definition

A fluorophore is a functional component of a molecule which can cause a molecule to be fluorescent by the means of absorbing energy of a specific wavelength and emitting energy at a different wavelength. Fluorescein isothiocyanate (FITC), rhodamine, and other derivatives are common fluorophores used for a variety of applications.

Fluorescent markers are specific molecules, like protein, which are covalently bound fluorophores that selectively bind to a functional group of the target for detection. The most commonly used fluorescent molecules are antibodies.

Cell markers are specified protein on the surface of every cell, called receptors, which can selectively bind or adhere to other "signaling" molecules. The biological uniqueness of the receptors and chemical properties of certain compounds are used to mark cells.

Cluster of differentiation (CD) molecules are markers on the cell surface, which can be recognized by specific sets of antibodies. Cluster of differentiation systems can be used to identify the cell type, stage of differentiation, and activity of a cell.

Multicolor flow cytometry is a technique in which cells or cell components (such as DNA) are stained with multiple fluorescent dyes to detect the fluorescence by laser beam illumination using up to 18 independently measurable colors for the identification and sorting of cells. Multicolor flow cytometry offers a platform to acquire detailed information on specific cells within a mixed population, which is critical to maximize the data that can be obtained from a small or limited sample.

Characteristics

Multicolor Flow Cytometry

A combination of fluorescence conjugated antibodies in flow cytometry is utilized for the determination and sorting of certain cell population.

Hematopoietic Stem Cell Markers

Hematopoietic stem cells (HSCs), which constitute 1:10,000 of cells in myeloid tissue, are multipotent stem cells that give rise to blood cell types of the myeloid and lymphoid lineages, including myeloid (monocytes and macrophages, neutrophils, basophils, eosinophils, erythrocytes, megakaryocytes/platelets, dendritic cells), and lymphoid lineages (T-cells, B-cells, NK-cells). In the embryo, hematopoietic stem cells are formed from the mesoderm during embryogenesis and are deposited in specific hematopoietic sites. In adults, HSCs are found in the bone marrow and peripheral blood following pretreatment with cytokines, such as G-CSF (granulocyte colony-stimulating factors), that induce cells to be released from the bone marrow compartment. Other sources of HSCs include the placenta and umbilical cord blood.

There is no single specific marker for stem cell. Generally, HSCs are characterized by the absence of lineage-specific marker expression and expression of a combination of cell markers.

Human HSCs are determined as $CD34^+$, $CD59^+$, $CD90/Thy1^+$, $CD38^{low/-}$, $c\text{-}Kit^{-/low}$, and Lin^-. Murine HSCs are determined as $CD34^{low/-}$, $Sca\text{-}1^+$, $CD90/Thy1^{+/low}$, $CD38^+$, $c\text{-}Kit^+$, and Lin^-(Geraerts and Verfaillie 2009). Alternative methods that allow for more efficient identification of stem cells, such as SLAM (Laje et al. 2010) family of cell surface molecules, are now emerging.

Mesenchymal Stem Cell Markers

Mesenchymal stem cells (MSCs) are multipotent stem cells which are capable of differentiating into a variety of cell types, such as osteoblasts, chondrocytes, and adipocytes. Traditionally, MSCs are found in the bone marrow. Alternatively, other tissues can be a more accessible source for mesenchymal stem cell isolation, including adipose tissue, cord blood, and peripheral blood.

Due to the lack of unique single marker for both human and murine MSCs so far, a combination of cell markers is commonly used to identify MSCs. Human MSCs are determined as $CD11c^-$, $CD14^-$, $CD31^-$, $CD34^-$, $CD45^-$, $CD117^-$, $CD29^+$, $CD44^+$, $CD90^+$, $CD73^+$, $CD105^+$, $CD106^+$, and $CD166^+$ (Delorme et al. 2006; Parekkadan and Milwid 2010; Aicher et al. 2011); Murine MSCs are determined as $CD34^-$, $CD45^-$, $CD44^+$, $CD90^+$, $CD105^+$, $CD117^-$ $SCA1^+$ (Schrepfer et al. 2007; Soleimani and Nadri 2009; Zhu et al. 2010).

Cell Cycle

Flow cytometry can be utilized in cell cycle analysis to distinguish different phases of the cell cycle.

After being permeablized by detergent like Triton X-100 or NP-40, or fixed with ethanol, cells are incubated with fluorescent DNA dyes for DNA quantification in single cell. Because fluorescent DNA dye also stains RNA, cells are usually treated with RNase A to remove RNAs. The fluorescence intensity of each stained cell at a certain wavelength correlates linearly with amount of DNA content in it. In this way, the G0/G1 phase, S phase, G2/M phase (DNA duplicates) can be distinguished by the intensity of fluorescent DNA dye according to the amount of DNA the cell contains.

Besides Propidium iodide, 7-Aminoactinomycin D (7AAD), DAPI, and Hoechst are frequently used as fluorescent DNA dye.

Cell Apoptosis

Annexin V affinity assay is a method in molecular biology that employs flow cytometry to quantify the number of cells undergoing apoptosis, which uses the protein Annexin V to tag apoptotic and dead cells.

The fluorescence conjugated Annexin V protein binds to negative charged phosphatidylserine (PS) on the membrane of cells undergoing apoptosis. By conjugating FITC to Annexin V it is possible to identify and quantitate apoptotic cells on a single-cell basis by flow cytometry.

When cells were stained with Annexin V (AV) and propidium iodide (PI) simultaneously, it is easy to distinguish intact cells (AV−PI−), early apoptotic (AV+PI−), and late apoptotic or necrotic cells (AV+PI+) (Muppidi et al. 2004).

Proliferation Assay

Cell proliferation may be assessed by flow cytometry by labeling cells with the dye CFSE, which readily crosses intact cellular membranes and irreversibly couples to both intracellular and cell surface proteins. When cells divide, the CFSE labeling is then distributed into the two daughter cells equally (Parish et al. 2009).

Cross-References

- ▶ Cell Cycle Analysis, Flow Cytometry
- ▶ Cell Sorting
- ▶ Flow Cytometry

References

Aicher WK, Buhring HJ et al (2011) Regeneration of cartilage and bone by defined subsets of mesenchymal stromal cells – potential and pitfalls. Adv Drug Deliv Rev 63(4–5):342–351

Darzynkiewicz Z, Juan G et al (2001) Determining cell cycle stages by flow cytometry. Curr Protoc Cell Biol Chapter 8: Unit 8 4

Delorme B, Chateauvieux S et al (2006) The concept of mesenchymal stem cells. Regen Med 1(4):497–509

Geraerts M, Verfaillie CM (2009) Adult stem and progenitor cells. Adv Biochem Eng Biotechnol 114:1–21

Laje P, Zoltick PW et al (2010) SLAM-enriched hematopoietic stem cells maintain long-term repopulating capacity after lentiviral transduction using an abbreviated protocol. Gene Ther 17(3):412–418

Muppidi J, Porter M et al (2004) Measurement of apoptosis and other forms of cell death. Curr Protoc Immunol Chapter 3: Unit 3 17

Parekkadan B, Milwid JM (2010) Mesenchymal stem cells as therapeutics. Annu Rev Biomed Eng 12:87–117

Parish CR, Glidden MH et al (2009) Use of the intracellular fluorescent dye CFSE to monitor lymphocyte migration and proliferation. Curr Protoc Immunol Chapter 4:Unit 4 9

Schrepfer S, Deuse T et al (2007) Simplified protocol to isolate, purify, and culture expand mesenchymal stem cells. Stem Cells Dev 16(1):105–107

Soleimani M, Nadri S (2009) A protocol for isolation and culture of mesenchymal stem cells from mouse bone marrow. Nat Protoc 4(1):102–106

Zhu H, Guo ZK et al (2010) A protocol for isolation and culture of mesenchymal stem cells from mouse compact bone. Nat Protoc 5(3):550–560

Fluorescent Probe

- ▶ Fluorochrome

Fluorochrome

Yi Zeng
Department of Pediatrics, University of Arizona, Tucson, AZ, USA

Synonyms

Fluorescent dye; Fluorescent probe

Definition

Fluorochromes are molecules capable of exhibiting fluorescence.

Cross-References

► Fluorescence Microscopy

References

Herman B (1998) Fluorescence microscopy, Microscopy handbooks. Bios Scientific, Oxford. ISBN 9780387915517
Huang K, Murphy RF (2004) From quantitative microscopy to automated image understanding. J Biomed Opt 9(5): 893–912, ISSN 1083–3668
MedlinePlus (2011) Medical encyclopedia. http://www.nlm.nih.gov/medlineplus/ency/encyclopedia_A.htm
Molecular Expressions (2011) Fluorescence microscopy. http://micro.magnet.fsu.edu/primer/techniques/fluorescence/fluorhome.html
Rost FWD (1991) Quantitative fluorescence microscopy. Cambridge University Press, Cambridge. ISBN 9780521394222
Spring KR (2003) Fluorescence microscopy. In: Driggers RG (ed) Encyclopedia of optical engineering, Dekker encyclopedias series. Marcel Dekker, New York
Wolf DE (2007) Fundamentals of fluorescence and fluorescence microscopy. In: Sluder G, Wolf DE (eds) Digital microscopy, vol 81, 3rd edn, Methods in cell biology. Academic, Amsterdam, pp 63–91
Wouters FS (2006) The physics and biology of fluorescence microscopy in the life sciences. Contem Phys 47(5): 239–255, ISSN 0010-7514

Fluorophore

Yi Zeng
Department of Pediatrics, University of Arizona, Tucson, AZ, USA

Definition

Fluorophore is the structural domain or specific region of a molecule that is capable of exhibiting fluorescence. Fluorochromes that are conjugated to a larger macromolecule through absorption or covalent bonds are termed "fluorophores."

Cross-References

► Fluorescence Microscopy
► Fluorescent Markers

References

Herman B (1998) Fluorescence microscopy, Microscopy handbooks. Bios Scientific, Oxford. ISBN 9780387915517
Huang K, Murphy RF (2004) From quantitative microscopy to automated image understanding. J Biomed Opt 9(5): 893–912, ISSN 1083–3668
MedlinePlus (2011) Medical encyclopedia. http://www.nlm.nih.gov/medlineplus/ency/encyclopedia_A.htm
Molecular Expressions (2011) Fluorescence microscopy. http://micro.magnet.fsu.edu/primer/techniques/fluorescence/fluorhome.html
Rost FWD (1991) Quantitative fluorescence microscopy. Cambridge University Press, Cambridge. ISBN 9780521394222
Spring KR (2003) Fluorescence microscopy. In: Driggers RG (ed) Encyclopedia of optical engineering, Dekker encyclopedias series. Marcel Dekker, New York
Wolf DE (2007) Fundamentals of fluorescence and fluorescence microscopy. In: Sluder G, Wolf DE (eds) Digital microscopy, vol 81, 3rd edn, Methods in cell biology. Academic, Amsterdam, pp 63–91
Wouters FS (2006) The physics and biology of fluorescence microscopy in the life sciences. Contem Phys 47(5): 239–255, ISSN 0010-7514

Flux Balance Analysis

Meghna Rajvanshi and Kareenhalli V. Venkatesh
Department of Chemical Engineering, Indian Institute of Technology Bombay, Powai, Mumbai, Maharashtra, India

Synonyms

Metabolic flux balancing

Definition

Flux Balance Analysis (FBA) is a method in ► metabolic pathway modeling to quantify a metabolic state of a cell. It is a constraint-based approach, which uses linear programming, subject to constrains imposed by the stoichiometry of the metabolic network, thermodynamics, and the measured rates (r_m). The fluxes of

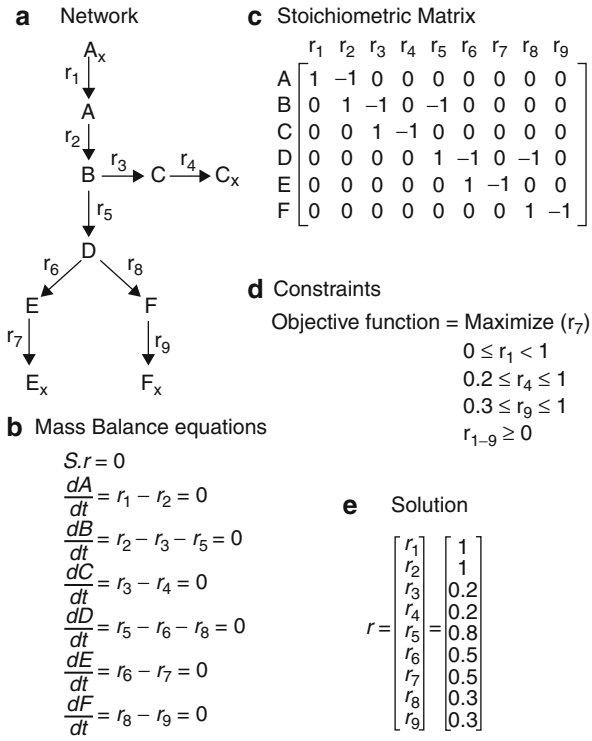
the metabolic network are computed under these constraints by optimizing an objective function under ▶ steady state operation of the metabolic network (Edwards et al. 2002). FBA is used when limited information about the accumulation rates, typically of an extracellular metabolite (i.e., measured rates, r_m), is known and the stoichiometric matrix with coefficients of all the unmeasured reactions (denoted by S_u) is noninvertible to provide a unique solution.

Characteristics

The first step in FBA is reconstruction of the metabolic network. Once all the internal and transport reactions are identified, dynamic mass balance equations for all the metabolites are formulated. These mass balances can also be represented in a matrix form representing the stoichiometric matrix (S), and the flux vector (r). FBA analyzes the metabolic network assuming a steady-state condition, therefore the dynamic mass balance equations are set to zero, as given below:

$$\frac{dc}{dt} = S.r = 0 \tag{1}$$

The above equation $(S.r = 0)$ forms the constraint for the linear optimization problem. Since, in most biological systems, the number of reactions (n) is usually more than the number of metabolites (m), that is, $(n > m)$, the system is underdetermined and has $(n - m)$ degrees of freedom and therefore, cannot be solved algebraically. In order to solve the system, additional constraints are specified, which constrict the solution space. These constraints are, as mentioned previously, thermodynamic (defining reversibility or irreversibility of the reactions) and capacities of enzymes and transporters (defines maximum uptake or reaction rates) (Edwards et al. 2002). Additionally, certain flux values are also specified, which are experimentally measured. These constraints define a range of feasible values in the solution space. However, to obtain a unique solution, the space is further constricted by an objective function. The solution thus obtained is specific to a metabolic state that optimizes for a specific objective. Therefore, solving an underdetermined system translates to an optimization problem for a defined objective function (Z), wherein linear programming methodologies may be applicable (Edwards et al. 2002).



Flux Balance Analysis, Fig. 1 Flux Balance Analysis (FBA) of a simple metabolic network. The solution is obtained for maximizing the accumulation rate of "E_x," that is, maximizing the rate, r_7

Mathematically, the maximization problem can be stated as:

$$\max Z = c.r \tag{2}$$

Subject to the constraints: $S.r = 0, r \geq 0,\ r \leq r_{max}$, $r_{m.min} \leq r_m \leq r_{m.max}$

Here, Z denotes the linear objective function and c is a row vector of weights (coefficients) on the fluxes "r" used to define an objective function. Weights indicate the contribution of reaction "r" toward an objective function. r_{max} is the maximum flux derived from an enzyme or it corresponds to the maximum transport capability of an enzyme. $r_{m.min}$ and $r_{m.max}$ are the minimum and maximum rates achievable for a particular reaction, respectively.

The methodology of FBA can be elucidated by solving an example by using a simple hypothetical metabolic network. Figure 1a illustrates an example of a hypothetical metabolic network. This system has nine reactions and ten metabolites. Four of these

metabolites (A_x, C_x, E_x and F_x) are external and the rest are internal metabolites. All reactions in this system are irreversible, out of which r_1, r_4, r_7, and r_9 are exchange reactions and r_2, r_3, r_5, r_6, and r_8 are internal reactions. For each of the internal metabolites, mass balance equations can be written as shown in Fig. 1b. These equations can be represented in the form of matrix (S), where rows of the matrix correspond to internal metabolites and column corresponds to reactions (Fig. 1c). A unique flux distribution for the network is obtained for an assumption that the system is optimized for the accumulation rate of metabolite "E_x." Further the solution is also dependent on additional constraints of the system under study, which are imposed by the stoichiometric coefficients of the metabolites, flux carrying capabilities of various reactions, irreversibility, and measured accumulations rates (Fig. 1d, e).

Since the solution to characterize metabolism is dependent on the optimization problem, defining a precise objective function is critical. The objective function should closely represent the cellular metabolism for a given condition. Typically, the basis of defining an objective function is the assumption that, due to evolutionary pressure, cells have evolved to an optimal behavior and therefore, the phenotypic state yields an optimal flux distribution. The most commonly used objective function is the maximization of growth or formation of biomass. Prediction of fluxes with this objective function has yielded results consistent with several experimental findings, such as 86% cases for *Escherichia coli*, 85% for *Pseudomonas aeruginosa*, 70% for *Leishmania major*, and 60% for *Helicobacter pylori* (Gianchandani et al. 2010). Other objective functions employed include optimizing ATP production (to determine optimal energy efficient condition), production of a metabolite, or rate of nutrient uptake (Raman and Chandra 2009).

Although FBA has been useful in quantifying the metabolic state consistent with experimental measurements, the formulation of the objective function is the limitation of this approach. Selection of an inappropriate objective function will lead to a solution which will not be an accurate representation of the cellular metabolism (Edwards et al. 2002). Apart from identification of a precise objective function, it is also essential to represent the objective function in a precise mathematical form. Further, FBA yields only one optimal solution though there might be other optimal or suboptimal solutions for a given set of constraints. However, prediction of the metabolic fluxes can be improved by introducing more number of measured fluxes in the analysis by which the search space is further constrained.

Applications of FBA

FBA is widely used for characterizing cellular metabolism and has given solutions which are consistent with experimental findings. In this regard, some of the most studied organisms are *Escherichia coli, Helicobacter pylori, Haemophilus influenzae, Saccharomyces cerevisae*, and *Methanosarcina barkeri*. FBA is extensively used in determining metabolic network properties, identifying redundancies (existence of alternate optimal solutions) (Papin et al. 2002) and robustness in the metabolic network. Robustness can be examined by introducing environmental perturbations and analyzing its effect with respect to optimal growth rate. Analysis of perturbations in the network by deleting or inserting genes are important as they yield information regarding the essentiality of the gene for survival of an organism and thereby allowing identification of potential drug targets (Raman et al. 2005). Metabolic engineering is another area where FBA is extensively used. It helps in identifying target genes, whose manipulation leads to improved production of metabolite of interest (Wang et al. 2006). Fongs et al. (2005) observed that strains engineered using findings of FBA behave suboptimally but after undergoing adaptive evolution, the cell can reach a state of metabolic optimality (Ibarra et al. 2002; Fong et al. 2005). FBA has also been used in analyzing the growth of an organism on different carbon sources (Covert and Palsson 2002). Recently, FBA has been applied in the area of medicine to investigate the metabolic network of human mitochondria and the effect of treatment on the metabolism. FBA has been applied to optimize the metabolism of cultured hepatocytes to be used in bio-artificial liver devices (Thiele et al. 2005). FBA has also been used to study plant metabolism, but the implementation of FBA is difficult due to the structural complexities in plants.

Thus, FBA is a powerful modeling approach for quantitative simulation of microbial metabolism. It assumes optimal cell behavior with respect to a known objective function. Its applicability to genome scale metabolic networks makes FBA popular in analyzing cellular physiology in such systems.

Cross-References

- ▶ FBA Analysis, Plant-Pathogen Interactions
- ▶ Network Modeling of Biochemical Transport Phenomena
- ▶ Pathway Modeling, Metabolic
- ▶ Optimization Algorithms for Metabolites Production
- ▶ Steady State

References

Covert MW, Palsson BO (2002) Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. J Biol Chem 277(31):28058–28064. doi:10.1074/jbc.M201691200

Edwards JS, Covert M, Palsson B (2002) Metabolic modelling of microbes: the flux-balance approach. Environ Microbiol 4(3):133–140

Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, Palsson BO (2005) In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid. Biotechnol Bioeng 91(5):643–648

Gianchandani EP, Chavali AK, Papin JA (2010) The application of flux balance analysis in systems biology. Wiley Interdiscip Rev Syst Biol Med 2(3):372–382

Ibarra RU, Edwards JS, Palsson BO (2002) *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature 420(6912):186–189

Papin JA, Price ND, Edwards JS, Palsson BO (2002) The genome-scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy. J Theor Biol 215(1):67–82

Raman K, Chandra N (2009) Flux balance analysis of biological systems: applications and challenges. Brief Bioinform 10(4):435–449. doi:10.1093/bib/bbp011

Raman K, Rajagopalan P, Chandra N (2005) Flux balance analysis of mycolic acid pathway: targets for anti-tubercular drugs. PLoS Comput Biol 1(5):e46

Thiele I, Price ND, Vo TD, Palsson BO (2005) Candidate metabolic network states in human mitochondria impact of diabetes, ischemia, and diet. J Biol Chem 280(12):11683–11695

Wang Q, Chen X, Yang Y, Zhao X (2006) Genome-scale in silico aided metabolic analysis and flux comparisons of *Escherichia coli* to improve succinate production. Appl Microbiol Biotechnol 73(4):887–894

Flux Control Analysis

- ▶ Metabolic Control Analysis

Flux Control Coefficient

Emma Saavedra and Rafael Moreno-Sánchez
Department of Biochemistry, National Institute for Cardiology "Ignacio Chávez", Mexico City, Mexico

Definition

It is the degree of control that each enzyme exerts on a metabolic pathway flux. A practical definition is the percentage of change in the pathway flux when a 1% change in the activity of a pathway enzyme is achieved. The corresponding written description is C^J_{ai}, where J is flux (or cellular function) and a is the activity of an enzyme i (or protein, transporter, or cellular process). The C^J_{ai} as well as the ▶ concentration control coefficients (see next description) are systemic properties of the enzymes when they are working together in the pathway or cellular network.

Cross-References

- ▶ Metabolic Control Theory

Fokker–Planck Equation

Ruiqi Wang
Institute of Systems Biology, Shanghai University, Shanghai, China

Definition

Fokker–Planck equations are a special type of master equation and are often used as a continuous approximation to master equations. By the Taylor expansion of master equation to order two, we obtain the Fokker–Plank equation.

The Fokker–Planck equation is beneficial in the sense that some theoretical analysis can be conducted. For example, for some simple cases, the equilibrium probability distribution can be obtained.

References

van Kampen NG (1981) Stochastic processes in physics and chemistry. Elsevier, Amsterdam
Risken H (1989) The Fokker-Planck equation. Springer, Berlin

Foreign Substance

► Xenobiotics

Formation of Phosphodiester Bond

► Phosphodiester Bond Formation

Formerly Known as the Physiome CellML Environment

► OpenCell

Forward and Inverse Parameter Estimation for Metabolic Models

I-Chun Chou, Zhen Qi, Melissa L. Kemp and Eberhard O. Voit
The Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA, USA

Definition

One of the most challenging bottlenecks of mathematical modeling is the identification of optimal parameter values with which the model matches experimental data well. This essay discusses distinct strategies for approaching this identification step.

Characteristics

Modeling Process

The generic process of modeling metabolic systems consists of six phases (Chou and Voit 2009):

1. Identification of network structure and regulation
2. Selection of a mathematical modeling framework
3. Estimation of parameter values
4. Model diagnostics
5. Model validation
6. Applications and uses of the model

The first phase is dedicated to identifying the network structure and regulation of the system. This phase relies on a combination of available information and assumptions or hypotheses, which are derived from the literature or from de novo experiments. Based on the available body of knowledge, the modeler needs to decide which components and interactions to include in the model and which to omit in order to keep the model manageable, while retaining the integrity of the system. Additional assumptions and simplifications are usually needed in order to fill gaps in information. The result of this model design phase is often visualized as a diagram with nodes for the components (metabolites) and arrows for interactions between them (fluxes). The second phase is dedicated to the choice of the best-suited mathematical framework and the corresponding formulation of a ► symbolic model, that is, a model in which no parameter values are specified yet. This process usually starts with converting the diagram of the system topology and regulation into mathematical equations. These may be linear or nonlinear, stochastic or deterministic, static or dynamic, and consist of explicit functions, differential equations, or difference equations (Voit et al. 2008). Metabolic systems are usually described with a set of differential equations that represent the dynamics of the system variables and are composed of sums and differences of the metabolic fluxes driving the system. Once the symbolic model is formulated, the next phase of parameter estimation consists of numerically configuring the symbolic model, which requires the determination of parameter values that render the model consistent with experimental observations. It is seldom expected that the model will exactly match all available data points, and the desired or required degree of consistency is therefore a matter of judgment. In addition to fitting the data, the numerical model needs to satisfy other characteristics inherent in biological systems, such as stability, robustness, and possibly different transient behaviors. The diagnosis of the model with respect to such features is followed by the validation of the model through testing against experimental data and biological, clinical,

pharmacological, or other information that had not been used for the design of the model. If the model has successfully gone through these design and testing phases, one may cautiously deem it appropriate for the purposes that initially led to the model design. In particular, one may use the model for explanations, predictions, and applications, the generation of new hypotheses, assistance in the design of new biological experiments, possibly the development of treatment strategies for diseases, and the manipulation and optimization of the model toward specific goals, such as the production of organics in metabolic engineering. The phasing of the modeling process may give the impression that modeling is straightforward. However, in most cases it is an iterative process requiring the repeated return to earlier phases (Voit et al. 2008).

The Challenge of Parameter Estimation

The most challenging among the phases of model development is usually the estimation of parameter values, especially when the investigated system is moderately large. Addressing this challenge successfully is crucial, because it is clear that the parameter values qualitatively and quantitatively characterize the metabolic model and distinguish it from others. Furthermore, most model analyses can only be executed when all parameter values are specified, and most predictions and many explanations made with the model are of a numerical nature that requires parameterization.

The development of parameter estimation methods is driven by the availability and characteristics of experimental data. Correspondingly, estimation methods can be very distinct, thereby reflecting the variety of experimental data types (Voit 2004; Goel et al. 2006). The currently available methods can be classified as:

1. Forward (bottom-up) approaches and
2. Inverse (top-down) methods using steady-state or time series data

Forward Estimation

Before molecular high-throughput data were available, essentially all metabolic models were developed according to the first strategy, and most models are still generated in this fashion today. Specifically, a model is designed based on *local* kinetic information that is used to formulate functions for individual biochemical processes, such as enzyme-catalyzed reactions and transport steps. The default choice for these functions or *rate laws* is typically a Michaelis–Menten (MM) or Hill function, or one of their generalizations; however, other options include power-law functions, lin-log representations, or more complex mathematical descriptions (Voit and Chou 2009). The computational aspects of estimating parameters for kinetic rate functions are rather simple, as these functions are explicit and usually contain only a few parameters. For instance, in the case of MM rate laws, the parameters can be estimated with methods of linear regression.

The data characterizing the kinetic properties of an enzyme or transporter are usually measured on isolated enzymes in vitro and account for optimal temperature and pH ranges, cofactors, modulators, and secondary substrates. Once sufficient information on individual rate processes has been assembled, the modeler attempts to merge this information into an integrative mathematical model of the entire system, which typically consists of ordinary differential equations.

Though this forward approach is theoretically straightforward, it has several disadvantages. First, it requires that a considerable amount of local kinetic information is at hand. Second, even if this information is available in the literature, it had often been obtained from different organisms or with experiments under various conditions. As a consequence, the ultimately emerging integrated model is often not internally consistent or does not match biological observations, thus mandating numerous rounds of refinements, restructuring, and reparameterization. These iterations are very labor intensive and time consuming and greatly benefit from a combination of biological and computational expertise which, however, is still rare.

Inverse Estimation

Instead of first representing one process at a time and subsequently merging all representations into a system of differential equations, it is also possible to infer quantitative information directly from data that characterize the entire system. These inverse approaches fall into two categories that correspond either to ▶ steady-state data or to *time series data*.

Steady-state data characterize a metabolic system under a condition where all concentrations have reached constant levels and all influxes and effluxes are in balance. The most common approaches for parameter estimation from such data use *stoichiometric analysis* (▶ Conservation Analysis) and ▶ flux balance analysis (FBA) (e.g., Palsson 2006). In both methods, the stoichiometry is assumed to be known,

some influxes and effluxes are measured as they enter or leave the system under steady-state conditions, and all other fluxes are inferred with computational methods. In addition, FBA accounts for physicochemical constraints on the fluxes in the system and furthermore assumes that the cell attempts to optimize an overall objective, such as maximal growth. This objective and all constraints are reformulated as a constrained linear optimization task, which permits an effective determination of a unique, optimal flux distribution that satisfies all constraints and maximizes the selected objective. These stoichiometrically based methods solely characterize a steady-state flux distribution and do not provide information on concentrations, regulation, or parameters associated with dynamic features of the system.

To a limited degree, parameter values can also be obtained from data *close to a steady state*. In this case, the data are obtained from experiments that measure the responses of a biological system to small perturbations (Sorribas and Cascante 1994). Such experiments may use biochemical inhibitors, artificial ligands, genetic mutations, or a variety of other methods.

The second category of inverse estimation approaches is very different from these methods at, or close to, a steady state. Namely, they are based on time series data that characterize the full dynamic response of a system to some stimulus, such as an environmental stress or the sudden availability of food. These types of data are becoming more prevalent, thanks to recent advancements in molecular high-throughput techniques at the genomic, proteomic, and metabolomic levels. At least in principle, these data have the capability of characterizing global responses in a dynamic manner, which subsequently permits the estimation of parameter values and even the identification of the topology and regulation of a system in a "top-down" manner.

The experimental tools that can generate dynamic data of metabolites include nuclear magnetic resonance (NMR), mass spectrometry (MS) coupled with high-performance liquid chromatography (HPLC), and flow cytometry. The experiments can be performed under many different conditions, such as different initial settings, various gene knock-outs or knock-downs, or different types of enzyme inhibition, and shed light on the system from different angles. Clear advantages of such "global" data include that their information content is rich, that they can be collected from the same organism, sometimes even in vivo, and that insights gained from their analysis are therefore as close to reality as is currently feasible.

The concept of these methods is intuitively simple: Use the chosen symbolic model and an optimization algorithm to determine values for all model parameters such that the dynamics of the model matches the observation data as closely as possible. The attraction of this approach is that time series data, if obtained from in vivo perturbation or stimulus–response experiments, directly or indirectly account for all processes associated with the reaction of the system to the perturbation.

Algorithms for Inverse Estimation

Many optimization algorithms have been proposed for inverse estimation tasks in biology. Most of them attempt to minimize the difference between the experimental data and a model response that is obtained per computer simulation. However, in spite of substantial efforts, none of these algorithms is truly satisfactory in realistic situations of moderately large systems or where time series data are sparse, incomplete, or corrupted by noise.

The most prominent algorithms for inverse tasks may be divided into three groups. The first group consists of *gradient-based*, *steepest-descent*, or *hill-climbing* methods. Best known among these are the Gauss–Newton and Levenberg–Marquardt algorithms, which are included in all major software packages of the field, such as Mathematica and Matlab. The second group consists of *stochastic search algorithms* (▶ Evolution Programming). These include *evolutionary computation* (EC) (▶ Stochastic Simulation Algorithm), ▶ simulated annealing (SA), *adaptive stochastic methods*, ▶ clustering methods, and other meta-heuristics, such as *ant colony optimization* (ACO) and ▶ particle swarm optimization (PSO). Some of these algorithms are biologically inspired and have been applied to parameter estimation tasks with the goal of finding global solutions, especially in the context of identifying the structures of gene regulatory networks (Moles et al. 2003). The best-known evolutionary method among these is the genetic algorithm (GA), which is now widely available in many variants that have proven useful and practical in a variety of biological applications. The third group of methods accounts for the fact that observed data are affected by random events. Thus, these approaches consider parameter estimation as a branch

of statistics and use statistical or machine-learning techniques as solution strategies. These techniques include maximum likelihood estimation, Markov chain Monte Carlo methods, and Kalman filtering.

One significant issue with time series analysis is that the differential equations of the model have to be solved thousands of times. Approximation methods have been developed to avoid this numerical solution step. While they save an enormous amount of computing time, they face their own challenges, which are not always easily overcome (Chou and Voit 2009; Voit and Almeida 2004; Ramsay et al. 2007).

Quality of Fit

Comparisons between parameter estimation algorithms are usually based on the goodness of their fits to experimental data, which is typically assessed as the sum of squared errors (SSE) between model results and the observed data (Voit 2011). However, the best fit should not be the only criterion, whether the estimation occurs in the forward or inverse direction. In many actual cases, the "best" model, as judged by the smallest SSE, actually tends to run into ▶ overfitting problems, which means that the model contains too many parameters. As a consequence, its extrapolation and prediction power with respect to untested conditions is low. Related to this issue is the observation that the same dataset can often be fitted by distinctly different parameter sets or even different model structures with similar accuracy. This issue of unidentifiability and sloppiness has received much attention in recent times (Gutenkunst et al. 2007; Srinath and Gunawan 2010).

The insistence on a good fit is not necessarily the best strategy due to the nature of biological phenomena. First, biological data are often affected significantly by inter-individual variability, as well as uncertainties due to slight variations in experimental conditions. As a consequence, biological observations are not as exactly replicable as experiments in physics. For instance, the K_M value is often seen as a genuine property of an enzyme, and in vitro characterizations are taken as true. However, the same enzyme might slightly vary between species and strains, and even a high degree of sequence homology may not prevent differences in the affinity between enzyme and substrate, which directly affects the K_M value. In other areas of science, the intrinsic variability among items can often be characterized with large numbers of measurements, but this strategy is not always feasible in biology. While variability is natural, biological phenomena are usually also very robust, so that modest alterations in parameter values often do not change their behavior. One contributor to this robustness is redundancy, which allows the compensation of failures in one component with changes in other components. This compensation directly translates into different sets of parameter values, which all perform with similar effectiveness and create biological sloppiness that may or may not be related to the sloppiness in optimized parameter sets. Thus, precise solutions, consisting of unique sets of parameter values, may not even be the ultimate criterion for a quality fit, and instead of aiming for a model with the smallest residual error, one might set as the goal to identify *all sets* of different parameterizations that are consistent with the data, within some acceptable error. Some of these may be discarded if they correspond to models with high sensitivities, a low degree of stability, or other undesirable properties. The remaining set of parameter values may turn out to be clustered tightly, consist of several distinct clusters, or be scattered throughout a large domain within the search space. In each case, the characteristics of this set, combined with comparative analyses of the candidate models, may identify one model as more likely than its alternatives or suggest hypotheses and critical laboratory experiments that may ultimately yield deeper insights into the system under investigation.

Cross-References

- ▶ Linear Regression
- ▶ Metabolic Networks, Reconstruction
- ▶ Optimization and Parameter Estimation, Genetic Algorithms
- ▶ Ordinary Differential Equation (ODE)
- ▶ Pathway Modeling, Metabolic

References

Chou I-C, Voit EO (2009) Recent developments in parameter estimation and structure identification of biochemical and genomic systems. Math Biosci 219:57–83

Goel G, Chou I-C, Voit EO (2006) Biological systems modeling and analysis: a biomolecular technique of the twenty-first century. J Biomol Tech 17:252–269

Gutenkunst RN, Casey FP, Waterfall JJ, Myers CR, Sethn AJP (2007) Extracting falsifiable predictions from sloppy models. Ann N Y Acad Sci 1115:203–211

Moles CG, Mendes P, Banga JR (2003) Parameter estimation in biochemical pathways: a comparison of global optimization methods. Genome Res 13:2467–2474
Palsson BØ (2006) Systems biology: properties of reconstructed networks. Cambridge University Press, Cambridge, UK
Ramsay JO, Hooker G, Cao J, Campbell D (2007) Parameter estimation for differential equations: a generalized smoothing approach. J Roy Stat Soc B 69:741–796
Sorribas A, Cascante M (1994) Structure identifiability in metabolic pathways: parameter estimation in models based on the power-law formalism. Biochem J 298: 303–311
Srinath S, Gunawan E (2010) Parameter identifiability of power-law biochemical system models. J Biotechnol 149(3):132–140
Voit EO (2004) The dawn of a new era of metabolic systems analysis. Drug Discov Today BioSilico 2:182
Voit EO (2011) What if the fit is unfit? In: Dehmer M, Emmert-Streib F, Salvador A (eds) Applied statistics for biological networks. Wiley, New York
Voit EO, Almeida JS (2004) Decoupling dynamical systems for pathway identification from metabolic profiles. Bioinformatics 20(11):1670–1681
Voit EO, Chou I-C (2009) Parameter estimation in canonical biological systems models. Int J Syst Synth Biol 1: 1–19
Voit EO, Qi Z, Miller GW (2008) Modeling complex biological systems, one step at a time. Pharmacopsychiatry 41(suppl 1): S78–S84

Forward-scattered Light (FSC)

Xiaojun Liu
Internal Medicine, The Second Hospital of Hebei Medical University, Shijiazhuang, Hebei, China

Definition

FSC is a measurement of mostly diffracted light. It is the part of the laser light that is detected just off the axis of the incident laser beam in the forward direction by a photodiode. It is proportional to cell-surface area or size. FSC provides a suitable method of detecting particles greater than a given size independent of their fluorescence and is, therefore, often used in immunophenotyping to trigger signal processing.

Cross-References

► Cell Sorting

Foundational Model of Anatomy

Mark A. Musen
Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA, USA

Definition

The most thorough, thoughtful, and comprehensive ► ontology of human anatomy in existence, initiated by Prof. Cornelius Rosse at the University of Washington in the 1990s.

Cross-References

► Protégé Ontology Editor

Founders Effect

Sabina Leonelli
ESRC Centre for Genomics in Society, University of Exeter, Exeter, Devon, UK

Definition

Founders effect is the narrowing of experimentation to a few well-studied organisms, usually portrayed as the opposite of bestowing resources on comparative research among organisms (Joergensen 2001 and Krebs 1975).

Cross-References

► Model Organism

References

Jørgensen CB (2001) August Krogh and Claude Bernard on basic principles in experimental physiology. Bioscience 51:59–61
Krebs HA (1975) The August Krogh Principle: 'for many problems there is an animal on which it can be most conveniently studied. J Exp Zool 194:221–226

Fourier Transform Infrared Microspectroscopy (FTIR)

Xiaohua Wu
Department of Pediatrics, Herman B Wells Center for Pediatric Research, Indiana University School of Medicine, Indianapolis, IN, USA

Definition

FTIR spectroscopy is a long-established and invaluable technique, which is based on the principle that molecules absorb mid-IR radiation, yielding richly structured IR absorption spectra.

Cross-References

► Spectroscopy and Spectromicroscopy

Frame Language

Mark A. Musen
Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA, USA

Definition

A knowledge representation based on prototype classes, where subclasses of each superclass specialize the superclass in some way. Frames are like objects in object-oriented programming. Frame classes have slots (or properties or attributes) that take on values when the frame is instantiated.

Cross-References

► Protégé Ontology Editor

Framework Region

► Framework Region (FR-IMGT)

Framework Region (FR-IMGT)

Marie-Paule Lefranc
Laboratoire d'ImmunoGénétique Moléculaire, Institut de Génétique Humaine UPR 1142, Université Montpellier 2, Montpellier, France

Synonyms

Framework region; FR-IMGT

Definition

A Framework region (FR-IMGT) is a region made of one or several beta strands, part of the framework of a V-DOMAIN (► Variable (V) domain), and delimited according to the ► IMGT unique numbering for V domain (Lefranc et al. 2003). There are four FR-IMGT in a V-DOMAIN: FR1-IMGT (beta strands A and B), FR2-IMGT (beta strands C and C′), FR3-IMGT (beta strands C″, D, E and F), and FR4-IMGT (beta strand G).

In a V-DOMAIN (V domain of the immunoglobulins (IG) or antibodies and T cell receptors (TR)), the first three FR-IMGT are part of the V-REGION (encoded by a ► variable (V) gene), whereas the FR4-IMGT corresponds to part of the J-REGION (encoded by a ► joining (J) gene).

By comparison, in a V-LIKE-DOMAIN (V domain of ► immunoglobulin superfamily (IgSF) other than IG and TR), the nine strands that correspond to the four FR-IMGT of a V-DOMAIN are usually encoded by one exon of a gene.

Amino acid positions of the FR-IMGT of a V-DOMAIN have always the same number according to the ► IMGT unique numbering for V domain (Lefranc et al. 2003). This allows to define, in an ► IMGT Collier de Perles, the lengths of the FR-IMGT, the anchor positions that support the complementarity determining regions (► Complementarity Determining Region (CDR-IMGT)) and the five conserved amino acids of a V-DOMAIN (by comparison, four for a V-LIKE-DOMAIN). These characteristics, based on the ► IMGT-ONTOLOGY concepts and managed in the ► IMGT® information system, are reported in Table 1.

Starting from amino acid sequences, the FR-IMGT lengths are obtained using the IMGT/DomainGapAlign

Framework Region (FR-IMGT), Table 1 Characteristics of the framework regions (FR-IMGT) of a V-DOMAIN

FR-IMGT	Beta strands	FR-IMGT lengths of a basic V-DOMAIN without gaps	Anchors and conserved amino acids[a]
FR1-IMGT	A, B	26	1st-CYS 23
			Anchor 26
FR2-IMGT	C, C′	17	Anchor 39
			CONSERVED-TRP 41
			Anchor 55
FR3-IMGT	C″, D, E, F	39	Anchor 66
			Conserved hydrophobic amino acid 89
			2nd-CYS 104 (Anchor 104)
FR4-IMGT	G	11	Anchor 118 (J-TRP or J-PHE in V-DOMAIN)[a]
Total		93	

[a]By comparison, the amino acid at position 118 is not conserved in a V-LIKE-DOMAIN. Otherwise, the anchors and conserved amino acids are identical between a V-DOMAIN and a V-LIKE-DOMAIN (▶ Variable (V) Domain).

tool that compares the V-DOMAIN with the closest germline V and J genes (http://www.imgt.org) (Lefranc 2009; Ehrenmann et al. 2010). The lengths of the individual beta strands (for both V-DOMAIN and V-LIKE-DOMAIN) are provided with amino acid changes according to the IMGT physicochemical classes (Pommié et al. 2004).

The lengths of the four FR-IMGT of a V-DOMAIN are shown between brackets and separated with dots. For a basic V-DOMAIN without gaps, the lengths of the FR-IMGT are [26.17.39.11] with a total of 93 positions in the ▶ IMGT Collier de Perles. For a human VH (V-DOMAIN of an IG-Heavy chain), the FR-IMGT lengths are [25.17.38.11] with a total of 91 amino acids, whereas for a human VL (V-DOMAIN of an IG-Light chain), the FR-IMGT lengths are [26.17.36.10] with a total of 89 amino acids.

Cross-References

- ▶ Complementarity Determining Region (CDR-IMGT)
- ▶ IMGT Collier de Perles
- ▶ IMGT Unique Numbering
- ▶ IMGT® Information System
- ▶ IMGT-ONTOLOGY
- ▶ Immunoglobulin Superfamily (IgSF)
- ▶ Joining (J) Gene
- ▶ Variable (V) Domain
- ▶ Variable (V) Gene

References

Ehrenmann F, Duroux P, Giudicelli V, Lefranc M-P (2010) Standardized sequence and structure analysis of antibody using IMGT®. In: Kontermann R, Dübel S (eds) Antibody engineering, Ch 2, vol 2, 2nd edn. Springer-Verlag, Berlin, Heidelberg, pp 11–31

Lefranc M-P (2009) Antibody databases and tools: The IMGT® experience. In: Zhiqiang An (ed) Therapeutic monoclonal antibodies: from bench to clinic, Ch 4 Wiley, Hoboken, New Jersey, pp 91–114

Lefranc M-P, Pommié C, Ruiz M, Giudicelli V, Foulquier E, Truong L, Thouvenin-Contet V, Lefranc G (2003) IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. Dev Comp Immunol 27:55–77

Pommié C, Levadoux S, Sabatier R, Lefranc G, Lefranc M-P (2004) IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. J Mol Recognit 17:17–32

Frequency

Tianshou Zhou
School of Mathematics and Computational Sciences, Sun Yet-Sen University, Guangzhou, Guangdong, China

Definition

For cyclical processes, such as rotation, oscillations, or waves, frequency is defined as a number of cycles per unit time. In physics and engineering disciplines, such as optics, acoustics, and radio, frequency is usually denoted by a Latin letter f or by a Greek letter ν (nu).

In SI units, the unit of frequency is the hertz (Hz), named after the German physicist Heinrich Hertz. 1 Hz means that an event repeats once per second. A previous name for this unit was cycles per second.

A traditional unit of measure used with rotating mechanical devices is revolutions per minute, abbreviated RPM. 60 RPM equals 1 Hz. The period, usually denoted by T, is the length of time taken by one cycle, and is the reciprocal of the frequency f:

$$T = \frac{1}{f}$$

The SI unit for period is the second.

Calculating the frequency of a repeating event is accomplished by counting the number of times that event occurs within a specific time period, then dividing the count by the length of the time period. For example, if 71 events occur within 15 s the frequency is:

$$f = \frac{71}{15\,\mathrm{s}} \approx 4.7(\mathrm{Hz})$$

If the number of counts is not very large, it is more accurate to measure the time interval for a predetermined number of occurrences, rather than the number of occurrences within a specified time. The latter method introduces a random error into the count of between 0 and 1, so on average half a count. This is called gating error and causes an average error in the calculated frequency of $\Delta f = 1/(2\ \mathrm{Tm})$, or a fractional error of $\Delta f / f = 1/(2\ f\ \mathrm{Tm})$, where Tm is the timing interval and f is the measured frequency. This error decreases with frequency, so it is a problem at low frequencies where the number of counts N is small.

Frequent Pattern Mining

Jesús Aguilar-Ruiz[1], Domingo Rodríguez -Baena[1] and Ronnie Alves[2]
[1]School of Engineering, Pablo de Olavide University, Escuela Politécnica Superior, Seville, Spain
[2]Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

Definition

Frequent Pattern Mining is a ▶ Data Mining subject with the objective of extracting frequent itemsets from a database. Frequent itemsets play an essential role in many ▶ Data Mining tasks and are related to interesting patterns in data, such as ▶ Association Rules. Some concepts are necessary in order to understand this definition:

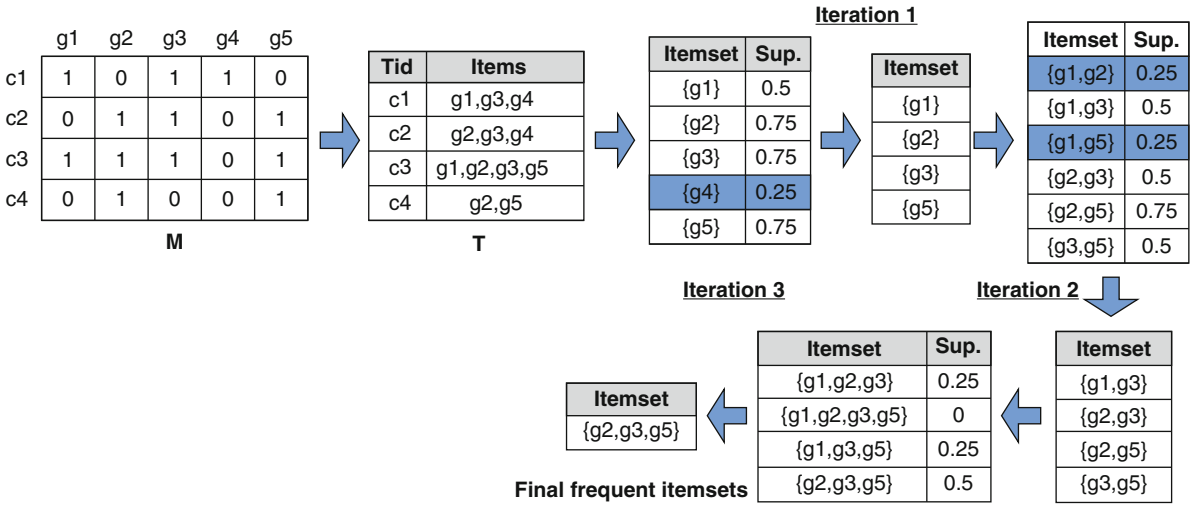
1. Transaction: Let $X = \{x_1, x_2, \ldots, x_m\}$ be a set of m elements called items and let $T = \{t_1, t_2, \ldots, t_n\}$ be a set of n subsets of items called transactions. Each transaction in T identifies a subset of items.
2. Frequent itemset: Given a set of items $X = \{x_1, x_2, \ldots, x_m\}$ and a set of transactions $T = \{t_1, t_2, \ldots, t_n\}$, a subset of X, S, is called a frequent itemset if S occurs in a percentage of all transactions in T that exceeds a threshold, named support.
3. Support: The support of an itemset Y, support (Y), is defined as the number of transactions in T which contain the itemset Y.

What exactly constitutes an item or a transaction depends on the application and on the type of information to be extracted. In a ▶ Gene Association Analysis context, the meaning of transaction is usually associated with "Over-expression," that is, only those "Over-expressed" genes will be understood to be included in the transaction. Equivalently, the term frequent itemset is related to frequent subset of genes.

Frequent Pattern Mining (FPM) techniques provide methods to extract automatically all the frequent itemsets from a dataset, being an extremely costly task (Alves et al. 2010). In ▶ Microarray data analysis, the specific ▶ Gene Expression dataset structure (thousands of genes against only hundreds of experimental conditions) increases the frequent itemsets mining process complexity. Due to this fact, developing efficient FPM techniques to be applied to ▶ Genomic studies has been an important challenge during the last years.

Next, in order to provide a better understanding of Frequent Pattern Mining methods, a simple APRIORI (*see the review paper of* Han et al. 2007 *for further information*) example is illustrated in Fig. 1.

Let M be a discretized matrix, where 1 and 0 mean "Over-expressed" and "Under-expressed," respectively. Table T represents the transactions and their items. The APRIORI algorithm generates, iteratively, candidate itemsets. In every iteration, the *support* of every candidate itemset is calculated, eliminating those itemsets with a *support* value under a threshold (set to 2/4 in this example). Based on the idea that an



Frequent Pattern Mining, Fig. 1 Example of the Apriori algorithm with support set to 2/4, that is, every itemset, to be considered as a valid candidate, has to appear in at least two of the four transactions

itemset is candidate if all its subsets are known to be frequent, the resulting itemsets are combined to create new candidate itemsets. The algorithm ends when no new candidate group can be generated. In Fig. 1 , after three iterations the final frequent itemset is composed of the genes $g2$, $g3$, and $g5$.

References

Alves R, Rodriguez-Baena D, Aguilar-Ruiz J (2010) Gene association analysis: a survey of frequent pattern mining from gene expression data. Brief Bioinform 11(2):210–224

Han J, Cheng H, Xin D, Yan X (2007) Frequent pattern mining: current status and future directions. Data Min Knowl Discov 15(1):55–86

Frequentist Approach

Larissa Stanberry
Bioinformatics and High-throughput Analysis Laboratory, Seattle Children's Research Institute, Seattle, WA, USA

Synonyms

Frequentist inference

Characteristics

In statistics, the goal is to make inference about parameters of the population. In frequentist inference, the parameters are constant, unknown values and data are viewed as repeatable random samples from the distribution. The parameters are estimated using point estimates and confidence intervals. Both point and interval estimators vary from sample to sample.

The point estimator is chosen with respect to its properties including unbiasedness, consistency, and efficiency. The unbiased estimator has its expectation equal to the population parameter, the consistent estimator converges in probability to the parameter value, and the relative efficiency of two estimators is defined by the ratio of their mean squared errors.

For example, consider data sampled from the univariate normal distribution with mean μ and variance $\sigma 2$. The sample mean, the sample median, and the midrange of a sample are the examples of a point estimator of the mean μ. All three estimators are consistent and converge to the true population mean as the sample size increases, but it is the sample mean that is efficient.

In the above example, the sample mean has desirable theoretical properties. A single number, however, gives little information as to how close

the estimated mean is to the true mean. A confidence interval gives an interval estimate of the population parameter. The length of the confidence interval indicates how well sample statistic estimates the population parameter. Confidence intervals are constructed at a certain confidence level $(1 - \alpha)$, where α is the type I error rate, i.e., the probability of rejecting null when it is true. The 95% confidence interval means that if one repeatedly samples data from the population, 95% of the corresponding confidence intervals would contain the true population parameter.

In the frequentist approach, the hypothesis testing about the parameters uses the distribution of the test statistics under the null hypothesis. When the exact distribution is not available, the hypothesis testing relies on large-sample theory to construct the approximate null distribution. Using the sampling distribution of the test statistic, the p-value is computed as the probability of observing a statistic as or more extreme than the given one. The decision is guided by comparing the p-value to the selected significance level α. The p-value should not be overinterpreted but rather it should be considered with respect to a given study design, sample size, and wider knowledge of the subject.

Frequentist Hypothesis Testing

▶ Hypothesis Testing

Frequentist Inference

▶ Frequentist Approach

FR-IMGT

▶ Framework Region (FR-IMGT)

FuGE

Peter Wilkinson[1] and Andrew R. Jones[2]
[1]Vaccine and Gene Therapy Institute, Port St. Lucie, FL, USA
[2]Institute of Integrative Biology, University of Liverpool, Liverpool, UK

Synonyms

FuGE-OM; Functional genomics experiment model

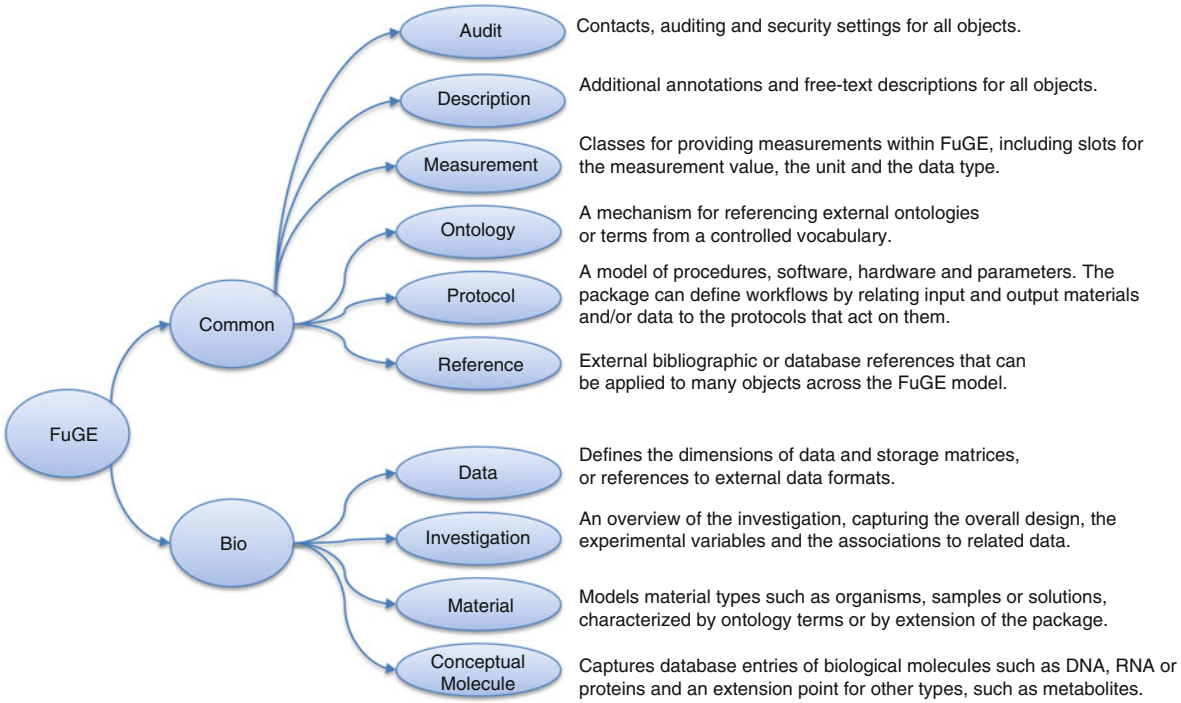
Definition

The Functional Genomics Experiment (FuGE) model is a set of computational resources, designed to facilitate rapid prototyping of new data exchange formats and corresponding relational database schemas for the life sciences. It comprises a Unified Modeling Language (UML) object model with software for automatically generating different implementations, including an XML Schema (XSD) for the exchange standard, a relational database schema, and a programming layer in Java to interface between the XSD and the database. Bioinformatics groups can develop extensions of FuGE through two modes – either by extending the object model in free graphical editing software, or by working directly with a cut-down version of the XSD (FuGE-light), to create new models for the biological domain of interest. FuGE is particularly useful for "omics" investigations and systems biology, where numerous heterogeneous data types may describe the biological system. The use of FuGE ensures that different data models have the same underlying core, allowing rapid development of new models and supporting software, and facilitating the integration of the resulting data.

Characteristics

FuGE Overview

The core resource in the FuGE project is an object model, comprising 10 packages (Fig. 1), each of which pertains to a particular type of data or metadata that is generically required for digitally representing any type of life sciences experiment (Jones et al. 2007). An example is the



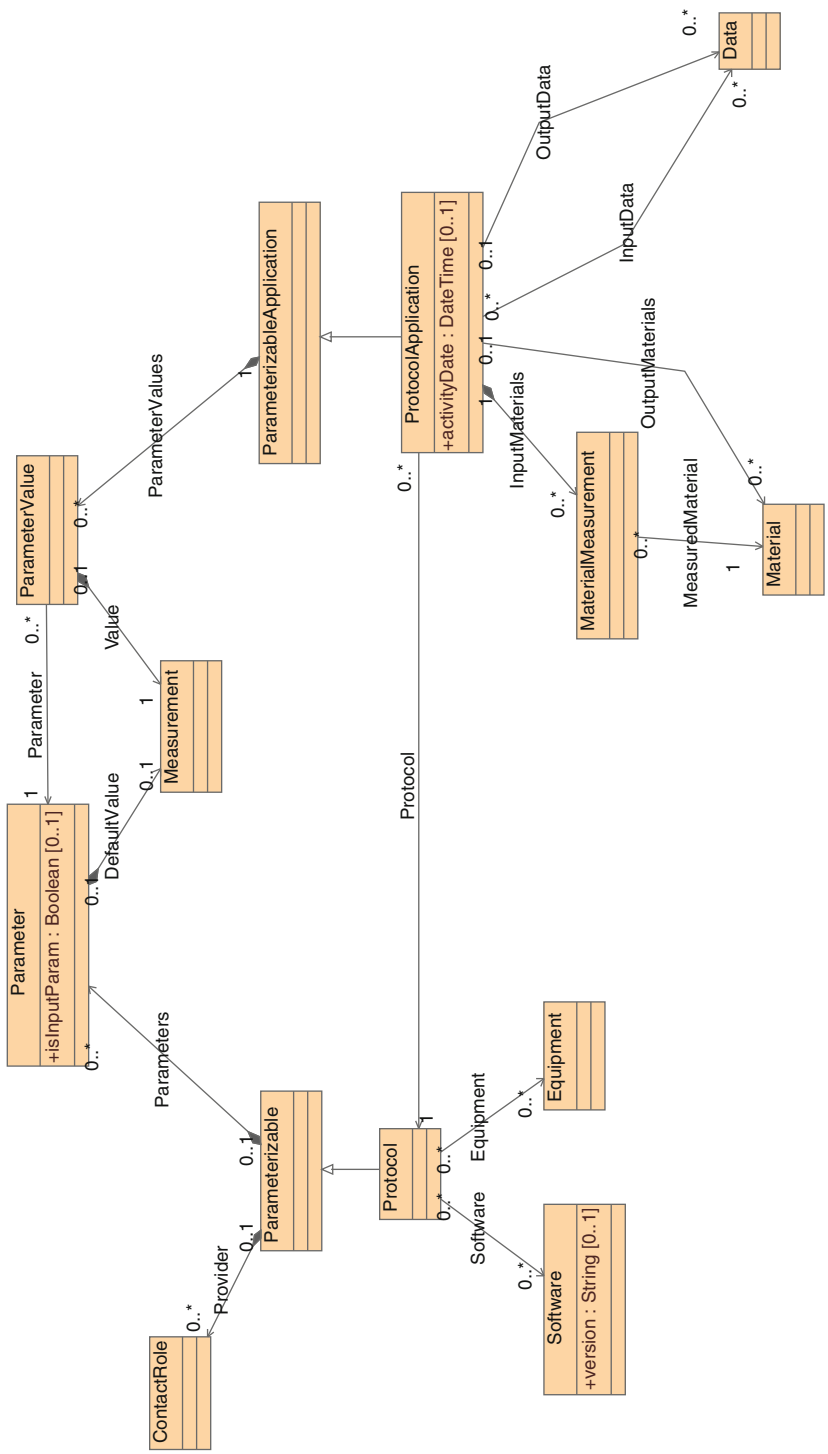
FuGE, Fig. 1 An overview of the FuGE package structure

Protocol package (Fig. 2), which contains a set of UML classes and diagrams showing the relationships between objects. The key components in the Protocol package are the *Protocol* class – a description of what should be done, for example, a standard operating procedure – and the *ProtocolApplication* class – a description of what was done, that is, with any runtime parameter values differing from the defaults defined in the Protocol. Similarly, the Data package contains structures for capturing data values in regular multidimensional matrices or as external files. *ProtocolApplication* is used to map between the input and output data (or biological samples) used at each stage of an experimental or data analysis pipeline, and thus can track the full audit trail for how the final results were generated through each stage in the process.

FuGE for Data Format Development

One of the primary uses of FuGE is to develop new data exchange formats or standards (Jones et al. 2009). A user interacts with the FuGE model with a UML editing tool or an XSD editing tool if working in the FuGE-light mode. Extensions are built to capture the data types specific to the experimental or analysis method. For example, extensions built for proteomics capture the essential parameters of protein separation techniques (Gibson et al. 2010) as defined by the associated minimum reporting guidelines documents – MIAPE (Taylor et al. 2007). In this way, the format developer does not have to recreate all the models for the basic core; FuGE provides all the elements for handling protocols, data files, samples, and ontologies. New data models can be developed quickly, and the associated project software can be used to create a relational database or format an editing software automatically (Belhajjame et al. 2008; Swertz et al. 2010).

Systems biology investigations typically use a range of experimental approaches to understand cellular processes. In the absence of using an extensible model, developing a new data format or exchange standard can be a challenging process, and resulting models for different domains tend to represent similar concepts (such as protocols and parameters) using different terminologies and different levels of detail, thus making data integration more difficult. Furthermore, some models do not allow for audit trails to capture all the processes that have been applied to the sample or data file.



FuGE, Fig. 2 A component of the Protocol package, showing how parameter values are provided by a ProtocolApplication

Adoption of the FuGE methodology for model development should encourage developers to capture this level detail where appropriate, which will help with unambiguous interpretation of the results by the diverse community of potential consumers of the data.

FuGE for Database Development

Another use of FuGE is to develop an SQL or object-based database schema for data storage. Database systems are designed to meet specific project requirements and are generally purpose built for optimized query execution or data entry. Fundamentally, the data model is the most critical aspect of system design and function, and the model should reflect "real world" objects and their relationships to ensure durability that can outlast any application, many of which are not known when the system is first put into production. These assertions are important in addressing the needs of systems biology investigations because investigations are large, lengthy, multi-institutional, and often need to be integrated with one another, and, as such, require long-term data persistence, protocol management, and data exchange. FuGE is a "real world" model of functional genomics experiments and can thus be used as a reference to derive a suitable database persistence layer. SyMBA (http://symba.sourceforge.net/), for example, provides a persistence layer directly based on the FuGE XML Schema. FuGE can also be used to create an efficient, relatively generic, SQL-based schema that can be queried based on description or ontology types as consistent with the FuGE model, as described at the project website (http://fuge.sourceforge.net/).

FuGE for Pipeline Management

Systems Biology experiments also provide another challenge in terms of analysis execution. Using ad hoc analysis, similar challenges to data persistence emerge that include a reduced ability to audit and manage large amounts of data and reporting. The FuGE *Protocol* Package provides a suitable model to describe in silico protocols and applications of protocols. For example, it is straightforward to describe an analysis workflow as a pipeline by using the *ProtocolApplication* to describe the sequence of steps with their respective parameter values, inputs, and outputs. Furthermore, the results of such a workflow can be moved to a persistence layer, and exchanged later on.

The FuGE model is a real world description of omics experiments that also provides a means for tractable management of high-throughput systems biology investigations. FuGE captures the necessary object abstractions that can be implemented across three layers of data management: (1) data persistence, (2) in silico and bench protocol management, and (3) data exchange.

References

Belhajjame K, Jones AR, Paton NW (2008) A toolkit for capturing and sharing FuGE experiments. Bioinformatics 24:2647–2649

Gibson F, Hoogland C, Martinez-Bartolomé S, Medina-Aunon JA, Albar JP, Babnigg G, Wipat A, Hermjakob H, Almeida JS, Stanislaus R, Paton NW, Jones AR (2010) The gel electrophoresis markup language (GelML) from the proteomics standards initiative. Proteomics 10:3073–3081

Jones A, Miller M, Aebersold R, Apweiler R, Ball C, Brazma A, Degreef J, Hardy N, Hermjakob H, Hubbard S, Hussey P, Igra M, Jenkins H, Julian R, Laursen K, Oliver S, Paton N, Sansone S-A, Sarkans U, Stoeckert C, Taylor C, Whetzel P, White J, Spellman P, Pizarro A (2007) The functional genomics experiment model (FuGE): an extensible framework for standards in functional genomics. Nat Biotech 25:1127–1133

Jones AR, Lister AL, Hermida L, Wilkinson P, Eisenacher M, Belhajjame K, Gibson F, Lord P, Pocock M, Rosenfelder H, Santoyo-Lopez J, Wipat A, Paton NW (2009) Modeling and managing experimental data using FuGE. OMICS: A J Integr Biol 13:239–251

Swertz M, Velde KJ, Tesson B, Scheltema R, Arends D, Vera G, Alberts R, Dijkstra M, Schofield P, Schughart K, Hancock J, Smedley D, Wolstencroft K, Goble C, de Brock E, Jones A, Parkinson H, members of the Coordination of Mouse Informatics, R, Genotype-To-Phenotype, C, Jansen R (2010) XGAP: a uniform and extensible data model and software platform for genotype and phenotype experiments. Genome Biol 11:R27

Taylor CF, Paton NW, Lilley KS, Binz P-A, Julian RK, Jones AR, Zhu W, Apweiler R, Aebersold R, Deutsch EW, Dunn MJ, Heck AJR, Leitner A, Macht M, Mann M, Martens L, Neubert TA, Patterson SD, Ping P, Seymour SL, Souda P, Tsugita A, Vandekerckhove J, Vondriska TM, Whitelegge JP, Wilkins MR, Xenarios I, Yates JR, Hermjakob H (2007) The minimum information about a proteomics experiment (MIAPE). Nat Biotech 25:887–893

FuGE-OM

▶ FuGE

Function

▶ Function, Distributed

Function Type

- ▶ IMGT-ONTOLOGY, FunctionType
- ▶ IMGT-ONTOLOGY, SpecificityType

Function, Biological

Arno Wouters
Department of Philosophy, Erasmus University Rotterdam, Rotterdam, The Netherlands

Definition

The activity, role, value or purpose of a part, activity, or trait of an organism.

Characteristics

Terms like "function," "functions," and "functional" are used in many different ways. The 2005 edition of the *New Oxford American Dictionary* gives the following as the first meaning of "function": "an activity or purpose natural to or intended for a person or thing" with "Vitamin A is required for good eye function" as an example. This definition is suitable as a general characterization of the term "function" and at the same time it contains the seeds of many confusions about the notion of biological function, especially because it talks about "activity or purpose" and"natural to or intended for."

The idea that biological function is somehow related to purposes and the idea that there can be natural purposes in addition to intended ones has been a source of inspiration for philosophical discussion. In the 1950s and 1960s, philosophers of science, of a logical positivist inclination, searched for ways to define the notion of biological function without appeal to purpose. Since the 1980s, many philosophers think that evolutionary theory provides us with a notion of natural purpose that can be used to develop a naturalized account of purposes, norms, and meaning in the philosophy of mind and language. According to these "etiological theories," it is the natural purpose and, for that matter, the "proper function" of a trait of an organism to produce the effects for which that trait was maintained in the process of natural selection in the (possibly recent) past of that organism's population (Wright 1976; Millikan 1989; Neander 1991). In the philosophical debate that emerged in reaction to these theories, many different understandings of function and functional explanations have been developed (see Wouters [2005] and Garson [2008] for overviews).

In biology, the connotation of "function" is usually not purpose but activity, in a broad sense of that term, including "what it does," "how it works," and "how it is used." For example, "functional morphology" is typically defined as the study of the form of organisms and their parts in relation to their activity and use. The many articles yielded by a Google Scholar search on "structure and function" typically discuss both the way in which a part of an organism is built (its structure) and the way it works (its function). Within this broad sense of function as activity, two uses of the term function can be distinguished: function as activity in a stricter sense (what it does and how it works) and function as biological role (how it is used).

"*Function as activity*" refers to what a system does by itself (in abstraction of its effects on its environment) and the way it works – internally (e.g., the way in which the activity is generated) or externally (e.g., the order of its changes). The notion of function as what it does is typically used to distinguish form (or structural) characteristics from functional characteristics. The form characteristics of a system concern its appearance (shape, volume, color, pattern, texture, etc.), structure (composition, size, and spatial arrangement of the parts, e.g., amino acid sequences), and statics (hardness, weight, mass, etc.); the functional characteristics of a system concern its activity (frequency, order, velocity, momentum, reaction rates, oxygen consumption, kinetic energy, etc.). For example, talk of "functional homology" might refer to a common pattern in muscle movement, whereas talk of "structural homology" might refer to a common pattern in the spatial arrangement of the muscles. The notion of function as how it works is typically used to make comparisons. For example, when it is said that the heart's ventricle functions as a pressure pump and the atrium as a suction pump, one compares the way in which these two systems work.

The notion of *function as biological role* refers to the role of a system in enabling life. In general, role functions concern the role of a system or activity in

bringing about an organized characteristic of an encompassing system (a role function of a brake is to enable the driver to stop the car because stopping the car is how the brake contributes to the car's organized ability to transport people). The *biological* role of a part of an organism is the role of that part in bringing about the organism's state of being alive. For example, the main biological role of the glycolysis is the production of ATP because that is how the glycolysis contributes to an organism's ability to stay alive. Note that role functions are positions in an organization rather than measurable properties.

Ascriptions of biological roles are the handle to understand life. Just as it is possible to explain how a company works by means of an organization chart that outlines the tasks of the different functionaries and departments and the way in which they interact, the ability of an organism to stay alive can be explained by outlining the roles the different organ systems play in bringing about the living state. The ability of each organ system to perform its biological role, in turn, can be explained by describing the roles the different parts of that system play in bringing about that ability, and so on, until a level is reached at which the relevant subsystems can be explained in terms of the physical and chemical characteristics of the molecules that make up that subsystem (cf. Cummins 1975). Such an organization chart provides a unifying framework for biology that relates detailed studies of specific mechanisms at different levels to the general aim of understanding life.

Yet, another use of the term "function" stems from behavioral biology. In this area of study, "function" often refers to the advantages of behaving in one way rather than another. More generally, the notion of *function as biological advantage* (also called "survival value," "adaptive value," or "biological value") is used to refer to the way in which a certain trait influences the life chances of an organism in a certain environment as compared to other traits that might replace it. An advantage of a trait in a certain environment is an ability resulting from that trait due to which the life chances of organisms with that trait are higher than the life chances would be of organisms in which that trait were replaced by another one (Canfield 1964; Bigelow and Pargetter 1987).

Advantage articulations compare organisms with a certain trait with similar organisms in which that trait is replaced by another one (or removed). The hypothetical organisms with which the real organisms are compared need not be real. Quite often, a comparison is made between a real organism and a hypothetical organism that cannot possibly exist and the point of the comparison is precisely that: to show that it cannot exist (because it lacks an essential ability).

Advantages differ from role functions in many ways. Advantages are abilities to solve certain problems, not positions in an organization. Advantages are, unlike role functions, relative to an environment and to the traits used for comparison. In addition, role functions are typically attributed to parts or activities, whereas advantages are effects of traits (i.e., of the properties of systems or activities, including the presence of certain items or the performance of certain activities). It is, for example, the biological role of the heart (a part of an organism) to pump the blood around, whereas pumping blood by means of a heart (a trait) is advantageous relative to pumping blood by means of beating blood vessels (the trait for comparison) in environments with certain types of prey and predators because this allows for faster oxygen transport (an ability resulting from the presence of a heart), which allows the organism to be more active and, hence, to escape from predators or to catch prey in situations where an organism with beating blood vessels would not be able to do so (more distal abilities resulting from that trait).

Functional biology can be defined as the study of how living systems (organisms) and their parts work. Functional biologists are concerned with two kinds of explanations that deal with synchronic relations between the different parts and activities of organisms and the environment in which they live: mechanistic explanations (also called "causal explanations") and functional explanations (also called "ecological explanations" or "design explanations"). The ascription of role functions is central to explanations of both kinds (see ▶ Explanation in Biology).

Mechanistic explanations address questions about how a certain biological role is performed (e.g., "how does the glycolysis generate ATP?"), by describing a mechanism that produces the behavior that enables that system to perform this role. Because of their concern with biological roles, mechanistic explanations in biology are sometimes called "functional explanations" or "functional analyses" (especially by philosophers) (Cummins 1975). This kind of explanation is discussed in the entry on mechanistic explanation.

Functional explanations address questions about why a biological role is performed the way it is (e.g., "why do many pathways that generate ATP start by activating their substrate?") by pointing to the advantages of performing the role in that way rather than in some conceivable alternative way (Wouters 2007). This kind of explanation is often called "functional explanation" (especially by biologists) because it is concerned with the advantages of certain forms of organization rather than with the question of how those forms are brought about. It is discussed in the entry on functional explanation (▶ Explanation, Functional).

Biological roles also play an important role in certain explanations in evolutionary biology (the study of the history and dynamics of lineages of organisms), especially in adaptation explanations. Adaptation explanations (also called "selection explanations") are evolutionary explanations that explain certain characteristics of a population as the result of past interaction in that population between variants that differed in their fitness. If a certain trait evolved because the fitness of past variants having that trait was higher than that of their competitors lacking that trait because the presence of that trait improved the performance of a certain role function, one might say that the trait evolved as an adaptation for performing that role function. For this reason, adaptation explanations are sometimes called "functional explanations" (especially by philosophers) (Brandon 1990). This kind of explanation is discussed in the entry on evolutionary explanation (▶ Explanation, Evolutionary).

Cross-References

▶ Explanation in Biology
▶ Explanation, Evolutionary
▶ Explanation, Functional

References

Bigelow J, Pargetter R (1987) Functions. J Philos 84:181–196
Brandon RN (1990) Adaptation and environment. Princeton University Press, Princeton
Canfield JV (1964) Teleological explanation in biology. Br J Philos Sci 14:285–295
Cummins R (1975) Functional analysis. J Philos 72:741–765
Garson J (2008) Function and Teleology. In: Sarkar S, Plutynski A (eds) A companion to the philosophy of biology. Blackwell, Cambridge, pp 525–549
Millikan RG (1989) In defense of proper functions. Philos Sci 56:288–302
Neander K (1991) The teleological notion of 'function'. Australa J Philos 69:454–468
Wouters AG (2005) The function debate in philosophy. Acta Biotheor 53:123–151
Wouters AG (2007) Design explanation: determining the constraints on what can be alive. Erkenntnis 67:65–80
Wright L (1976) Teleological explanations: an Etiological analysis of goals and functions. University of California Press, Berkeley

Function, Distributed

Jonathan F. Davies
Fondazione Bruno Kessler, Trento, Italy

Synonyms

Delocalized; Function

Definition

A distributed function is a feature of mechanistic explanations of some of the properties of complex systems (▶ Complex System). It is a component causal-role ▶ function that cannot be localized onto a readily identified component structure. A function may be distributed across multiple nonlinearly interacting elements or processes, across the whole system or even transcend systemic boundaries.

Characteristics

Central strategies in the mechanistic explanation of the properties and behaviors of biological systems are decomposition and functional localization. Decomposition depends on the assumption that the behavior of a system is a product of a set of subordinate functions, and that the interactions between the functional elements are minimal and can be handled additively (Bechtel and Richardson 2010). Functional localization is achieved when a structural decomposition of the system – in simple cases achieved through observation, and in more complex cases by the identification, of regions of maximal causal interaction, bounded with

regions of low interactivity (see ▶ Top-down Decomposition of Biological Networks) – yields identifiable components that are held to be responsible for these functions (▶ Functional Modules and Complexes). The full explanation is achieved by the localization of a complete set of bottom-level component functions (activities) in particular structures (entities) possessing the requisite characteristics or capacities that enable them to be the bearer of these functions (Machamer et al. 2000). Although it is not necessary to assume that a single component (in the sense of an individual, spatially localized entity or structure identified as a component of the system) is responsible for a specific activity, this assumption is often made, if only as a first approximation. Even in cases where components interact (a feature of what Herbert Simon (1996) called near-decomposable systems), the primary explanatory burden rests on the intrinsic (context-independent) properties of the components. Interactions are usually modeled in a linear, additive fashion and organizational factors are relegated to a secondary role, as constraints.

A well-known example of this explanatory strategy in molecular biology is Jacob and Monod's (1961) operon model of ▶ gene regulation. They posited regulatory genetic elements as a solution to the problem of accounting for changes in gene expression. The details of this model include the functionally defined ▶ repressor. This consists of a protein that exhibits certain structural features allowing it to bind to the operator and so prevent the expression of the structural genes. The shape and molecular constitution of the regulatory protein ensures that it is able to perform the bottom-level activities (in this case, geometrical/mechanical and chemical bonding) that constitute an instance of gene regulation. Other component functions (structural genes, ▶ Promoter, operator, etc.) are performed by identified structures each constituted in such a way as to make their individual contribution to the behavior of the system intelligible. The adequacy of the explanation is a result of the gross systemic behavior being the sum of the linear sequence of sub-tasks (functions), which in turn are the sum of their component sub-tasks until we reach the lowest level of entities and activities of concern to the field (Machamer et al. 2000).

Standard strategies for the development of this kind of mechanistic explanation often involve the intuitive parsing of the systemic behavior into component functions giving rise to a plausible ▶ mechanism sketch or schema (Craver 2007). Competing schemas are then tested, for example, through the use of targeted ▶ perturbation experiments. Successful explanations are achieved when functional decompositions map onto structural decompositions. Problems arise, however, when this mapping is not achieved or when the perturbation of components yields counterintuitive results. Some features of biological systems do not appear to be analyzable into functionally discrete components and the functionally relevant properties of many components are context sensitive to some degree. Bechtel and Richardson (2010) identify systems in which the structural components seem to perform tasks that would not appear in an intuitively plausible functional decomposition of the system. This makes the localization of component functions onto identified discrete structures impossible and has given rise to the suggestion that the emergent behavior of the system is inexplicable in mechanistic terms (▶ Emergence).

Kauffman's (1993) explanation of the stability of ▶ gene regulatory networks on the basis of their connectivity is an extreme example of an explanation of a systemic feature that does not make use of functional localization. His network model consists of simple nodes, each node being in one of two states (on, off). The interactions between the nodes are simple activation or repression, so making transitions between successive states of the network Boolean operations (▶ Boolean Networks). Kauffman argues that networks of this sort exhibit behavior that encounters stable cycles (▶ Stability) even in the face of perturbations and that a gene network characterized by a similar architecture will be robust. The explanatory properties for the ▶ robustness of the network are distributed across the entire system and the intuitively satisfying parsing of the systemic behavior into functionally discrete components is not possible. Any explanatory properties of individual nodes are dependent on their relations to other nodes rather than on their intrinsic properties; for example, a ▶ hub is a hub in virtue of its high number of connections and if it is functionally significant, it is so because of these connections.

A less radical example of a process of the distribution of a functional component of a biological system is the notion of a eukaryotic gene. Genes are made up of non-contiguous exons (▶ Exon) interspersed with non-coding introns (▶ Intron), which, to complicate matters further, are utilized in differing ways in different developmental processes. The localization of the functional unit ("gene") onto a discrete molecular structure

(i.e., a contiguous section of DNA) does not appear to be possible. This, and further complications with regard to the physiological roles of the gene products, has led Moss (2003) to argue that the functional unit (called "gene-P" to signify its connection to phenotype) cannot be mapped onto any recognizable molecular structure (called "gene-D" to signify status as a developmental resource). The phenotypic (functional) role of any particular molecular gene (gene-D) is not fixed by its intrinsic structural properties, but is rooted in features distributed throughout a dynamic network of regulatory and developmental resources. The extent of this distribution will depend on the particular process and the context in which it is taking place.

It is argued that biological systems are characterized by a spectrum from highly localized to radically distributed functional elements and that neither the "reductionistic" (▶ Reduction) localization of bottom-level component functions in discrete structures nor the "holistic" (▶ Holism) representation of system dynamics constitute adequate explanatory strategies (Krohs and Callebaut 2007). Instead a plurality of explanatory approaches, which depend on the nature of the subject matter and the interests of the investigator, should be pursued (Mitchell 2003). In systems biology, this means dealing with hierarchically organized dynamic systems and networks (see ▶ Hierarchy and ▶ Organization), the features of which are explicable in terms of combinations of network and system dynamics and the properties of components structures, sub-networks, and pathways (Kitano 2002). Addressing such ▶ complexity, non-linearity, and heterogeneity is only possible with huge quantities of data across the full range of omic levels and the capacity to analyze these data in parallel. Traditional techniques for localizing functions through, for example, single gene knockout experiments cannot deal with robust systems or organizational and dynamic properties. One suggestion is to attempt simulations in silico that integrate network and dynamical systems models with molecular data concerning known components (Kitano 2002). Another suggestion, which relates directly to the problems of explaining systems containing distributed functional components, is to focus on the "unbiased" structural decomposition of the system. Making use of quantitative methods for delineating network modules, through the identification of local maxima of interaction, avoids the need for the intuitive decomposition of the system into functional components (Krohs and Callebaut 2007). The secondary ascription of functional roles to individual modules, combinations of modules, and other network features (e.g., connectivity, small world architecture; see ▶ Small-World Property and ▶ Functional Modules and Complexes) can then be attempted by coordinated perturbation experiments, conducted in vitro or in silico (or even in vivo) on the basis of molecular data and theoretical work.

The assumption underlying these methodological suggestions is that functionality in biological systems is not always readily localizable onto discrete structural components. An individual function (which is defined in relation to the concerns of the investigator) may be located onto a single molecular component, a discrete structure, a network module or complex, a chemical pathway, a dynamic process, an architectural feature of the whole system or may even interfere in the interaction between the system and its environment. One of the novel features of systems biology is thought to be that it facilitates the empirical investigation into the extent of distribution of any particular function in cases in which our intuitions and mental capacities let us down. Even in cases in which the function bearer is a dynamic process (so involving multiple structural components over a period of time), it may be possible to construct a simulation that accounts for its role in the overall system and ascertain to what extent it is spatiotemporally distributed. It remains to be seen to what extent the computational and theoretical tools available to systems biologists will overcome the obstacles to achieving such ambitious aims. However, the developments in the capacity of systems biologists to account for the behaviors of complex biological systems are already throwing new light on the roles of decomposition and functional localization and encouraging philosophers of science to think again about the limits of mechanistic explanatory strategies.

Cross-References

- ▶ Boolean Networks
- ▶ Complex System
- ▶ Complexity
- ▶ Emergence
- ▶ Exon
- ▶ Function
- ▶ Functional Modules and Complexes
- ▶ Gene Regulation
- ▶ Gene Regulatory Networks
- ▶ Holism

- ▶ Hub
- ▶ Intron
- ▶ Mechanism
- ▶ Organization
- ▶ Perturbation
- ▶ Promoter
- ▶ Reduction
- ▶ Repressor
- ▶ Robustness
- ▶ Small-World Property
- ▶ Stability
- ▶ Top-down Decomposition of Biological Networks

References

Bechtel W, Richardson RC (2010) Discovering complexity: decomposition and localization as strategies in scientific research. MIT Press, Cambridge

Craver CF (2007) Explaining the brain: mechanisms and the mosaic unity of neuroscience. Clarendon Press, Oxford

Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 3:318–356

Kauffman S (1993) The origins of order: self organization and selection in evolution. Oxford University Press, Oxford

Kitano H (2002) Looking beyond the details: a rise in system-oriented approaches in genetics and molecular biology. Curr Genet 41:1–10

Krohs U, Callebaut W (2007) Data without models merging with models without data. In: Boogerd F et al (eds) Systems biology: philosophical foundations. Elsevier, Amsterdam, pp 181–213

Machamer P, Darden L, Craver CF (2000) Thinking about mechanisms. Phil Sci 67:1–25

Mitchell SD (2003) Biological complexity and integrative pluralism. Cambridge University Press, Cambridge

Moss L (2003) What genes can't do. MIT Press, Cambridge

Simon H (1996) The sciences of the artificial. MIT Press, Cambridge

Functional

Marie-Paule Lefranc
Laboratoire d'ImmunoGénétique Moléculaire, Institut de Génétique Humaine UPR 1142, Université Montpellier 2, Montpellier, France

Definition

"*Functional*" is a ▶ leafconcept of the "▶ *FunctionalityType*" concept of identification (generated from the ▶ IDENTIFICATION Axiom) of ▶ IMGT-ONTOLOGY, the global reference in ▶ immunogenetics and ▶ immunoinformatics (Giudicelli and Lefranc 1999; Lefranc et al. 2004, 2005, 2008; Duroux et al. 2008), built by IMGT®, the international ImMunoGeneTics information system® (http://www.imgt.org) (▶ IMGT® Information System). "*Functional*" identifies, whatever the molecule type (▶ MoleculeType), the functionality of ▶ Molecule_EntityType leafconcepts in *undefined* or *germline* configuration (▶ Configuration Type), whose coding region has an open reading frame without stop codon and for which there is no described defect in the splicing sites, recombination signals (▶ Recombination Signal (RS)) and/or regulatory elements.

"Functional" is one of the three leafconcepts (the other two being "▶ ORF" and "▶ Pseudogene") that identify the functionality of Molecule_EntityType leafconcepts in *undefined* configuration (▶ Configuration Type) conventional genes (▶ Conventional Gene) and immunoglobulin (IG) and T cell receptor (TR) constant (C) genes (▶ Constant (C) Gene), or in *germline* configuration (▶ Configuration Type) (IG and TR variable (V), diversity (D), and joining (J) (▶ Variable (V) Gene, ▶ Diversity (D) Gene, ▶ Joining (J) Gene) genes *before* DNA rearrangements) (Giudicelli and Lefranc 1999; Lefranc et al. 2004; Duroux et al. 2008).

Cross-References

- ▶ Configuration Type
- ▶ Constant (C) Gene
- ▶ Conventional Gene
- ▶ Diversity (D) Gene
- ▶ FunctionalityType
- ▶ IMGT-ONTOLOGY
- ▶ IMGT-ONTOLOGY, IDENTIFICATION Axiom
- ▶ IMGT-ONTOLOGY, Leafconcept
- ▶ IMGT® Information System
- ▶ Immunogenetics
- ▶ Immunoinformatics
- ▶ Joining (J) Gene
- ▶ Molecule Entity Type
- ▶ MoleculeType
- ▶ Open Reading Frame (ORF)
- ▶ Pseudogene
- ▶ Recombination Signal (RS)
- ▶ Variable (V) Gene

References

Duroux P, Kaas Q, Brochet X, Lane J, Ginestoux C, Lefranc M-P, Giudicelli V (2008) IMGT-Kaleidoscope, the Formal IMGT-ONTOLOGY paradigm. Biochimie 90:570–583

Giudicelli V, Lefranc M-P (1999) Ontology for immunogenetics: the IMGT-ONTOLOGY. Bioinformatics 12:1047–1054

Lefranc M-P, Giudicelli V, Ginestoux C, Bosc N, Folch G, Guiraudou D, Jabado-Michaloud J, Magris S, Scaviner D, Thouvenin V, Combres K, Girod D, Jeanjean S, Protat C, Yousfi MM, Duprat E, Kaas Q, Pommié C, Chaume D, Lefranc G (2004) IMGT-ONTOLOGY for immunogenetics and immunoinformatics. In Silico Biol 4:17–29

Lefranc M-P, Clément O, Kaas Q, Duprat E, Chastellan P, Coelho I, Combres K, Ginestoux C, Giudicelli V, Chaume D, Lefranc G (2005) IMGT-Choreography for immunogenetics and immunoinformatics. *In Silico* Biol 5:45–60

Lefranc M-P, Giudicelli V, Regnier L, Duroux P (2008) IMGT, a system and an ontology that bridge biological and computational spheres in bioinformatics. Brief Bioinform 9:263–275

Functional Differential Equations

▶ Dynamical Systems Theory, Delay Differential Equations

Functional Enrichment Analysis

Jiguang Wang
Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, Beijing, China

Definition

For a given cluster of biological elements such as gene, functional enrichment analysis is to compare the enrichment of any type of biologically relevant labels such as gene ontology terms in these genes to that in the background. Hypergeometric distribution or binomial distribution is usually applied to give the *p*-value for whether the enrichment is significantly different in the given gene cluster and in the background. So the choice of background and the cutoff value are important for the reporting result. There have been several available tools for application, e.g., FatiGO (Al-Shahrour et al. 2004), g:profiler (Reimand et al. 2007), etc.

References

Al-Shahrour F, et al. FatiGO: a web tool for finding significant associations of gene ontology terms with groups of genes. Bioinformatics. 2004;20:578–80.

Reimand J, Kull M, Peterson H, Hansen J, Vilo J. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res. 2007;35(Web Server Issue):W193–200.

Functional Genomics Experiment Model

▶ FuGE

Functional GO and Pathway-Based Network

▶ Functional/Signature Network Module for Target Pathway/Gene Discovery

Functional Interaction Network

▶ Organelle and Functional Module Resources

Functional Model

▶ Process-based Model

Functional Modules and Complexes

Xianwen Ren
Key Laboratory of Systems Biology of Pathogens, Ministry of Health, Institute of Pathogen Biology, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Definition

Proteins act in concert in cells. Some form complexes in which member proteins bind to each other stably to

act as a whole functional unit, e.g., ribosomes. Complexes are expected to be predicted from protein-protein interaction networks through identifying functional modules.

Functional module refers to a set of proteins that are densely connected within themselves but sparsely connected with the rest in the biological molecular network.

Cross-References

► Pathway, Functional Units

References

Spirin V, Mirny AL. Protein complexes and functional modules in molecular networks. Proc Natl Acad Sci USA. 2003;100(21):12123–8

Functional/Signature Network Module for Target Pathway/Gene Discovery

Shipra Agrawal[1] and M. R. Satyanarayana Rao[2]
[1]BioCOS Life Sciences Pvt. Limited, Institute of Bioinformatics and Applied Biotechnology, Bangalore, Karnataka, India
[2]Chromatin Biology Laboratory, Molecular Biology and Genetics Unit, Jawaharlal Nehru Center for Advanced Research, Bangalore, Karnataka, India

Synonyms

Co-expression network; Expression signature; Functional GO and pathway-based network; Genomics network

Definition

High-throughput molecular biology data is often analyzed and represented today through molecular networks. The network of interacting proteins/genes is an undirected network, where proteins/genes are represented by nodes and the interaction between them are represented by edges. The multiple nodes and corresponding edges in a network together form modules/subnetworks that manifest high internal similarity/correlation (i.e., modular networks). These modules form an independent connected component with each other, while they are sparsely connected with the rest of the network (Erten et al. 2009). The network modules carrying biologically meaningful information are termed as functional modules. A general concept on identifying signature and functional module is described below.

Characteristics

Identification of Signature and Functional Network Modules

Modular networks are analyzed to identify signature/function network modules from larger molecular networks. It involves topological analysis of the networks to find the maximally scoring regions in the network, followed by mapping the results to the true function categories like those present in the protein function annotation databases, e.g., the Gene Ontology (GO) and pathway databases. Network topology is the layout pattern of interconnections of the various elements like nodes and edges. The functional annotation of signature modules is based on their gene/protein enrichment analysis from GO and pathway database. The gene or protein enrichment analysis of networks is carried out by mapping network genes to known pathways and gene ontology terms to determine which pathways/GO terms are overrepresented in a given set of genes.

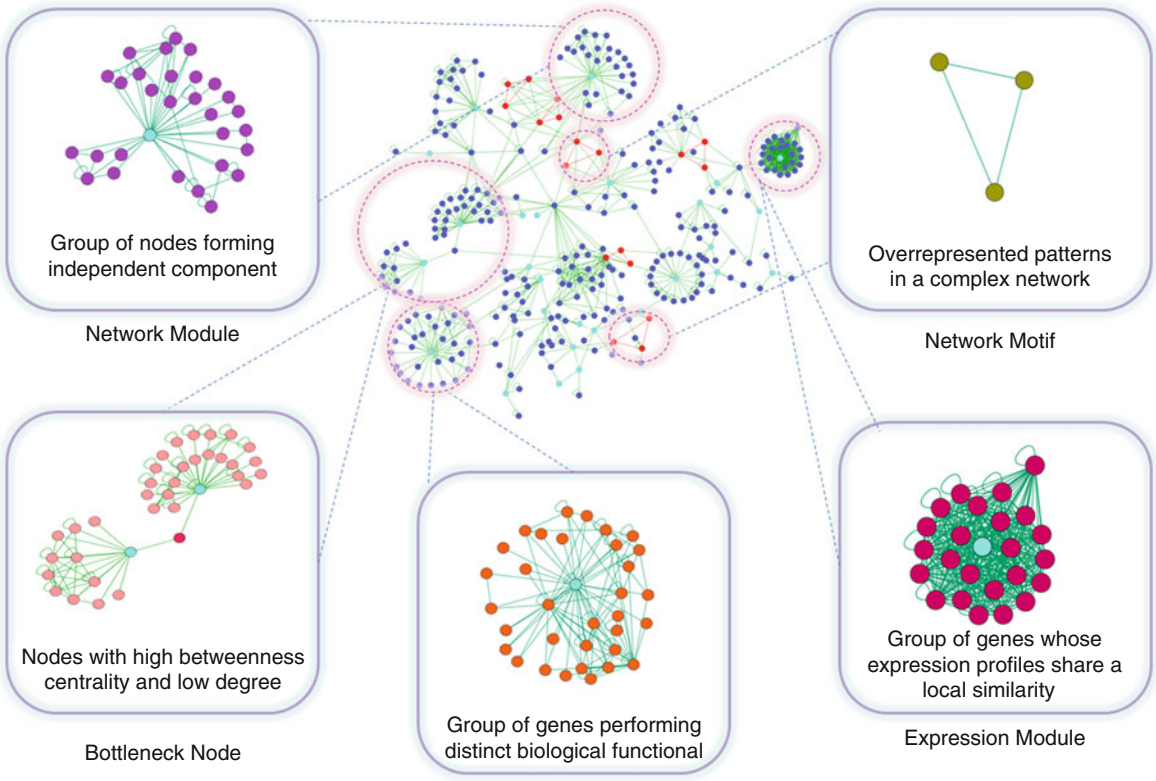
Functional modules are significantly enriched in known target genes and can be used as fingerprints to identify genes relevant to some specific biological functions or diseases or to discover new potential drug targets (Wong et al. 2008). The system level expression network is processed to identify upregulated/downregulated signature and functional expression network modules. For example, identification of downregulated synaptic vesicle module for brain disorders that includes Alzheimer's disease, bipolar disorder, Schizophrenia, and glioblastoma (Suthram et al. 2010) (refer Expression Network Module Box for details).

Expression Network Modules

Expression modules consist of clusters of genes whose expression profiles share a local similarity or coordinate expression. The local similarity/co-expression across genes is measured using statistical correlation parameters, i.e., Pearson's Correlation Coefficient. The genes with very high co-expression levels constitute a *cluster/module*, which usually indicates that they have a common biological function or share a common physiological condition. The condition-specific co-expression information provides clues to the dynamic features of these network modules. For example, dilated cardiomyopathy, which is a leading cause of heart failure, has been well studied using expression network modules. These networks have facilitated identification of putative biomarkers or therapeutic targets for heart failure and the underlying molecular mechanism of dilated cardiomyopathy (Lin et al. 2010)

The modular structure of complex biological networks also facilitates identification of biologically relevant gene hubs, bottleneck nodes, network motifs, and biomarker nodes, which are described below.

1. Gene Hubs
   In a molecular network, nodes (gene/protein) which are highly connected to other nodes are called gene hubs/network hubs. The gene hubs have a high degree of connectivity with their neighboring genes. The latest method for scoring a functional hub is based on the role the nodes play in providing connectivity among genes or proteins of interest relative to their role in the global network. They are centrally important for the cellular functions and tend to be essential and conserved in a network module. In the context of disease pathways, hubs may represent potential drug targets (Wuchty 2004; Levy and Siegal 2008). Scientists working on common diseases have reported several important drug targets by identifying important functional hubs across larger molecular networks derived from high-throughput expression as well as protein complex data.
2. Bottleneck Nodes
   Bottleneck nodes are those, which have a higher betweenness (i.e., "bottleneck-ness") and lower degree of connectivity with neighboring genes/proteins. Betweenness is one of the most important topological properties of a network. It measures the number of shortest paths (the shortest distance between two nodes) between nodes going through a certain node. Therefore, nodes with the highest betweenness control most of the information flow in the network, representing the critical points of the network (Fig. 1). They act as important links between modules in protein interaction networks, just as bridges connecting two important hubs/modules. Bottlenecks control the major information flow in a network and correspond to the dynamic components of the interaction network. They are observed to be significantly less well co-expressed with their neighbors. They have been found to be present in the yeast interactome in abundance (Yu et al. 2007).
3. Network Motifs
   Network motifs are patterns of a larger and more complex network. They are overrepresented node connectivity patterns and recur frequently in a given network than expected at random. They may represent autoregulation or feed-forward loops in a regulatory network and indicate functional and evolutionary constraints in a network (Lee et al. 2002). The transcription networks of well-studied microorganisms are made up of a small set of network motifs. These motifs are also found in protein modification network and interactions between neuronal cells and signaling networks. The specific ways (i.e., autoregulation or feed-forward loops) in which the network motifs are wired together describes the dynamics of each individual motif. These patterns exhibit a robust dynamical stability across a complex biological network (Fig. 1).
4. Biomarker Nodes
   Biomarkers are biomolecular signatures, which are detectable and measurable and can be used to study the diagnosis/prognosis of disease or therapy. Under normal conditions, they may be present at basal levels in the cell. However, if the amount of these molecules changes, they may indicate response to therapies, complications or diseases. Biomarker nodes can be identified by studying



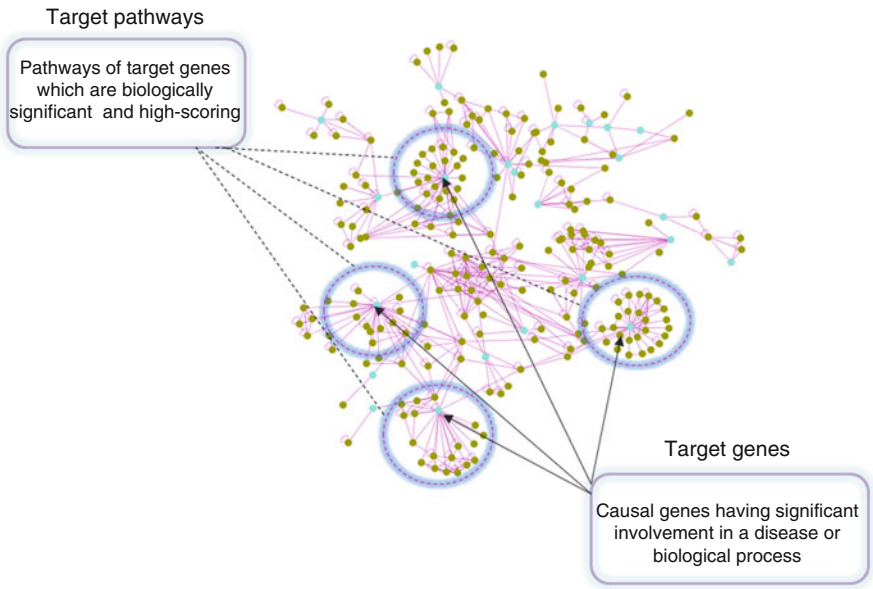
Functional/Signature Network Module for Target Pathway/Gene Discovery, Fig. 1 A complex biological network and functionally important network elements. Topological analysis of a system level biological network leads to the identification of network modules, network motifs, and bottleneck nodes. If the network is constructed from co- expression (expression similarity data), its statistical analysis leads to identification of expression module. The expression modules or network module can be functionally annotated to have a biological function, which is finally termed as a functional module

gene network models having disease-associated gene expression profiles and biofluid proteomes (Schiess et al. 2009). A gene, which is coordinately and consistently connected to critical disease causal candidate genes and pathways are termed "biomarker nodes." A highly connected biomarker node sometimes targets multiple related diseases; e.g., AZGP1 gene is a biomarker node for cardiac hypertrophy, idiopathic cardiomyopathy, and idiopathic thrombocytopenic purpura, (Dudley and Butte 2009).

Mapping Pathway/Gene Target From Networks

Protein–protein interaction and transcriptional regulatory networks can be used to identify putative biomarkers, target genes, target pathways, etc. The target pathways and genes are more reliably detected from expression profiling–based perturbation experiments. The statistically significant functional gene/protein networks can be further analyzed and annotated to map biologically important pathways and gene hubs as targets. This section is focused on identification of perturbation targets, target pathways/genes, using network biology approach.

1. Perturbation Targets from Network Biology
   The network biology approach can be used to compare gene expression profiles of drug-treated or diseased-cell population with those of the normal cells. This helps in identifying the target genes and pathways that are directly affected by the perturbation experiments, e.g., genes that are deregulated in some disease pathways or genes whose products can be used as potential targets of some drug (Mani et al. 2008, Karlebach and Shamir 2008). Topologically important networks are processed for gene ontology and pathway annotations to



Functional/Signature Network Module for Target Pathway/Gene Discovery, Fig. 2 Identification of biologically significant target pathways/genes from a complex biological network

define the effect of perturbation experiment and subsequently to identify the targets of molecular perturbation from global expression profile data.

2. Target Pathways
   Biologically significant pathways can be identified by performing functional annotation and pathway enrichment analysis of molecular networks. The high-scoring and conserved pathways can be used in disease-related studies, i.e., for understanding the pathogenesis of disease, identifying important deregulated metabolic and signaling networks and subsequently discovering potential drug targets for the disease (Fig. 2) (Zheng and Christina 2009; Zhou and Wong 2009). Further analysis of biologically significant pathways and their cross talk across networks leads to identification of the most critical pathways in a complex disease mechanism that could also serve as potential key target pathways. For example, human Toll-like receptor signaling network has been used to establish the important control points through pathway cross talk studies. The analysis identified potential candidates for inhibitory mediation of TLR signaling with respect to their specificity and potency (Li et al. 2008).
3. Target Genes
   A system level network biology approach can be used to identify novel target genes of interest which are found to have significant involvement in some disease, biological process, or an abnormality. This involves constructing networks and studying the kinetic interactions of a gene/protein with other genes/proteins in a network (Fig. 2) (Dezso et al. 2009; Gilchrist et al. 2006). The discovery of target pathways and genes is interrelated. A researcher can always narrow down the list of target genes from the list of target pathways discovered from modular networks.

References

Dezso Z, Nikolsky Y, Nikolskaya T, Miller J, Cherba D, Webb C, Bugrim A (2009) Identifying disease-specific genes based on their topological significance in protein networks. BMC Syst Biol 3:36

Dudley JT, Butte AJ (2009) Identification of discriminating biomarkers for human disease using integrative network biology. Pac Symp Biocomput 27–38

Erten S, Li X, Bebek G, Li J, Koyutürk M (2009) Phylogenetic analysis of modularity in protein interaction networks. BMC Bioinformatics 10:333

Gilchrist M, Thorsson V, Li B, Rust AG, Korb M, Roach JC, Kennedy K, Hai T, Bolouri H, Aderem A (2006) Systems biology approaches identify ATF3 as a negative regulator of toll-like receptor 4. Nature 441(7090):173–178

Karlebach G, Shamir R (2008) Modelling and analysis of gene regulatory networks. Nature Reviews Molecular Cell Biology 9:770–780

Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne JB, Volkert TL,

Fraenkel E, Gifford DK, Young RA (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. Science 298(5594):799–804
Levy SF, Siegal ML (2008) Network hubs buffer environmental variation in *Saccharomyces cerevisiae*. PLoS Biol 6(11):e264
Lin CC, Hsiang JT, Wu CY, Oyang YJ, Juan HF, Huang HC (2010) Dynamic functional modules in co-expressed protein interaction networks of dilated cardiomyopathy. BMC Syst Biol 4:138
Mani KM, Lefebvre C, Wang K, Lim WK, Basso K, Dalla-Favera R, Califano A (2008) A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas. Mol Syst Biol 4:169
Schiess R, Wollscheid B, Aebersold R (2009) Targeted proteomic strategy for clinical biomarker discovery. Mol Oncol 3(1):33–44
Suthram S, Dudley JT, Chiang AP, Chen R, Hastie TJ, Butte AJ (2010) Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets. PLoS Comput Biol 6(2):e1000662
Wong DJ, Liu H, Ridky TW, Cassarino D, Segal E, Chang HY (2008) Module map of stem cell genes guides creation of epithelial cancer stem cells. Cell Stem Cell 2(4):333–344
Wuchty S (2004) Evolution and topology in the yeast protein interaction network. Genome Res 14(7):1310–1314
Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M (2007) The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. PLoS Comput Biol 3(4):e59
Zheng L, Christina C (2009) Systems biology for identifying liver toxicity pathways. BMC Proc 3(Suppl 2):S2
Zhou X, Wong ST (2009) Computational systems bioinformatics and bioimaging for pathway analysis and drug screening. Proc IEEE Inst Electr Electron Eng 96(8):1310–1331

Functionality Type

▶ FunctionalityType

FunctionalityType

Marie-Paule Lefranc
Laboratoire d'ImmunoGénétique Moléculaire, Institut de Génétique Humaine UPR 1142, Université Montpellier 2, Montpellier, France

Synonyms

Functionality type

Definition

FunctionalityType is a concept of identification (generated from the ▶ IDENTIFICATION Axiom) of ▶ IMGT-ONTOLOGY, the global reference in ▶ immunogenetics and ▶ immunoinformatics (Duroux et al. 2008), built by IMGT®, the international ImMunoGeneTics information system® (http://www.imgt.org) (▶ IMGT® Information System), that allows to identify, whatever the molecule type (▶ MoleculeType) (gDNA, cDNA, mRNA, or protein), the type of functionality of a ▶ Molecule_EntityType leafconcept (Giudicelli and Lefranc 1999; Lefranc et al. 2004; Duroux et al. 2008).

The "FunctionalityType" concept comprises five leafconcepts, divided into two categories, according to the configuration type (▶ Configuration Type) of the Molecule_EntityType leafconcept.

Three leafconcepts, ▶ *functional*, *ORF* (open reading frame), and ▶ *pseudogene*, identify the functionality of Molecule_EntityType leafconcepts in *undefined* configuration (conventional genes (▶ Conventional Gene) and immunoglobulin (IG) and T cell receptor (TR) constant (C) genes (▶ Constant (C) Gene) or in *germline* configuration (IG and TR variable (V), diversity (D) and joining (J) (▶ Variable (V) Gene, ▶ Diversity (D) Gene, ▶ Joining (J) Gene) genes *before* DNA rearrangements).

Two leafconcepts, ▶ *productive* and *unproductive*, identify the functionality of Molecule_EntityType leafconcepts in *rearranged* or *partially-rearranged* configuration (IG and TR entities *after* DNA rearrangements, and by extension fusion entities resulting from translocations, and hybrid entities obtained by biotechnology molecular engineering).

Cross-References

▶ Configuration Type
▶ Constant (C) Gene
▶ Conventional Gene
▶ Diversity (D) Gene
▶ Functional
▶ IMGT-ONTOLOGY
▶ IMGT-ONTOLOGY, IDENTIFICATION Axiom
▶ IMGT-ONTOLOGY, Leafconcept
▶ IMGT-ONTOLOGY, Unproductive
▶ IMGT® Information System

- ▶ Immunogenetics
- ▶ Immunoinformatics
- ▶ Joining (J) Gene
- ▶ Molecule Entity Type
- ▶ MoleculeType
- ▶ Open Reading Frame (ORF)
- ▶ Productive
- ▶ Pseudogene
- ▶ Variable (V) Gene

References

Duroux P, Kaas Q, Brochet X, Lane J, Ginestoux C, Lefranc M-P, Giudicelli V (2008) IMGT-Kaleidoscope, the formal IMGT-ONTOLOGY paradigm. Biochimie 90:570–583

Giudicelli V, Lefranc M-P (1999) Ontology for immunogenetics: the IMGT-ONTOLOGY. Bioinformatics 12:1047–1054

Lefranc M-P, Giudicelli V, Ginestoux C, Bosc N, Folch G, Guiraudou D, Jabado-Michaloud J, Magris S, Scaviner D, Thouvenin V, Combres K, Girod D, Jeanjean S, Protat C, Yousfi Monod M, Duprat E, Kaas Q, Pommié C, Chaume D, Lefranc G (2004) IMGT-ONTOLOGY for immunogenetics and immunoinformatics. In Silico Biol 4:17–29

Functional-Structural Plant Modeling

Gerhard Buck-Sorlin
Institut de Recherche en Horticulture et Semences, UMR 1345 (INRA/Agrocampus-ouest/Université d'Angers) Agrocampus Ouest, Angers, France

Synonyms

Virtual plant modeling

Definition

Functional-structural plant modeling (FSPM) refers to a paradigm for the description of a plant by creating a (usually object-oriented) computer model of its structure and selected physiological and physical processes, at different hierarchical levels: organ, plant individual, canopy (a stand of plants), and in which the processes are modulated by the local environment. Structure comprises the explicit topology (connection between organs) and geometry (orientation, inclination, and shape) of the organs and the plant. At the individual level, this is also referred to as plant architecture. An FSPM may consider a change in organ and plant structure in time, thereby simulating growth, extension, and branching processes of a given plant. This type of FSPM is referred to as dynamic. A static FSPM, on the other hand, only considers an unchanging structure, which is used as a model input in order to explain spatial heterogeneity in physiological processes.

Physiological and physical processes considered usually comprise essential, basic, and characteristic functions, such as ▶ photosynthesis, growth (biomass accumulation and organ extension in length and diameter), and branching, with physiological processes depending on environmental factors such as temperature, radiation, CO_2 content of the air, and relative humidity. These processes (or functions) are usually implemented at the level of each organ and then require local environmental parameters. The explicit consideration of the geometry and topology of structural elements at the organ level distinguishes FSPM from its predecessor, ▶ process-based models.

In an FSPM, a feedback relation exists between the structure and certain functions: A given structure (static or resulting from application of rules) can modulate the local output of processes (e.g., self-shading of leaves diminishing locally intercepted light for ▶ photosynthesis); on the other hand, all structures are built and maintained by processes (e.g., growth of new biomass fed by assimilates resulting from photosynthesis).

Characteristics

According to the FSPM paradigm, a plant responds to its environment by adaptive modification of processes constituting its physiology (e.g., ▶ photosynthesis) and plant architecture (e.g., bud break or dormancy, growth, development, morphogenesis), thereby explicitly addressing the feedbacks between structure and function. In addition, such feedbacks can be implemented and verified at different levels, e.g., locally at the organ scale and globally at the plant or canopy scale.

A typical reaction of the plant to a change in environment or to an intervention (cutting, bending, etc.) by management or in the course of an experiment will thus be the relatively rapid readjustment of

physiological functions and a somewhat slower reaction in terms of structural adaptation by growth of new structures or active shedding of existing structures (Vos et al. 2010).

Usually, the dynamics in an FSPM is simulated using rules (for growth, extension, and branching) always starting from a growing tip (meristem) which produces ▶ phytomers (consisting of a leaf, a node, an axillary bud, and an internode). The axillary bud can itself produce a phytomer (with a meristem on top) to form a new shoot; this is referred to as branching. For each of the organs of a phytomer, specific dynamics for extension or growth in biomass apply, which can be described using nonlinear functions of time. If such a dynamics follows a logistic function, then it is characterized by three parameters: time of onset of process, time of maximum rate, and maximum dimension. Thus, using a rule-based notation (e.g., a ▶ Lindenmayer system) to describe these processes, two types of rules are usually sufficient:

(a) For the formation of a new ▶ phytomer from a (terminal or axillary) meristem:

$$M \overset{conditions}{\rightarrow} IN[L][M]M$$

where M symbolizes the meristem and I, N, L the internode, node, and leaf, respectively; $\rightarrow$ designates a replacement of the left-hand side symbol (meristem M) by the string of symbols on the right hand-side, i.e., the organs constituting a phytomer, plus another meristem at the end of the string, thereby ensuring that the rule can be applied over and over again (as long as certain conditions are fulfilled); the square opening $[$ and closing $]$ brackets represent the beginning and end of a branch.

(b) For the dynamics of growth or extension:

$$O(l) \overset{conditions}{\Rightarrow} O(l + dl)$$

where O is one of $\{I, N, L, \ldots\}$ (can also be a fruit or flower), l is its dimension (e.g., length), and $\Rightarrow$ denotes a rule in which the dimension of an already existing organ (resulting from the application of rule (a)) is updated by adding a fraction dl to the dimension l (specified on the right-hand side of the rule). Expressed as a rate, dl/dt would be the derivative of the function employed to describe the growth or extension dynamics of the organ in question.

History

The concept of Functional-Structural Plant Modeling arose in the 1990s from the desire to link existing ▶ process-based models of crops with spatially explicit representations of plants (usually represented in a rule-based manner as ▶ Lindenmayer systems) in order to consider, in a mathematical model, interactions between plant structure and functioning (Sievänen et al. 1997). Early representatives of FSPMs were restricted to one aspect of plant functioning and showed how the selected function was affected by morphology (e.g., light interception, assimilate allocation, xylem sap flow). In these models, information was frequently unidirectional (from structure to function). Furthermore, due to the lack of conventions among modelers, these models were very specific and difficult to reuse or to transfer to other plant systems.

During a second wave, FSPMs were designed to be more modular, exhibiting submodels contributed by several authors and considering more than one aspect of functionality at the same time. Examples for this type of model are LIGNUM (Perttunen et al. 2001) or Greenlab (Guo et al. 2006). New features included in these models were, e.g., bidirectional information flow between structure and function (Hemmerling et al. 2008); processes taking place at more than one hierarchical level, e.g., phytohormone biosynthesis and transport (Buck-Sorlin et al. 2005); or the fluxes of phloem and xylem within the tree (Lacointe and Minchin 2008). Recent simulation studies addressed also aspects of agricultural pest management, the impact of biomechanics on tree architectural development, and more refined models of light distribution in canopies, including the effects of light quality on growth regulation (Vos et al. 2010). FSPMs have been created for crop plants such as the cereals maize, wheat, rice, and barley, but also for trees (Vos et al. 2010). In most of these FSPMs, the structure was modeled dynamically, from an initial meristem, but there exist also static FSPMs in which the structure is fixed (Vos et al. 2010).

Elements for Conception and Construction of a Functional-Structural Plant Model

The establishment of an FSPM starts with the observation of plant morphology, i.e., topology (arrangement of ▶ phytomers in shoots, with lateral shoots up to a maximum branching order) and geometry, including the biometry and morphology of the different plant organs at different positions within

the plant. As with every model, it is wise to determine the boundaries, the scale, and the elements of the system to be modeled: Usually, the scope of an FSPM is not more than a dozen plant individuals; the scales are that of the (sub)organ, shoot, individual and canopy level, and the elements considered are the organs constituting the phytomer, plus collective entities (the "plant" as the sum of all shoots, the shoot as the sum of its phytomers, etc.). Sometimes, other elements characteristic of a crop production system are simulated, such as a greenhouse with lamps or a patch of soil. Key topological parameters for an FSPM are maximum branching order, maximum rank of a phytomer within a shoot of a given branching order, or rank of lowest flower. Using rank and order as parameters for M in equation (a), those measured maximum values are used to restrict the production of phytomers by the meristems (see phytomer for a definition of rank and order). For each considered plant element, a number of (physical or) physiological processes (growth, extension, respiration, ▶ photosynthesis, etc.) are defined. This will yield an object-oriented model prototype.

In order to parameterize the dimensions and orientations of organs (e.g., growth and extension models (equation (b)), a database with lengths, diameters, areas, and angles (phyllotaxis, divergence) of representative organs needs to be established. Statistical analysis (regression) ideally returns a relationship between an organ dimension and its topology (rank, order). The same relation can be established between a physiological function (e.g., leaf photosynthesis) and topology. Methods used to establish a biometric database are manual measurement using a ruler and caliper, digitizing (using a stylus or laser scanning), and image analysis. After calibration of the FSPM, it is tested using a number of scenarios and then validated in the usual way by reparameterization with a calibration data set (measured at organ scale) and comparison with output variables, usually measured at canopy or plant scale.

FSPM Algorithms, Languages, and Software

Languages and formalisms employed for FSPM are usually following the rule-based or object-oriented paradigm, often in combination with the procedural paradigm. The most widespread rule-based formalisms are ▶ Lindenmayer systems and ▶ relational growth grammars. Dedicated languages for FSPM are L + C (Prusinkiewicz et al. 2007) and XL (Hemmerling et al. 2008); in addition, some FSPMs have been devised using a standard programming language (e.g., C, C++, Java, Simula). Platforms used for FSPM include L-Studio, GroIMP, and OpenAlea.

Cross-References

- ▶ Biological System Model
- ▶ Calibration
- ▶ Complex System
- ▶ Data Integration
- ▶ Ecological Modeling
- ▶ Feedback Regulation
- ▶ Function, Biological
- ▶ Functional
- ▶ Graphical Model
- ▶ Hierarchical Structure
- ▶ Interdisciplinarity
- ▶ Markov Chain
- ▶ Markov Process
- ▶ Model Validation
- ▶ Modularity
- ▶ Monte Carlo Simulation
- ▶ Ordinary Differential Equation (ODE)
- ▶ Paradigm
- ▶ Phenomenological vs Physiological Modeling
- ▶ Plant Systems Biology
- ▶ Rule
- ▶ Rule-Based Methods
- ▶ Spatiotemporal Pattern Formation
- ▶ Time-Course
- ▶ Topology and Toponomics

References

Buck-Sorlin G, Kniemeyer O, Kurth W (2005) Barley morphology, genetics and hormonal regulation of internode elongation modelled by a relational growth grammar. New Phytol 166:859–867

Guo Y, Ma Y, Zhan Z, Li B, Dingkuhn M, Luquet D, de Reffye P (2006) Parameter optimization and field validation of the functional-structural model Greenlab for maize. Ann Bot 97:217–230

Hemmerling R, Kniemeyer O, Lanwert D, Kurth W, Buck-Sorlin G (2008) The rule-based language XL and the modelling environment GroIMP illustrated with simulated tree competition. Funct Plant Biol 35:739–750

Lacointe A, Minchin PEH (2008) Modelling phloem and xylem transport within a complex architecture. Funct Plant Biol 35:772–780

Perttunen J, Nikinmaa E, Lechowicz MJ, Sievänen R, Messier C (2001) Application of the functional-structural tree model LIGNUM to sugar maple saplings (*Acer saccharum* Marsh) growing in forest gaps. Ann Bot 88:471–481
Prusinkiewicz P, Karwowski R, Lane B (2007) The L + C plant modeling language. In: Vos J, Marcelis LFM, de Visser PHB, Struik PC, Evers JB (eds) Functional-structural plant modelling in crop production. Springer, Dordrecht, pp 27–42
Sievänen R, Mäkelä A, Nikinmaa E, Korpilahti E (1997) Special issue on functional-structural tree models, preface. Silva Fennica 31:237–238
Vos J, Evers JB, Buck-Sorlin G, Andrieu B, Chelle M, de Visser PHB (2010) Functional–structural plant modelling: a new versatile tool in crop science. J Exp Bot 61:2101–2115

Functor

Baltasar Trancón y Widemann
Ecological Modelling, University of Bayreuth, Bayreuth, Germany

Synonyms

Category homomorphism; Category map

Definition

A functor is a mapping that respects the structure of a *category* (Lawvere and Schanuel 1997; Adámek et al. 2004). A functor F from category $\mathcal{C}$ to category $\mathcal{D}$ consists of the following:

- A mapping that associates with every object $X \in \mathcal{C}$ an object $F(X) \in \mathcal{D}$
- A mapping that associates with every arrow $f \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$ an arrow $F(f) \in \mathrm{Hom}_{\mathcal{D}}(F(X), F(Y))$
  Such that
- Identities are mapped to identities

$$F(\mathrm{id}_X) = \mathrm{id}_{F(X)} \tag{1}$$

- Compositions are mapped to compositions

$$F(g \circ f) = F(g) \circ F(f) \tag{2}$$

where the left composition is in $\mathcal{C}$ and the right composition is in $\mathcal{D}$.

A consequence of the properties of a functor $F : \mathcal{C} \to \mathcal{D}$ is that it can be applied to all parts of a diagram in category $\mathcal{C}$, yielding a valid diagram in category $\mathcal{D}$. Furthermore, the resulting diagram will commute if the original commutes.

Many mathematical constructions that relate entities of different nature can be understood as functors between suitable categories. Typical examples are (tensor) products or the mapping from Lie groups to Lie algebras. Functors also play a crucial role in universal algebra and coalgebra, where they specify expression languages for the construction and deconstruction of structured elements, respectively (Jacobs and Rutten 1997).

References

Adámek J, Herrlich H, Strecker GE (2004) Abstract and concrete categories. Wiley, New York. http://katmat.math.uni-bremen.de/acc/acc.htm, online edition
Jacobs B, Rutten J (1997) A tutorial on (co)algebras and (co) induction. EATCS Bull 62:222–259
Lawvere W, Schanuel S (1997) Conceptual mathematics: a first introduction to categories. Cambridge University Press, Cambridge/New York

Fuzzy Logic

Eyke Hüllermeier, Thomas Fober and Marco Mernberger
Philipps-Universität Marburg, Marburg, Germany

Synonyms

Fuzzy set theory

Definition

The term "fuzzy logic" is used with different meanings in the literature. In a narrow sense, it refers to a branch of mathematical logic, where it is studied as a special type of multivalued logic, that is, a logic with more than two truth degrees (Hajek 1998). In a wider (and more common) sense, fuzzy logic is used as an umbrella term for a collection of methods, tools, and

techniques for constructing intelligent systems that, by virtue of the very idea of *partiality of truth*, are capable of handling, processing and exploiting uncertain, imprecise, and incomplete information. These methods build on the key concept of a fuzzy set, as introduced by the founder of fuzzy logic, Lotfi A. Zadeh, in his seminal paper (Zadeh 1965). Fuzzy sets formalize the idea of graded class membership, according to which an element can partially belong to a set. In conjunction with generalized logical (set-theoretical) operators and derived notions like a fuzzy relation, the concept of a fuzzy set can be developed into a generalized set theory, which in turn provides the basis for generalizing theories in different branches of (pure and applied) mathematics as well as fuzzy set-based approaches to intelligent systems design, encompassing methods for information processing, decision making, optimization, and data analysis.

Characteristics

The notion of *truth* is commonly considered as a *bivalent* concept: Logically, a proposition is either true or false, but nothing in-between. This conception, which pervades modern science and thinking, has a long-standing tradition in Western philosophy, and manifests itself in standard mathematical theories, notably logic and set theory. And admittedly, formal systems based on bivalent logic (including theories of uncertainty based on such systems, like probability theory) have proved extremely useful in the scientific terrain, where they paved the way for the amazing success of the exact and engineering sciences in the last century.

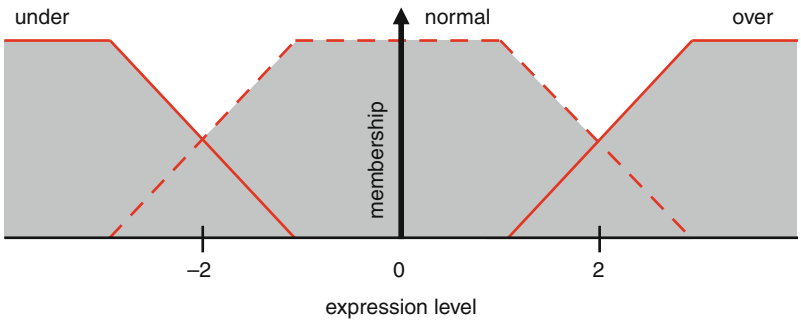
In many other less exact fields of science, however, ranging from the biological and life sciences over legal practice to the modeling of cognitive processes and human intelligence, the bivalence of truth can be called into question. In fact, it was already noticed by Bertrand Russell in 1923 that "All traditional logic habitually assumes that precise symbols are being employed. It is therefore not applicable to this terrestrial life, but only to an imagined celestial existence" (Russell 1923). Roughly speaking, this is because of the vagueness and ambivalence of the concepts dealt with in these fields: for the intension of these concepts, there is rarely a precise extension (in the sense of a set of real objects belonging to that concept) in the real word. For example, what is a "short DNA molecule"? Biologists have a vague though sufficiently clear idea of this concept, without using a precise definition in terms of an exact upper bound on the number of base pairs (bp) or the length in μm. Given such bounds, a proposition like "DNA molecule XYZ is small" would be either true or false; likewise, the truth degree of a natural language proposition like "Einstein was born around noon" could be determined in a unique way, given a precise definition of the meaning of "around noon" in terms of a time interval (e.g., 12 o'clock $\pm$ 15 min) and the possibility to pinpoint the exact time of a birth.

Needless to say, this way of adapting human thinking and language to conventional logic and set theory would be neither desirable nor useful. Instead, fuzzy logic offers an approximation in the other direction: logic and set theory are generalized, so as so enable a more faithful mathematical modeling of human conception. The core idea in this regard is the notion of a fuzzy set, which allows for partial membership and soft class boundaries (Pedrycz and Gomide 2007).

Fuzzy Sets

A fuzzy subset A of a reference set $\mathbb{X}$ is identified by a so-called *membership function*, often denoted $\mu_A(\cdot)$, which is a generalization of the characteristic function $\mathbb{I}_A(\cdot)$ of an ordinary set $A \subset \mathbb{X}$. For each element $x \in \mathbb{X}$, this function specifies the degree of membership of x in the fuzzy set; it can be interpreted as the truth degree of the proposition that $x \in A$. Usually, membership degrees $\mu_A(x)$ are taken from the unit interval [0,1], i.e., a membership function is an $X \rightarrow [0, 1]$ mapping. In principle, however, more general membership scales (such as ordinal scales or complete lattices) can be used. We denote by $\mathbb{F}(\mathbb{X})$ the set of all fuzzy subsets of $\mathbb{X}$.

Fuzzy sets are often used for discretizing numerical attributes in a "soft" manner, taking advantage of their ability to model "non-sharp" boundaries between classes. Thus, they serve as an interface between the original numerical scale and a symbolic scale comprised of the (natural language) terms associated with the fuzzy sets. For example, in ▶ gene expression analysis, one typically distinguishes between normally expressed, under-expressed, and over-expressed genes. This classification is made on the basis of the expression level of the gene (a normalized numerical value), by using



Fuzzy Logic, Fig. 1 Fuzzy partition of the gene expression level with a "smooth" transition between under-expression, normal expression, and over-expression; membership functions are shown as red lines

corresponding thresholds. For example, a gene is often called over-expressed if its expression level is at least twofold increased. Needless to say, a precise threshold of that kind is arbitrary to some extent and implies an unnatural sudden jump from completely over-expressed to not at all over-expressed. Figure 1 shows a fuzzy partition of the expression level with a "smooth" (and arguably less arbitrary) transition between under-, normal, and over-expression, formalized in terms of fuzzy sets with trapezoidal membership functions. For instance, according to this formalization, a gene with an expression level of at least 3 is definitely considered over-expressed, below 1 it is definitely not over-expressed, but in-between, it is considered over-expressed to a certain degree.

Generalized Logical Connectives

To operate with fuzzy sets in a formal way, fuzzy set theory offers generalized set-theoretical resp. logical connectives (like in the classical case, there is a close correspondence between set theory and logic). Especially, important in this regard is a class of operators called *triangular norms* or *t-norms* for short (Klement et al. 2002).

- A t-norm $\otimes$ is a $[0,1] \times [0,1] \to [0,1]$ mapping which is associative, commutative, monotone increasing (in both arguments) and which satisfies the boundary conditions $\alpha \otimes 0 = 0$ and $\alpha \otimes 1 = \alpha$ for all $0 \leq \alpha \leq 1$. Well-known examples of t-norms include the minimum $(\alpha, \beta) \mapsto \min(\alpha, \beta)$, the product $(\alpha, \beta) \mapsto \alpha\beta$, and the Łukasiewicz t-norm $(\alpha, \beta) \mapsto \max(\alpha + \beta - 1, 0)$.
  A t-norm naturally qualifies as a generalized logical *conjunction*. Moreover, it can be used to define the *intersection* of fuzzy subsets $A, B \in \mathbb{F}(\mathbb{X})$ as follows: $\mu_{A \cap B}(x) = \mu_A(x) \otimes \mu_B(x)$ for all $x \in \mathbb{X}$. In a quite similar way, the *Cartesian product* of fuzzy sets $A \in \mathbb{F}(\mathbb{X})$ and $B \in \mathbb{F}(\mathbb{Y})$ can be defined: $\mu_{A \times B}(x, y) = \mu_A(x) \otimes \mu_B(y)$ for all $(x, y) \in \mathbb{X} \times \mathbb{Y}$.
- The logical *disjunction* can be generalized analogously, namely by means of a t-conorm $\oplus$. If $\otimes$ is a t-norm, then $\oplus$ defined by $\alpha \oplus \beta = 1 - (1 - \alpha) \otimes (1 - \beta)$ is a t-conorm. Well-known examples of t-conorms include the maximum $(\alpha, \beta) \mapsto \max(\alpha, \beta)$, the algebraic sum $(\alpha, \beta) \mapsto \alpha + \beta - \alpha\beta$, and the Łukasiewicz t-conorm $(\alpha, \beta) \mapsto \min(\alpha + \beta, 1)$. A t-conorm can be used for defining the union of fuzzy sets: $\mu_{A \cup B}(x) = \mu_A(x) \oplus \mu_B(x)$ for all $x \in \mathbb{X}$.
- A generalized *implication* $\rightsquigarrow$ is a $[0,1] \times [0,1] \to [0,1]$ mapping which is monotone decreasing in the first and monotone increasing in the second argument, and which satisfies the boundary conditions $\alpha \rightsquigarrow 1 = 1$, $0 \rightsquigarrow \beta = 1$, $1 \rightsquigarrow \beta = \beta$. (Apart from that, additional properties are sometimes required.) Implication operators of that kind, such as the Łukasiewicz implication $(\alpha, \beta) \mapsto \min(1 - \alpha + \beta, 1)$, are especially important in connection with the modeling of fuzzy rules.

The Extension Principle

Apart from basic logical connectives, fuzzy logic offers a number of tools for generalizing and "fuzzifying" existing theories and methods. One of these tools is the so-called *extension principle*, which allows for extending a mapping $f : \mathbb{X}^n \to \mathbb{Y}$ to a fuzzy mapping $F : \mathbb{F}(\mathbb{X})^n \to \mathbb{F}(\mathbb{Y})$ in a generic way. F accepts fuzzy subsets of $\mathbb{X}$ as input and, correspondingly, produces fuzzy subsets of $\mathbb{Y}$ as output. More specifically, for fuzzy subsets $A_1, \ldots, A_n$ as input, the output $B = F(A_1, \ldots, A_n)$ is a fuzzy subset of $\mathbb{Y}$ with membership function

$$\mu_B(y) = \sup_{y=f(x_1,\ldots,x_n)} \min\{\mu_{A_1}(x_1), \ldots, \mu_{A_n}(x_n)\}$$

The supremum operator in this expression is playing the role of a generalized existential quantifier. Thus, $\mu_B(y)$ can be interpreted as the truth degree of the following proposition: $\exists(x_1, \ldots, x_n) \in \mathbb{X}^n : (\forall i \in \{1, \ldots, n\} : (x_i \in A_i)) \wedge (y = f(x))$. Or, stated differently, y belongs to the fuzzy output $F(A)$ insofar as there exist $x_1, \ldots, x_n$ that belong, respectively, to $A_1, \ldots, A_n$ and are mapped to y.

Fuzzy Inference

Fuzzy sets can be used to formalize vague and imprecise knowledge. For example, a basic proposition of the form "X is A", where X is a variable with domain $\mathbb{X}$ and A a fuzzy subset of $\mathbb{X}$, can be understood as a flexible ▶ constraint on the value of X: those values $x \in \mathbb{X}$ with $\mu_A(x) = 0$ are excluded, while all other values x are considered as possible to some degree, namely to the degree $\mu_A(x)$; in particular, those x with $\mu_A(x) = 1$ are declared as fully plausible.

In principle, fuzzy ▶ inference can then be realized by combining and propagating constraints of that type, using suitable logical operators as well as projection and extension operators for fuzzy relations. Fuzzy rule-based ▶ inference can be seen as an important special case. Here, the idea is to express knowledge about the (functional) dependency between attributes in the form of a set of IF–THEN rules:

$$R_i : \text{IF}(X_1 \text{ is } A_1^{(i)}) \text{ AND } (X_1 \text{ is } A_2^{(i)}) \text{ AND} \ldots \text{AND } (X_m \text{ is } A_m^{(i)}) \text{ THEN } (Y \text{ is } B^{(i)})$$

As an example, consider a rule like "If gene X is over-expressed and gene Y is under-expressed, then gene Z is over-expressed". A rule of that kind can be seen as a soft ▶ constraint that partly excludes some value combinations $(x_1, x_2, \ldots, x_m, y) \in \mathbb{X}_1 \times \mathbb{X}_2 \times \ldots \times \mathbb{X}_m \times \mathbb{Y}$. Given a fuzzy specification of the input attributes X_j in terms of fuzzy subsets $A_j (j = 1, \ldots, m)$, the output produced by the fuzzy rule system consisting of n rules $R_i (i = 1, \ldots, n)$ is a fuzzy subset B of $\mathbb{Y}$ such that

$$\mu_B(y) = \sup_{x \in \mathbb{X}_1 \times \ldots \times \mathbb{X}_m} \min\left(\mu_A(x), \min_{i=1,\ldots,n} (\mu_{A^{(i)}}(x) \rightsquigarrow \mu_{B^{(i)}}(y))\right),$$

where $\mu_{A^{(i)}}(x) = \mu_{A_1^{(i)}}(x_1) \otimes \mu_{A_2^{(i)}}(x_2) \otimes \ldots \otimes \mu_{A_m^{(i)}}(x_m)$. for a t-norm $\otimes$, and $\rightsquigarrow$ is a fuzzy implication.

Applications

Fuzzy methods have not only been developed for ▶ knowledge representation and information processing, but also in many other branches of applied mathematics, including optimization, decision making, statistics, and data analysis. Moreover, a plethora of concrete applications have been developed in different fields, ranging from fuzzy control systems over flexible querying in databases to medical expert systems.

In bioinformatics and systems biology, fuzzy methods have been used with the aim to capture the intrinsic fuzziness of terms, concepts, and relationships in the biological sciences, and advantages of fuzzy sets and fuzzy logic for analyzing biological data and modeling biological systems are becoming more and more recognized. As a concrete example, we mention the use of fuzzy ▶ clustering as a data analysis tool for inferring homogeneous groups of related biological entities (like genes) in a data-driven way; obliging to biological reality, these groups may have soft boundaries and are allowed to overlap (Dembélé and Kastner 2003; Möller-Leveta et al. 2005). Apart from this, fuzzy methods have been applied to many other problems in this field; for an overview of recent advances, see (Xu et al. 2008; Jin and Wang 2009).

Cross-References

- ▶ Clustering
- ▶ Constraint
- ▶ Gene Expression
- ▶ Inference
- ▶ Knowledge Representation

References

Dembélé D, Kastner P (2003) Fuzzy C-means method for clustering microarray data. Bioinformatics 19(8):973–980

Hajek P (1998) Metamathematics of fuzzy logic. Kluwer, Dordrecht

Jin Y, Wang L (eds) (2009) Fuzzy systems in bioinformatics and computational biology. Springer, Berlin
Klement EP, Mesiar R, Pap E (2002) Triangular norms. Kluwer, Dordrecht
Möller-Leveta CS, Klawonn F, Choc KH, Yina H, Wolkenhauer O (2005) Clustering of unevenly sampled gene expression time-series data. Fuzzy Set Syst 152(1):49–66
Pedrycz W, Gomide F (2007) Fuzzy systems engineering: toward human-centric computing. Wiley, New York
Russell B (1923) Vagueness. Aust J Psych Phil 1:84–92
Xu D, Keller JM, Popescu M (2008) Applications of fuzzy logic in bioinformatics. World Scientific, Singapore
Zadeh LA (1965) Fuzzy sets. Inform Control 8(3):338–353

Fuzzy Set Theory

► Fuzzy Logic