E

Early Embryonic Oscillator

► Cell Cycle of Early Frog Embryos

EBI Genome Resources

Akos Dobay[1] and Maria Pamela Dobay[2]
[1]Institute of Evolutionary Biology and Environmental Studies (IEU), University of Zurich, Zurich, Switzerland
[2]Department of Physics, Ludwig-Maximilians University, Munich, Germany

Synonyms

EMBL genome database

Definition

The European Bioinformatics Institute (EBI; http://www.ebi.ac.uk) is the main European repository of nucleotide sequence data. EBI exchanges the collected data with the DNA Data Bank of Japan (DDBJ; http://www.ddbj.nig.ac.jp) and the National Center for Biotechnology Information Database (NCBI; http://www.ncbi.nlm.nih.gov) on a daily basis. The three databases are part of the International Nucleotide Sequence Database (INSD; http://www.insdc.org) consortium, whose function is to ensure the integrity of the shared information. Apart from the genome resources, EBI also has resources for protein sequences and structures, as well as access to sequences from patents.

Characteristics

The European Bioinformatics Institute, an outstation of the European Molecular Biology Laboratory (EMBL; http://www.embl.org), was created in 1992 to respond to the requirements for organizing data from sequencing efforts, notably those originating from the genome projects (Brooksbank et al. 2003). Located on the Welcome Trust Genome Campus in Hinxton near Cambridge (UK), the EBI web portal constitutes the European node of the INSD consortium (Brooksbank et al. 2003). Other members are the DDBJ and the NCBI. Information on new entries and updates are exchanged between these databases on a daily basis. The EBI provides access to more than 60 databases containing high-quality annotated information and raw data that range from nucleotide sequences, whole-genome projects, to protein sequences, protein structures, and protein interactions. The data can be browsed and analyzed using various tools hosted on the site, and that are also available for download. The EBI constitutes the primary repository in Europe for sequence data.

DNA and RNA Sequence Databases Available from EBI

The main nucleotide sequence database is the European Nucleotide Archive (ENA). ENA is built from several databases: the EMBL Nucleotide Sequence Database (Kulikova et al. 2007) or EMBL-Bank created in early 1980; the Sequence Read Archive (SRA), a repository set up in 2009 for raw data obtained from next-generation sequencing platforms; and the European Trace Archive (ETA), composed of raw data from electrophoresis-based sequencing machines. The EBI also provides access to whole genome databases across the

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

phylum Chordata through the project Ensembl (http://www.ensembl.org). Ensembl was launched in 1999 as a joint collaboration between EBI and the Welcome Trust Sanger Institute (Flicek et al. 2011). Ensembl Genomes (http://www.ensemblgenomes.org) is an extension of the Ensembl project to non-vertebrate species such as bacteria, protists, fungi, plants, and metazoa.

Protein Sequence Databases and Functional Information Available from EBI

The Universal Protein Database (UniProt; http://www.uniprot.org), a consortium created in 2002, is a joint effort between the EBI, the Swiss Institute of Bioinformatics (SIB; http://www.isb-sib.ch), and the Protein Information Resource (PIR; http://pir.georgetown.edu, The UniProt Consortium 2008). UniProt is comprised of six different services:

1. The UniProt Knowledgebase (UniProtKB) consists of the two sections: the UniProtKB/Swiss-Prot for curated protein information and the UniProtKB/TrEMBL for automated annotation.
2. The UniProt Reference Clusters (UniProt/UniRef) databases, which contain clustered sets of sequences from the UniProtKB.
3. The UniProt Archive, a nonredundant protein database. Sequence retrieval is done through a unique protein identifier (UPI).
4. The UniProt Metagenomic and Environmental Sequences (UniProt/UniMES) database, which is a repository specifically developed for ▶ Metagenomics.
5. The UniProtKB-GOA, which provides high-quality gene ontology (GO; http://www.geneontology.org) annotations to proteins in UniProtKB/Swiss-Prot and UniProtKB/TrEMBL.
6. The UniProtKB Sequence/Annotation Version Archive (UniSave), which is a repository of UniProtKB/Swiss-Prot and UniProtKB/TrEMBL entry versions. InterPro, an integrated documentation resource used for the classification and automatic annotation of proteins and genomes, can be used to add in-depth annotation, including GO terms, to protein signatures in UniSave.

Other Databases Available from EBI

The ArrayExpress Archive is a database of microarray data for gene expression and functional genomics in accordance with the Minimum Information About a Microarray Experiment (MIAME) recommendations of the international Microarray Gene Expression Data (MGED; http://www.mged.org) Society. The EBI also provides access to patent data resources. These resources include the biology-related abstracts of patent applications, chemical compounds available from the Chemical Entities of Biological Interest (ChEBI; http://www.ebi.ac.uk/chebi), protein sequences of the European Patent Office (EPO; http://www.epo.org), the Japan Patent Office (JPO; http://www.jpo.go.jp), the Korean Intellectual Property Office (KIPO; http://www.kipo.go.kr) and the United States Patent and Trademark Office (USPTO; http://www.uspto.gov), as well as patent equivalents.

Cross-References

- ▶ DDBJ Genome Resources
- ▶ Metagenomics
- ▶ NCBI BioProject Genome Resources

References

Brooksbank C, Camon E, Harris MA, Magrane M, Martin MJ, Mulder N, O'Donovan C, Parkinson H, Tuli MA, Apweiler R, Birney E, Brazma A, Henrick K, Lopez R, Stoesser G, Stoehr P, Cameron G (2003) The European Bioinformatics Institute's data resources. Nucl Acids Res 31:43–50

Flicek P, Amode MR, Barrell D, Beal K, Brent S, Chen Y, Clapham P, Coates G, Fairley S, Fitzgerald S, Gordon L, Hendrix M, Hourlier T, Johnson N, Kähäri A, Keefe D, Keenan S, Kinsella R, Kokocinski F, Kulesha E, Larsson P, Longden I, McLaren W, Overduin B, Pritchard B, Riat HS, Rios D, Ritchie GRS, Ruffier M, Schuster M, Sobral D, Spudich G, Tang AY, Trevanion S, Vandrovcova J, Vilella AJ, White S, Wilder SP, Zadissa A, Zamora J, Aken BL, Birney E, Cunningham F, Dunham I, Durbin R, Fernández-Suarez XM, Herrero J, Hubbard TJP, Parker A, Proctor G, Vogel J, Searle SMJ (2011) Ensembl 2011. Nucl Acids Res 39:D800–D806

Kulikova T, Akhtar R, Aldebert P, Althorpe N, Andersson M, Baldwin A, Bates K, Bhattacharyya S, Bower L, Browne P, Castro M, Cochrane G, Duggan K, Eberhardt R, Faruque N, Hoad G, Kanz K, Lee C, Leinonen R, Lin Q, Lombard V, Lopez R, Lorenc D, McWilliam H, Mukherjee G, Nardone F, Pastor MPG, Plaister S, Sobhany S, Stoehr P, Vaughan R, Wu D, Zhu W, Apweiler R (2007) EMBL nucleotide sequence database in 2006. Nucl Acids Res 35:D16–D20

The UniProt Consortium (2008) The universal protein resource (UniProt). Nucl Acids Res 36:D190–D195

ECM

► Extracellular Matrix

Ecological Explanation

► Explanation, Functional

Ecological Modeling

Michael Hauhs
Ecological Modeling, University of Bayreuth, Bayreuth, Germany

Definition

Ecology studies organisms in their relationship to their living and nonliving environment. This relationship can either be conceptualized as a function of some internal structure or as characteristic behavior at the respective external interface. Models can be used to formalize these options. There are several scientific versions of the notion of models as exemplified by the usages in mathematics, physics, or engineering. Nothing has to be a model and anything can be a model (Mahr 2009), even an organism, e.g., ► model organism. This definition deals with models of organisms or ecosystems. The distinction between conceptual and computational models has been useful. In the following context of ► artificial life (AL), the discussion will be restricted to models which can be expressed or implemented on a computer. Hence, in ecological modeling, the empirical phenomena derived from scientific study as in ecosystem science (on the one hand) or in land-use traditions as in agriculture, forestry, or nature conservation (on the other hand) are linked with the theoretical results and engineering potential of computer science. In other words, the definition above does not restrict ecological modeling neither to a subfield of ecology nor to a subfield of computer science.

Organism is the central notion of biology and also ecology. Characterizing and expressing its key features require biological language and notions. One such key feature is that all life on earth is embedded into one single evolution history as indicated by the shared content in DNA. Defining organisms in relation to the embedding context, i.e., their environment including history, appears as easier than in relation to their constituents. No reconstruction of life from nonliving building blocks is yet possible. These features also set the stage for viewing living systems as *open* with respect to their environment. Since the times of Uexküll (Buchanan 2009), the question has been posed but not settled to what extent relationships between organisms and *their* environment are appropriately described by mechanisms or whether the role of life constitutes in turn another modeling relationship. Hence, proper definitions of organism, environment, and modeling may turn out to be recursive.

Environmental models (and *environmental* modeling) can be regarded as a field closely related to ecological modeling. An environment can only be defined relative to a system or interface of which it serves as environment. In environmental models, the key role of organisms is relaxed for characterizing the potential behavior of the environment. In this case, abiotic systems serve as a reference for the environment.

Characteristics

Ecological Modeling Classically

Environmental modeling was envisioned by John von Neumann as a key application area when computing machinery was introduced. He argued that computing would be critical in areas such as meteorology where results are relevant for society; they solve numerically a so far unsolvable problem and hence are capable of demonstrating the power of computing to a wide audience (McGuffie and Henderson-Sellers 2005).

The subsequent success of computer models in operational weather forecasting (prediction) has aptly demonstrated his vision. Computer models transformed weather prediction from an art into scientifically based technology. In this area, ► process-based models could be shown to outperform empirical models (McGuffie and Henderson-Sellers 2005). The success of environmental models has been used as a blueprint for other modeling approaches, especially those in the field of ecological modeling or climate models (Edwards 2010).

It is often assumed that computer models can become equally important when studying the phenomena of life. However, this puts the above definition of ecological modeling into a dilemma: On the one hand,

many ecologists regard ecosystems and living systems almost by definition as ▶ complex systems. Also, the atmosphere might be perceived as complex, but unlike large scale geo-systems, the complexity of life has no formalization and appears as such on all scales. Features of living systems and especially behavior such as reproduction or evolution are often described as the result of ▶ emergence or interaction (Oyama et al. 2000), a behavior towards the environment. Further, symptoms of unsettled theoretical and conceptual issues in biology can be found under the notion of holism; even the notion and meaning of causality is still contested (Laland et al. 2012).

On the other hand, each computer program runs on a mathematical machine, and any ecological models implement implicitly a formal approach to ecology, and any ▶ artificial life model implements a formal approach to biology. This implied theory can be seen as a hypothetical core to a theory about life and can be interpreted as reductionism. The goal of reductionist ecological and much of AL modeling is to follow the successful examples set by environmental modeling above. This, however, implies that ecological and AL models attempt serving holism and reductionism at the same time.

Models in ecology appear as either practically (empirically) relevant or as methodological rigorous but rarely both at the same time (Peters 1991). This leaves the field of ecological modeling with two opposite points of orientation: mainly empirical with a pragmatic focus on usefulness or mainly theoretical/speculative while remaining untestable.

A promising modeling approach to the ability of organisms of behaving strategically and deciding due to an actual environment is ▶ agent based modeling. This modeling technique is a new rapidly expanding field not only of ecological modeling (Grimm and Railsbeck 2005) but also in other fields such as sociology or economy. It is, however, still characterized by an unsettled relationship to theory (O'Sullivan and Haklay 2000) and a strained relationship with empirical studies (Clifford 2007).

Within the life sciences, ecological modeling has been perceived as an *applied* science. The theoretical background of the models being applied stems from physics. In such a reductionist perspective, the relationship of an organism or ecosystem with an environment is modeled as a function of state (e.g., implied by a natural law). In a holistic perspective, the relationship can be modeled as emergent behavior observed at an interface, e.g., between an ecosystem and its abiotic environment. Computers have been routinely used as (reductionistic) generators of virtual (living) behavior. However, it has not been widely recognized that computer science also offers theoretical tools studying the "design of behavior." Theoretical computer science could accordingly be summarized by a corresponding slogan "Computers, as they could be."

Ecological Modeling for Artificial Life

An area where such a link seems natural, but has also not been fully established so far, is Artificial Life research. (▶ Artificial Life) has constituted itself with the slogan "Life as it could be" (Langton 1989), in the sense of, could be artificially generated. Models in AL have been focused at the reductionist approach in which living behavior is sought to be generated artificially from abstract building blocks. The commentary of the results as "something is missing" by Brooks (2001) still largely holds. AL seems to share with ecology the strange relationship between methodological rigor and practical relevance. Using this analogy would imply focusing in a parallel manner on attempts "formalizing holism," i.e., studying life as virtual (rather than artificial) behavior with methods from computer science. So far, however, there has been little contact in the research agendas of AL and ecological modeling as indicated by the few respective entries in AL or ecological modeling conferences.

Rosen proposed to formalize modeling relations between organisms and environment in categorical language (Rosen 1991). In the language of (▶ Mathematical Structure, Category) theory, there are two modeling paradigms available which encompass the above dilemma: Firstly, a functional modeling paradigm which has been adopted from physics and which provides a constructive interpretation to the slogan: "Life as it could be artificially and symbolically *constructed* on a computer." It derives the environmental relationship of life as a function of state. Secondly, there is an interactive modeling paradigm which characterizes life at the interface with its environment as emergent behavior. It can be adopted from computer science and provides a *classificatory* interpretation to the slogan: "Life as it could *behave* virtually on a computer." In the related field of computational intelligence, the latter approach has become famous by the so-called Turing test. A correspondent behavioral classification of life has been proposed (Cronin et al. 2006).

Since the behavior-based definition of intelligence by A. Turing, theoretical computer science has come up with a unifying approach of formalizing behavior of systems (Rutten 2000). Categorically, this approach can be abstracted as (▶ Coalgebra). The two modeling approaches above are categorical duals of each other. However, only one of them has dominated the fields of AL and ecological modeling so far, i.e., the former physical one. The second approach exists as a theoretical possibility: In this perspective, the relationships of living system to their context and how they behave towards their environment become the defining features of life. The corresponding states observable at interfaces are the mere (epistemic) signatures of this behavior rather than their material causes.

Cross-References

- ▶ Agent-Based Modeling
- ▶ Artificial Life
- ▶ Coalgebra
- ▶ Complex System
- ▶ Emergence
- ▶ Mathematical Structure, Category
- ▶ Model Organism
- ▶ Process-based Model

References

Brooks R (2001) The relationship between matter and life. Nature 409:409–411
Buchanan B (2009) Onto-ethologies: the animal environments of Uexkull, Heidegger, Merleau-Ponty, and Deleuze. State University of New York Press, New York
Clifford NJ (2007) Models in geography revisited. Geoforum 39:675–686
Cronin L, Krasnoger N, Davis BD, Alexander C, Robertson N, Steinke JHG, Schroeder SLM, Cooper G, Gardner PM, Kholbystov AN, Siepmann P, Whitaker BJ, Marsh D (2006) The imitation game – a computational chemical approach to recognizing life. Nat Biotechnol 24(10):4
Edwards PN (2010) A vast machine: computer models, climate data, and the politics of global warming. MIT Press, Cambridge
Grimm V, Railsbeck SF (2005) Individual-based modeling and ecology. Princeton University Press, Princeton
Laland KN, Sterelny K, Odling-Smee J, Hoppitt W, Uller T (2012) Cause and effect in biology revisited: is Mayr's proximate-ultimate dichotomy still useful? Science 334:1512–1516
Langton C (1989) Artificial life. Addison-Wesley, Reading
Mahr B (2009) Die informatik und die logik der modelle. Informatik Spektrum 32(3):228–249
McGuffie K, Henderson-Sellers A (2005) A climate modelling primer. Wiley, London
O'Sullivan D, Haklay M (2000) Agent-based models and individualism: is the world agent-based? Environ Plann A 32:1409–1425
Oyama S, Griffiths PE, Gray RD (2000) Cycles of contingency – developmental systems and evolution. MIT Press, Cambridge
Peters RH (1991) A critique for ecology. University Press, Cambridge
Rosen R (1991) Life itself – a comprehensive inquire into the nature, origin, and fabrication of life. Columbia University Press, New York
Rutten JJMM (2000) Universal coalgebra: a theory of systems. Theor Comp Sci 249(1):3–80

Edge Betweenness

▶ Edge Betweenness Centrality

Edge Betweenness Centrality

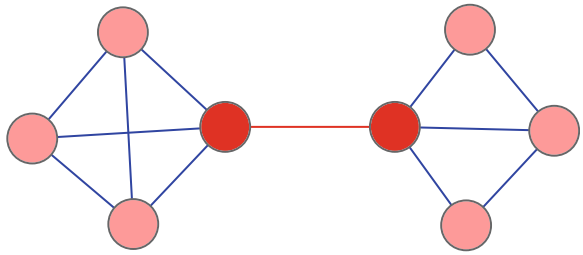
Long Jason Lu[1] and Minlu Zhang[2]
[1]Division of Biomedical Informatics, Cincinnati Children's Hospital Research Foundation, Cincinnati, OH, USA
[2]Department of Computer Science, University of Cincinnati, Cincinnati, OH, USA

Synonyms

Edge betweenness

Definition

The edge betweenness centrality is defined as the number of the shortest paths that go through an edge in a graph or network (Girvan and Newman 2002). Each edge in the network can be associated with an edge betweenness centrality value. An edge with a high edge betweenness centrality score represents a bridge-like connector between two parts of a network, and the removal of which may affect the communication between many pairs of nodes through the shortest paths between them. Figure 1 illustrates an example of eight nodes in a network, and the red



Edge Betweenness Centrality, Fig. 1 An edge with a high edge betweenness centrality value. In the network, the *red* edge between two *red* nodes has the highest betweenness centrality score among all edges

edge between two red nodes has a high edge betweenness centrality value of 16. The removal of this edge will result in a partition of the network into two densely connected subnetworks.

Cross-References

- ▶ Biological Applications of Network Modules
- ▶ Modules in Networks, Algorithms and Methods

References

Girvan M, Newman ME (2002) Community structure in social and biological networks. Proc Natl Acad Sci USA 99(12):7821–7826

Edge Ontology

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

Synonyms

Arrow ontology

Definition

Inspired by ▶ gene ontology, edge ontology provides a hierarchical vocabulary of terms for describing interactions between biological molecules. Different types of relationships between pathway components can be represented by edge ontology. Lu et al. (2007) provide a prototype of an edge ontology that divides edge into four levels: direction, type, subtype, and specification. Edge ontology is quite a novel concept still far from being completed.

References

Lu LJ, Sboner A, Huang YJ, Lu HX, Gianoulis TA, Yip KY, Kim PM, Montelione GT, Gerstein MB. Comparing classical pathways and modern networks: towards the development of an edge ontology. Trends Biochem Sci. 2007;32(7):320–331.

Efficiency

Olga Vitek
Department of Statistics, Department of Computer Science, Purdue University, West Lafayette, IN, USA

Definition

Efficiency is a property of experimental design. It is inversly proportional to the variance of the data-derived estimates of the unknowns. Smaller variation leads to higher efficiency.

Cross-References

- ▶ Designing Experiments for Sound Statistical Inference

EFM

- ▶ Elementary Mode

Eigen Decomposition

- ▶ Principal Component Analysis (PCA)

Eigenvalue

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

Definition

The eigenvectors of a square matrix are the nonzero vectors that, after being multiplied by the matrix, either remain proportional to the original vector (i.e., change only in magnitude, not in direction) or become zero. For each eigenvector, the corresponding eigenvalue is the factor by which the eigenvector changes when multiplied by the matrix. The prefix eigen- is adopted from the German word "eigen" for "own" in the sense of a characteristic description. The eigenvectors are sometimes also called characteristic vectors. Similarly, the eigenvalues are also known as characteristic values.

The mathematical expression of this idea is as follows: if A is a square matrix, a nonzero vector v is an eigenvector of A if there is a scalar λ (lambda) such that:

$$Av = \lambda v$$

The scalar λ is said to be the eigenvalue of A corresponding to v. An eigenspace of A is the set of all eigenvectors with the same eigenvalue together with the zero vector. However, the zero vector is not an eigenvector.

These ideas often are extended to more general situations, where scalars are elements of any filed, vectors are elements of any vector space, and linear transformations may or may not be represented by matrix multiplication. For example, instead of real numbers, scalars may be complex numbers; instead of arrows, vectors may be functions or frequencies; instead of matrix multiplication, linear transformations may be operators such as the derivative from calculus. These are only a few of countless examples where eigenvectors and eigenvalues are important.

In such cases, the concept of direction loses its ordinary meaning, and is given an abstract definition. Even so, if that abstract direction is unchanged by a given linear transformation, the prefix "eigen" is used, as in eigenfunction, eigenmode, eigenface, eigenstate, and eigenfrequency. Eigenvalues and eigenvectors have many applications in both pure and applied mathematics. They are used in matrix factorization, in quantum mechanics, and in many other areas.

Elasticity Coefficient

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

Definition

It is a quantitative value of the ability of a pathway enzyme (or group of enzymes) to change its rate or activity when changes in its ligands (substrates, products, activators, inhibitors) are varied during the pathway function. It is written as ${\varepsilon^{ai}}_X$ where a is the rate or activity of a pathway enzyme i and X is any of its ligands. The elasticities are positive for ligands that increase the enzyme rate whereas they are negative for those that decrease it. The elasticity coefficients are intrinsic properties of the enzymes since they only depend on its particular kinetic capabilities.

Cross-References

▶ Metabolic Control Theory

Electronic Health Record Systems

Jesus Bisbal
Departament de Tecnologies de la Informació i les Comunicacions, Universitat Pompeu Fabra, Barcelona, Spain

Definition

An electronic health record system is the software system which provides the functionality needed to

create, manage, and exploit the information contained in a set of ▶ electronic health records Bisbal and Berry 2011; Dick et al. 1997.

References

Bisbal J, Berry D (2011) An analysis framework for electronic health record systems: interoperation and collaboration in shared healthcare. Methods Inf Med 50(2):180–189

Dick RS, Steen EB, Detmer DE (eds) (1997) The computer-based patient record: an essential technology for health care, Revised editionth edn. Institute of Medicine, National Academy Press

Electronic Health Records

Jesus Bisbal
Departament de Tecnologies de la Informació i les Comunicacions, Universitat Pompeu Fabra, Barcelona, Spain

Synonyms

Computer-based patient records (CPR); Electronic healthcare records (EHCR); Electronic medical records (EMR); Electronic patient records (EPR)

Definition

An electronic health record (EHR) is the digital representation of health-related information about an individual. It provides a unified view over all relevant information, irrespective of where and how this data is created or managed. Its primary purpose is to improve the delivery of health care, including, but not limited to, increased efficiency and reduction of human errors (Kohn et al. 2000).

It must be differentiated from an ▶ electronic health record system, which is the software system which provides the functionality needed to create, manage, and exploit the information contained in a set of electronic health records.

Characteristics

Scope

The scope of an electronic health record refers to the range of information it contains. Ideally, it should provide a longitudinal, lifelong "cradle to grave" record of all kinds of information about an individual (Grimson 2001). Typically, however, this is limited to noncontiguous periods of life and includes information like demographics, medical history, medication and allergies, immunization status, laboratory test results, and radiology images.

Advances in clinical medicine and biomedical research impact on clinical practice. This, in turn, increases the expectations of caregivers on the scope of information to be included within an electronic health record. For example, genetic data is not yet commonly included in most systems, but it may be a requirement in the near future if therapies are routinely based on such information.

The adoption of electronic health records is a long and costly process, both at the technological as well as at the organizational levels, and there is still no sufficient scientific evidence that this process is indeed cost-effective (Chaudhry et al. 2006; Price Water House Coopers 2007) from the point of view of the health-care providers.

Electronic Health Records Organization

Health care generates large volumes of information, which must be appropriately structured and accessed to ensure its efficient exploitation during care delivery. There are several different ways in which the contents of an electronic health record can be organized (Bisbal and Berry 2011). A very common approach in existing systems is the episode-based EHR, in which information is indexed along the time line according to the interactions between a patient and the care provider(s). This organization emphasizes the administrative side of patient information.

Alternatively, information can be structured according to the problem-oriented EHR paradigm (Weed 1968), in which patient information is indexed by the list of clinical "problems" suffered by the subject. Such problems can include, for example, risk factors, symptoms, signs, and fully diagnosed clinical conditions. This provides a clinically-oriented view of patient information, often preferred by caregivers and commonly available in primary care, but not yet widely used in secondary/tertiary care.

Finally, more flexible approaches acknowledge the complexity of the health-care domain and the variety of user types that interact with an electronic health record system. No a priori record structure is assumed in this case; thus the organization of the information in the record is defined by each caregiver, according to his/her specific needs. For this reason, it has sometimes been referred to as neutral EHR organization (Bisbal and Berry 2011). This approach is technologically more challenging to realize. It can be achieved through the so-called two-level modeling paradigm (Garde et al. 2007). Its first level, the reference model, is a predefined set of very abstract classes and aggregation rules that provide the flexibility to model any information concept. Its second level, the ▶ Archetypes, add semantics and constraints to the potential instances of the reference model, in order to ensure that the desired semantics are captured by the clinical concepts defined by those archetypes.

Electronic Health Records Construction

Several mechanisms have been devised to create the "unified view" of all health-related information expected to be provided by an electronic health record (Bisbal and Berry 2011). The first and technically most feasible approach, referred as consolidated EHR, stores all this data in a repository considered part of the EHR system itself. It is the approach most commonly used in practice, due to its conceptual simplicity and because its development relies on sound database methodologies. It suffers, however, from a significant limitation. It assumes that the EHR system's own data repository will be the only source from which data is delivered to the caregiver. However, clinical decisions are commonly taken based on multisource information (e.g., centralized medical history, plus departmental specialized investigations), and the electronic health-care record should ideally unify access to all these sources.

As an alternative, it is also possible to provide the so-called federated EHR (Grimson et al. 1998). This approach provides a true view of the underlying data, querying the underlying data sources which store all required information, without requiring any changes to existing system (Sheth and Larson 1990). Thus, it leverages the investment made in legacy systems (Bisbal et al. 1999), in contrast to the consolidated approach. Another advantage of the federated approach is that data is never out-of-date, as it is at all times taken from its original data source. The approach is not without its technical difficulties, however. It must address syntactic and semantic integration issues, just like any other approach (except for the unlikely case of a green field site). It may also suffer from performance limitations, due to repeated queries to remote database servers and data integration (transformation) processes that must be performed "on the fly."

There is a middle ground between the consolidated and federated approaches, referred to as materialized EHR, in which all or part of the EHR data may be materialized in a local data repository (Hull and Zhou 1996).

Interoperability of Electronic Health Records

An Institute of Medicine's 2001 report (IOM 2001) identified four stages of health organization evolution in the progression from autonomous modes of working to fully coordinated shared care with a high degree of specialization and expertise as follows:

- Stage 1: highly fragmented practice with individuals functioning autonomously and little specialization.
- Stage 2: referral networks of loosely structured multidisciplinary teams.
- Stage 3: more patient-centered team-based care but focusing primarily on needs and intentions of health-care professionals with some decision support but little integration of health information.
- Stage 4: shared multidisciplinary care, *evidence-based and patient-centric practice* with strong service coordination between practices and with good quality practices and performance measures.

The latest stage of this evolution characterizes the health care which is hoped to be provided in most industrialized countries. Patient care is shared among several independent caregivers, and a health condition commonly requires collaboration between several clinicians. The complete realization of this shared health-care scenario clearly requires integrating all the relevant information (Dick et al. 1997), potentially from several independent institutions. In this context, interoperability between electronic health records systems becomes essential. This refers to the ability to exchange information between systems (Noy 2004). Several levels of interoperability can be defined (Bisbal and Berry 2011). Functional interoperability, for example, requires that communicating systems agree on the exact content, structure, and meaning of the information to be exchanged. Remarkable efforts like the HL7 industry standard have initially pursued this approach. However, without a precise definition of the semantics of this information,

interoperability between systems cannot be guaranteed (Iakovidis et al. 2007). In that respect, the combination of standardized terminologies (e.g., ▶ ontologies) together with domain concept definitions (i.e., ▶ archetypes, as described above) will raise the level of interoperability. This leads to semantic interoperability (▶ Semantic Web, Interoperability), which ensures that the information being exchanged is computer-processable. For that reason, international efforts (Eichelberg et al. 2005), like CEN's 13606 and HL7's RIMv3, aim to standardize how domain concepts (i.e., ▶ Archetypes, ▶ Templates) should be defined in order to achieve this level of interoperability.

Secondary Uses of Electronic Health Records

In addition to improving health care delivery, electronic health records also hold the promise of accelerating the outcomes and efficiency of biomedical research. The aggregation of vast amounts of patient-level information can, for example, be the basis upon which to test models/theories about disease mechanisms and treatment outcomes, as is done in systems biology and the virtual physiological human (Kohl and Noble 2009), or decrease patient recruitment cycle times for clinical trials.

The realization of this secondary use requires that electronic health records are in wide spread use. In addition, the level of ▶ interoperability between these systems needs to be sufficiently high as to facilitate the integration of related sources, in order to efficiently exploit this wealth of information.

References

Bisbal J, Lawless D, Wu B, Grimson J (1999) Legacy information systems: issues and directions. IEEE Software 16(5):103–111

Bisbal J, Berry D (2011) An analysis framework for electronic health record systems: interoperation and collaboration in shared healthcare. Methods Inf Med 50(2):180–189

Chaudhry B, Wand J, Wu S, Maglione M et al (2006) Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 144(10):742–752

Dick RS, Steen EB, Detmer DE (eds) (1997) The computer-based patient record: an essential technology for health care (revised edition). Institute of Medicine, National Academy, Washington, DC

Eichelberg M, Aden T, Riesmeier J, Dogac A, Laleci GB (2005) A survey and analysis of electronic healthcare record standards. ACM Comput Surv 37(4):277–315

Grimson J, Grimson W, Berry D, Stephens G, Felton E, Kalra D et al (1998) A CORBA-based integration of distributed electronic healthcare records using the synapses approach. IEEE Trans Inf Technol Biomed 2(3):124–138

Grimson J (2001) Delivering the electronic healthcare record for the 21st century. Int J Med Inform 64:111–127

Garde S, Hovenga E, Buck J, Knaup P (2007) Expressing clinical data sets with openehr archetypes: a solid basis for ubiquitous computing. Int J Med Inform 76(3):334–341

Hull R, Zhou G (1996) A framework for supporting data integration using the materialized and virtual approaches. ACM SIGMOD Record 25(2):481–492

Iakovidis I, Dogac A, Purcarea O, Comyn G, Laleci GB (2007) Interoperability of eHealth systems – selection of recent EU's research programme developments. In: Proceedings of the international conference eHealth: combining health telematics, telemedicine, biomedical engineering and bioinformatics to the edge, Berlin

IOM (2001) Commitee on quality of healthcare in america institute of medicine. Crossing the quality chasm: A new health system for the 21st century. National Academic Press

Kohn LT, Corrigan JM, Donaldson MS (eds) (2000) To err is human: building a safer health system. National Academic Press, Washington, DC

Kohl P, Noble D (2009) Systems biology and the virtual physiological human. Mol Syst Biol 5:292

Noy NF (2004) Semantic integration: a survey of ontology-based approaches. ACM SIGMOD Record 33(4):65–70, Special section on semantic integration

Price Water House Coopers (2007) The economics of IT and hospital performance. Price-Water House Coopers' Health Research Institute, Boston

Sheth AP, Larson JA (1990) Federated database systems for managing distributed, heterogeneous, and autonomous databases. ACM Comput Surv 22(3):183–236, Special issue on heterogeneous databases

Weed LL (1968) Medical records that guide and teach. New Eng J Med 278(11): 593–599 and 278(12):652–657. Published also in MD Comput 10(2):100–114, March–April 1993

Electronic Healthcare Records (EHCR)

▶ Electronic Health Records

Electronic Medical Records (EMR)

▶ Electronic Health Records

Electronic Patient Records (EPR)

▶ Electronic Health Records

Elementary Flux Mode

▶ Elementary Mode

Elementary Mode

Sang Yup Lee
Department of Chemical and Biomolecular Engineering and Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea

Synonyms

EFM; Elementary flux mode; ELMO

Definition

Elementary modes are a minimal set of pathways and all possible steady-state flux distributions that the network inherently can achieve (Schuster et al. 2002). This set of vectors can be calculated from the stoichiometric matrix of a biochemical network using convex analysis, and there exists a unique set of elementary modes for a given network.

1. Each elementary mode is a minimal set of pathways that consists of the minimum number of reactions necessary to exist as a functional unit. This property is called "non-decomposability" or "genetic independence." If any reaction in an elementary mode is eliminated, the resulting elementary mode cannot operate as a functional unit.
2. The elementary modes are the set of all routes through a given network corresponding with property.

References

Schuster S, Hilgetag C, Woods JH, Fell DA (2002) Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. J Math Biol 45(2):153–181

ELMO

▶ Elementary Mode

Elongation Complex

▶ Transcription Elongation Complex

Elongation Factor

▶ Transcription Elongation Factor

Elongation Phase of Translation

▶ Translation Elongation

EMBL Genome Database

▶ EBI Genome Resources

Embryological Explanation

▶ Explanation, Developmental

Embryonic Induction

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

Definition

Before the rise of genetics, Hans Spemann in 1901 already emphasized the role of inducers at the tissular scale, a role which can be played by many chemical

substances, as Joseph Needham recognized later (*Order and life*), leading to the "paradox of an apparently nonspecific stimulus eliciting a specific developmental response" (Jacobson and Sater 1988, p. 341). Inducers trigger the development of sets of cells, morphogenetic fields, likely to develop into specific cell fates. The neural crest, for example, is a transitory structure which among other things gives rise to the nervous system. Morphogenetic fields can until some stage be induced into other fates by proper induction as it has been shown by Spemann and Mangold's sea urchin experiments. This proves the plasticity proper to early development.

The experimental investigation of the nature of inducers as well as the relation between morphogenetic fields and organizers has been a main endeavor of embryology in the pre-molecular biology era. Current developmental theory investigates the genetic mechanisms of induction in terms of transcription factors, etc., and the developmental genes, such as homeogenes, underpinning morphogenetic fields (De Robertis et al. 1991).

Inducers cannot induce anything except in a "competent" tissue, that is, a tissue which is already prepared to deal with them and respond to them. This has been again shown about the induction of the central nervous system, which actually involves more requisites than the sole induction by a specific center in the mesoderm – as one usually thought – but also signals expressed in the ectoderm already from the blastula stage, without which neural induction is inefficient (Kuroda et al. 2004). Therefore, developmental processes require both induction and "competence" (as a capacity to respond to specific signals) (Dawid 2004), which is a modern version of the complementarity between epigenesis (inducers) and preformism (competence) as two poles in developmental explanations.

Cross-References

▶ Explanation, Developmental

References

Dawid I (2004) Organizing the vertebrate embryo – a balance of induction and competence. PlosBio 2(5):0579–0583

De Robertis E, Morita E, Cho K (1991) Gradient fields and homeobox genes. Development 112:669–678

Jacobson A, Sater A (1988) Features of embryonic induction. Development 104:341–359

Kuroda H, Wessely O, De Robertis E (2004) Neural induction in Xenopus: requirement for Ectodermal and Endomesodermal signals via Chordin, Noggin, b-Catenin, and Cerberus. PlosBio 2(5):623–634

Emergence

Ulrich Krohs
Department of Philosophy, University of Münster, Münster, Germany

Definition

Emergence is the appearance of novel properties, structures, patterns, and processes in a system, which are absent from the isolated components of the system. Concepts of emergence vary in demanding the novel features also being unpredictable from knowledge of the components and their possible interactions (*weak emergence*), or being irreducible (▶ Reduction) to properties of the components (*strong emergence*). Just novelty, without unpredictability and irreducibility, is sometimes called *nominal emergence*.

Characteristics

Taxonomy and History

The concept of emergence captures the view that the whole is more than the sum of its parts. When G.H. Lewes introduced the term in 1875, it was linked to the view of irreducibility of the systems level to the level of components. This metaphysical view of emergence was most influentially put forward by Broad (1923) and, in a less strict version, by Alexander (1920) (see McLaughlin 1992). Today, the concept comes differentiated into several more specific versions (see Bedau 2003), not all of which require irreducibility. The concept of emergence has thus lost its strong metaphysical implications and is used today even within strictly physicalist frameworks.

Nominal Emergence

Nominal emergence requires novelty, as any other concept of emergence, but neither unpredictability nor

unexplainability or irreducibility. A systemic entity is novel if does not occur – is not instantiated without – a system of a particular sort. For example, a stripe pattern of a fabric emerges n*ominally*, since no smallest component of the fabric, even if colored, can be striped. Similarly, an electric dipole displays electrical properties that no simple carrier of an elementary charge can have. So the property of being a dipole emerges nominally in a system of spatially separated charges. Notwithstanding, we can easily predict the nominally emergent property from more basic knowledge about the system and its components.

The concept of nominal emergence is the least demanding member of this family of concepts, but consequently, it captures the least interesting cases of emergence.

Weak or Epistemic Emergence

The concept of weak emergence takes emergence as unpredictability or unexplainability of higher-level properties. It is thus an epistemological concept. In this perspective, emergence is theory relative, because predictability of a systemic property obviously depends on the theories available for prediction, as does explainability. What is emergent today, may no longer be emergent tomorrow, given that a theory becomes available that allows for prediction or explanation of the property in question. Regulatory properties of a feedback loop were *weakly* emergent when first observed. Nowadays, cybernetics and systems theory allow predicting, in principle, such loops and their systemic effects. So, there was a state of knowledge where feedback loops were weakly emergent, while it is nowadays merely nominally emergent.

Weak emergence is scientifically and metaphysically unproblematic. It makes no presupposition about anything falling in principle out of the scope of scientific explainability. It is rather the manifestation of the limitation of our knowledge at any given time.

The concept may be strengthened by claiming that unpredictability and unexplainability in a particular case do hold for principle reasons and will not vanish with any possible future development of theorizing. Theory relativity is thus replaced by absolute unpredictability or unexplainability. However, it is unclear how such in-principle unexplainability could be proven. Consequently, the absolute, not theory-relative concept of weak emergence may not have any application.

Strong or Metaphysic Emergence

Strong emergence, in contrast, is an ontological concept, which is meant to describe ontological or conceptual irreducibility of an emergent property to properties and configuration of components. The case was made for the emergence of the mind, or of mental properties, on complex neuronal structures (Broad 1923). It is all but clear both, whether mental properties do emerge as claimed, and, if so, whether they are irreducible to neuronal properties.

The concept of strong emergence conceptualizes the idea that a complex system brings entities into being which are irreducible not only for the reason of the actual or in-principle limitedness of the human mind, but because of their categorical difference to anything which can possibly be linked theoretically to the components of the system. This case was what the British emergentists had in mind when introducing the concept. The human mind was thought to emerge from neuronal structures, being of a quality that is irreducible to any interplay of material entities.

The Dynamical Criterion for Emergence

A fourth concept of emergence avoids the spookiness of strong, metaphysical emergence without making the contingent status of theory building at a certain time the criterion for emergence. This concept relies on a dynamical criterion for emergence. According to this concept, any behavior that is either the result of a bifurcation where a structural instability leads to a shift in dynamical form, or that establishes a new system level (Hooker 2011). The latter is the case, e.g., in phase transitions to the solid state. The former occurs, e.g., in flocking behavior and in a bifurcation that establishes a limit cycle from a singularity. The concept of dynamical emergence thus relies on novelty alone and avoids any claim about unexplainability or unpredictability, while avoiding at the same time focusing on the more or less trivial cases of nominal emergence.

Emergence and Causality

Unfortunately, emergence is often tried to be cashed out in terms of causality. The higher-level property is said to cause the systems components to show certain behavior. This view of so-called downward causation (▶ Interlevel Causation) is flawed by a major misconception of what a causal relation is. Let us take first the opposite view: components making up a system. The behavior of the system depends on, and

is perhaps determined by, the behavior of the components. Despite this determination, the relation is not a causal one. It is constitutional instead: The components in their relation to each other constitute the system. Their behavior *is* at the same time part of the behavior of the system. To test this intuition, we may adopt a view of causality as transfer of energy or other conserved parameters. We see immediately that the components cannot transfer energy to the system of which they are proper parts, since their energy is already included in the energy of the system. So the relation is not causal. Causal processes are diachronic, constitution is synchronic. "Upward causality" is synchronic and thus constitution rather than diachronic causality. Analogously, a causal relation of "downward causation" does not exist, though obviously a part behaves differently whether it is isolated or integrated in a system. However, the system does not transfer energy to its parts; instead, parts exchange energy with other parts. The system merely constrains what the parts can do – and even this is not the precise description of what is going on. In fact, the interacting parts constrain each other – which is most conveniently described as systems behavior. 'Consisting of parts' and 'being constituted by' are synchronic relations, while the influence between the components is diachronic. Thus, only the latter is a causal relation. Emergence is not a diachronic, causal, but a synchronic, constitutional relation. This does not rule out that emergent properties may have causal effects, e.g., on other high-level systems (Stephan 1997).

Emergence and Reduction

Whether emergence and reduction are opposites or even necessarily linked to each other depends on the concepts or emergence and reduction used. Weak or epistemic emergence obviously excludes theory reduction, since it is defined exactly via irreducibility. On the other hand, it is perfectly compatible with ontologic reduction, namely, the thesis that nothing consists of more than its proper parts. In contrast, strong emergence in its metaphysical understanding is opposed to and thus incompatible with ontologic reduction. It is also incompatible with theory reduction. The claim is that there are higher-level entities and consequently high-level concepts which cannot be translated into the concepts of the lower level. Emergence according to the dynamical criterion, finally, matches perfectly well with ontological as well as with theory reduction. Nothing with regard to the emergent dynamics or level needs to remain unexplained, and no other than low-level entities are assumed to make up the system.

Cross-References

- ▶ Causality
- ▶ Complexity
- ▶ Mechanism
- ▶ Reduction
- ▶ Supervenience

References

Alexander S (1920) Space, time, and deitiy. Macmillan, London
Bedau MA (2003) Downward causation and autonomy in weak emergence. Princ Rev Int Epistemol 6:5–50
Broad CD (1923) Scientific thought. Humanities Press, New York
Hooker C (2011) Conceptualizing reduction, emergence and self-organization in complex dynamical systems. In: Hooker C (ed) Philosophy of complex systems. Elsevier, Amsterdam, pp 195–222
McLaughlin B (1992) The rise and fall of British emergentism. In: Beckermann A, Flohr H, Kim J (eds) Emergence or reduction? Essays in the prospects of nonnaturalist physicalism. de Gruyter, Berlin, pp 49–93
Stephan A (1997) Armchair arguments against emergentism. Erkenntnis 46:305–314

Enabling Conditions

Marie I. Kaiser and Andreas Hüttemann
Department of Philosophy, University of Cologne, Cologne, Germany

Synonyms

Manifestation conditions; stimulus conditions

Definition

Enabling conditions are those conditions, which are – according to the simple conditional analysis (SCA) – necessary and sufficient for the occurrence of the manifestation (Choi and Fara 2012).

Cross-References

► Disposition

References

Choi S, Fara M (2012) Dispositions. In: Zalta EN (ed) The Stanford Encyclopedia of Philosophy (Spring 2012 Edition), http://plato.stanford.edu/archives/spr2012/entries/dispositions/

Endoreplication

Orsolya Kapuy and Béla Novák
Department of Biochemistry, Oxford Centre for Integrative Systems Biology, University of Oxford, Oxford, UK

Definitions

Endoreduplication: A type of over-replication. More than one copy of DNA is created without intervening chromosome segregation. This mechanism causes new rounds of complete S phases without mitosis and each replication is a discrete event. Replication origins are not activated more than once in each S phase. The DNA content is increasing by doubling (2N, 4N, 8N, etc.) (Arias and Walter 2007; Porter 2008).

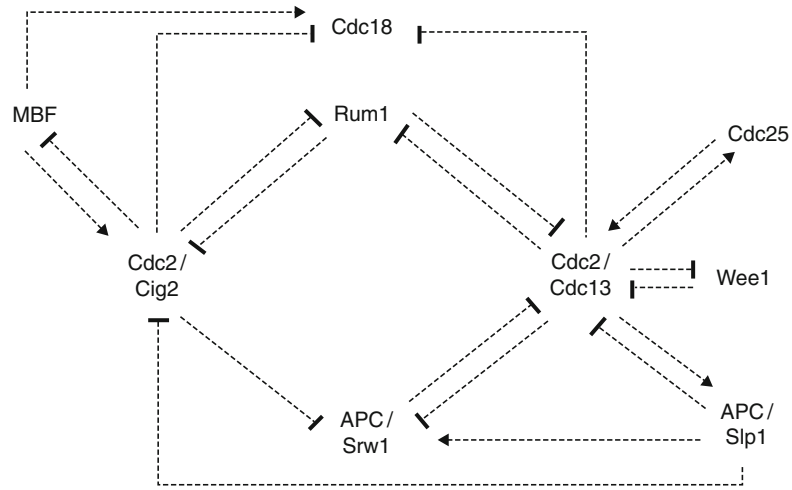
Re-replication: A type of over-replication. Replication origins undergo more than one initiation event per S phase leading to increasing DNA content. There are overlapping replications resulting onion-like structures with intermediate N values (e.g., between 2N and 4N) (Arias and Walter 2007; Porter 2008).

Characteristics

One of the most important roles of the mitotic cell cycle machinery is to maintain the alternation of S and M phases through generations. The cell has to duplicate its full set of chromosomes, and then the two copies have to be distributed equally to the daughter cells at mitosis. DNA replication is controlled by a two-step mechanism to make sure that one and only one S phase occurs per each cell cycle. The two separated steps are the following: first, essential components of the DNA replication machinery have to be loaded onto the replication origins. This licensing step can only be accomplished when the CDK activity is low (i.e., G1 phase of the cell cycle). The licensed origins can initiate replication during the next step when CDK activity is increasing and promotes replication by phosphorylation. An important consequence of the increasing CDK activity is the inhibition of a new round of licensing. Therefore, CDK activity inhibits the first step (licensing), but activates the second step (initiation) in replication control. Accordingly, the replication origin can only be relicensed in the subsequent cell cycle after the CDK activity dropped to zero. The CDK activity drop occurs at the end of the mitosis and this allows one and only one DNA replication per cell cycle (Morgan 2006).

Eukaryotic cells employ two different CDKs to drive their mitotic cycles. The S-phase and M-phase CDKs peak during DNA replication and mitosis, respectively. Both CDKs inhibit the relicensing of the replication origin and this inhibition avoids over-replication under normal conditions. Over-replication means that the cell breaks the rule of alternating DNA replication and mitosis, and initiates new replication without undergoing mitosis. This can happen in the absence of M-phase CDK, when the periodic oscillation of S-phase CDK is able to drive full rounds of complete S phases. This type of over-replication is called endoreduplication. Another possible way for DNA over-replication is when the replication origins fire more than once per S phase thereby breaking the "once and only once DNA synthesis" rule. This type of over-replication is called re-replication. Large increase in the DNA content with some exceptions is lethal for the cell (Morgan 2006; Arias and Walter 2007; Porter 2008).

The molecular requirements of over-replication mechanisms were first characterized in fission yeast by the Nurse's lab. Therefore, in this entry we are focusing on some well-studied examples of endoreduplication and re-replication in fission yeast. The Cdc2/Cdc13 and the Cdc2/Cig2 complexes represent the mitotic and the S-phase CDKs in this organism. Both complexes are inhibited by the Rum1 stoichiometric inhibitor in G1 phase and both Cdc13 and Cig2 levels are downregulated by APC/Slp1 at the end of mitosis. The APC/Srw1 promotes Cdc13 degradation in G1 phase of the cell cycle. The G2/M transition is controlled by both Wee1 kinase and Cdc25 phosphatase regulating the inhibitory Tyr-phosphorylation of Cdc2/Cdc13. The synthesis of



Endoreplication, Fig. 1 The main regulatory interactions in the fission yeast cell cycle regulatory network controlling DNA replication and mitosis. Arrows and blocked end lines represent stimulation of synthesis/activation and degradation/inhibition, respectively

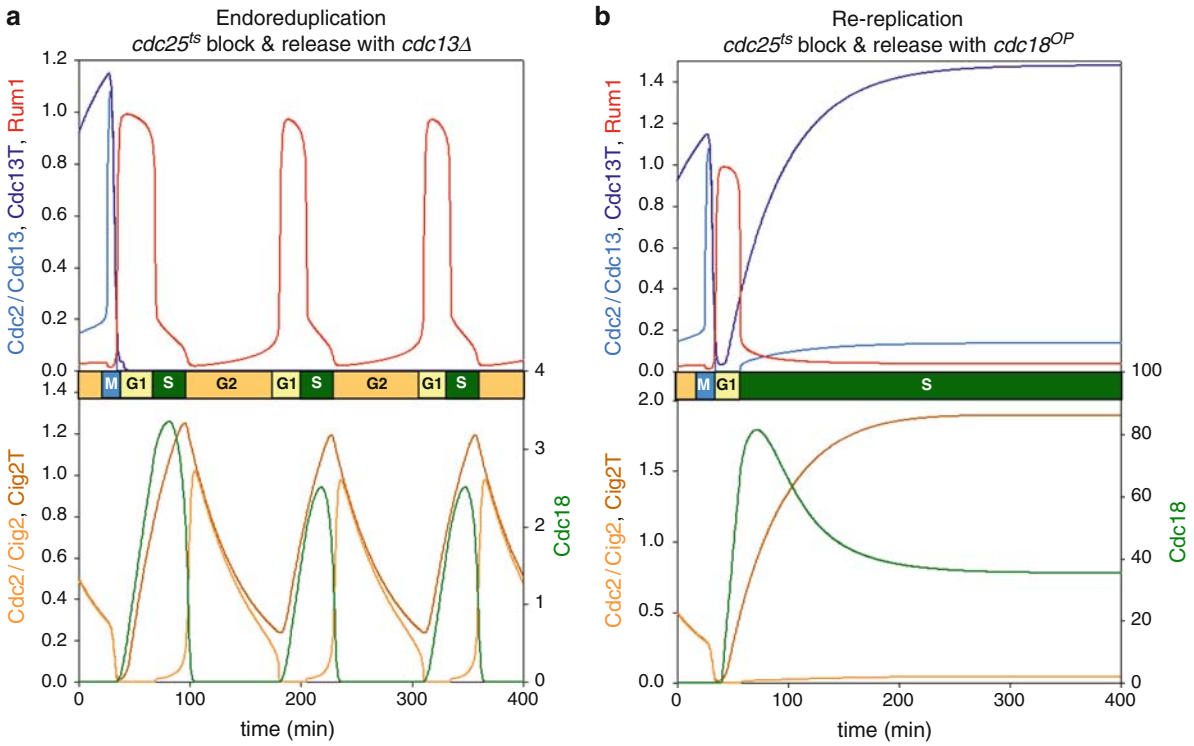
S-phase cyclin is regulated by a transcription factor complex (MBF), which also assists the activation of the licensing factor, Cdc18, in fission yeast. The degradation of Cdc18 is promoted by CDK-dependent phosphorylation. Both Cdc2/Cdc13 and Cdc2/Cig2 control the activity of their regulators through feedback loops (Fig. 1). Depletion of Cdc10 (an MBF subunit) blocks the cell in G1 phase, while *cdc18Δ* enters into mitosis without DNA replication resulting in a "cut" phenotype (Nishitani and Nurse 1995; Sveiczer et al. 2004; Morgan 2006).

Endoreduplication

The first example for endoreduplication in fission yeast came during studying Cdc2 kinase, the catalytic subunit of both S- and M-phase CDKs (Broek et al. 1991). *Cdc2*^{ts} mutants blocked in post-replicative G2 phase of the cycle were nitrogen starved and heat-shocked (49°C) for a short time in order to increase protein degradation. When cells were released from the *cdc2*^{ts} block, they were re-replicating their DNA rather than entering into mitosis. This suggested that cells "forgot" their cell cycle stage as a consequence of these treatments. Later experiments have shown that the "memory" molecule lost is the Cdc2/Cdc13 complex responsible for mitotic initiation and progression. When the Cdc13 expression was turned off by germinating spores deleted for *cdc13* (*cdc13Δ*), cells underwent through multiple round of DNA replications (Hayles et al. 1994).

In the absence of Cdc2/Cdc13, cells cannot enter into mitosis but the activity of the S-phase CDK (Cdc2/Cig2) oscillates and drives G1-S-G2-G1 endoreduplication cycles. The Cig2 cyclin level is low in G1, while its inhibitor Rum1 is high which allows licensing of replication origins in G1. The increasing level of Cdc2/Cig2 complex downregulates Rum1 after reaching a threshold and S phase is initiated. However, the high Cig2 level cannot be maintained because Cdc2/Cig2 turns off its own transcription factor (MBF). This effect results in a decrease of Cig2 level in G2 phase and makes possible the re-accumulation of Rum1 in the absence of Cdc2/Cdc13. The rise of Rum1/Cig2 ratio resets the cells back to G1 where relicensing of origins can take place. The boundary between G2 and G1 phases are experimentally not defined yet, but multiple rounds of DNA replication are well separated (Hayles et al. 1994; Novak and Tyson 1997; Kiang et al. 2009). *Cdc13Δ* cells undergo periodic relicensing shown by *cdc13Δcdc10*^{ts} and *cdc13Δrum1Δ* double mutants. *Cdc13Δ* cells are arrested in G1 after depletion of the MBF component, because the synthesis of both the licensing factor and Cig2 is blocked. In contrast, Rum1 depletion stops the endoreduplication in G2. The *Cdc13ΔCig2Δ* double mutant cannot re-replicate either because Cig2 is required to initiate S phase (Hayles et al. 1994; Novak and Tyson 1997; Kiang et al. 2009).

Overexpression of the CDK inhibitor (Rum1) leads to a phenotype similar to *cdc13Δ* (Moreno and Nurse 1994). After upregulation of Rum1 level, the first G1 phase is much longer than in wild-type, because the stoichiometric inhibitor has to be eliminated before the cell can enter into S phase. Rum1 inhibits both Cdc2/Cig2 and Cdc2/Cdc13 complexes, but it is a weaker inhibitor for the S-phase CDK than for the Cdc2/Cdc13. Therefore, the high Rum1 level creates



Endoreplication, Fig. 2 Time course results of two different types of over-replication. The relative activity/level of the main regulatory components are plotted on the "y" axis. The particular cell cycle phases are indicated on the figure. (**a**) Endoreduplication. Fission yeast cells are blocked and released from *cdc25*^{ts} then Cdc13 was deleted from the G1-S-G2 phases are alternating without mitosis resulting in the proper doubling of the DNA content. (**b**) Re-replication. Fission yeast cells are blocked and released from *cdc25*^{ts} then Cdc18 licensing factor was overexpressed. The cells enter S phase and have a continuous DNA replication without any relicensing period

a complete block of mitosis while Cdc2/Cig2 can be activated periodically and triggers S phases (Moreno and Nurse 1994; Novak and Tyson 1997).

Re-replication

Overexpression of the Cdc18 licensing factor results in another type of over-replication called re-replication. Cdc18 plays an important role in the initiation of DNA replication and its protein level is peaking during S phase. Overproduction of Cdc18 protein is able to spoil the control of origin firing. Upregulation of Cdc18 level by thiamine repressible *nmt1* promoter blocks mitosis and cells show enlarged nucleus (Fig. 2b). Cdc18-induced over-replication is clearly different than endoreduplication. The FACS profiles are much more smeared referring to incomplete rounds of DNA replications or multiple replication initiations at origins rather than G1-S-G2-G1 type oscillation. The large amount of Cdc18 keeps the cell in the S phase and promotes the re-initiation of the replication origins over and over again. This continuous DNA replication does not require any Cdc10 function (*cdc18*^{OP}*cdc10*^{ts} double mutant re-replicates) suggesting that cells are not passing through Start during re-replication. Even more Cdc18-driven over-replication does not depend on any further protein synthesis (Cdc18 overexpression followed by cycloheximide addition does not block re-replication) (Nishitani and Nurse 1995).

Controlling Over-Replication in Higher Eukaryotes

Higher eukaryotes have some extra pathways which have to work coordinately to block more than one DNA replication per cycle. A protein called geminin and the Cul4-Ddb1^{Cdt2} pathway can suppress over-replication by inhibiting the licensing factors. Emi1 directly can regulate APC activity negatively. However, the essential element of the system is again CDK, and its level clearly defines that the cell could go to M phase or should be blocked in an over-replication state (Arias and Walter 2007; Porter 2008).

Cross-References

- ▶ Cdk Inhibitors
- ▶ Cell Cycle Checkpoints
- ▶ Cell Cycle of Mammalian Cells
- ▶ Cell Cycle, Fission Yeast
- ▶ Cell Cycle Modeling, Differential Equation
- ▶ Cell Cycle, Physiology
- ▶ Cyclins and Cyclin-Dependent Kinases
- ▶ DNA Replication
- ▶ Mitosis

References

Arias EE, Walter JC (2007) Strength in numbers: preventing rereplication via multiple mechanisms in eukaryotic cells. Genes Dev 21:497–518

Broek D, Bartlett R, Crawford K, Nurse P (1991) Involvement of p34cdc2 in establishing the dependency of S phase on mitosis. Nature 349:388–393

Hayles J, Fisher D, Woollard A, Nurse P (1994) Temporal order of S phase and mitosis in fission yeast is determined by the state of the p34cdc2-mitotic B cyclin complex. Cell 78:813–822

Kiang L, Heichinger C, Watt S, Bähler J, Nurse P (2009) Cyclin-dependent kinase inhibits reinitiation of a normal S-phase program during G2 in fission yeast. Mol Cell Biol 29:4025–4032

Moreno S, Nurse P (1994) Regulation of progression through the G1 phase of the cell cycle by the rum1+ gene. Nature 367:236–242

Morgan D (2006) The cell cycle – principles of control. Oxford University Press, UK

Nishitani H, Nurse P (1995) p65cdc18 plays a major role controlling the initiation of DNA replication in fission yeast. Cell 83:397–405

Novak B, Tyson JJ (1997) Modeling the control of DNA replication in fission yeast. Proc Natl Acad Sci USA 94:9147–9152

Porter AC (2008) Preventing DNA over-replication: a Cdk perspective. Cell Div 3:3

Sveiczer A, Tyson JJ, Novak B (2004) Modelling the fission yeast cell cycle. Brief Funct Genomic Proteomic 2:298–307

Enhancer

Jianhua Ruan
Department of Computer Science, University of Texas at San Antonio, San Antonio, TX, USA

Definition

An enhancer is a short segment of DNA that can be bound by proteins (transcription activators) to enhance transcription levels of specific genes. An enhancer can be located either upstream or downstream to the transcriptional start site, and does not need to be immediately next to the genes it regulates.

Ensembl

Jingky Lozano-Kühne
Department of Public Health, University of Oxford, Oxford, UK

Definition

A database that provides genomic information on human genome, other vertebrates, and model organisms (Hubbard et al. 2002).

References

Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, Durbin R, Eyras E, Gilbert J, Hammond M, Huminiecki L, Kasprzyk A, Lehvaslaiho H, Lijnzaad P, Melsopp C, Mongin E, Pettett R, Pocock M, Potter S, Rust A, Schmidt E, Searle S, Slater G, Smith J, Spooner W, Stabenau A, Stalker J, Stupka E, Ureta-Vidal A, Vastrik I, Clamp M (2002) Ensembl genome database project. Nucleic Acids Res 30(1):38–41

Ensemble

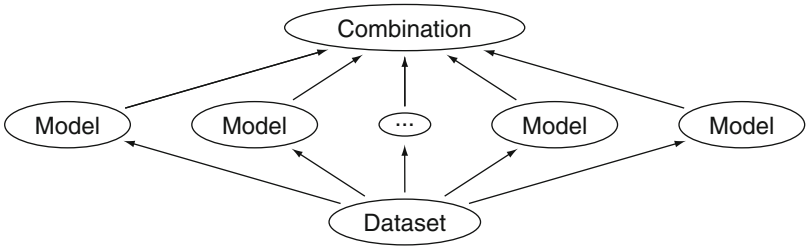
Celine Vens
Department of Computer Science, Katholieke Universiteit Leuven, Leuven, Belgium

Synonyms

Committee; Multiple classifier system

Definition

An ensemble is a collection of learning models (▶ Model Validation, Machine Learning), whose



Ensemble, Fig. 1 Schematic representation of ensembles

outputs are combined to form the output of the ensemble. Ensembles can be used for several machine learning tasks, such as predicting class labels (▶ Classification), predicting numeric outcomes (▶ Regression), or ▶ clustering. Therefore, the individual models can be combined in several ways.

Characteristics

An ensemble learning algorithm constructs a set of base learners and produces an output by combining the outputs of these base learners, by Fig. 1. While ensembles have been proposed for unsupervised learning (▶ Learning, Unsupervised) tasks, such as ▶ clustering, they are especially popular for supervised learning (▶ Learning, Supervised) problems, where individual predictions can be combined by taking a (weighted or unweighted) vote for ▶ classification tasks, and by taking the average in the case of ▶ regression.

Ensemble techniques are popular prediction methods because of their excellent generalization performance. A necessary and sufficient condition for an ensemble of classifiers to be more accurate than any of its individual members is that the classifiers are accurate and diverse (Hansen and Salamon 1990). An accurate classifier does better than random guessing on new examples. Two classifiers are diverse if they make different errors on new data points. The underlying principle is that these uncorrelated errors get corrected by the voting or averaging procedure. Dietterich (1997) gives a detailed explanation why ensembles can improve the predictive performance of their base classifiers.

Ensembles can be constructed either by using a different learning algorithm in each iteration, or by using the same algorithm, but training it in different ways. The latter can be accomplished by introducing some diversity, e.g., by manipulating the training set, by manipulating the descriptive attributes or the target attribute (in the case of prediction tasks), by injecting randomness into the base learning algorithms, or by a combination of several of these techniques. Many different ensemble learning techniques have been proposed. The most well-known include ▶ bagging (Breiman 1996) and boosting (Freund and Shapire 1997), which both manipulate the training examples.

Ensemble classifiers have been applied in several bioinformatics domains, such as protein folding (Shen and Chou 2006), cancer classification (Tan and Gilbert 2003), and gene function prediction (Guan et al. 2008).

Cross-References

- ▶ Bagging
- ▶ Classification
- ▶ Clustering
- ▶ Learning, Supervised
- ▶ Learning, Unsupervised
- ▶ Model Testing, Machine Learning
- ▶ Regression

References

Breiman L (1996) Bagging predictors. Mach Learn 24(2):123–140
Dietterich TG (1997) Machine-learning research. AI Magazine 18(4):97–136
Freund T, Shapire RE (1997) A decision-theoretic generalization of online learning and an application to boosting. J Comput Syst Sci 55(1):119–139
Guan Y, Myers CL, Hess DC, Barutcuoglu Z, Caudy AA, Troyanskaya OG (2008) Predicting gene function in a hierarchical context with an ensemble of classifiers. Genome Biol 9(Suppl 1):S3
Hansen L, Salamon P (1990) Neural network ensembles. IEEE Trans Patt Anal Mach Intell 12:993–1001
Shen HB, Chou KC (2006) Ensemble classifier for protein fold pattern recognition. Bioinformatics 22(14):1717–1722
Tan AC, Gilbert D (2003) Ensemble machine learning on gene expression data for cancer classification. Appl Bioinform 2(3):S75–S83

Entity Disambiguation

▶ Term Disambiguation, Text Mining

Entity Identification

▶ Entity Mention Normalization

Entity Mention Normalization

Jörg Hakenberg
Department of Computer Science and Department of Biomedical Informatics, Arizona State University, Tempe, AZ, USA

Synonyms

Entity identification; Grounding

Definition

In ▶ text mining and data integration, entity mention normalization refers to the process of mapping a (biomedical) ▶ named entity, such as a gene, disease, or species, to a unique identifier provided by an authority. For genes, such authorities are NCBI's EntrezGene (http://www.ncbi.nlm.nih.gov/sites/entrez/?db=gene), HUGO for human genes (http://www.genenames.org/), MGI for murine genes (http://www.informatics.jax.org/); for diseases, identifiers could map to UMLS concepts (http://www.nlm.nih.gov/research/umls/).

Challenges for entity mention normalization arise from synonymity, where an entity can be referred to by many names; and in particular from homonymity, where one and the same name can refer to multiple entities. As an example, "Fas" can refer to the "TNF receptor superfamily, member 6" gene, as well as to "fatty acid synthase." In addition, entity mention normalization has to map the gene to the correct organism, as both are known for various species.

Cross-References

▶ Named Entity Recognition
▶ Text Mining
▶ Text Mining, Tools
▶ Word Sense Disambiguation

Entrez

Jingky Lozano-Kühne
Department of Public Health, University of Oxford, Oxford, UK

Definition

Entrez is the integrated, text-based search and retrieval system used at the National Center for Biotechnology Information for the major databases, including PubMed, Nucleotide and Protein Sequences, Protein Structures, Complete Genomes, Taxonomy, and others (NCBI 2011).

References

National Center for Biotechnology Information (2011) Databases. World Wide Web. http://www.ncbi.nlm.nih.gov/Database/. Accessed 24 May 2011

Entrez Genome Project

▶ NCBI BioProject Genome Resources

Entropy

Daniel Polani
Adaptive Systems Research Group, School of Computer Science, University of Hertfordshire, Hatfield, UK

Synonyms

Entropy, Information Theory

Definition

Uncertainty (or *Entropy*, in the sense of Shannon's information theory) is a quantitative measure for the uncertainty of the outcome of a random experiment modeled by a probabilistic variable. Given a probabilistic variable X with outcome set $\mathcal{X}$ and probability p, the entropy $H(X)$ of X is defined as:

$$H(X) := -\sum_{x \in \mathcal{X}} p(x) \log\, p(x), \tag{1}$$

with the convention that $0 \log 0 \equiv 0$. The logarithm is often taken with basis 2, and the outcome is then expressed in *bits*.

Extreme Cases: The entropy is never negative. It is 0 if X has only one outcome x', that is, $p(x') = 1$ for exactly the outcome x', and $p(x) = 0$ for any other outcome x. The entropy $H(X)$ reaches its maximal value of $\log |\mathcal{X}|$ (with $|\mathcal{X}|$ the number of possible outcomes) if all possible outcomes x have the equal probability $p(x) = \frac{1}{|\mathcal{X}|}$.

Entropy of Joint Variables: For joint probabilistic variables X and Y, the entropy is given by

$$H(X,Y) = -\sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p(x,y) \log p(x,y)$$

Conditional Entropy: For joint probabilistic variables X and Y, one defines the *conditional entropy* of Y given a particular outcome x by:

$$H(Y|X = x) = -\sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x). \tag{2}$$

From this, one defines the conditional entropy of Y given X as the average over all conditional entropies given particular outcomes x:

$$H(Y|X) = \sum_{x \in \mathcal{X}} p(x) H(Y|X = x). \tag{3}$$

Characteristics: The entropy measures the uncertainty about the outcome of the experiment or observation X in advance. In the special case of only one outcome, there is no uncertainty about the outcome, and thus the entropy vanishes. In the other special case, all outcomes are equally probable. This corresponds to the case that no outcome can be a priori favored to any other outcome, and thus the entropy, that is, uncertainty is maximal.

Maximum Entropy Principle: For an experiment where there is no prior knowledge or expectation on the outcome, one assumes a probability of maximal uncertainty, that is, maximum entropy. If nothing is known, one therefore will assume equiprobable probabilities on the outcomes. This generalizes to the case when there is prior knowledge. In that case, one seeks a probability distribution that achieves the maximum possible entropy value while respecting this prior knowledge (Jaynes 1957). This constitutes the least biased prior on the probability distribution that is consistent with the prior knowledge.

References

Jaynes ET (1957) Information theory and statistical mechanics. Phys Rev 106(4):620–630

Entropy, Information Theory

▶ Entropy

Entropy Reduction

▶ Information

Entry into Mitosis

▶ Cell Cycle Transitions, G2/M

Environment

C. Vyvyan Howard
Centre for Biomedical Sciences, University of Ulster, Coleraine, UK

Definition

The environment can be defined as the totality of the biological and nonbiological influences that act upon

an organism, a population of organisms, or an ecological system and that can influence survival. Biological factors include the organisms, their food, and their interactions. Nonbiological factors can be divided into physical and chemical and can act in combinations. Purely physical influences include all electromagnetic radiations, temperature, and meteorology. Purely chemical influences include soil constitution, water and air quality, and chemical pollution. Organisms can respond to environmental changes by evolutionary adaptations.

Cross-References

▶ Cancer and Environmental Influences

Environmental Genome

▶ Metagenome

Environmental Genomics

▶ Metagenomics

Environmental Metabolomics

▶ Metametabolomics

Environmental Proteomics

▶ Metaproteomics

Environmental Transcriptomics

▶ Metatranscriptomics

Enzyme Databases

▶ Metabolic Networks, Databases

Enzyme-Ligand Interaction Databases

▶ Specialized Metabolic Component Databases

Epifluorescence Microscope

▶ Fluorescence Microscopy

Epifluorescence Microscopy

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

Definition

Epifluorescence microscopy is a method of fluorescence microscopy that is widely used in life sciences. The excitatory light is passed from above (or, for inverted microscopes, from below), through the objective lens and then onto the specimen instead of passing it first through the specimen.

Cross-References

▶ Spectroscopy and Spectromicroscopy

Epigenetic Modification

▶ Post-Replication Modification

Epigenetic 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

Epigenetic regulation is the inherited regulation of phenotype or gene expression caused by nongenetic mechanisms other than underlying DNA sequence.

Cross-References

▶ Gene Regulation

Epigenetics

Vani Brahmachari and Shruti Jain
Dr. B. R. Ambedkar Center for Biomedical Research, University of Delhi, Delhi, India

Synonyms

Gene regulation

Definition

The term *Epigenetics* was coined by C. H. Waddington in 1942. Ptashne and Gann (2002) defined it as "...a change in the state of expression of a gene that does not involve a mutation, but that is nevertheless inherited in the absence of the signal (or event) that initiated the change." The Greek prefix *epi* meaning "on the top of" implies that the epigenetic features override the effect of the nucleotide sequence of the gene. Epigenetics impacts gene expression without changing the nucleotide sequence of the gene, which can be maintained through cell division. Epigenetic markings can be considered as signatures for the transcriptional state of the gene.

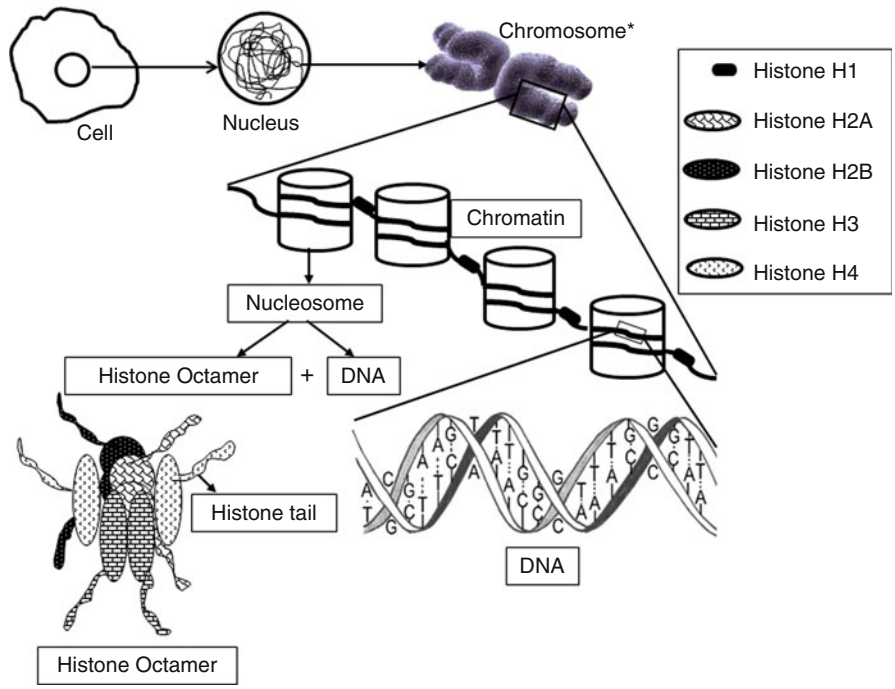
Characteristics

Epigenetics operates through ▶ post-replication modification of DNA and the ▶ post-translational modification of ▶ chromatin proteins. Unlike DNA sequence which can vary between individuals, epigenetic variations can be detected not only between individuals, but also between tissues and as a function of time in terms of development and aging of an organism. Epigenetics provides an interface between genetic constitution (genotype) of an organism and the environment.

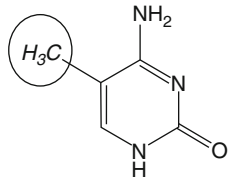
DNA in the nucleus of eukaryotic cells is packed compactly through interaction with specific nuclear proteins called ▶ histones, leading to the formation of the basic units of chromatin called ▶ nucleosomes (Fig. 1). Chromatin is the template for transcription and the consequent interaction of proteins that regulate gene expression. Though nucleosomes are present all along the ▶ genome, the level of compaction differs between regions of the genome. This brings about variation in accessibility of the genome to regulators of transcription. The more compact regions of chromatin are less transcriptionally active and they constitute ▶ heterochromatin; while the more open and active regions form the ▶ euchromatin. The epigenetic marking can be either on DNA or on the histones.

Epigenetic Marks on DNA

The DNA generally consists of the four bases adenine (A), guanine (G), cytosine (C), and thymine (T). These can be modified by covalent linkage of methyl groups at specific positions within their ring structure, after replication of DNA during cell division. The most common modification seen in the eukaryotic cells is the methylation of C at the fifth carbon, leading to the formation of 5-methyl cytosine (Fig. 2). The cytosine residues with guanine as the 3′ neighbor, in the dinucleotide CpG are methylated. This methylation status can be detected by ▶ methylation-sensitive restriction endonucleases or by direct sequencing following ▶ bisulfite conversion (Frommer et al. 1992). Such analysis has revealed tissue-specific methylation patterns, the unmethylated DNA being associated with actively transcribing genes and methylated DNA with silenced genes. The unmethylated CG sites, in regions



Epigenetics, Fig. 1 Diagrammatic representation of packaging of DNA in the nucleus. Chromatin as shown represents the "Bead on a string" structure. *The chromosome shown is the structure chromosomes assume during cell division and are referred to as metaphase chromosomes. The tail of histone H4 is marked, as shown, and similar extensions called histone tails are present in all the histone monomers and these are most frequently modified by acetylation/methylation/phosphorylation as described in the text

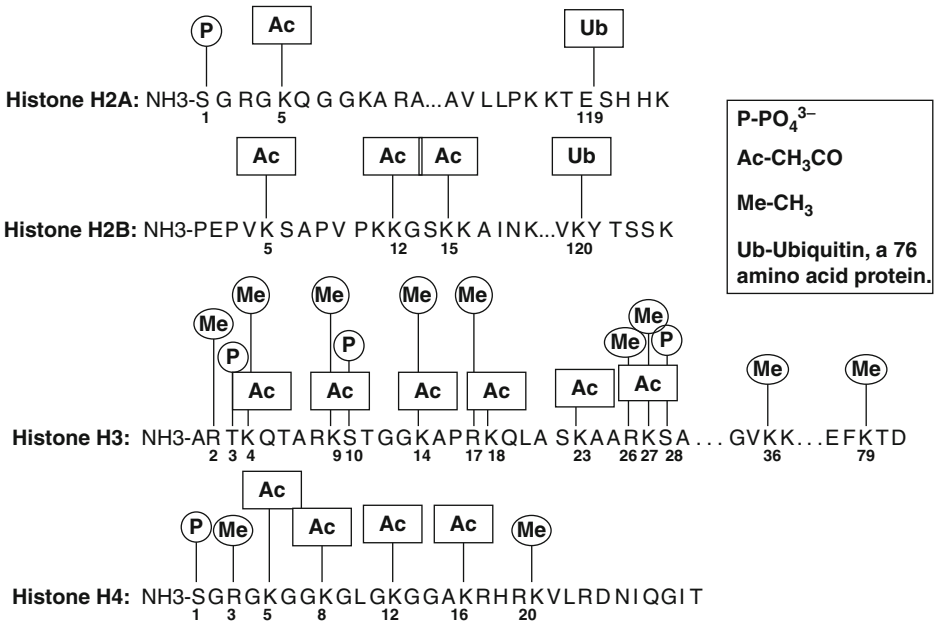


Epigenetics, Fig. 2 Structure of 5-methyl cytosine. The methyl group added after replication is marked

known as *CpG islands* are observed in constitutively expressing genes. The methylated cytosine inhibits the *trans*-acting factors from binding to the regulatory regions producing a closed chromatin structure. Alternatively, an inhibitory protein like methylated cytosine binding protein 2 (▶ MeCP2) can also bind to methylated DNA to produce closed chromatin structure and hence, regulate the expression of gene (Bogdanović and Veenstra 2009). Therefore, undermethylation is associated with active transcription of the gene. The palindromic nature of CG dinucleotide ensures the maintenance of methylation after DNA replication, thus, satisfying the "memory" component of the epigenetic mechanism. As the replication fork moves, the newly synthesized DNA strand, carrying hemi-methylated CpG dinucleotides, is methylated by DNMT (DNA methyltransferase) resulting in conserving the DNA methylation pattern on the genome through cell division. However, demethylation of the DNA can occur only as a consequence of failure to methylate the C residues in CG dinucleotides on the newly synthesized DNA strand. Presently there is no direct evidence for the existence of enzymes that can remove methyl group from cytosine residue in the DNA (DNA demethylase).

Epigenetic Marks on Chromatin Proteins

The ▶ histone proteins in the chromatin structure are also important targets of epigenetic modifications. There are organisms like the yeast (*Saccharomyces cerevisiae*) and the fruitfly (*Drosophila melanogaster*) where DNA methylation has not been detected so far. However histone modifications are known in almost all eukaryotes. Histone modifications are carried out after translation of the protein; hence they are post-translational modifications (Latchman 2005). Based on the three-dimensional structure of histones, the core and the N-terminal tails are distinguished from each other (Fig. 1). Post-translational modifications are known to occur both on the core and the N-terminal tails. Most often the lysine and arginine residues in the histones undergo covalent modifications like



Epigenetics, Fig. 3 Potential sites of post-translational modifications of histone tails of nucleosomes. The alphabets after -NH3 are the single letter symbols for amino acids. The chemical nature of the modifications is shown. *Ac* acetylation, *Me* methylation, *P* phosphorylation, *Ub* ubiquitination (Jaskelioff and Peterson 2003). The amino acids most frequently modified are S-serine, K-lysine, E-glutamic acid, R-arginine, and T-tyrosine. The numbers below the amino acids indicate their position in the amino acid sequence of the respective histone

acetylation, methylation, phosphorylation, ubiquitination, and sumolyation (Fig. 3). These post-translational modifications in turn can be activating or silencing marks, which flag the recruitment of different *trans*-acting regulatory factors that generally occur as protein complexes. Apart from post-translational modifications, variant histones like H2A.Z and H3.3 coded by different genes in the genome are also known to be associated with active genes (Henikoff et al. 2004) and H1b with silenced genes (Lee 2004).

Each modification of histones is abbreviated to indicate the histone modified, the chemical nature of the modification, and the amino acid modified and its position with in the histone protein. For example, histone H3 tri-methylated at 27th lysine residue from its N-terminal end is represented conventionally as H3K27me3. The repressive and the activating marks on the histones are listed in Table 1.

Linking Epigenetic Modifications to Gene Expression State

The DNA methylation, brought about by DNA methyl transferases (DNMT), plays a key role in converting the chromatin structure to either inactive or active state. One of the mechanisms is that the methylated-CpG binds to specific proteins resulting in a compact closed chromatin structure occluding the DNA from transcription factors, and thus repressing gene expression and 5-methyl cytosine binding protein 2 (MeCP2) is one such protein as described earlier. These proteins can recruit various other histone-modifying enzymes like ► histone deacetylases and other complexes. Therefore, conversion of a genomic region into a transcriptionally inactive state is a multistep process with different proteins participating in the process in different genomic contexts.

The next most important repressive epigenetic regulation is brought about by post-translational histone modifications like deacetylation, ubiquitination, sumolyation, and methylation at specific amino acid residues of different histones. Histone deacetylation, brought about by histone deacetylase, is an important step in converting an actively transcribing region into a repressed state. The proteins acting as repressors in fact bring about repression by recruiting histone deacetylase to the desired genomic location. Deacetylation results in the positive charge on the lysine residue of the histone molecule, leading to a more compact chromatin structure and transcription repression. Similarly, addition of small

Epigenetics, Table 1 Histone modifications and functions

Histone modification	Histone (residues modified)	Functional regulated
Acetylation	H3(K9,K14,K18,K56)H4(K5,K8,K13,K16)	Transcriptional activation
	H3K56	DNA repair
	H4K16	Chromosome condensation
Methylation	H3 (K4,K36,K79), H3 (R17,R23), H4(R3)	Transcriptional activation
	H3 (K9,K27), H4(K20)	Transcriptional repression
	H4K20	DNA repair
Phosphorylation	H3S10	Transcriptional activation
	H2AS129, H4S1	DNA repair
	H3S10, H3T3	Chromosome condensation
Ubiquitylation	H2AK119, H2BK20	Transcriptional activation
	H3, H4, H2A	DNA repair
Sumoylation	H4, H2A, H2B	Transcriptional repression

proteins like ▶ ubiquitin or SUMO (small ubiquitin-related modifier) on an internal lysine residue in the histone is associated with repression of gene expression.

The epigenetic modification can be reversed to activate a silenced gene. Actively transcribing regions are devoid of DNA methylation. The conversion of a repressed gene with methylated cytosine residues into an active unmethylated state has to be done with DNA synthesis either during replication or repair and MeCP2-like proteins are phosophorylated leading to their disassociation from DNA and opening up of chromatin structure to facilitate entry of transcription machinery.

The relationship between histone modifications and active gene expression is direct as in the case of acetylation. Acetylation of specific lysine residues K9, 14, 18, 23, 27 and K5, 8, 12, 16 in H3 and H4, respectively, leads to loss of positive charge on lysine residues and weakens the association between histones and DNA, resulting in open chromatin state allowing the entry of transcription activators. However, violation of this correlation is observed in many cases, apparently repressive epigenetic marking on active genes which are explained by indirect effects. For example, ubiquitination is a repressive mark, but ubiquitination of H2B acts as an activating mark as it is accompanied with methylation on lysine residues at position 4 and 79. The phosphorylation of serine residues in histone H3 leads to activation of gene expression, and also enhances the ability of histone acetyl transferases (HATs) to acetylate lysine 14. On the other hand, dephosphorylation and deacetylation at lysine 10 and 14 of histone H4 are signatures for repression, but they are known to promote methylation at position 9 of H4 and hence some of the potentially active genes bear these modifications. The term "bivalent mark" has been used to describe such states (Bernstein et al. 2006). Thus the signature for expression state of genes in differentiated cells is the outcome of a combination of epigenetic modifications.

The observation of different modifications of histones within a given region of the genome has given rise to the concept of "Histone code" akin to the genetic code. The concept first put forward by Jenuwein and Allis (2001) suggests that the pattern of histone modifications in the genome generates a highly complex pattern that can be correlated with transcriptional status of the gene. The post-translational modifications of the histones confers a pattern of flagging the basic unit of chromatin in the nucleus to signal either the recruitment of transcription activating complexes like the ▶ trithorax complexes or silencing complexes like the ▶ polycomb complexes.

The memory component of epigenetics has a complex mechanism. It is observed that the epigenetic marks are always faithfully transmitted through mitosis, and sometimes, through meiosis like in cases of ▶ genomic imprinting and ▶ X-chromosome inactivation. The mechanisms of transcriptional memory are yet to be completely understood.

Cross-References

- ▶ Bisulfite Conversion
- ▶ Chromatin
- ▶ Epigenetics, Drug Discovery
- ▶ Euchromatin
- ▶ Genome
- ▶ Genomic Imprinting
- ▶ Heterochromatin
- ▶ Histones

- ▶ MeCP2
- ▶ Methylation-Sensitive Restriction Endonucleases
- ▶ Nucleosomes
- ▶ Polycomb Complexes
- ▶ Post-Replication Modification
- ▶ Post-translational Modifications
- ▶ Trithorax Complexes
- ▶ Ubiquitin and SUMO
- ▶ X Chromosome Inactivation

References

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125(2):315–326

Bogdanović O, Veenstra GJC (2009) DNA methylation and methyl-CpG binding proteins: developmental requirements and function. Chromosoma 118(5):549–565

Frommer M, Mcdonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Nat Acad Sci USA 89:1827–1831

Henikoff S, Furuyama T, Ahmad K (2004) Histone variants, nucleosome assembly and epigenetic inheritance. Tren Genet 20(7):320–326

Jaskelioff M, Peterson CL (2003) Chromatin and transcription: histones continue to make their marks. Nat Cell Biol 5:395–399

Jenuwein T, Allis CD (2001) Translating the histone code. Science 293:1074–1080

Latchman DS (2005) Gene regulation: a eukaryotic perspective. Taylor & Francis, London

Lee H (2004) Msx1 cooperates with histone H1b for inhibition of transcription and myogenesis. Science 304:1675–1678

Ptashne M, Gann A (2002) Genes and signals. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Epigenetics, Drug Discovery

Vani Brahmachari and Shruti Jain
Dr. B. R. Ambedkar Center for Biomedical Research, University of Delhi, Delhi, India

Synonyms

Drug metabolism; Drug targets, Epigenetics; Epigenome; Gene regulation; Pharmacogenomics

Definition

The response to treatment varies widely in patients with reference to beneficial effects as well as adverse reactions. It is established that there is a genetic basis for this variability, and ▶ pharmacogenomics is the study of these underlying genetic variations at the whole genome level. The outcome of drug treatment is a consequence of multiple steps culminating in the interaction between the therapeutically active molecule and the target. The genes coding for drug transporters and metabolizers and those coding for targets of drug action are important players in the final outcome of treatment for diseases. Therefore, the mechanisms that influence expression of the genes involved in ▶ pharmacokinetics and pharmocodynamics are important in drug discovery and disease treatment.

In addition to genetic variations, epigenetic marks on these classes of genes will impact drug metabolism. In the systems biology approach to drug discovery and optimization of existing drugs, it is important to study not only the genetic variations but also epigenetic variations between populations and individuals.

Characteristics

The discovery of novel drug targets and drugs to target them, as well as, discovering novel drugs for known targets are the aspects of drug discovery relevant for consideration in this section. Significance of epigenetic effects in the context of drug discovery could be in the following contexts:

1. Epigenetic changes in the genome under disease conditions as novel targets for treatment
2. Epigenetic variation in genes involved in drug metabolism in turn resulting in altered response to drugs
3. Drug molecules bringing about alteration in epigenetic profile of tissues and, thus, affecting gene expression

Epigenetic Changes in Disease Conditions

The tools, presently available, have made it possible to analyze the epigenome. *Epigenome* is the epigenetic marks both on DNA and the histones (▶ Epigenetics) at the whole genome level. Epigenome analysis has been carried out in many disease contexts specially cancers and significant differences in DNA methylation profile between normal and cancer cells from the same

individual are reported in literature (Lo and Sukumar 2008). Such data can be obtained by different approaches: (1) using methylation-sensitive restriction endonucleases (▶ Epigenetics) or (2) ▶ chromatin immunoprecipitation (ChIP) with antibodies directed against 5-methyl cytosine (▶ Epigenetics). These methods applied on the total genome of the cells under analysis will yield a pool of methylated DNA sequences which are sequenced to derive information on the DNA methylation pattern of the whole genome. It is observed that the genome of cancer cells have higher methylation (hypermethylation) than normal cells. Since DNA methylation leads to repression, there will be inappropriate gene silencing in cancer cells compared to normal cells. Pharmacological intervention by incorporating ▶ 5-azaCytosine, the non-methylatable analogue of cytosine has been shown to reverse gene silencing in cancer cells (Gilbert et al. 2004). Incorporation of this analogue also inhibits the activity of DNA methyl transferases (DNMT). The molecules which are used to alter the epigenome are designated as *epigenetic drugs* and are currently in use against hematological cancers like myelodysplastic syndrome (Gilbert et al. 2004). However, one has to consider the lack of selectivity of such analogues at the cell and the gene level which would have undesirable effects.

The pattern of histone modifications on the whole genome of cancer cells has also been compared with that of normal cells using ChIP approach, which has shown that there is lower acetylation of histones in cancer cells. Acetylated histones that are present in actively transcribed genes are deacetylated by the action of enzymes called histone decetylases (HDAC, ▶ Epigenetics). Deacetylation will lead to transcription repression, thus, altering gene expression profile of cancer cells compared to normal cells. Therefore the effort is to use inhibitors of histone deacetylation as therapeutic agents. Phenylbutyrate and buphenyl are inhibitors of histone deacetylases. There are two HDAC inhibitors approved for clinical application in cutaneous T-cell lymphoma (CTCL; Grant et al. 2010; Mercurio et al. 2010).

Epigenetic Influence on Drug Metabolism Genes

What Is Drug Metabolism?

Drugs are metabolized by complex biochemical pathways dedicated for metabolism of ▶ xenobiotics. Pharmacological compounds are metabolized by the same enzyme system, leading to the conversion of either a therapeutically active drug molecule into a therapeutically inactive form or a therapeutically inactive drug (referred to as ▶ prodrug) into a therapeutically active form. The enzymes involved in drug metabolism can be divided into two categories based on the reaction they catalyze: *Phase I* comprising of oxidizing, reducing, and hydrolyzing enzymes; *Phase II* comprising of methylating, sulphonating, and acetylating enzymes. Two most studied examples of drug-metabolizing enzymes are cytochrome P450 monooxygenase system and alcohol dehydrogenase, *ADH* which are oxidizing enzymes and cytosolic N-acetyltransferase (*NAT2*) which has acetylating activity. Various isoforms of cytochrome P450 monooxygenase system are encoded by genes *CYP1A1*, *CYP2A6*, *CYP2C9*, and *CYP2E1*.

Where and When Are They Active?

The principal organ of action of these genes is the liver as it is the first organ to be perfused by drugs absorbed in the gastrointestinal tract and hence, it has very high concentrations of the drug-metabolizing enzymes. Other sites include the kidney, gut, lung, and the skin.

Genetic Variations in the Drug Metabolism Genes

The genes coding for these enzymes vary between individuals both at genetic and epigenetic level resulting in variable drug response between individuals. The ▶ genetic polymorphisms in drug-metabolizing genes have given rise to distinct subgroups in the population that differ in drug response. Polymorphisms in ▶ microsatellite repeats and ▶ single nucleotide polymorphisms (SNPs) in genes coding for drug transporters, metabolizing enzymes, or drug targets lead to their functional variability. Thus, within a given population, individuals can be characterized as hypermetabolizers, poor metabolizers, and non-metabolizers. Genetic polymorphisms in drug-metabolizing enzymes may have more subtle, yet clinically important consequences for interindividual variability in drug response.

For example, SNPs in the gene *NAT2* can distinguish populations into two distinct subgroups: "slow acetylators" and "fast acetylators." *NAT2* codes for the enzyme N-acetyltransferase type 2 that conjugates substrates like ▶ isoniazid with an acetyl group, which is subsequently eliminated from circulation. The slow acetylators, therefore, have decreased ability to metabolize this class of drugs by NAT2 (Roden and George 2002). Similarly *CYP1A* gene polymorphisms can group individuals in to distinct subgroups. Further, the

combination of genetic polymorphism in genes coding for different drug-metabolizing enzymes like *NAT2* and isoforms of *CYP1A* determines the drug response profile of an individual. For example, increased *CYP1A*, coupled with slow acetylation results in reduced anticancer activity of amonafide (Ratain et al. 1996).

Epigenetic Influence on Drug Metabolism Genes

The role of epigenetic modifications in drug-metabolizing enzymes is well established in certain cases, for example, drug transporters like ▶ ATP-binding cassette B1 transporter (*ABCB1*) and solute carrier (*SLC22A8*) and *CYP1A*. The hypermethylation of CpG islands of *CYP1A1* involved in the metabolism of carcinogens and polycyclic aromatic hydrocarbons in cell lines like MCF-7 (human breast carcinoma) and HeLa (human cervical adenocarcinoma) results in its repression (Hirota et al. 2008). The acetylation of histone H3 is another chromatin remodeling modification which is well characterized in transcriptional activation of *ABCB1* gene (Hirota et al. 2008).

Drugs Modulating Epigenetic Profile

Apart from the drugs recently being developed as epigenetic drugs mentioned above, the case of volproic acid as an epigenetic drug is unique and creates a new paradigm to be considered in a system biology approach to drug discovery. Drugs currently in clinical use for certain diseases lead to epigenetic modulation of genes in addition to their effect on their target; they can have wider effect on the transcriptional profile of the target cells. For example, if a drug induces hypermethylation and thus, inactivation of the gene involved in its metabolism, it can lead to the accumulation of the drug and therefore toxicity. On the other hand, if the drug leads to hypomethylation it can result in higher expression of the gene. Therefore drugs directed at other targets can cause epimutations (alteration in epigenetic marking) leading to altered gene expression, without changing the DNA sequence. One of the drugs known to have such an effect is valproic acid (VPA, 2-propylpentanoic acid). It is widely used as an antiepileptic drug and for the therapy of bipolar disorders. Its action was initially found to be primarily related to neurotransmission and modulation of intracellular pathways. More recently, it is recognized as an antineoplastic agent as well, acting on cell growth, differentiation, and ▶ apoptosis. It has been demonstrated that valproic acid directly inhibits histone deacetylases. This property is important for neuroprotective mechanisms in several neurodegenerative diseases, in addition to its role as antiepileptic and antimalignancy drug (Monti et al. 2009).

Genetic and epigenetic variation in genes between individuals is an important issue to be considered in drug discovery. In addition, the nontarget effect of drugs not only on protein activity but also in terms of their epigenetic effects being recognized recently, demands particular attention and cannot be ignored. Interestingly similar to drug efficacy differences, adverse effects can be also negligible or absent in the background of the genetic and epigenetic variation profile of different populations.

Cross-References

- ▶ 5-azaCytosine
- ▶ Apoptosis
- ▶ ATP-Binding Cassette B1 Transporter
- ▶ Chromatin Immunoprecipitation
- ▶ Epigenetics
- ▶ Genetic Polymorphisms
- ▶ Isoniazid
- ▶ Microsatellite Repeats
- ▶ Pharmacogenomics
- ▶ Pharmacokinetics and Pharmacodynamics
- ▶ Phase I Enzymes
- ▶ Phase II Enzymes
- ▶ Prodrug
- ▶ Single Nucleotide Polymorphisms
- ▶ Xenobiotics

References

Gilbert J, Gore SD, Herman JG, Carducci MA (2004) The clinical application of targeting cancer through histone acetylation and hypomethylation. Clin Cancer Res 10:4589–4596

Grant C, Rahman F, Piekarz R, Peer C, Frye R, Robey RW, Gardner ER, Figg WD, Bates SE (2010) Romidepsin: a new therapy for cutaneous T-cell lymphoma and a potential therapy for solid tumors. Expert Rev Anticancer Ther 10:997–1008

Hirota T, Takane H, Higuchi S, Ieiri I (2008) Epigenetic regulation of genes encoding drug-metabolizing enzymes and transporters; DNA methylation and other mechanisms. Curr Drug Metab 9:34–38

Lo P-K, Sukumar S (2008) Epigenomics and breast cancer. Pharmacogenomics 9(12):1879–1902

Mercurio C, Minucci S, Pelicci PG (2010) Histone deacetylases and epigenetic therapies of hematological malignancies. Pharmacol Res 62:18–34

Monti B, Polazzi E, Contestabile A (2009) Biochemical, molecular and epigenetic mechanisms of valproic acid neuroprotection. Curr Mol Pharmacol 2(1):95–109
Ratain MJ, Mick R, Janisch L, Berezin F, Schilsky RL, Vogelzang NJ, Kut M (1996) Individualized dosing of amonafide based on a pharmacodynamic model incorporating acetylator phenotype and gender. Pharmacogenetics 6(1):93–101
Roden DM, George AL Jr (2002) The genetic basis of variability in drug responses. Nat Rev Drug Discov 1:37–44

Epigenome

▶ Epigenetics, Drug Discovery

Epistasis

Christoph Adami
Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI, USA

Synonyms

Mutation interactions

Definition

In genetics, epistasis refers to the situation where the effect of one gene on a particular ▶ mendelian trait is modified by another gene. Such genes are said to be epistatically interacting when producing the trait. More generally, epistasis can refer to mutations within a single gene that interact epistatically in conferring ▶ fitness to the gene. In that case, the effect of one mutation on fitness depends on the identity of another mutation within the same gene. Epistasis is related to but distinct from pleiotropy, where a single mutation or gene affects multiple traits.

Cross-References

▶ Fitness
▶ Mendelian Traits

Epitope

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

Synonyms

Antigenic determinant

Definition

"*Epitope*" is a concept of the "▶ *SpecificityType*" 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). "*Epitope*," or "antigenic determinant," identifies the part of the antigen (Ag) or of the peptide/major histocompatibility (pMH) that is recognized by the ▶ Paratope, or antigen-binding site, of the variable (V) domains (▶ Variable (V) Domain) of an immunoglobulin (IG) or antibody or of a T cell receptor (TR), respectively.

The "Epitope" concept has two leafconcepts (▶ IMGT-ONTOLOGY, Leafconcept):

- the "B cell epitope" identifies the part of the Ag that is recognized by the paratope of a specific IG or antibody
- the "T cell epitope" identifies the part of the pMH that is recognized by the paratope of a specific TR.

An "epitope" is "conformational" or "linear," based on its structure and its interaction with the paratope.

A "conformational" epitope (or "discontinuous" epitope) comprises amino acids that are close in the three-dimensional (3D) structure but distant in the primary amino acid sequence. Most B cell epitopes are conformational.

A "linear" epitope (or "contiguous" epitope) comprises amino acids that are in continuity in the primary amino acid sequence. T cell epitopes are usually

identified as "linear" when referring to the processed peptide (p) presented in the groove (G) domains (▶ Groove (G) Domain) of the major histocompatibility (MH) proteins. However, in IMGT-ONTOLOGY, the "T cell epitope" concept is identified as "discontinuous" as it comprises amino acids of the peptide and amino acids of the helices of the MH that bind to the TR V domains (▶ Variable (V) Domain).

Analysis of the paratope/epitope interactions in IG/Ag and TR/peptide/MH (TR/pMH) complexes with three-dimensional (3D) structures are available in IMGT/3Dstructure-DB (http://www.imgt.org) (Kaas et al. 2008).

"Epitope" is one of the two major concepts characteristics of the IG/Ag and TR/pMH interactions, "epitope" is on the antigen side, whereas the other one, "▶ Paratope," is on the antigen receptor (IG and TR) side.

Cross-References

- ▶ Chain Type
- ▶ Complementarity Determining Region (CDR-IMGT)
- ▶ Groove (G) Domain
- ▶ IMGT Collier de Perles
- ▶ IMGT Unique Numbering
- ▶ IMGT-ONTOLOGY
- ▶ IMGT-ONTOLOGY, IDENTIFICATION Axiom
- ▶ IMGT-ONTOLOGY, Leafconcept
- ▶ IMGT-ONTOLOGY, SpecificityType
- ▶ IMGT® Information System
- ▶ Immunogenetics
- ▶ Immunoinformatics
- ▶ Paratope
- ▶ Systems Immunology, Data Modeling and Scripting in R
- ▶ Variable (V) Domain

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

Kaas Q, Duprat E, Tourneur G, Lefranc M-P (2008) IMGT standardization for molecular characterization of the T cell receptor/peptide/MHC complexes. In: Schoenbach C, Ranganathan S, Brusic V (eds) Immunoinformatics. Immunomics reviews, series of Springer Science and Business Media LLC, Springer, New York, USA, Chap. 2, pp 19–49

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

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

Equilibrium

- ▶ Life Span, Turnover, Residence Time
- ▶ Lymphocyte Population Kinetics
- ▶ Steady State

Equilibrium Probability Distribution

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

Synonyms

Steady-state probability distribution

Definition

In the stochastic framework, it is difficult to define an equilibrium configuration in terms of the number of molecules as in a deterministic framework because any reaction, which changes the number of molecules, takes place in probability. In deterministic differential equations, an equilibrium means that there is no net change in the number or concentration of molecules. In contrast, for a stochastic model, there is an equilibrium probability distribution, i.e., a set of probabilities that

the system has certain number of molecules, even though the number of molecules may change. Generally, it is difficult to get the equilibrium probability distribution of a master equation analytically. However, for some simple cases, it can be obtained by recursion relations or Fokker–Planck equations.

References

Hasty J, Pradines J, Dolnik M, Collins JJ (2000) Noise-based switches and amplifiers for gene expression. Proc Natl Acad Sci 97:2075–2080

Reddy VTN (1975) On the existence of the steady state in the stochastic Volterra-Lotka model. J Stat Phys 13:61–64

eRF

▶ Release Factor, Translation

Ergodicity

M. Carmen Romano[1,2] and Ian Stansfield[2]
[1]Institute for Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen, UK
[2]Institute of Medical Sciences, University of Aberdeen, Aberdeen, UK

Definition

A system is said to be ergodic if the time average of the variable describing it, is equal to its ensemble average. Suppose that we have measured a signal $\left\{x_t^{S_m}\right\}$ at different points in time $t = 1,\ldots, T$, where S_m denotes the underlying system. The time average of the signal is then denoted by $\left\langle x_t^{S_m}\right\rangle_t = \frac{1}{T}\sum_{t=1}^{T} x_t^{S_m}$. Now assume that instead of measuring the signal from one single system S_m at different time points, we have a whole ensemble $\{S_m\}$, $m = 1,\ldots, M$ of identical systems, and we measure the signal $\left\{x_t^{S_m}\right\}$ at a fixed time point t for every system S_m, from which we then calculate the ensemble average $\left\langle x_t^{S_m}\right\rangle_m = \frac{1}{M}\sum_{m=1}^{M} x_t^{S_m}$. Then, if $\left\langle x_t^{S_m}\right\rangle_t = \left\langle x_t^{S_m}\right\rangle_m$, the system is said to be ergodic (Gardiner 2002).

Cross-References

▶ Stochastic Modeling of Translation Elongation and Termination

References

Gardiner CW (2002) Handbook of stochastic methods. Springer, Berlin

Error of Type I and Type II

Martin Carrier
Department of Philosophy, Bielefeld University, Bielefeld, Germany

Definition

Type I error, also known as false positive, concerns the mistaken acceptance of an incorrect assumption. Type II error, also known as false negative, designates the erroneous rejection of a true account.

Cross-References

▶ Non-empirical Values

ESCEC

▶ Experimental Standard Conditions of Enzyme Characterizations

Euchromatin

Vani Brahmachari and Shruti Jain
Dr. B. R. Ambedkar Center for Biomedical Research, University of Delhi, Delhi, India

Synonyms

Open chromatin

Definition

The actively transcribing regions of the genome which is in an open chromatin state is called euchromatin. The nomenclature "euchromatin," in contrast to ▶ heterochromatin, was initially given based on the lighter staining of the chromatin in these regions with dyes that bind to DNA, when the nuclei are observed under the microscope. Euchromatin corresponds to gene-rich regions and active transcription. It is also enriched in epigenetic marks associated with active transcription.

Cross-References

▶ Epigenetics
▶ Heterochromatin

Eukaryotic Translation Initiation Factor Interactions

Tao You[1], George M. Coghill[2] and Alistair J.P. Brown[3]
[1]Computational Biology, Innovative Medicines, AstraZeneca, Macclesfield, Cheshire, UK
[2]Institute for Complex System and Mathematical Biology, University of Aberdeen, Aberdeen, UK
[3]Institute of Medical Sciences, University of Aberdeen, Aberdeen, UK

Synonyms

Eukaryotic translation initiation pathway

Definition

Eukaryotic translation initiation is the process that allows a ribosome to assemble at the start codon on a messenger RNA. A set of proteins known as eukaryotic translation initiation factors (eIFs) orchestrates this process along with the two ribosomal subunits and initiator tRNA. Some of the events in this process happen in a specific temporal order. The cooperativity among protein interactions gives rise to some defined preinitiation complexes.

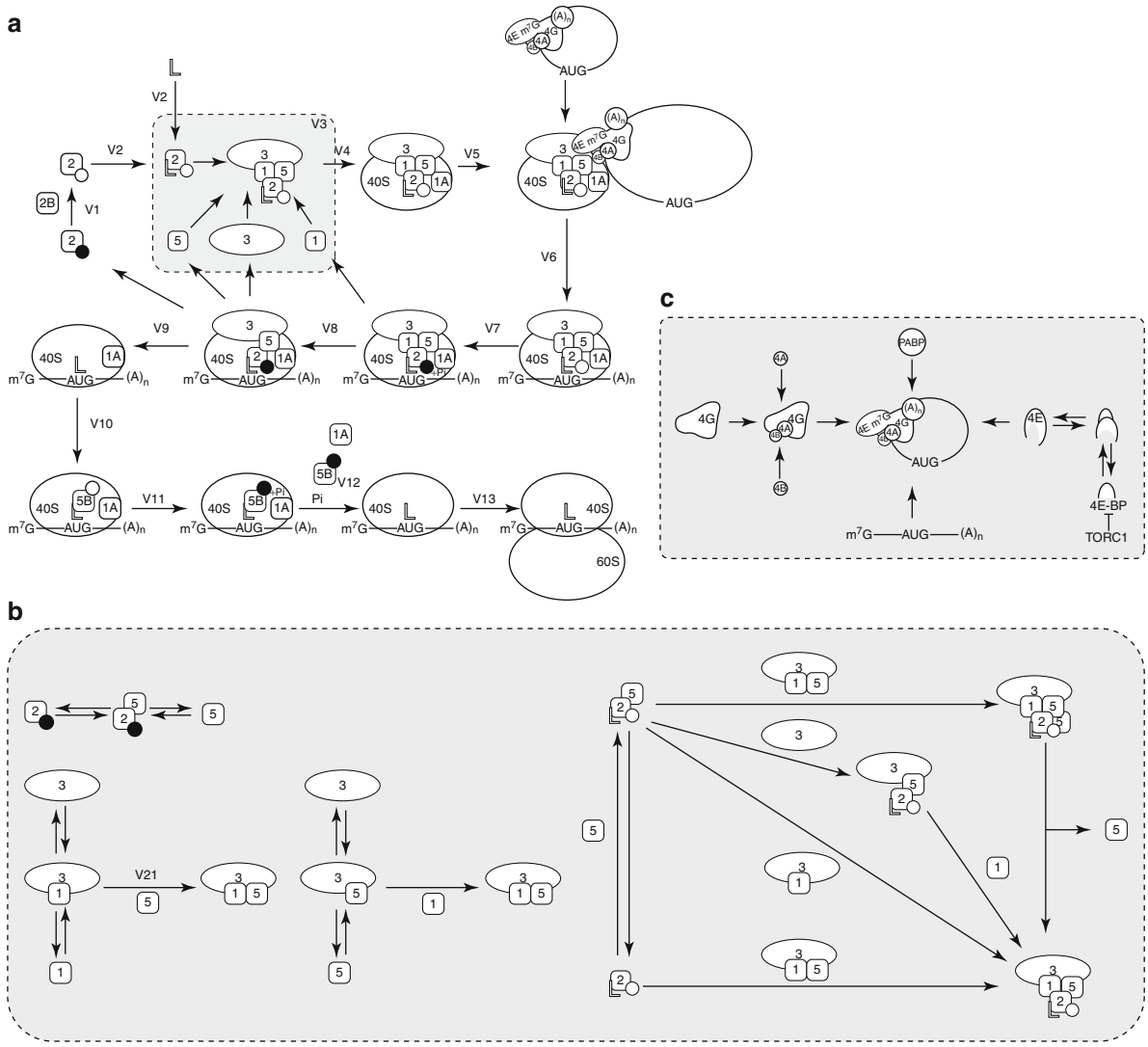
Characteristics

Eukaryotic Translation Initiation Process

In eukaryotic cells, eIF2 preferentially binds GDP. The guanine nucleotide exchange factor eIF2B replaces this GDP to GTP, but only on the unphosphorylated form of eIF2 (Fig. 1a: v_1), so as to comprise the ternary complex (i.e., TC, comprising eIF2, GTP, and initiator methionyl-tRNA, Met-tRNA$_i$). This ternary complex is essential for translation initiation (Fig. 1a: v_2). The interaction between eIF1A and the 40S ribosomal subunit promotes the dissociation of a ribosome and ensures that the 40S ribosomal subunit is open to receive other eIFs. Subsequently, eIF1, eIF3, eIF5, and ternary complex are recruited to the 40S ribosomal subunit thereby generating the 43S preinitiation complex, in which the initiator tRNA Met-tRNA$_i$ is held close to the ribosome's P site (Sonenberg and Hinnebusch 2009). This may happen via binding of the individual factors to the 40S ribosomal subunit in separate steps. On the other hand, these factors may interact through cooperative binding. There is evidence to suggest that these factors load onto the 40S ribosomal subunit in the form of a multifactor complex (Fig. 1a: v_3, v_4, Asano et al. 2000). Also various partial complexes have been isolated. Therefore, the multifactor complex may form via different routes, as shown in Fig. 1b (You et al. 2010).

eIF4E interacts with the 5′ cap structure of an mRNA (beginning of an mRNA), and poly-A binding protein (PABP) binds the poly-A structure at the 3′ end (mRNAs end), as depicted in Fig. 1c. eIF4G circularizes the mRNA by binding both eIF4E and PABP. eIF4G also recruits the mRNA helicase eIF4A and its activating partner eIF4B, which removes mRNA secondary structures that might impede the movement of the ribosome during scanning (below). This entire complex is known as eIF4F (Fig. 1c).

The 43S preinitiation complex loads onto the 5′ end of an mRNA via multiple interactions between eIF3/eIF5 and eIF4E/eIF4B complexes (Fig. 1a: v_5). Subsequently, the complex moves down the mRNA in a 3′ direction, seeking the 5′ proximal start codon, in a process known as scanning (Fig. 1a: v_6). eIF5 catalyzes the hydrolysis of GTP bound to eIF2, which is in equilibrium with GDP plus phosphate (Fig. 1a: v_7). Base pairing between the anticodon loop of the initiator tRNA Met-tRNA$_i$ and the start codon leads to a change in the conformation of the 43S preinitiation



Eukaryotic Translation Initiation Factor Interactions, Fig. 1 Schematic diagram of eukaryotic translation initiation

complex. This triggers release of eIF1 in a mechanism involving eIF1A and eIF5 (Fig. 1a: v_8, Lorsch and Dever 2010). This leads to release of the phosphate, GDP, eIF2, and eIF5, and commits the 40S ribosomal subunit in an immobilized closed conformation over the start codon (Fig. 1a: v_9, Sonenberg and Hinnebusch 2009).

Next, eIF5B·GTP is recruited to the 40S ribosomal subunit (Fig. 1a: v_{10}). The 60S ribosomal subunit joins the complex after the hydrolysis of this eIF5B-bound GTP (Fig. 1a: v_{11}) and the release of the phosphate, eIF1A and eIF5B (Fig. 1a: v_{12}, v_{13}). Upon completion of the translation initiation process, a ribosome is poised at the start codon ready for translation elongation (Sonenberg and Hinnebusch 2009).

Multifactor Complex

As described above, assembly of the 43S preinitiation complex relies on a multifactor complex. The cooperative binding between the individual factors that comprise multifactor complex means a partial complex has higher affinity with the binding partners compared with the individual components of the partial complex. This stimulation favors the formation of a complete

multifactor complex, and hence ensures the proper assembly of a functioning 43S preinitiation complex. The multifactor complex contains equimolecular amounts of the individual factors. However, these eIFs are not equimolar in vivo (von der Haar and McCarthy 2002). The optimal operation of the translation initiation apparatus relies on the complex interactions among the eIFs, and overexpression of a specific eIF does not necessarily lead to higher protein synthesis activities. For instance, overexpression of eIF5 sequesters eIF2·GDP in a nonproductive eIF5·eIF2·GDP complex, thereby reducing protein synthesis (Singh et al. 2006; You et al. 2010).

eIFs Control Protein Synthesis

According to metabolic control theory (▶ Metabolic Control Analysis), control over a multistep process is distributed among the enzymes that catalyze each step, and there is no specific rate-limiting step. Hence, control over protein synthesis is not exerted upon the activity of only one eIF. Instead, such control is distributed among different factors. A further complication is that the topology of the network that gives rise to the multifactor complex is not linear. To characterize the control of protein synthesis by a specific eIF in a small perturbation regime, a local flux control coefficient is introduced:

$$C_{\mathrm{eIF}}^{J} = \frac{\mathrm{eIF}}{J} \cdot \frac{\partial J}{\partial \mathrm{eIF}}, \qquad (1)$$

where J denotes the protein synthesis rate. This control coefficient is a normalized value, and it illustrates how sensitive the protein synthesis is to a small perturbation around a particular eIF activity, while keeping the activity of other eIFs constant.

Experimental work demonstrates that some of the eIFs exert a biphasic mode of control. C_{eIF}^{J} is small at small reduction in eIF, indicating that protein synthesis is relatively robust against small decrease in eIF activities. Below a certain threshold, C_{eIF}^{J} assumes a higher value. This means that protein synthesis becomes sensitive to the activity of an eIF once this eIF is reduced below certain value (Sangthong et al. 2007).

Computational modeling has provided additional insights. Significantly, modeling predicts that eIF2 has the highest impact on protein synthesis rate, while eIF1 is among those factors with the least impact on protein synthesis rate (Dimelow and Wilkinson 2009; You et al. 2010). eIF2 binds initiator tRNA Met-tRNA$_i$ as well as the GTP whose hydrolysis is central to start codon recognition. Therefore, reducing eIF2 activity would directly affect the levels of the multifactor complex and consequently the 43S preinitiation complex. Therefore, sufficient levels of eIF2 expression are important for efficient translation. On the other hand, eIF1 has a critical role in start codon recognition. Evolution might have led to the maintenance of high eIF1 activities in vivo to ensure that mRNA translation is not impeded at the initiation step by this factor.

Modes of Control in Translation Initiation

4E-BP (eIF4E binding protein) competitively binds eIF4E, and inhibits its interaction with eIF4G (Fig. 1c). In mammals the activity of 4E-BP is negatively regulated by mTORC1 (mammalian target of rapamycin complex 1). When nutrients are limiting, mTORC1 activity is lowered. This increases the activity of 4E-BP and eventually reduces protein synthesis via inhibition of eIF4E (Sonenberg and Hinnebusch 2009).

In addition, the activities of eIF2 and eIF2B are negatively regulated in response to stress conditions (see ▶ Translational Control of GCN4). A reduction in eIF2 activity leads to global translational arrest, while some specific mRNA species gain higher translation activity in mechanism involving upstream open reading frames.

Cross-References

▶ Translational Control of GCN4

References

Asano K, Clayton J, Shalev A, Hinnebusch AG (2000) A multifactor complex of eukaryotic initiation factors, eIF1, eIF2, eIF3, eIF5, and initiator tRNA(Met) is an important translation initiation intermediate in vivo. Genes Dev 14(19):2534–2546

Dimelow RJ, Wilkinson SJ (2009) Control of translation initiation: a model-based analysis from limited experimental data. J R Soc Interface 6(30):51–61

Lorsch JR, Dever TE (2010) Molecular view of 43S complex formation and start site selection in eukaryotic translation initiation. J Biol Chem 285(28):21203–21207

Sangthong P, Hughes J, McCarthy JE (2007) Distributed control for recruitment, scanning and subunit joining steps of translation initiation. Nucleic Acids Res 35(11):3573–3580
Singh CR, Lee B, Udagawa T, Mohammad-Qureshi SS, Yamamoto Y, Pavitt GD, Asano K (2006) An eIF5/eIF2 complex antagonizes guanine nucleotide exchange by eIF2B during translation initiation. EMBO J 25(19):4537–4546
Sonenberg N, Hinnebusch AG (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell 136(4):731–745
von der Haar T, McCarthy JE (2002) Intracellular translation initiation factor levels in *Saccharomyces cerevisiae* and their role in cap-complex function. Mol Microbiol 46(2):531–544
You T, Coghill GM, Brown AJP (2010) A quantitative model for mRNA translation in *Saccharomyces cerevisiae*. Yeast 27(10):785–800

Eukaryotic Translation Initiation Pathway

► Eukaryotic Translation Initiation Factor Interactions

Event Extraction Corpora

► Text Corpora, Molecular Event Extraction

Evidence-based Medicine

Ravi Iyengar
Department of Pharmacology and Systems Therapeutics, Mount Sinai School of Medicine, New York, NY, USA

Definition

Utilizing existing clinical trials, studies, and meta-analysis, as well as biological understanding to determine treatment choice and effectiveness.

Cross-References

► Systems Pharmacology

Evolution of Elongation Factor 1 Alpha

Kazuki Saito and Koichi Ito
Graduate School of Frontier Sciences, University of Tokyo, Chiba, Tokyo, Japan

Definition

Elongation factor-Tu (EF-Tu, bacterial)/eukaryotic elongation factor 1α (eEF1α also known as eEF1A) is a critical factor for protein synthesis and is highly conserved across all three domains of life. EF-Tu/eEF1α is a principal member of the homologous translational GTPase family. Among the members of the family, eEF1α-related factors, eRF3, HBS1, and EFL, are the particular interests of this entry. eRF3, a GTPase subunit of the polypeptide chain release factors, and HBS1, an mRNA quality control protein, are closely related to EF-Tu/eEF1α, thus comprising members of the EF1α family. eRF3 is completely and HBS1 is almost completely conserved among all eukaryotes. Their evolution evokes attention to the molecular mimicry between tRNA and protein factors. The discovery of EFL in some eukaryotes disrupts the seemingly perfect phylogenetic conservation of EF-Tu/eEF1α among all domains of life.

Characteristics

The Translational GTPase Superfamily

The translational GTPase superfamily consists of protein factors that share highly conserved motifs in the G-domain as well as in adjacent regions. Translational GTPases include protein factors that stimulate GTPase hydrolysis activity on the large subunit of the ribosome and play essential roles in the initiation, elongation, and termination of protein synthesis. Some factors within the translational GTPase family do not appear to be involved in translation but are engaged in the maintenance or biogenesis of ribosomes instead. The members of this family are largely divided into subfamilies according to their similarities with EF-Tu/eEF1α, EF-G/eEF2, and IF2/eIF5B (Table 1, see ► Translation Initiation and ► Translation Elongation). Typically, the genome of individual species encodes many members of EF-Tu/eEF1α and EF-G/eEF2 families whereas IF2/eIF5B is

Evolution of Elongation Factor 1 Alpha, Table 1 The translational GTPase family of eubacteria and eukaryotes

Ancestor type	Factor name		
	Eubacteria	Eukaryotes	Function
IF2	IF2	eIF5B	Initiation
EF-Tu	EF-Tu	eEF1α and/or EFL	Elongation
	CysN[a]	n.i.*	Adenosine-5′-phosphosulfate synthesis
	n.i.	eRF3[b]	Termination
	n.i.	HBS1[b]	mRNA surveillance
	n.i.	Ski7[b]	mRNA surveillance
	n.i.	eIF2gamma[c]	Initiation
	SelB[a]	EEFSEC	Incorporation of selenocysteine
EF-G	EF-G	EF2	Elongation
	RF3[a]	n.i.	Termination
	LepA[a]	n.i.	Unclear
	Tet[a]	n.i.	Tetracycline resistance
	TypeA[a]	n.i.	Unclear
	n.i.	SNU114[d]	mRNA splicing

*n.i. indicates "not identified"
For derived factors, references are as follows:
[a]Margus et al. 2007,
[b]Atkinson et al. 2008,
[c]Keeling et al. 1998,
[d]Fabrizio et al. 1997

encoded by a single gene. Presumably, in relation with numbers of genes, EF-Tu/eEF1α and EF-G/eEF2 have many more derivatives than IF2/eIF5B. Interestingly, EF-Tu/eEF1α has more derivatives in eukaryotes, and EF-G/eEF2 has more derivatives in bacteria (Fig. 1). Also interestingly, with the exception of SelB (▶ Translational Control by Cis RNA Elements, Bacteria; ▶ Selenocysteine-specific EF-Tu/eEF1α) specific EF-Tu/eEF1α (Keeling et al. 1998), derivatives of these two factors are unique to the different domains of life (Table 1 and Fig. 1).

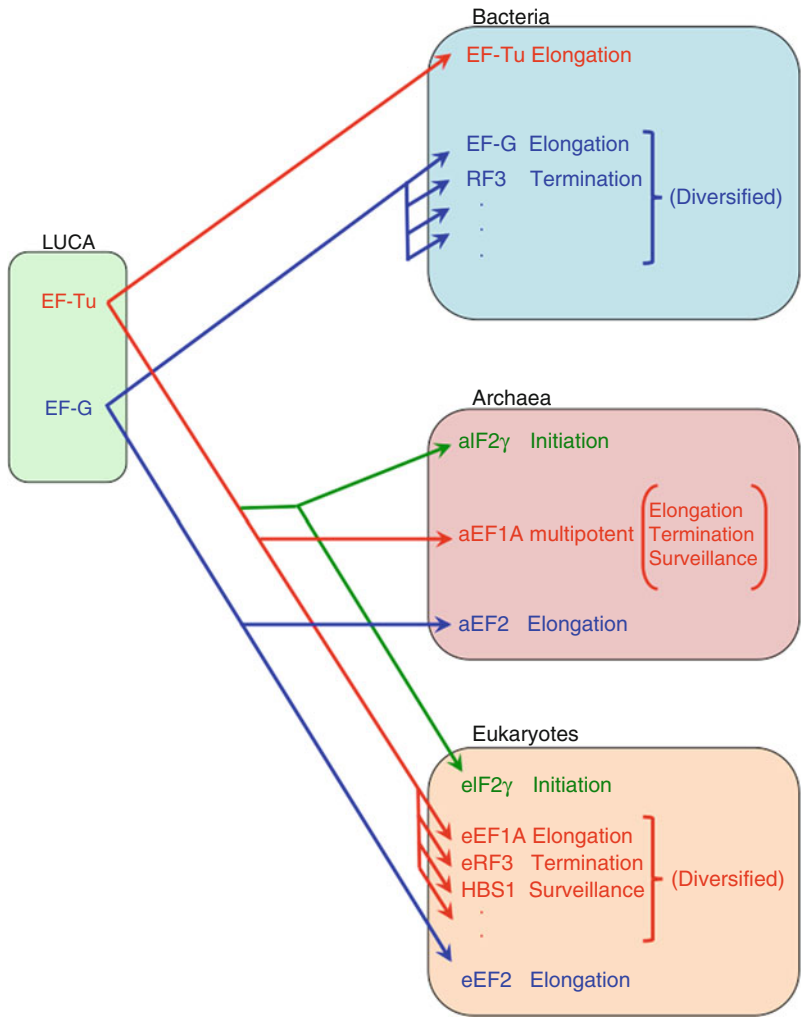
EF-Tu/eEF1α: The Key GTPase Elongation Factor

eEF1α is the eukaryotic ortholog of the bacterial elongation factor EF-Tu. The fundamental function of eEF1α is equivalent to that of EF-Tu. In polypeptide elongation cycles, eEF1α and eEF2, another GTPase that is the ortholog of bacterial EF-G, sequentially bind the ribosome. EF-Tu/eEF1α forms a complex with an aminoacyl-tRNA (aa-tRNA) in a GTP-binding state and delivers the aa-tRNA to the ribosomal A site. In the ribosomal A site, upon cognate base-pairing interaction between the codons of mRNA and anticodons of tRNA, EF-Tu/eEF1α hydrolyzes GTP, releases the aa-tRNA, and then exits the ribosome. These processes lead to the relocation of the aminoacyl-CCA stem of the aa-tRNA toward the peptidyl transferase center of the ribosome and result in elongation of the peptide chain. Subsequently, GTP-bound EF-G/eEF2 enters the ribosome and translocates the peptidyl-tRNA along with mRNA toward the P site, leaving the A site ready for the next round of reactions catalyzed by EF-Tu/eEF1α (Agirrezabala and Frank 2009). The molecular functions of EF-Tu/eEF1α and EF-G/eEF2 are the essential characteristics of protein synthesis, and these two proteins are almost (see below) fully conserved among all living organisms.

EFL

Despite the critical role of eEF1α in protein synthesis, there are certain eukaryotic species, such as green algae and chromalveolates, that seem to lack eEF1α orthologs and instead encode an elongation factor-like protein (EFL). Although the amino acid sequence of EFL is clearly divergent from that of eEF1α, the sequences are similar enough to suggest that EFL and eEF1α are functionally equivalent. In most cases, the distribution of EFL and eEF1α among species is mutually exclusive. Only one species has so far been identified to possess both eEF1α and EFL, and no species has been found which possess neither of them. This mutual exclusiveness makes EFL different from the other EF1α family members, eRF3 and HBS1.

The distribution of EFL is not vertical but is unevenly scattered in a wide range of phylogenic trees, i.e., close relatives of organisms with EFL have eEF1α instead of EFL in multiple branches. This situation suggests distribution of EFL by lateral gene transfer. Direct evidence for lateral gene transfer of EFL has been reported in the phylogenetic analyses of EFL in Rhizaria and algae (Kamikawa et al. 2008). Despite this, it is not clear if lateral transfer has occurred in other phylogenetic groups containing species with EFL. The gene-loss scenario is entirely possible (Cocquyt et al. 2009). Namely, the common eukaryotic ancestor was assumed to have both EFL and eEF1α, of which EFL was lost in most lineages. Further analysis is required to elucidate the evolutionary processes that underlie the distributions of EFL and eEF1α within and among different taxonomic groups.

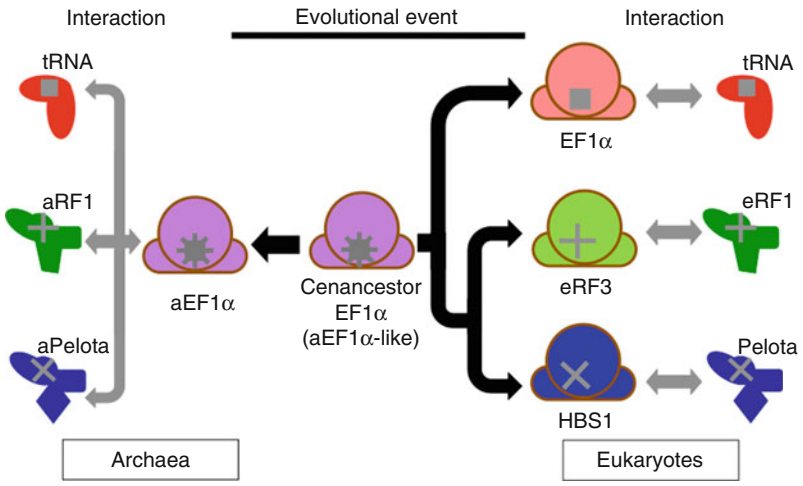


Evolution of Elongation Factor 1 Alpha, Fig. 1 *Evolutionary relationship of EF1α and EF2 family proteins found in the three domains of life on Earth.* LUCA, the last universal common ancestor

The EF1α Family Members – eRF3 and HBS1 – and Mimicry of the EF-Tu/eEF1α-aa-tRNA Complex

Two family members of eEF1α in eukaryotes have similar modes of action as eEF1α. Instead of forming complexes with RNA factors like tRNA, these proteins form complexes with protein factors that mimic the structure and function of tRNA. One of these proteins is the class II eukaryotic release factor (eRF3) that forms a complex with class I eukaryotic release factor (eRF1), which mimics tRNA in functional as well as structural aspects to catalyze the decoding of stop codons in a way similar to tRNAs. Another family member, HBS1, forms a complex with Dom34, which is partly homologous to eRF1. The HBS1–Dom34 complex resembles the eEF1α-aa-tRNA complex in appearance, and this complex is thought to function in mRNA surveillance or mRNA quality control (Doma and Parker 2006). However, the precise molecular mechanism underlying this putative function is not well understood.

The tRNA mimicking factors, eRF1 and Dom34, are conserved in archaea as well as in eukaryotes. However, their interacting GTPases, eRF3 and HBS1 respectively, have not been identified in archaeal genomes despite extensive research (Atkinson et al. 2008). A study has recently shown that canonical archaeal EF1α can form functional tRNA mimicking complexes with any of the archaeal eRF1 or Dom34 factors (Saito et al. 2010). Archaeal EF1α (aEF1α) may thus function as a versatile carrier protein for



Evolution of Elongation Factor 1 Alpha, Fig 2 *The specialized eukaryotic EF1α and the related proteins eRF3 and HBS1 and the multipotent archaeal EF1α. Black arrows* indicate the evolutionary events and *gray arrows* indicate interactions. Archaeal and eukaryotic EF1α and eukaryotic eRF3 and HBS1 presumably share a cenancestor that functioned in elongation, termination, and mRNA surveillance. In archaea (toward the *left*), aEF1α remains a versatile translational GTPase. In eukaryotes (toward the *right*), the cenancestoral EF1α differentiated into multiple proteins with specific functions and interacting partners

tRNA and tRNA-like protein adaptors. The cenancestoral EF1α for archaea and eukaryotes is thought to function equally in translation termination and mRNA surveillance, as well as in elongation, by switching interacting partners as in the current archaea. This suggests that a gene duplication/triplication event may have occurred in the cenancestral eukaryote and that the resulting gene copies subsequently diversified with their respective binding partners to facilitate differential regulation (Figs. 1 and 2).

Bacterial RF3 is Not an EF1α Family Member

Unlike eRF3 or aRF3, the bacterial class 2 release factor, RF3, is not an EF1α family member. Instead, it is an EF2 family member, a striking example of convergent evolution (Kisselev et al. 2003) (Table 1). This explains marked difference in mechanisms underlying eRF3 versus RF3 functions during translation termination (► Translation Termination).

Cross-References

► Selenocysteine
► Translation Elongation
► Translation Initiation
► Translation Termination
► Translational Control by CIS RNA Elements, Bacteria

References

Agirrezabala X, Frank J (2009) Elongation in translation as a dynamic interaction among the ribosome, tRNA, and elongation factors EF-G and EF-Tu. Q Rev Biophys 42:159–200

Atkinson GC, Baldauf SL, Hauryliuk V (2008) Evolution of nonstop, no-go and nonsense-mediated mRNA decay and their termination factor-derived components. BMC Evol Biol 8:290

Cocquyt E, Verbruggen H, Leliaert F, Zechman FW, Sabbe K, De Clerck O (2009) Gain and loss of elongation factor genes in green algae. BMC Evol Biol 9:39

Doma MK, Parker R (2006) Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. Nature 440:561–564

Fabrizio P, Laggerbauer B, Lauber J, Lane WS, Lührmann R (1997) An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2. EMBO J 16:4092–4106

Kamikawa R, Inagaki Y, Sako Y (2008) Direct phylogenetic evidence for lateral transfer of elongation factor-like gene. Proc Natl Acad Sci USA 105:6965–6969

Keeling PJ, Fast NM, McFadden GI (1998) Evolutionary relationship between translation initiation factor eIF-2gamma and selenocysteine-specific elongation factor SELB: change of function in translation factors. J Mol Evol 47:649–655

Kisselev L, Ehrenberg M, Frolova L (2003) Termination of translation: interplay of mRNA, rRNAs and release factors. EMBO J 22:175–182

Margus T, Remm M, Tenson T (2007) Phylogenetic distribution of translational GTPases in bacteria. BMC Genomics 8:15

Saito K, Kobayashi K, Wada M, Kikuno I, Takusagawa A, Mochizuki M, Uchiumi T, Ishitani R, Nureki O, Ito K (2010) Omnipotent role of archaeal elongation factor 1 alpha (EF1{alpha}) in translational elongation and termination, and quality control of protein synthesis. Proc Natl Acad Sci USA 107:19242–19247

Evolution of Eukaryotic Translation Initiation Factors

Katsura Asano
Molecular, Cellular, and Developmental Biology Program, Division of Biology, Kansas State University, Manhattan, KS, USA

Definition

The current repertoire of eukaryotic translation initiation factors (eIFs) evolved on the basis of the set of initiation factors found in archaea (aIFs) and the addition of new proteins belonging to different protein superfamilies including those specifically found in eukaryotes. In this entry, the evolutionary relationship and origin of these eIFs are described.

Characteristics

Of ten eIFs involved in eukaryotic translation initiation, two of them, eIF1A and eIF5B are universally conserved (Fig. 1a) (for initiation factors, see ▶ Translation Initiation and its Table 1). eIF5B is a member of translational GTPase superfamily and its bacterial counterpart is IF2 (▶ Evolution of Elongation Factor 1 Alpha). eIF1A is an OB-fold protein homologous to bacterial IF1; OB-fold proteins include many ribosomal proteins and are generally characterized as nucleic acid-binding proteins (Theobald et al. 2003).

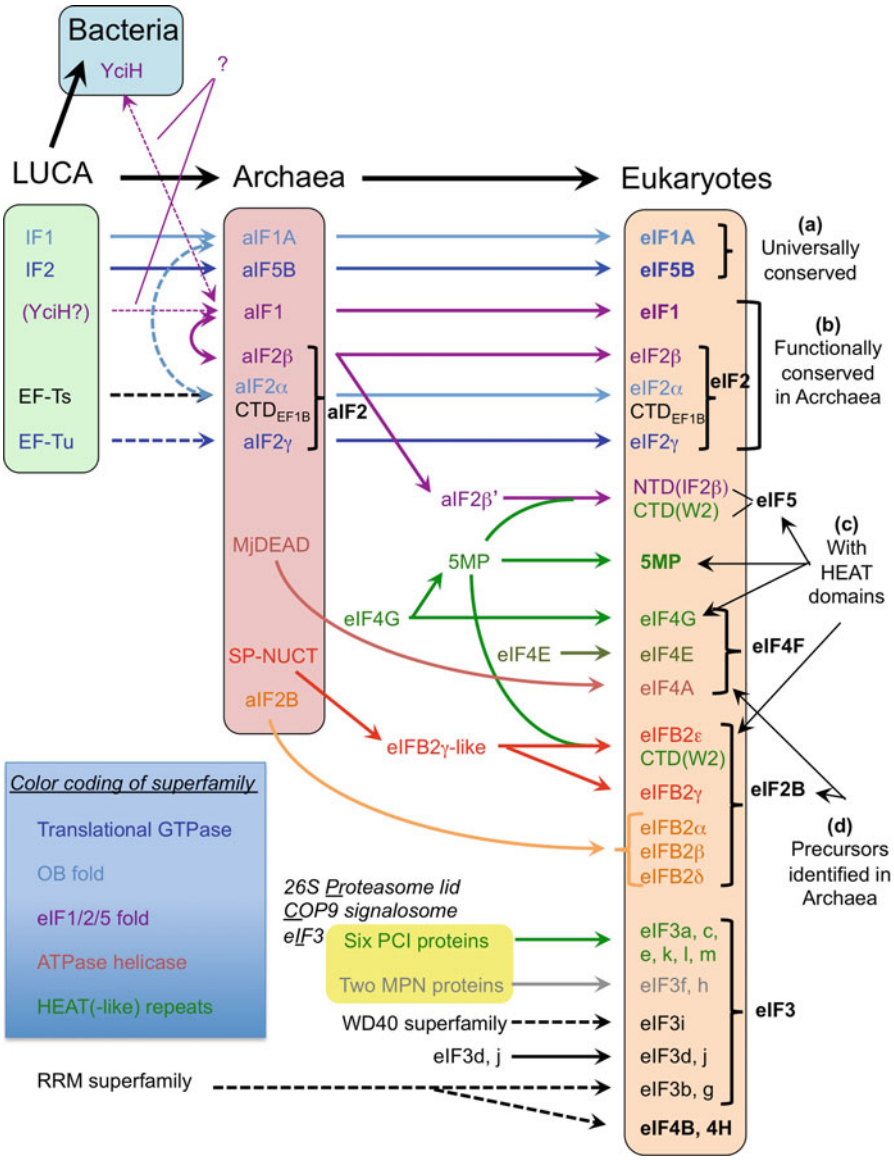
Evolution of eIF1 and eIF2 Conserved in Archaea

eIF1, eIF1A, and eIF2 directly interact with the 40S subunit, the eukaryotic ribosomal small subunit (SSU), prevent the SSU reassociation with the large subunit, and load Met-tRNA$_i^{Met}$ onto the P-site (▶ Translation Initiation). The direct partner of Met-tRNA$_i^{Met}$ is eIF2 made of α, β, and γ subunits. The GTP-binding subunit, a/eIF2γ, is a close structural homologue of EF-Tu/eEF1A (Schmitt et al. 2002) (Fig. 1b) (▶ Evolution of Elongation Factor 1 Alpha). Similar to the latter, GTP-bound eIF2 forms a ternary complex (TC) with Met-tRNA$_i^{Met}$ off the ribosome and delivers the aa-tRNA to it. The Ser-51 of eIF2α is the conserved site of phosphorylation by various stress-activated kinases (▶ Translation Initiation, ▶ Translational Control of GCN4). Interestingly, e/aIF2α is a fusion between the N-terminal OB-fold domain, containing Ser-51, and the C-terminal domain (CTD) similar to eEF1Bα, the catalytic subunit of eEF1B homologous to EF-Ts (▶ Translation Elongation) (Ito et al. 2004). The aIF2α-CTD interacts with the domain II of aIF2γ, just as eEF1Bα binds the domain II of eEF1A during the guanine nucleotide exchange reaction, although their orientations relative to the domain II of the GTP-binding protein are different (Yatime et al. 2007). The evolutionary relationship between the OB-fold domains of e/aIF1A and e/aIF2α is yet unclear.

eIF1 and eIF2β-CTD have a similar eIF1/2/5-fold, suggesting their common origin (Conte et al. 2006). Because a/eIF1's function in start site fidelity appears to be fundamental to the ribosome function, it is tempting to hypothesize that this was the precursor of the proteins of this family. In agreement with this, some bacteria encode an eIF1 relative, termed YciH (named in *E. coli*), which can replace eIF1 function in mammalian reconstitution assays (Lomakin et al. 2006). Based on this result, Lomakin et al. proposed that eIF1/YciH was the major fidelity factor in the common ancestor and replaced with the newer protein, IF3, in bacterial lineages. However, since only a few bacterial species encode YciH (Kyrpides and Woese 1998), it cannot be ruled out that this was transferred from Archaea by horizontal gene transfer ("?" in Fig. 1).

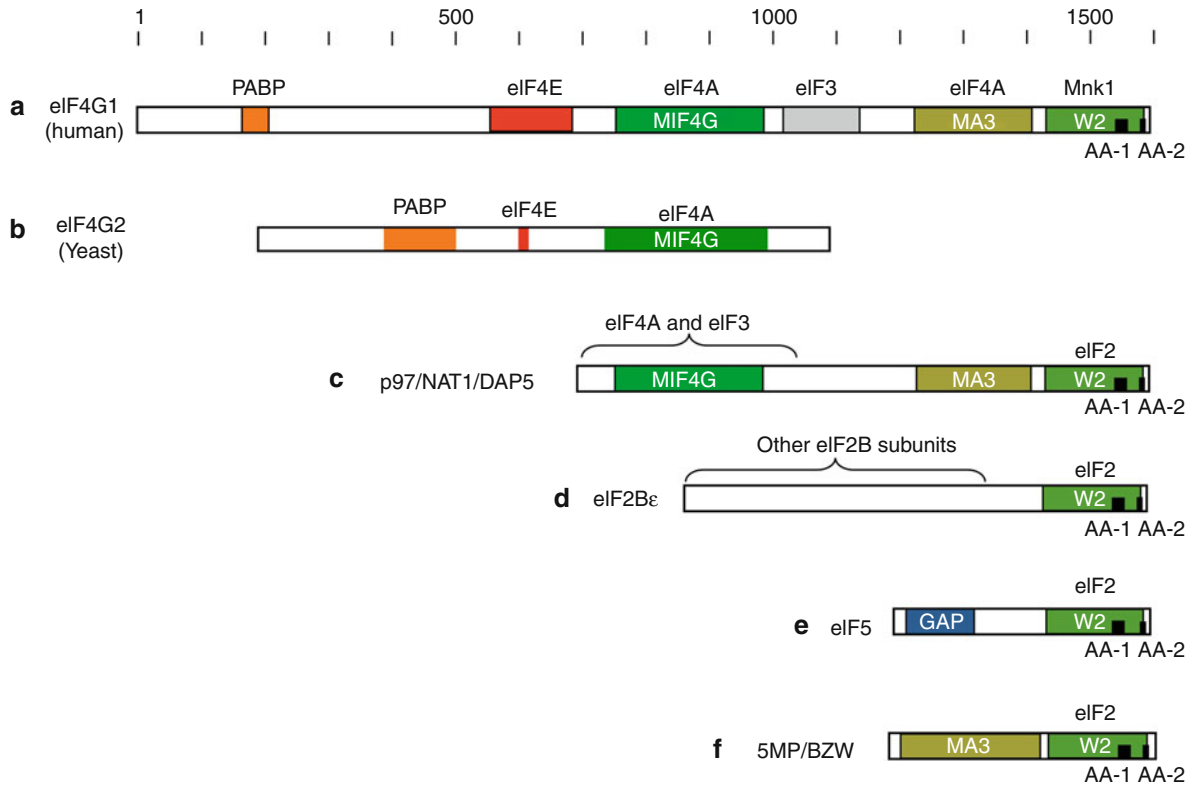
Evolution of HEAT-Domain-Containing Factors and Regulators

The mRNA-binding factor, eIF4G, in most metazoan species carry three HEAT domains, MIF4G/HEAT1, MA3/HEAT2, and W2/HEAT3 as eIF4A- and eIF4E kinase-binding sites (Fig. 2a). eIF4G identified in other species, such as plants and fungi, lack the W2 domain or MA3 and W2 domains together (see Fig. 2b for



Evolution of Eukaryotic Translation Initiation Factors, Fig. 1 *Evolutionary relationship between translation initiation factors found in the universal ancestor (LUCA, last universal common ancestor), archaea and eukaryotes*. Each *vertical column* with a distinct color represents the group of proteins in the hypothetical genome of LUCA (*left column*), or the prototypical genome of archaea (*middle*) or eukaryotes (*right*). Proteins belonging to the same protein superfamily are described in the same color as defined in the *blue box* at *bottom left*. Proteins connected by *solid horizontal arrows* are homologous proteins. *Dotted horizontal arrows* connect proteins under the same superfamily but without direct homology. Eukaryotic factors conserved universally (**a**), functionally conserved with archaeal factors (**b**), with HEAT domains (**c**), and with their precursors identified in archaea (**d**) are indicated by *bracket* or *thin arrow*. Proteins, which appeared during eukaryotic evolution, are listed between the columns representing archaea and eukaryotes. RRM superfamily is found in all three domains of life, so this is listed to the left of the archaeal column. eIF2α, eIF2Bε, and eIF5 are proposed to be generated by fusion between two distinct domains, and their origins are listed separately by the color code and two arrows that merge together as defined in the *blue box* (except eIF2α-CTD$_{EF1B}$ derived from EF-Ts/eEF1Bα)

yeast eIF4G). Marintchev and Wagner identified a significant homology between each of the three distinct HEAT domains of the metaozan eIF4G and those of the nuclear cap-binding protein 80 (CBP80) (Marintchev and Wagner 2005). CBP80 is the larger subunit of *nuclear* cap-binding complex (CBC). This is

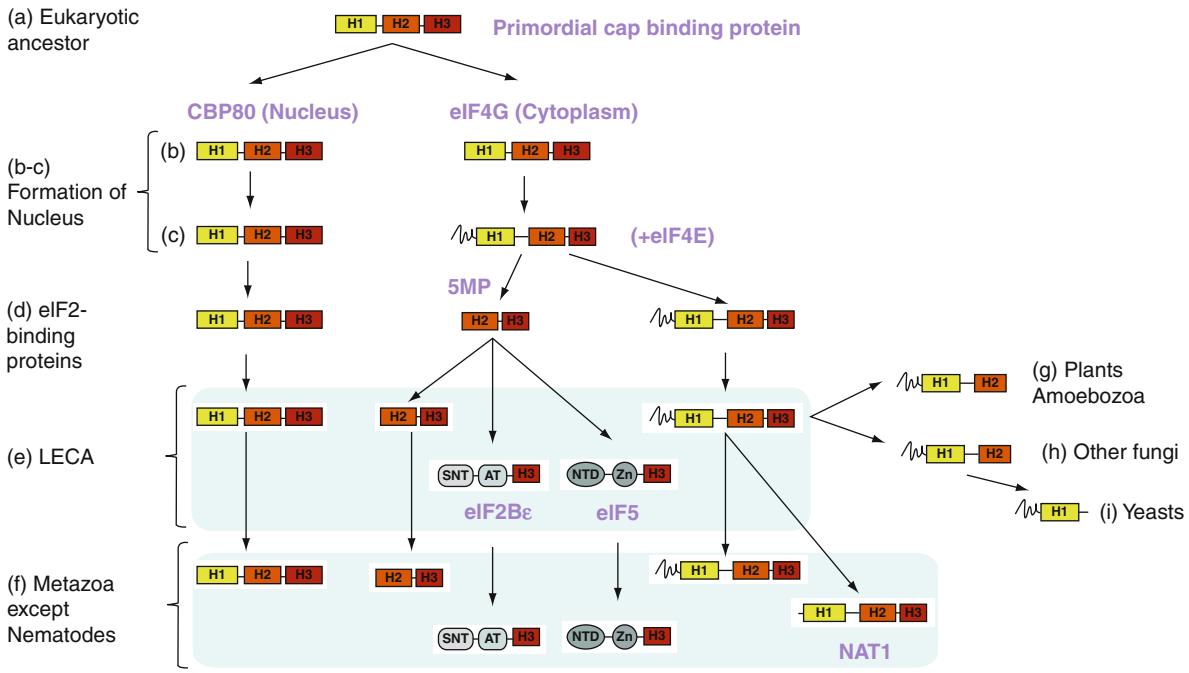


Evolution of Eukaryotic Translation Initiation Factors, Fig. 2 *HEAT domain-containing translation initiation factors and regulators*. Primary structures of human eIF4G1, yeast (*S. cerevisiae*) eIF4G2, human p97/NAT1/DAP5, human 5MP1/BZW2, and yeast eIF2Bε and eIF5 are drawn to scale with *filled boxes* indicating segments known to interact with their partners listed across the *top*. *Bracket indicates* an approximate area of interaction with indicted partners. *Green boxes* indicate HEAT domains. The C-terminal HEAT domain is the W2 domain with *short thick lines* representing the location of AA-boxes 1 and 2 (AA-1, AA-2, respectively)

a highly conserved protein and this protein from all eukaryotes has the three HEAT domain structure, in contrast to eIF4G. The smaller subunit CBP20 of CBC, carrying a common and ancient RNA recognition motif (RRM), directly binds the mRNA cap. The cytoplasmic cap-binding protein, eIF4E, has a distinct unique fold, suggesting that the cytoplasmic complex (eIF4F) is a newer invention in terms of the cap-binding function. Together, these findings led Marintchev and Wagner to propose that the metazoan eIF4G was the first HEAT-domain-containing translation factor, which evolved by gene duplication of primordial cap-binding protein (Fig. 3a–b). The extant form of eIF4G was assumed to have evolved when it acquired a long N-terminal unstructured extension containing the binding sites for eIF4E and PABP (Fig. 3c; see Fig. 2a–b for eIF4E and PABP binding sites). The diversity of eIF4G structure in eukaryotes is understood as having lost the W2/HEAT3 domain, or subsequently the MA3/HEAT3 domain in yeast, by intron or stop codon mutations (Fig. 3g–i).

eIF5 and eIF2Bε, the GTPase activating protein (GAP) and the catalytic subunit of the guanine nucleotide exchange factor (GEF) for eIF2, carry the W2 domain at their C-termini as eIF2-binding sites (Fig. 2d, e) (▶ Translation Initiation). Translational regulators NAT1/p97/DAP5 and eIF5-mimic protein (5MP)/BZW also carry HEAT domains, including W2, as found in metazoan eIF4G (Fig. 2c, f; reviewed in Singh et al. 2011). Because the W2 domains in these proteins are distinct from the HEAT3 of CBP80, lacking the last HEAT repeat found in the latter, these W2 domains are likely to be derived from the archaic eIF4G with the three HEAT domains.

Because NAT1 family is limited to Metazoa and is more similar to eIF4G proteins therein than those from



Evolution of Eukaryotic Translation Initiation Factors, Fig. 3 *The model of the evolution of HEAT domain-containing translation initiation factors and regulators.* The proposed course of evolution of HEAT domain-containing proteins, starting with the primordial cap-binding protein with three HEAT domains (three boxes representing HEAT1, HEAT2, and HEAT3), is depicted in each hypothetical step of eukaryotic evolution. (**a**) The early eukaryotic ancestor without a nucleus contained a single cap-binding protein. (**b**) Formation of the nucleus allowed the evolution of nuclear and cytoplasmic versions by gene duplication. (**c**) The cytoplasmic version obtained N-terminal extension to bind eIF4E and became eIF4G. (**d**) eIF5-mimic protein (5MP) evolved by truncation of the second eIF4G copy and became eIF2-binding translational regulator. (**e**) eIF5 and eIF2Bε acquired the W2 domains from 5MP in the last eukaryotic common ancestor (LECA). (**f**) NAT1 evolved from eIF4G more recently in metazoa (**f**). Because eIF4G has transferred the function carried by W2 to 5MP, this domain of eIF4G tends to be lost during the course of eukaryotic evolution (**g**–**i**)

outside (Marintchev and Wagner 2005), this protein evolved in the ancestor of Metazoa (Fig. 3f). On the other hand, nearly universal distribution of 5MP with MA3 and W2 domains in eukaryotes including the primitive species *Giardia lamblia* (Marintchev and Wagner 2005) argues for its ancient origin and suggests that this was the evolutionary precursor for the W2 domains of eIF5 and eIF2Bε (Fig. 3d–e). Importantly, this protein lacks the MIF4G domain essential for eIF4A binding. Instead, this has the W2 domain, which serves as the eIF2-binding site for all the extant W2 domains except the eIF4G-W2 domain (reviewed in Singh et al. 2011). Thus, the original function of 5MP might have been to bind eIF2 and regulate its guanine nucleotide binding, similar to the GDP dissociation inhibition activity found for yeast eIF5-W2 and the adjacent linker peptide (Jennings and Pavitt 2010).

The emergence of eIF5 and eIF2Bε is understood as the gene fusion of a duplicated copy of aIF2β (eIF1/2/5 fold) and the ancestor of the unique protein eIF2Bγ (see below), respectively, with the C-terminal portion of 5MP (Figs. 1c and 3d–e).

Evolution of eIF4A and eIF2B

Proteins with the DExD/H-box are ATP-dependent RNA helicases, which have diversified to carry out a variety of biological processes in eukaryotes (Tanner and Linder 2001). These include transcription, ribosome biogenesis, RNA processing and export, and translation initiation and termination. eIF4A is the major DExD/H-box RNA helicase involved in translation initiation. Although eIF4A requires the association with eIF4G for its activity, some of the DExD/H helicases, such as yeast Ded1p and human DHX29 are

involved in translation initiation, but do not require or bind eIF4G for their activities. Interestingly, archaea contain a prototypical member of this family, MjDEAD (named in *M. jannaschii*). The original function of MjDEAD in archaeal biology is unknown.

About half of the sequenced archaeal genomes encode a protein termed aIF2B more similar to eIF2Bα/β/δ than distinct archaeal enzymes with similar folds (RBPI, ribose-1,5-bisphoshate isomerase, and MTNA, methylthioribose-1-phosphate isomerase). aIF2B copurifies with aIF2α and a sugar-phosphate nucleotidyltransferase-like protein (SP-NUCT) from *T. kodakaraensis* (Dev et al. 2009). Interestingly, bioinformatics analysis indicates that eIF2Bγ/ε has motifs found in SP-NUCT. Because the SP-NUCT does not carry the W2 domain essential for GEF catalysis, this archaeal protein would rather be homologous to non-catalytic subunit eIF2Bγ. These results suggest that aIF2B and the SP-NUCT originally evolved as eIF2-binding proteins, which might regulate eIF2 in response to sugar or nucleotide binding. The extant eIF2B is hypothesized to have arisen by gene triplication and duplication, respectively, of aIF2B and SP-NUCT, heteropentamer formation by their gene products and a subsequent gene fusion between a SP-NUCT copy and 5MP, generating the catalytic subunit eIF2Bε (Fig. 1d).

Evolution of eIF3 and RRM-containing Factors

Human eIF3 has 13 distinct subunits (a-m), of which six of them contain a PCI domain whereas two of them contain a MPN domain (Zhou et al. 2008). Six PCI proteins and two MPN proteins are also found in the 26S proteasome lid and the COP9 signalosome deNEDDylase complexes (yellow box in Fig. 1). Cryo-EM analysis of the 26S proteasome lid suggests that the PCI and MPN proteins form a ring structure. Thus, it appears that these motifs were used to build a large scaffold. Besides these subunits, eIF3 contains a WD40 subunit (eIF3i) (WD40 family is uniquely eukaryotic and forms a β-propeller structure), two RRM subunits (eIF3b and g) and two other subunits (eIF3d and j) unique to eIF3. As expected for a reaction involving multiple mRNA binding steps, additional factors eIF4B and eIF4H contain an RRM. A structural homology between CBP20 (see above), the cap-binding subunit of CBC, and eIF4H has been suggested (Marintchev and Wagner 2005), but its significance in the mechanism of translation initiation is unknown.

Unstructured Segments of eIFs Defining Functions Unique to Eukaryotic Initiation

Similar to other eukaryotic proteins, more than half, on average, of the entire length of each eIF polypeptide is unfolded, yet carrying important functions. Examples include the N-terminal segment of eIF4G containing PABP- and eIF4E-binding sites and the N-terminal segment of eIF2β carrying conserved lysine-rich segments (K-boxes) mediating eIF2 binding to eIF5 and eIF2Bε (▶ Eukaryotic Translation Initiation Factor Interactions). eIF1A, the universally conserved factor, also carries eukaryote-specific N-terminal and C-terminal tails, which play important roles in regulating ribosome conformation in response to AUG selection (Saini et al. 2010). Thus, the additional proteins or protein segments found in eukaryotic initiation systems confer mRNA recruitment by unique terminal modifications (m^7G-cap and polyA tail), the accuracy in start codon selection and (via phospho-eIF2-regulated eIF2:eIF2B interaction) the translational regulation unique to this domain of life.

Cross-References

- ▶ Eukaryotic Translation Initiation Factor Interactions
- ▶ Evolution of Elongation Factor 1 Alpha
- ▶ Translation Elongation
- ▶ Translation Initiation
- ▶ Translational Control of GCN4

References

Conte MR, Kelly G, Babon J, Sanfelice D, Youell J, Smerdon SJ, Proud CG (2006) Structure of the eukaryotic initiation factor 5 reveals a fold common to several translation factors. Biochemistry 45:4550–4558

Dev K, Santangelo TJ, Rothenburg S, Neculai D, Dey M, Sicheri F, Dever TE, Reeve JN, Hinnebusch AG (2009) Archaeal aIF2B interats with eukaryotic translation initiation factors eIF2alpha and eIF2Balpha: implications for aIF2B function and eIF2B regulation. J Mol Biol 392:701–722

Ito T, Marintchev A, Wagner G (2004) Solution structure of human initiation factor eIF2alpha reveals homology to the elongation factor eEF1B. Structure 12:1693–1704

Jennings MD, Pavitt GD (2010) eIF5 has GDI activity necessary for translational control by eIF2 phosphorylation. Nature 465:378–381

Kyrpides NC, Woese CR (1998) Universally conserved translation initiation factors. Proc Natl Acad Sci USA 95:224–228

Lomakin IB, Shirokikh NE, Yusupov MM, Helln CUT, Pestova TV (2006) The fidelity of translation initiation: reciprocal activities of eIF1, IF3 and YciH. EMBO J 25:196–210
Marintchev A, Wagner G (2005) eIF4G and CBP80 share a common origin and similar domain organization: implications for the structure and function of eIF4G. Biochemistry 44:12265–12272
Saini AK, Nanda JS, Lorsch JR, Hinnebusch AG (2010) Regulatory elements in eIF1A control the fidelity of start codon selection by modulating tRNAiMet binding to the ribosome. Genes Dev 24:97–110
Schmitt E, Blanquet S, Mechulam Y (2002) The large subunit of initiation factor aIF2 is a close structural homologue of elongation factors. EMBO J 21:1821–1832
Singh CR, Watanabe R, Zhou D, Jennings MD, Fukao A, Lee B-J, Ikeda Y, Chiorini JA, Fujiwara T, Pavitt GD et al (2011) Mechanisms of translational regulation by a human eIF5-mimic protein. Nucl Acids Res 39:8314–8328
Tanner NK, Linder P (2001) DExD/H Box RNA helicases: from generic motors to specific dissociation functions. Mol Cell 8:251–262
Theobald DL, Mitton-Fry RM, Wuttke DS (2003) Nucleic acid recognition by OB-fold proteins. Annu Rev Biophys Biomol Struct 32:115–133
Yatime L, Mechulam Y, Blanquet S, Schmitt E (2007) Structure of an archaeal heterotrimeric initiation factor 2 reveals a nucleotide state between the GTP and the GDP states. Proc Natl Acad Sci USA 104:18445–18450
Zhou M, Sandercock AM, Fraser CS, Ridlova G, Stephens E, Schenauer MR, Yokoi-Fong T, Barsky D, Leary JA, Hershey JWB et al (2008) Mass spectrometry reveals modularity and a complete subunit interaction map of the eukaryotic translation factor eIF3. Proc Natl Acad Sci USA 105:18139–18144

Evolution of Metabolism, Amino Acid Biosynthesis Pathways

Georgina Hernández-Montes[1], Dagoberto Armenta-Medina[2] and Ernesto Pérez-Rueda[2]
[1]Departamento de Medicina Molecular y Bioprocesos, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México
[2]Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México

Definition

Metabolical Networks and Evolution Models

In classical terms, *metabolism* has been defined as a global process through which living systems acquire and use free energy to perform a wide diversity of functions. This process couples exergonic reactions of nutrient oxidation to the endergonic processes required to maintain the vital state of the cell, such as mechanical work, active transport, and biosynthesis of molecules. Therefore, the origin and evolution of enzyme-driven metabolism are based on the idea of gene duplication, followed by divergence. In this regard, two main hypotheses have been proposed to explain how the metabolic pathways actually originated: The ▶ retrograde hypothesis suggests that, in the case where a substrate tends to be depleted, gene duplication can provide an enzyme capable of supplying the exhausted substrate, giving rise to homologous enzymes catalyzing consecutive reactions. In other words, if a compound A was essential for the survival of primordial cells, when A became depleted from the primitive soup, this should have imposed a selective pressure allowing the survival and reproduction of those cells that were become able to perform the transformation of a chemically related compound "B" into "A" catalyzed by enzyme "a" that would have led to a simple, one-step pathway. On the other hand, the ▶ patchwork hypothesis postulates that duplication of genes encoding primitive and promiscuous enzymes (capable of catalyzing various reactions) allows each descendant enzyme to specialize in one of the ancestral reactions. These hypotheses could explain the actual anatomy of the metabolical pathways. Recently, another mechanism associated to the metabolical growth suggests the existence of ▶ alternative routes *or alternologs*, defined as branches or enzymes that, proceeding via different metabolites, converge in a common end product. These alternative branches contribute to genetic buffering similar to gene duplication. It has been proposed that the origin of alternative branches is closely related to different environmental metabolite sources and life styles among species (Conant and Wagner 2003; Hernandez-Montes et al. 2008; Horowitz 1945).

Characteristics

The study of the evolution of metabolism is central to understanding the adaptive processes of cellular life, the emergence of high levels of organization (multicellularity), the diversity of cellular organization in the three major domains of life, Archaea, Bacteria,

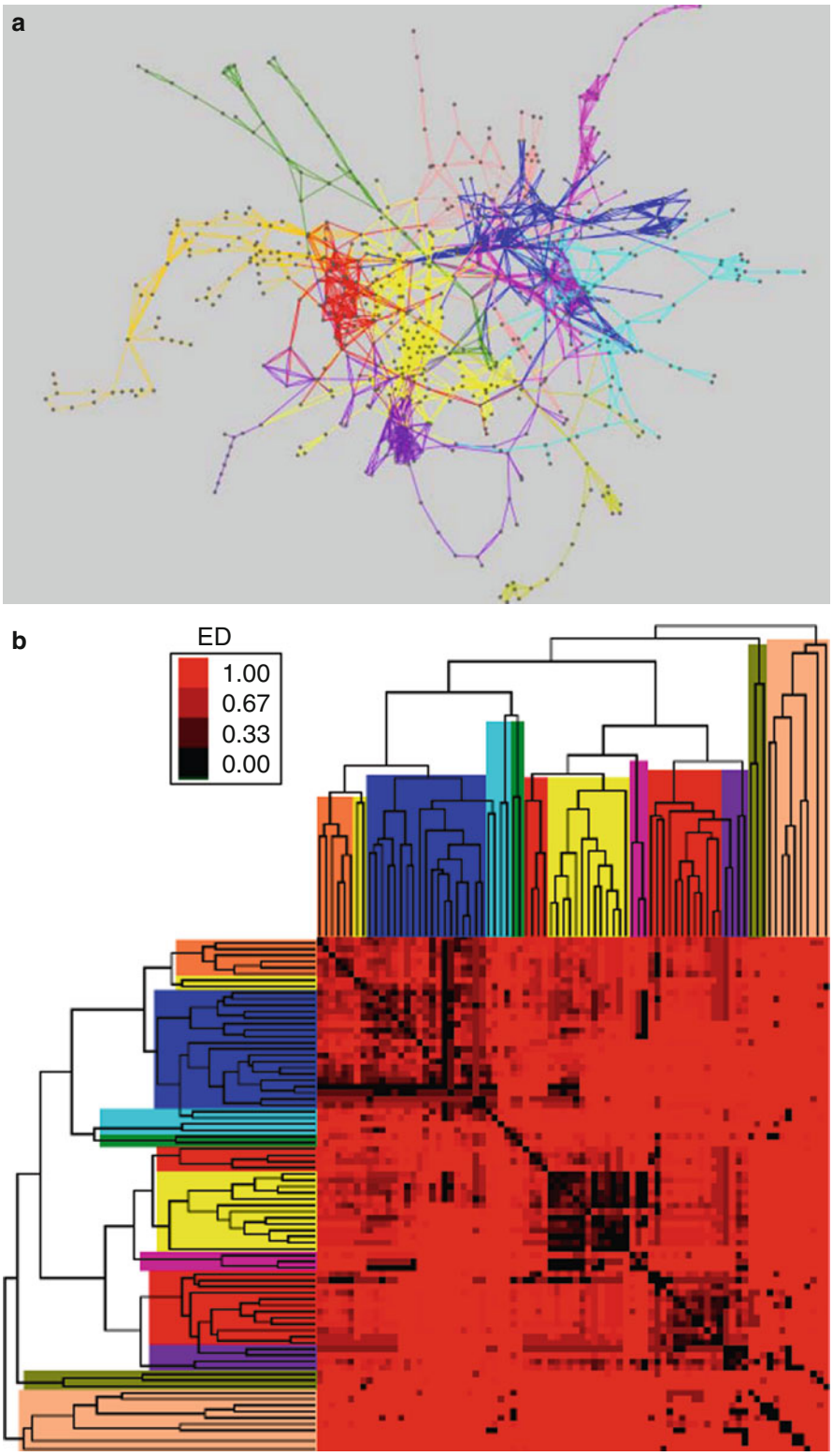
and Eucarya and the complexity of the living world (Caetano-Anolles et al. 2007). At present the large-scale information derived from genomics and proteomics studies has allowed the development of databases devoted to understand the metabolic processes, such as KEGG and MetaCyc. The information contained in these databases can be analyzed from a network perspective (Diaz-Mejia et al. 2007) giving a more integrative view of cellular functioning and for the global understanding of genomes properties and dynamics. In this regard, cell metabolism can be considered as one of the most ancient biological networks, where the nodes represent substrates and/or enzymes and the edges represent the relationships among them (Fig. 1). From this perspective, the study of metabolic networks has focused on describing topological properties such as the existence of functional modules, giving special relevance to the clustering and motif formation, and showing the existence of similar attributes to the small-world and scale-free networks (Barabasi and Oltvai 2004). A small-world network is a type of graph in which most nodes are not neighbors of one another, but most nodes can be reached from every other by a small number of steps, and where the time required for a spreading of perturbation is close to the theoretically possible minimum for a graph with the same number of nodes and edges. This property allows metabolism to react faster to perturbations. In addition, the scale freeness property (few nodes called "hubs," having many more connections than the most of others nodes in the network) is related with the robustness of a metabolic networks, where the fraction of nodes with low connectivity in a small-world network is much higher than the fraction of hubs, in consequence the probability of deleting an important node is very low. Another relevant feature of metabolic networks is its modularity, where each module is a discrete entity of elementary components (enzymes and/or substrates) that perform a certain task, separable from the functions of other modules. The elements of each module are similar to each other and may have been subjected to the same evolutionary process, such as the amino acid biosynthesis pathways and lipid and nucleotide metabolisms (Fig. 1a).

Metabolic Evolution by Recruitment of Modules

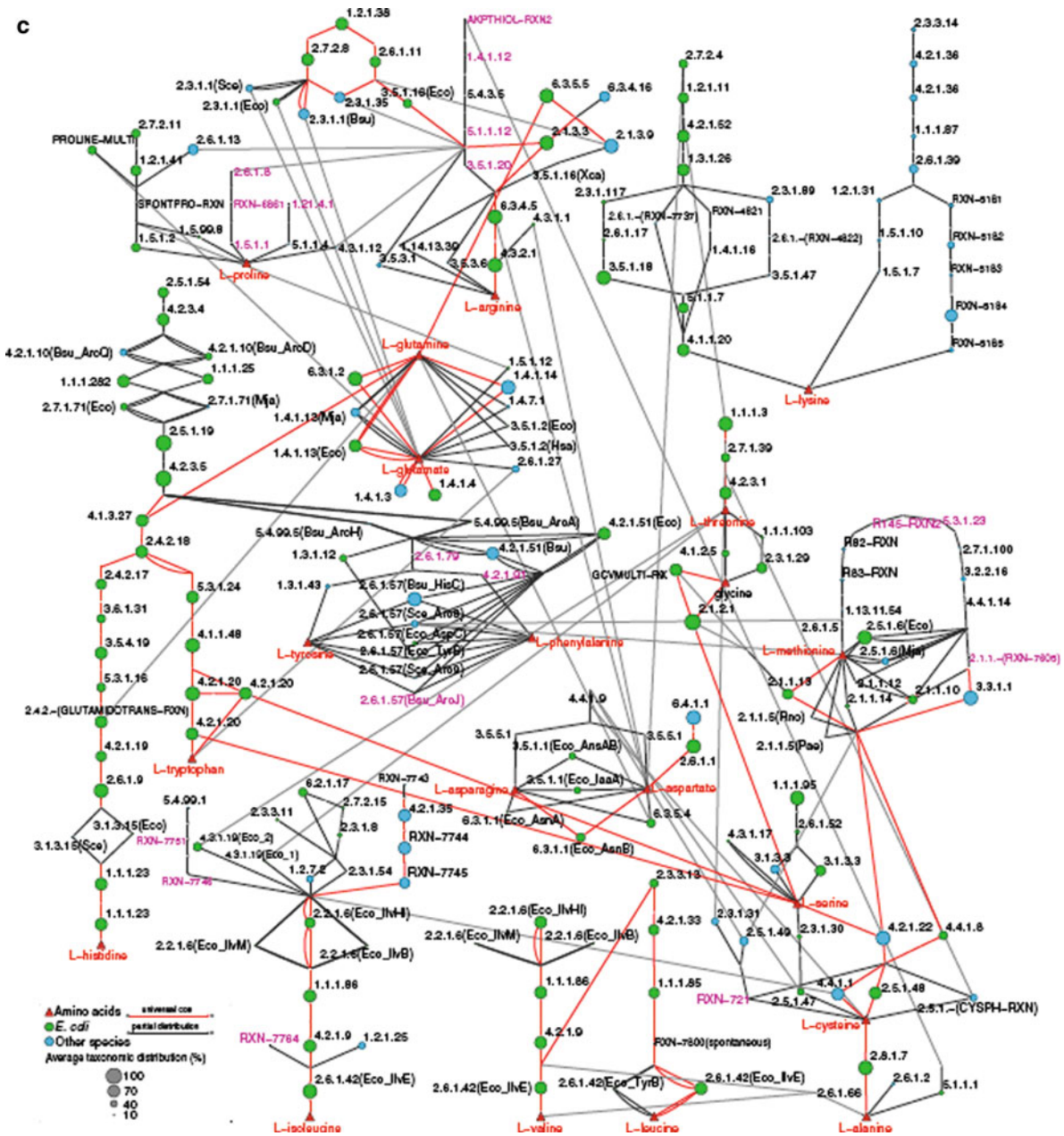
Based on a network-based approach, an increased retention of duplicates for enzymes catalyzing consecutive reactions has been identified. As a consequence, metabolic networks exhibit a high retention of duplicates within functional modules, and a preferential biochemical coupling of reactions. This retention of duplicates may result from the biochemical rules governing substrate-enzyme-product relationships. The retention of duplicates between chemically dissimilar reactions is, however, also greater than expected by chance as suggested Diaz-Mejia et al. (2007). Finally, a significant retention of duplicates as groups, instead of single pairs, has been also reported. In brief, the duplication events were evaluated among modules using a hierarchical clustering algorithm. A paired measure of evolutionary distance (ED) for all-against-all metabolic pathways was then calculated. *In brief, the ED is a measure of the retention of duplicates between pathways within and between modules, and is calculated as follows: (ED) = A′/(A′ + AB), where A′ is the number of enzymes of the smaller pathway (pA) without homologs in the second pathway (pB). AB is the number of enzymes of pA with homologs in pB. At one extreme, when all the enzymes of pA have homologs in pB, the evolutionary distance converges on 0. In contrast, when the two pathways share no homologs, the value of evolutionary distance converges on 1* (Diaz-Mejia et al. 2007). This analysis shows that metabolic pathways of the same module tend to have a lower ED, suggesting a greater retention of duplicates within modules than among them. Based on these data, the capability of metabolic networks to grow modularly by gene duplication may be a consequence of two factors: the closeness between reactions, and the kind of substrate(s) participating within each module (Fig. 1b).

Retention of Duplicates as Groups and Single Entities

It was observed that a high frequency of duplicates were retained as groups (those defined here as pairs of consecutive reactions), instead of single entities. Based on this approach, it has been determined that enzymes catalyzing chemical similar reactions (CSRs, or those reactions that share a common enzymatic E.C. number (preferentially the first two digits). Therefore, despite having similar reaction chemistry, enzymes could exhibit different substrate specificities) in the fatty-acid metabolism have been originated by gene duplication (Diaz-Mejia et al. 2007). Thus, an ancestral pathway catalyzed both fatty-acid degradation and biosynthesis could have originated in a first step.



Evolution of Metabolism, Amino Acid Biosynthesis Pathways, Fig. 1 (continued)



Evolution of Metabolism, Amino Acid Biosynthesis Pathways, Fig. 1 The metabolism can be analyzed as a network, where enzymes are nodes and substrates are edges, allowing the evaluation of the influence of network modularity on the retention of duplicates: (**a**) A hierarchical clustering showing the modules in metabolic networks (*in colors*) (**b**) Metabolic pathways (branches in the trees) within and across modules were compared using a measure of evolutionary distance (ED). Modules comprising related branches are indicated by color as in (**a**). A value of ED closer to zero (*the darker squares*) implies a greater retention of duplicates between the two given pathways. From this tree the amino acid pathways (**c**) exhibit a large fraction of duplicated events. Posterior analyses showed that the amino acid biosynthetic network exhibits diverse alternative pathways, defined as alternative or alternolog, whereas ancient pathways are more connected among them

The direction of this ancestral pathway would be dependent on the acyl carriers and fatty acids available, evidencing the existence of two different pathways. Diaz-Mejia et al. and Light et al. (Diaz-Mejia et al. 2007; Light et al. 2005) using similar approaches evaluated the generality of this observation, using an all-against-all comparison of the enzymes catalyzing consecutive CSRs. They identified that around 15% of enzymes have at least one homolog in a metabolic pathway. Of these, two thirds are retained as isolated duplicates and a third is retained as groups, for instance, the amino acid metabolic pathways. Alternatively, an estimation of retention of duplicates in metabolism using an enzyme-centric network approach showed significant results highlight the influence of both distances apart in the network and chemical similarity of reactions on the retention of duplicates. Specifically, an increased retention of duplicates between consecutive reactions was observed and showed that this bias can be partially attributed to the preferential biochemical coupling of reactions. In addition, these analyses show a significant retention of duplicates acting on both CSRs and chemical different reactions (CDRs, defined by the mismatch in the first two digits of the sequence enzymatic number. Therefore, despite having different reaction chemistry, enzymes could have similar substrate specificities), supporting the idea that gene duplication is important in generating innovations as well as metabolic variants. A synergy between closeness in the network and chemical similarity between reactions explains the high retention of duplicates between consecutive CSRs, suggesting that duplicates are significantly retained as groups that can be extended to several series of reactions.

New Insights on the Evolution of Metabolic Pathways

The modern perspective of metabolic processes has shown that evolutionary studies must include not only phylogenetic relationships among enzymes, but also the influence of some topological properties of metabolic networks. In evolutionary terms, one can assume that the universal occurrence of some pathways and branches in modern species would suggest that they existed in the last common ancestor (LCA). The evolution of these pathways and the emergence of diverse homologues reflect an increased metabolic diversity as a consequence of increasing genome size, protein structural complexity, and selective pressures in changing environments. In the evolution of amino acid biosynthesis, for instance, diverse studies have suggested that they could be among the earliest metabolic compounds and that their growth has been affected by gene duplication. Based on this analysis, a core of 11 pathways that probably occurred in the last common ancestor was identified using a combination of genomic tools with a network approach (Hernandez-Montes et al. 2008). Figure 1c. One of the most important results from this analysis is that diverse enzymes synthesizing 16 out of the 20 standard amino acids were identified in the three cellular domains. It is important to highlight that the full biosynthetic routes for these 16 amino acids, such as tryptophan, arginine, and proline, among others amino acid pathways appeared early in evolution. Therefore, when a network approach is used to identify functional relationships among all pathways (Fig. 1c), the universal branches are connected to each other, allowing the possibility that they feedback and they complete a minimal set of enzyme-driven reaction for the biosynthesis of amino acids. In addition, pathways, as lysine, methionine, and valine, among others, are partially distributed across the three domains of life, with branches identified in the LCA. This data suggest that partial branches may be intimately connected to the main branches identified as the core, previously described. In addition, in the same analysis, the authors identified the contribution of paralogs (duplicated genes within an organism to occupy two different positions in the same genome, then copies evolve new functions, even if these are related to the original one), analogs (genes that perform the same function but their sequences do not have a common evolutionary history and therefore come from independent origins) and alternologs (see Definition section) in the evolution of biosynthesis of amino acids. Eleven out of the 20 amino acid biosynthetic routes revealed an important contribution of paralogs to the generation of diversity, 8 out of the 20 amino acids routes show contribution of analog enzymes, while alternologs routes participate in 9.

Concluding Remarks

It is well accepted that biological networks exhibit similar topological properties, such as their scale-free behavior and hierarchical modularity. Based on recent studies,

diverse authors have suggested that the origin and evolution of networks must consider not only the topological properties they share but also those that differentiate them (Artzy-Randrup et al. 2004). In this regard, the modeling of metabolic networks by including the preferential biochemical coupling of reactions explains more appropriately how metabolism evolves. A more detailed analysis looking at other functional constraints, such as metabolite similarity and binding versus catalytic enzyme properties, as well as massive gene duplications and horizontal gene transfer, could increase our understanding of the influence of metabolic versatility in the evolution of species. In this regard, Shuster et al. (Schuster et al. 2008) showed coevolution of increasing specific group-transfer metabolite coenzymes with their specific enzymes, in a group-transfer network, selected for growth rate, suggesting a coupling between enzyme evolution and cell level fitness subject to reasonable environmental conditions. Finally, genetic and genomic data have revealed that genetic dynamics, such as horizontal gene transfer and gene duplications, in all organisms has also influenced the evolution of metabolic processes, and given evidence that older enzymes are more highly connected (duplicated and diverged enzyme may preserve past reagents).

References

Artzy-Randrup Y, Fleishman SJ, Ben-Tal N, Stone L (2004) Comment on "network motifs: simple building blocks of complex networks" and "superfamilies of evolved and designed networks". Science 305:1107; author reply 1107

Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5:101–113

Caetano-Anolles G, Kim HS, Mittenthal JE (2007) The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. Proc Natl Acad Sci USA 104:9358–9363

Conant GC, Wagner A (2003) Asymmetric sequence divergence of duplicate genes. Genome Res 13:2052–2058

Diaz-Mejia JJ, Perez-Rueda E, Segovia L (2007) A network perspective on the evolution of metabolism by gene duplication. Genome Biol 8:R26

Hernandez-Montes G, Diaz-Mejia JJ, Perez-Rueda E, Segovia L (2008) The hidden universal distribution of amino acid biosynthetic networks: a genomic perspective on their origins and evolution. Genome Biol 9:R95

Horowitz NH (1945) On the evolution of biochemical syntheses. Proc Natl Acad Sci USA 31:153–157

Light S, Kraulis P, Elofsson A (2005) Preferential attachment in the evolution of metabolic networks. BMC Genomics 6:159

Schuster S, Pfeiffer T, Fell DA (2008) Is maximization of molar yield in metabolic networks favoured by evolution? J Theor Biol 252:497–504

Evolution of Regulatory Circuits

Xiaojuan Sun and Jinzhi Lei
Zhou Pei-Yuan Center for Applied Mathematics, Tsinghua University of Beijing, Beijing, China

Definition

Evolution is the genetic change in a population of species over generations. The certain changes that aid survival and reproduction will be selected in successive generations of population. Evolution with environmental changes contributes ▶ adaptation, a process in which an organism becomes better suited to its habitat. The gradual modification of transcription circuits over evolutionary time scale is an important source of diversity of life (Tuch et al. 2008).

Evolution of regulatory circuits occurs through modifications either in transcription factors or in promoter structure. In the evolution of bacterial regulator circuits, for example, there are four classes of changes (Perez and Groisman 2009): (1) transcriptional rewiring whereby the promoters of orthologous genes in related species differ in the presence or absence of a binding site(s) for a conserve transcription factor(s), (2) embedding of horizontally acquired genes under regulation of an ancestral transcription factor, (3) restructuring of the promoters controlled by a transcription factor, and (4) modifications in the transcription factors.

References

Perez JC, Groisman EA (2009) Evolution of transcriptional regulatory circuits in bacteria. Cell 138:233–244

Tuch BB, Li H, Johnson AD (2008) Evolution of eukaryotic transcription circuits. Science 319:1797–1799

Evolution Programming

Feng-Sheng Wang and Li-Hsunan Chen
Department of Chemical Engineering, National Chung Cheng University, Chiayi, Taiwan

Synonyms

Evolutionary algorithm; Evolutionary computation

Definition

Evolution programming is a population-based stochastic optimization technique that is used in some mechanisms in artificial intelligence by simulated biological evolution to determine a global optimal solution. The algorithm employs real-coded variables and, in its original form, it relied on mutation as the search operator. It starts with a set of finite state machines (FSMs) $FSM \equiv \{FSM_t | FSM_t(Input) - Output_t\}$ and a fitness function to evaluate how well each finite state machine survives. The evolution drops the FSMs with lower grades in fitness function and replaces them with new FSMs. The new FSMs are generated from the survivors with small differences, or called mutation. The evolution repeats until a FSM with required grades in fitness function is found or the limit number of iterations is reached.

Characteristics

Pseudocode :

```
t=0
Initpopulation FSM(0) // initialize random population of finite state machines
Evaluate FitFSM(0)    // evaluate fitness of each FSM
Do
    FSM(t+1)=survive(FSM(t)), mutate(survive(FSM(t)))
                      // keep the FSMs with good fitness
                      // generate new FSMs from the survivors with some
perturb
    t=t+1
While maximum iterations or required fitness is not attained
```

Example. Evolutionary programming has been applied to determine network models of gene regulatory networks using synthetic and gene expression data from DNA microarrays (Alina et al. 2010; Martin et al. 2005).

References

Alina S, Heather JR, Martin C (2010) Comparison of evolutionary algorithms in gene regulatory network model inference. BMC Bioinformatics 11(59):1471–2105

Martin S, Thomas H, Werner D, Johannes M, Niall P (2005) Reverse-engineering gene-regulatory networks using evolutionary algorithms and grid computing. J Clin Monit Com 19(4):329–337

Evolution Programs

▶ Optimization and Parameter Estimation, Genetic Algorithms

Evolutionary Algorithm

▶ Evolution Programming
▶ Genetic Algorithms

Evolutionary Algorithm, Transcription Regulatory Network Construction

Xiujun Zhang
Institute of System Biology, Shanghai University, Shanghai, China

Synonyms

Genetic algorithm

Definition

Evolutionary algorithms (EAs) are a group of newly developed heuristic optimization algorithms, which are inspired by natural selection and evolution in biological systems. EAs differ from other traditional optimization techniques, and they aim to find the optimal solution from a "population" of solutions rather than from a single initial solution so that EAs can avoid being trapped into the local optimal solutions to some extent.

In general, an EA consists of four operations, including reproduction, mutation, recombination, and selection, and all these operations are repeated until the algorithm converges to a certain point with some criteria satisfied. In an EA, each candidate solution is represented as a chromosome. In each step of EA, the chromosomes compete against each other and those representing poor solutions will be kicked out before next step. To evaluate the chromosomes, a fitness function is defined as the objective function. The solutions

with high "fitness" are treated as good solutions, and will be "reproduced" from last step and "recombined" with other solutions by swapping parts of the chromosome. Furthermore, the chromosomes are "mutated" with small changes made to the elements of the chromosomes. Some new chromosomes are generated through recombination and mutation operations. Only those chromosomes with high fitness will enter next step.

Evolutionary algorithms have been successfully applied to various fields, such as pattern recognition, biomedical engineering, and transcription regulatory network construction. Especially, evolutionary algorithms can be used to calculate the ligand-receptor binding affinities (▶ Binding Affinity) (Morris et al. 1998). The transcription regulations can be represented as a network, where nodes represent proteins or genes and edges represent direct regulatory interactions (e.g., the binding of a transcription factor to the promoter of its target genes). The transcription regulatory network (TRN) describes the TF-gene interactions as a graph or network, and gene expression is regarded as a function of regulatory inputs specified by interactions between proteins and DNA (Blais and Dynlacht 2005). Transcription regulation has influence on most physiochemical activities in a cell (Thomas 1998; Albert and Barabási 2002). Therefore, it is important to construct transcription regulatory network. Compared with other methods for constructing transcription regulatory networks, the evolutionary algorithms are able to infer smaller regulatory networks with good sensitivity and precision. However, much more computation power is required to infer a large regulatory network from the experimental data.

Characteristics

The initial population of chromosomes that represent the regulatory relationships between genes can be produced randomly or according to user-defined restriction. Sequentially, these chromosomes can be interpolated by recombination and mutation. Then fitness function (▶ Fitness Function) is employed to evaluate these recombined and mutated chromosomes and only those chromosomes with high fitness are selected to reproduce the next generation population. One important issue in EA is the design of the fitness function which is the criteria for evolution. In transcription regulatory network construction, the fitness function can be the difference between the observed gene expression data and the calculated expression data. Another possible fitness function is the Akaike's information criterion (AIC) (▶ Akaike's Information Criterion (AIC)) which is a measure of the goodness of fitting an estimated statistical model with the data and can be used as the fitness function in evolutionary algorithm–based transcription regulatory network construction (Shin and Iba 2003).

The General Steps of EA

The general steps of evolutionary algorithm are listed below:

1. Initialize a population of chromosomes.
2. Apply recombination and mutation to generate new chromosomes.
3. Evaluate each chromosome according to the fitness function.
4. Select chromosomes to reproduce and replace the old chromosomes.
5. Repeat steps 2–4, until the specified numbers of generations are completed or special condition is satisfied.

An Example

Here, an example is used to explain the workflow to infer the transcription regulatory network by employing evolutionary algorithm. Suppose there is a microarray dataset with n genes measured at m time points. The expression of genes at time point $t + 1$ should be the regulatory result of the expression of regulators at time point t. The regulatory relationship of every gene with $i = 1, 2, \cdots, n$ and $t = 1, 2, \cdots, m$ can be expressed as:

$$a_{1i}x_{1t} + a_{2i}x_{2t} + \cdots + a_{ni}x_{nt} = x_{it+1} \qquad (1)$$

where a_{ji} is the regulatory strength of gene j on gene i, x_{jt} is the expression data of gene j at time point t, and x_{it+1} is the expression data of gene i at time point $t + 1$.

Since it is not known which genes regulate gene i, all the genes are possible candidate regulators for gene i. The regulatory strength a_{ji} can tell which genes regulate gene i, and those genes with $a_{ji} \neq 0$ are possible regulators. Since there are usually only a few measurements (i.e., m) but with many variables (i.e., n) in the model, it is not easy to solve the problem. An initial population of N solutions can be randomly generated.

The difference between the true expression data and the calculated data is regarded as a fitness function which can be used to evaluate the chromosomes. For each chromosome k, its fitness is defined as follows:

$$Fn = \min_{1 \le k \le N} \left\{ \left| x_{it} - x_{it}^{k} \right| \right\} \quad (2)$$

Cross-References

- ▶ Akaike's Information Criterion (AIC)
- ▶ Binding Affinity
- ▶ Fitness Function

References

Albert R, Barabási AL (2002) Statistical mechanics of complex networks. Rev Mod Phys 74(1):47–97
Blais A, Dynlacht BD (2005) Constructing transcriptional regulatory networks. Genes Dev 19:1499–1511
Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 19:1639–1662
Shin A, Iba H (2003) Construction of genetic network using evolutionary algorithm and combined fitness function. Genome Inform 14:94–103
Thomas R (1998) Laws for the dynamics of regulatory networks. Int J Dev Biol 42:479–485

Evolutionary Algorithms

Ettore Mosca, Ivan Merelli and Luciano Milanesi
Institute for Biomedical Technologies – CNR (Consiglio Nazionale delle Ricerche), Segrate, Milan, Italy

Definition

In artificial intelligence, evolutionary algorithms (EA) are generic population-based metaheuristic optimization algorithms (Eiben and Smith 2003). In the context of EAs, each candidate solution of an optimization problem plays the role of an individual of a finite population of individuals, and is associated to a cost function (designated as fitness function) which defines the quality of the solution. In order to identify better and better solutions, EAs apply some operators inspired by evolution of living organisms (mutation, recombination, and selection) to the population of individuals.

There are different types of EAs, such as Genetic Algorithms, Evolutionary Programming and Evolution Strategies, which differ mainly by the encoding of the individuals and the operators used during the evolution.

References

Eiben AE, Smith JE (2003) Introduction to evolutionary computing (natural computing series). Springer, Berlin

Evolutionary Capacity

- ▶ Evolvability, Generalized Biology

Evolutionary Computation

- ▶ Evolution Programming

Evolvability

Marie I. Kaiser and Andreas Hüttemann
Department of Philosophy, University of Cologne, Cologne, Germany

Definition

Two definitions of evolvability can be distinguished (Love 2003): (1) Evolvability_{U} is understood as the ability of a population to respond to natural selection. This understanding of evolvability is especially found in quantitative genetics. (2) A more restricted definition of evolvability_{R} was introduced to explain differential evolutionary success. Contrary to the notion of fitness, evolvability_{R} is an exclusively population-level trait. Evolvability_{U} is the capacity of a population for

evolutionary change and by this a function of the amount of variation available on which natural selection can act.

Cross-References

► Disposition

References

Love A (2003) Evolvability, dispositions, and intrinsicality. Philos Sci 70:1015–1027

Evolvability, Generalized Biology

Chrystopher L. Nehaniv
Royal Society/Wolfson Bio Computation Research Laboratory, School of Computer Science, University of Hertfordshire, Hatfield, UK

Synonyms

Capacity to evolve; Evolutionary capacity

Definition

Evolvability in generalized biology refers to the capacity of a Darwinian evolutionary system to evolve fit individuals. The evolvability of different instances of Darwinian evolution can be quantified as the capacity of the evolutionary system (including mechanisms of heredity, interaction with the environment and other entities, etc.) to produce individuals fitter than any that have yet existed (Altenberg 1994), or alternatively, fitter than any currently existing (Nehaniv 2005).

In addition, sometimes "evolvability" refers to the capacity to evolve adaptive, complex structure often in an open-ended way in which the parameters of evolutionary space, such as the genotype-phenotype relationship, are neither fixed nor circumscribed at the outset.

Characteristics

In biology and other realms, evolvability is a concept applying to any Darwinian evolutionary system (Wagner & Altenberg 1996). Here, "Darwinian evolutionary system" can be understood axiomatically (e.g., Nehaniv 2005): A *(generalized) Darwinian evolutionary system* is taken to be any dynamical system in which a population of *individuals* (a term that will not be defined here, cf. (Michod 1999)) is characterized as satisfying certain axioms: individuals give rise somehow to other individuals in a manner exhibiting (1) heredity, (2) variability, (3) differential reproductive success (depending at least in part on inherited characteristics), and (4) finite resources and a turnover of generations. The classical reasoning of Darwin and Wallace (1858) then shows that in such a setting there always is a struggle for existence that can lead to increasingly fit individuals.

However, many instances of Darwinian evolution can be studied or constructed and these can vary greatly in their capacity to produce fitter individuals. Instances where the concept of evolvability can be applied include evolving biological populations on earth; populations of computer programming as in genetic programming; candidate solutions to optimization problems encoded as binary strings as in typical genetic algorithms; or as vectors of real numbers in evolution strategies; or populations of living or life-like entities elsewhere in the universe (cf. astrobiology); evolving populations of machines, RNA molecules, pharmaceuticals under design; artificial genetic regulatory networks, or forms of artificial life; or other as yet undiscovered forms of life.

Evolution of Evolvability

The origin and maintenance of evolvability in Darwinian evolution is a topic inviting controversy since evidently evolvability is not a property of an individual on which natural selection can act. However, relating evolvability to lineage selection (e.g., in any of the above-mentioned instances of Darwinian examples), one may attempt to treat a lineage as an abstract, higher-level individual in the Darwinian evolution axioms. Populations whose "individuals" are (possibly overlapping) lineages, lead to consideration of the struggle for existence among lineages, and this may lead to investigation of the evolution of evolvability in terms of a different, higher level Darwinian dynamic.

Limitations

In more general settings, such as in discussions of "software evolution" or the "evolution" of culture, of behavior, or of artifacts, one can try to analyze these systems in Darwinian terms and attempt to investigate their evolvability; however, since in these realms it is difficult to identify well-defined "individuals," these concepts are challenging to apply (Nehaniv et al. 2006). Nevertheless, one can identify "descent with modification," and persistence and spread analogous to survival and fecundity of individuals. However, *fitness* is then also difficult to define in such settings when there are no clearly defined individuals, so the capacity to produce fitter variants is similarly vague.

Cross-References

- ▶ Artificial Evolution
- ▶ Artificial Life
- ▶ Evolvability
- ▶ Fitness
- ▶ Gene Regulatory Networks
- ▶ Genetic Algorithms

References

Altenberg L (1994) The evolution of evolvability in genetic programming. In: Kinnear KE (ed) Advances in genetic programming. MIT Press, Cambridge, MA, pp 47–74

Darwin C, Wallace A (1858) On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. J Proc Linn Soc Zool 3:45–62

Michod RE (1999) Darwinian dynamics: evolutionary transitions in fitness and individuality. Princeton University Press, Princeton

Nehaniv CL (2005) Self-replication, evolvability, and asynchronicity in stochastic worlds. In: Lupanov OB, Kasim-Zade OM, Chaskin AV, Steinhöfel K (eds) *Proceedings of the 3rd symposium on stochastic algorithms, foundations and applications (Symposium stochastische algorithmen, grundlagen und anwendungen) – SAGA'05, Moscow, 20–22 Oct 2005*. Lecture Notes in Computer Science, vol 3777, Springer, 2005, pp 126–169. Corrected version on-line at URL http://homepages.feis.herts.ac.uk/~comqcln/nehaniv-SAGA05-withcorrections.pdf

Nehaniv C, Hewitt H, Christianson B, Wernick P (2006) What software evolution and biological evolution don't have in common. In: Proceedings of the 2nd international IEEE workshop on software evolvability (SE'06). IEEE Computer Society Press, Philadelphia, pp 58–65

Wagner GP, Altenberg L (1996) Complex adaptations and the evolution of evolvability. Evolution 50(3):967–976

Ex Vivo Imaging

▶ Live Cell Imaging

Exact Test for Independence

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

Definition

Fisher's exact test evaluates the hypothesis of independence between two categorical random variables.

Characteristics

Fisher's Exact Test for 2 × 2 Tables

Consider a 2 × 2 contingency table summarizing the observed values of two random binary variables X and Y

	0	1	Total
0	n_{00}	n_{01}	$n_{00} + n_{01}$
1	n_{10}	n_{11}	$n_{10} + n_{11}$
Total:	$n_{00} + n_{10}$	$n_{01} + n_{11}$	n

If X and Y are independent, then the probability of observing any particular combination of frequencies given marginal totals is given by the hypergeometric distribution:

$$p(t) = P(n_{00} = t) = \frac{(t + n_{01})!(n_{10} + n_{11})!(t + n_{10})!(n_{01} + n_{11})!}{n!t!n_{01}!n_{10}!n_{11}!}$$

When the marginal totals are fixed, n_{00} determines the remaining cell counts. For 2 × 2 contingency tables, independence is equivalent to the unity odds ratio, i.e.,

$$\frac{n_{00}n_{11}}{n_{01}n_{10}} = 1$$

Departure of the odds ratio from 1 represents evidence of association between the two variables. When marginal totals are fixed, tables with larger n_{00} have larger odds ratio providing evidence in favor of the alternative hypothesis. The p-value of the test is given by the probability $P(n_{00} \geq t^*)$, where t^* is the observed value of n_{00} (Agresti 2002).

Fisher's Tea Tasting Experiment

Fisher's colleague claimed that when testing the tea, she could distinguish the order in which milk and tea were added to the cup. To test her claim, Fisher devised a small tasting experiment in which four cups had milk added first and another four cups had tea poured first. The cups were presented in random order for tasting. A colleague was asked to predict which four cups had milk added first. The table below shows the results

Truth/Guess	Milk	Tea	Total
Milk	3	1	4
Tea	1	3	4
Total:	4	4	8

Truly distinguishing the order of pouring as opposed to simply guessing corresponds to testing whether odds ratio is greater than one. Under the null hypothesis, the observed table has a probability $p(3) = 0.229$. The p-value is given by the probability $P(n_{00} \geq 3) = P(n_{00} = 3) + P(n_{00} = 4) = 0.243$. Hence, there is no evidence to confirm that predictions were more accurate than a simple guess. The sample size, however, is rather small, so the lack of association cannot be deduced with certainty. Fisher in reality was convinced of his colleague's ability.

Testing Two-sided Alternatives

To test for one-sided alternative, we can order the tables by their odds ratios or by their first cell frequency n_{00}. The resulting p-value will be exactly the same, irrespective of the choice of the ordering criteria.

In practice, the two-sided tests are a lot more common. However, when testing a two-sided alternative, different criteria may result in different p-values. To test for a two-sided alternative, one can calculate the p-value by summing over $P(n_{00}) = t$ for all t with $p(t) \leq p(t*)$. Another criterion sums $p(t)$ for tables that are farther from the null, i.e., $P(|n_{00} - E(n_{00})| \geq |t^* - E(n_{00})|)$. The criteria can provide different results due to discreteness and potential skewness.

Conclusions

Fisher's test of independence for 2×2 tables is applicable for small sample sizes. The test uses exact distributions rather than large-sample approximations. Note that in small samples it is not always possible to achieve an exact significance level because the hypergeometric distribution is discrete and the p-value can assume only a few values. Consequently, the Fisher's test is conservative as the true type I error would be less than the pre-specified one.

Cross-References

▶ Fisher's Test

References

Agresti A (2002) Categorical data analysis. Wiley, Hoboken

Exon

Luiz O. F. Penalva
Department of Cellular and Structural Biology,
Greehey Children's Cancer Research Institute,
University of Texas Health Science Center,
San Antonio, TX, USA

Definition

Exons are short (50–200 bp in length) RNA sequences that are found in pre-mRNAs. Exons contain the sequences found in a mature mRNA. When a gene is transcribed, the precursor mRNA is composed of both introns and exons. After the removal of introns by the spliceosome, the precursor mRNA matures into mRNA which is then exported into the cytoplasm. This mature mRNA can now be translated into protein.

Cross-References

▶ Post-transcriptional Regulatory Networks

Expectation

▶ Prognosis

Expectation-Maximization Algorithm

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

Definition

An expectation-maximization (EM) algorithm is a method for finding maximum likelihood estimates of parameters in statistical models where the available data set is incomplete. The EM algorithm consists of two steps, firstly guessing a probability distribution over completions of missing data given the current model (known as the E-step) and secondly reestimating the model parameters using these completions (known as the M-step).

Given a likelihood function L(θ; X, Z), where θ is the parameter vector, X is the observed data, and Z represents missing values, the maximum likelihood estimate (MLE) is determined by the marginal likelihood of the observed data L(θ; x), which is often intractable. Therefore, the EM algorithm maximizes the expectation of the log-likelihood function, based on the observed samples and the current estimate of θ. The two steps of the algorithm are:

E-step: At the (t)th step of the iteration, where θ(t) is available, compute the expected value of the log-likelihood function.

$$Q(\theta|\theta(t)) = E[\log\ L(\theta|X,Z)|X,\theta], \qquad (1)$$

M-step: Compute the next (t + 1)th estimate of θ by maximizing $Q(\theta|\theta(t))$, that is,

$$\theta(t+1) : \frac{\partial Q(\theta|\theta(t))}{\partial \theta} = 0, \qquad (2)$$

References

Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J Roy Stat Soc B. 1977;39(1):1–38. JSTOR 2984875. MR0501537.

Experiment

Jutta Schickore
Department of History and Philosophy of Science, Indiana University, Bloomington, IN, USA

Definition

A procedure, method, or setting of devices and tools designed to manipulate objects and study their features and working.

Characteristics

For the most part of the twentieth century, philosophy of science focused on scientific theories and conceptual foundations of science, with a special emphasis on physics. Only in the last 30 years or so, experiments and biology have become themes for philosophical and historical analysis. In the following, I present some key general issues and questions for philosophy of experiment as well as a number of methodological, epistemological, and ethical challenges that arise specifically from experimenting with living, evolving things: What are the distinctive features and epistemic roles of experiments? What do experiments tell us about nature? When is an experimental result valid and informative? What ethical and sociopolitical concerns may arise from biological experimentation?

Analytic Descriptions of Biological Experiments and Their Epistemic Roles in Science

One of the key tasks for philosophy of biological experimentation is the identification of the distinctive features of biological experiments, their aims, and the nature of experimental reasoning. Most of this work has been done in response to the kind of philosophy of science that dominated the early twentieth century. Until the late twentieth century, philosophy of science was almost exclusively concerned with physics, and for the most part with physical theories, their structure, and the dynamics of theory change. Experimental practice was often regarded as part of the "context of discovery" and thus not amenable to philosophical analysis. Also, for a good portion of the twentieth century, philosophers generally recognized only one

epistemic role for experimental evidence: It serves as test to support or reject theories or to decide between competing theories. Arguably, it was the biological sciences that helped challenge the received view about the nature of scientific experimentation and the roles of experiments in science. Historians' and philosophers' increased attention to biology in the 1980s and 1990s highlighted the diversity of scientific practices and thus the limitation of physics-based philosophy. They showed that experiments play a number of epistemic roles in science. Three conceptions have been particularly fertile for the study of experiments: "discovering mechanisms," "experimental systems," and "exploratory experimentation."

The notion that biological research can be described as "discovering mechanisms" has been very influential for recent thought about experimentation. This approach has been developed to characterize the nature of experimental reasoning in biology, thereby challenging the view that reasoning practices are outside the scope of philosophical analysis. The basic idea originates in the study of molecular biology. Molecular biology does not have the kind of grand theory based around a set of laws or a set of mathematical models that is familiar from the physical sciences. Rather, general knowledge in the field takes the form of descriptions of the interaction of entities and activities in a biological system (such as mechanisms of DNA replication or protein synthesis). Several philosophers of biology have argued that these mechanisms – specific collections of entities and their distinctive activities – are the fundamental unit of scientific discovery and scientific explanation (Machamer et al. 2000). Experiments serve to identify the components and interactions of the mechanisms. This conception has proven valuable not only in molecular biology but in a wide range of special sciences, such as neuroscience.

A very fruitful analytic term to describe biological experiments is Hans-Jörg Rheinberger's notion "experimental system," a term that is taken directly from experimental biologists' parlance. An experimental system is the working unit of both scientists and epistemologists. The basic idea is that investigators never perform isolated experiments but series of interlocking experimental trials, and that the major driving force for modifications and changes in the experimental settings and for reorientations of questions and goals is not a preconceived theory or hypothesis but the behavior of the system itself. Following François Jacob, Rheinberger emphasizes that experimental systems are "machines to create the future." Because they have the power to produce unexpected and surprising outcomes, experimental systems influence the direction of research. Rheinberger's concept "experimental system" thus redirects the focus of the analysis from the role of experiments as tests for well-developed theories to the role of experimental systems as generators of new knowledge. The point of Rheinberger's conception of experimental systems is that it explicitly acknowledges the material, technical, and institutional setting of experiments. The experimental system with all its technical, instrumental, institutional, social, and epistemic aspects is the site of knowledge generation. This notion has made it easy for historians and sociologists to connect to Rheinberger's work, while philosophers have been slower to take it up.

The third new conception that has become fruitful for philosophy of experiment more generally is "exploratory experimentation" (see also "▶ Exploratory Experimentation"). Episodes of experimentation that can be characterized as "exploratory" have been identified in the history of physics as well as biology. The term highlights a mode of experimentation where explicit theoretical principles, guidelines, or concrete theoretical expectations are absent. In other words, exploratory experimentation is the opposite of experimentation for the purpose of testing well-developed theories. To say it with Richard Burian, exploratory experimentation is "data driven," not "hypothesis driven" (Burian 2007). Large-scale sequencing technologies are a striking example of exploratory research: These technologies generate hundreds of results, which can be considered robust even though the theoretical interpretation remains unclear.

In recent studies of biological experimentation, the once-common view that the epistemic role of experiments is to be a "test" of theories has thus come under attack from various angles. Of course, biological experiments *may* function as tests, but they may also fulfill a number of other functions. They may be motors for the reorientation of research activity, aids for discovering mechanisms, and devices for creating robust results where no theory is involved.

Reality and Reliability

General philosophy of scientific experimentation has concentrated on two issues: the reality of the objects of

our theories and experiments and the reliability of experimental results. The long-standing debate in philosophy of science about realism and antirealism concerns the interpretations of the relation to reality of scientific theories and the entities that these theories are about, such as electrons and quarks. The philosophical question is the following: Do these things really exist in the world, or are they merely theoretical constructs (useful fictions, conceptual tools) that scientists use to explain observations and experimental results? Like physics and chemistry, biological research also concerns processes, phenomena, and objects that cannot be directly observed, such as genes, neurons, and proteins. What reasons do we have for believing that the objects and phenomena of biological thought really exist?

In his influential book *Representing and Intervening*, Ian Hacking makes this debate a topic for philosophy of experiment. Hacking argues that the dispute between realists and antirealists can be brought to a close if we take into account that experimenters often use experimental objects to manipulate others – particle accelerators, for instance, utilize subatomic particles to interfere with other particles. In these cases, Hacking claims, one may assume that the particles that are used as tools are real. Biologists base their interpretations and explanations of biological things and phenomena on the data and images produced by instruments in often highly complex experimental settings. But like physicists, they routinely interfere with the research objects and assess their results by assessing the effects of their interventions. Like physicists, they thus have good reason to assume that the subvisible things and phenomena they investigate really exist.

But even if we follow Hacking and adopt a realist stance vis-à-vis experimental objects in biology, other questions arise. Experimental settings in the laboratory are artificial surroundings very different from natural habitats, and in this sense, experiments create effects that do not exist in nature. This raises issues about the scope of the knowledge generated by laboratory experiments. What, if anything, can such experiments tell us about objects and phenomena outside the laboratory?

This question is especially pressing in the context of biological experimentation because today much experimental work is carried out on a small number of model organisms. Model organisms are selected and standardized experimental objects (see "▶ Model Organism"). They are selected because they are very simple, readily available, hardy, with short gestation periods, and a large number of offspring, such as the nematode *C. elegans* or the fruit fly. These objects are studied as stand-ins for others. Features that have been uncovered in detail in one model organism are taken to be "exemplars" for others that cannot be easily studied.

Philosophical discussion has focused on a number of issues such as the sense in which model organisms are "models" (Ankeny 2001). Are they exemplars or abstractions, typical, ideal, homologous, or analogous models? And is it really appropriate to transfer the insights gained about model organisms to organisms in their natural habitat, especially given that in some cases, the laboratory environment has changed the model organisms currently being used to an extent that they no longer resemble organisms in the wild? Also, to what extent is it legitimate to transfer these insights to other, infinitely more complex organisms, especially humans, given that model organisms are usually chosen for practical purposes, such as easy tractability and general availability? Philosophers have addressed these questions through examining modeling practices and the usages of the information gained from experimenting with model organisms (Weber 2005, Chap. 6).

Another central task for philosophy of experimentation is the assessment of the empirical information generated in experiments. Experiments are complex affairs involving intricate settings and recalcitrant objects, and it is a struggle to get experiments to work. When is an experiment(al result) reliable and informative? Recent philosophers of experiment have put together sets of strategies that may be used to assess the validity of experimental results, including experimental checks and calibration; elimination of plausible sources of error and alternative explanations of the result; using an apparatus based on a well-corroborated theory; and statistical validation of the evidence (Franklin 1986).

Several philosophers of biology have argued that the concept of robustness plays a major part in the validation of experimental results. If multiple different experiments are robust, that is, if they agree in their outcomes, we may assume that the experimental result is reliable. In its most general characterization, the concept of robustness straddles epistemology and

ontology. Robustness is regarded as the decisive criterion in separating what is real from what is unreal and what is reliable from what is unreliable (Wimsatt 1981). In recent years, philosophers and historians of biology have probed a number of scientific episodes to establish whether robustness really is the main criterion for the assessment of experimental outcomes, and what other criteria have been used, and should be used, to evaluate experimental results. This discussion is still ongoing.

Sociopolitical and Ethical Dimensions of Biological Experimentation

Scientific experimentation raises important ethical and sociopolitical issues related to the use of experimentally generated knowledge and the responsibility of investigators. These are vast topics, and the discussions surrounding them have grown extremely complex and controversial, involving experts and arguments from ethics, politics, religion, and law. Here I can only touch upon some of the questions and issues that need to be addressed.

Most sensitive ethical questions arise of course in experiments with living beings, especially in biomedical contexts. What ethical codes are and should be applied in biological and biomedical experimentation? What are the limits of ethically responsible experimentation? On the most general level, experimenters are confronted with the question: When is it acceptable to expose some to risks of harm for the benefit of others? This question has been hotly debated ever since the first experiments were performed on humans and animals (cf., LaFollette and Shanks 1997). The general ethical principles that should guide human experimentation – the principles of beneficence, respect, and justice – are stipulated in official documents and policy statements, as are the regulations for the use of animals in research facilities. But the interpretation of these guidelines and their application to concrete cases are often extremely difficult. For a long time, it was common to assume that the appeal to ethical theory such as utilitarianism, deontology, or virtue ethics should guide the practice of bioethics. But philosophical discussions often proved unhelpful in concrete cases. In recent philosophical debates about the relation between (applied) bioethics and philosophical theory, many bioethicists and several philosophers have become skeptical of applying ethical theory to practical problems. Instead, they advocate starting with the problem at hand and clarifying the meaning and moral significance of key concepts such as harm and coercion, but without appeal to high-level ethical theory. Philosophers of experiment may contribute to these discussions through the analysis of potential conflicts between methodological standards of valid experimentation and ethical concerns about responsible experimentation.

Two fields of biomedical research in particular have been subject of heated ethical as well as political debates: stem cell research and the human genome project. Stem cell research aims to identify the mechanisms that govern cell differentiation. The ultimate goal is to turn stem cells into specific cell types that can be used for treating debilitating and life-threatening diseases and injuries. While this research promises to have extremely important therapeutic implications, it is also strongly contested because stem cells are obtained from human embryos. The very practice of experimenting raises intricate ethical questions. Is it ethical to destroy human embryos for the purposes of clinical research? Here, the discussions focus on the beginning of human life, the embryo's right to life and the relative values of embryonic life and the life of a mature human being. At present, each of these points is still under discussion.

The human genome project (HGP) – the mapping and sequencing of the human genome – has also generated much discussion among philosophers of science, bioethicists, political philosophy, and philosophy of law, but here the discussion has focused more on the application and use of the knowledge generated by the HGP. As the HGP assists scientists in the identification of genes and their functions, genetic testing made possible by the human genome project has raised a number of concerns and worries, for instance, about access to genetic information by insurers and employers, and the possibility of genetic discrimination. Prenatal genetic testing raises concerns about reproductive rights and eugenics. The rapid development of new technologies poses additional challenges to the discussions.

Finally, the commercialization of biomedical research raises challenging questions for philosophers of biology. Arguably, in recent decades, the influence on biomedical research of business and industry has led to changes in the ethical and methodological norms of biomedical practice (Krimsky 2003). It is an important task for philosophy of experiment to assess and

critique the impact on the integrity of biomedical experimentation of the privatization and commercialization of biomedicine.

Philosophy of experiment is still a vibrant branch of philosophy of science, and each of the aspects discussed in this entry invites further study. As the recent development of the field shows, philosophy of experiment has considerably broadened its scope, seeking to make contact with the biological sciences as well as history and sociology of science and thus to increase its social relevance.

Cross-References

- ▶ Exploratory Experimentation
- ▶ Model Organism

References

Ankeny R (2001) Model organisms as models: understanding the 'Lingua Franca' of the human genome project. Philos Sci (Proc) 68:S251–S261

Burian RM (2007) On MicroRNA and the need for exploratory experimentation in post-genomic molecular biology. Hist Philos Life Sci 29:285

Franklin A (1986) The neglect of experiment. Cambridge University Press, Cambridge

Hacking I (1983) Representing and intervening. Cambridge University Press, Cambridge

Krimsky S (2003) Science in the private interest. Has the lure of profits corrupted biomedical research? Rowman-Littlefield Publishing, Lanham

LaFollette H, Shanks N (1997) Brute science: dilemmas of animal experimentation. Routledge, London/New York

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

Rheinberger H-J (1997) Toward a history of epistemic things. Stanford University Press, Stanford

Weber M (2005) Philosophy of experimental biology. Cambridge University Press, Cambridge

Wimsatt WC (1981) Robustness, reliability, and overdetermination. In: Brewer MB, Collins BE (eds) Scientific inquiry and the social sciences. Jossey-Bass, San Francisco, pp 124–163

Experiment Design

- ▶ Model-based Experimental Design, Global Sensitivity Analysis

Experimental Design

Clemens Kreutz[1] and Jens Timmer[2,3,4]
[1]Institute for Physics, Freiburger Center for Data Analysis and Modeling (FDM), University of Freiburg, Freiburg, Germany
[2]Institute for Physics, University of Freiburg, Freiburg, Germany
[3]BIOSS Centre for Biological Signalling Studies and Freiburg Institute for Advanced Studies (FRIAS), Freiburg, Germany
[4]Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden

Definition

An *experimental design* or shortly a *design* is defined as the set of experimental conditions which can be chosen by the experimenter. In statistics, the experimental conditions are also called *independent variables* or *design variables*. In contrast, the measured variables are called *dependent variables* because the realizations depend on the independent variable, i.e., on the design, and on the system's behavior. The design $\mathcal{D}$ usually covers all variables of a model $F(\mathcal{D}, \theta)$ except the parameters θ.

In systems biology, an experimental design $\mathcal{D}$ usually specifies the choice of the external perturbations, the observables, as well as the number and time points of the measurements (Kreutz and Timmer 2009). In mathematical terms, the design $\mathcal{D} = \{t_i, u_i, o_i\}$ is the set of all measurement times t_i, perturbations u_i, and observables o_i for all data points i.

The procedure of finding optimally informative experimental designs, i.e., to choose the number of data points N as well as t_i, u_i, o_i for all measurements $i = 1, \ldots, N$, is termed optimal experimental design or design of experiments (DoE).

Cross-References

- ▶ Active Learning
- ▶ Experimental Design, Variability
- ▶ Input
- ▶ Observable
- ▶ Optimal Experimental Design

References

Kreutz C, Timmer J (2009) Systems biology: experimental design. FEBS J 276:923–942

Experimental Design, Variability

Roger Higdon
Seattle Children's Research Institute,
Seattle, WA, USA

Synonyms

Analysis of variance (ANOVA) tables; Components of variance; Sources of variability

Definition

Identifying and accounting sources of variability is one of the key aspects of statistical experimental design. These can correspond to experimental factors under study, subjects or biological specimens, samples for subjects, time points, technical replication of experimental protocols, and blocking or grouping of subjects or samples.

Characteristics

Sources of variability in the experimental design of biological study are often divided into two categories: biological variability (variability due to subjects, organisms, and biological samples) and technical variability (variability due measurement, instrumentation, and sample preparation). However, within biological and technical variability there may be many levels and sources of variability, including experimental factors, time points, subsampling, repeated measurements, and blocking. In order to design and analyze biological experiments, it is necessary to properly account for the different sources of variation (Box et al. 2005; Cox and Reid 2000). A useful tool for keeping track of the sources of variability is the analysis of variance (▶ ANOVA) table. Table 1 provides a simple example; it shows the sources of variation and the levels of replication (degrees of freedom) and indicates the level of variation. Terms higher up in the table contain variability from the terms below. Parentheses explicitly mark terms that are nested within others while crossed terms indicate interactions between different levels of variation. As can be seen in Table 1, the potential sources of variation in systems biology can get very large even in a relatively simple study.

Experimental Design, Variability, Table 1 Sample ANOVA table showing sources of variability for a medical study utilizing proteomics analysis

Source	DF
Hospitals	m – 1
Treatment	1
Patients (hospitals)	n(m – 1)
Treatment x hospitals	m – 1
Treatment x patients (hospitals)	n(m – 1)
Samples (patients)	kn(m – 1)
Treatment x samples (patients)	lkn(m – 1)
LC fractions (samples)	jlkn(m – 1)
Treatment x LC fractions (samples)	jkln(m – 1)
MS runs (LC fractions)	ijkln(m – 1)
Treatment x MS runs	ijkln(m – 1)
Total	N – 1 = 2mnklji – 1

Not replicating sources of variability can lead to confounding between different sources of variability. For example, in Table 1 if there is only one patient per treatment group and even if samples or MS runs are replicated, one will not be able to differentiate variability due to differences in treatments from the variability due to differences between patients. This would be an example of pseudo-replication. Even if sources of variability are replicated, insufficient replication can lead to poor power for statistical tests. This can occur even when many repeated runs of a sample are done with little replication of biological samples or subjects.

Sources of variability need to be accounted for in the analysis and experimental design of a study. Historically, R. A. Fisher was among the pioneers in utilizing statistical methods to account for variability in experimental design (Fisher 1918). There are several approaches to accounting for variability within the experimental design. Holding factors constant by utilizing standard protocols, the same instrumentation and settings, the same operator, and so on can eliminate the sources of variability. Using blocking can reduce other sources of variability. Blocking is grouping experimental samples within similar experimental or environmental conditions, for example, grouping by

time, geography, or batch. If all experimental factors are represented within the same block, comparisons can be made within the block; if not, experimental variability may become confounded with the blocking factor. Randomizing experimental factors to biological samples and biological samples to blocks can convert other sources of variation to random variation that can then be accounted for by the statistical model. Replication of biological samples and blocks can reduce the amount of biological variability.

There are a number of approaches that can be applied to the analysis stage to reduce or eliminate sources of variability. Normalization can reduce systematic and technical sources of variability that are part of the measurement process. The use of covariates such as measurements of the initial state of a biological sample such as weight or age can correct for variability in the initial conditions or other properties of biological samples.

There are examples of such approaches in systems biology (Kreutz and Timmer 2009). In relative expression analysis, multichannel microarray studies implicitly use the array as a blocking factor to eliminate the measurement variability between arrays (Churchill 2002). Other types of blocking are common in high-throughput studies: the day the samples were run, the cage the mice were kept, the instrument used, the lab that ran the assay, or the 96 well plate containing the sample. Normalization is commonly used in high-throughput data studies such as microarrays or proteomics to eliminate systematic variability. Mixed and multilevel models are an effective statistical analysis approach to deal with multiple sources of variability.

Cross-References

- ▶ Mixed and Multi-Level Models
- ▶ Relative Expression Analysis

References

Box GE, Hunter WG, Hunter JS, Hunter WG (2005) Statistics for experimenters: design, innovation, and discovery, 2nd edn. Wiley, New York

Churchill GA (2002) Fundamentals of experimental design for cDNA microarrays. Nat Genet 32:490–495, Supplement

Cox DR, Reid NM (2000) The theory of design of experiments. Chapman & Hall/CRC, Boca Raton

Fisher R (1918) Studies in crop variation. I. An examination of the yield of dressed grain from Broadbalk. J Agr Sci 11:107–135

Kreutz C, Timmer J (2009) Systems biology: experimental design. FEBS J 276(4):923–942

Experimental Methods for Studying Immune Signaling Network Dynamics

▶ Immune Pathway Dynamics, Biological Methods

Experimental Organism

▶ Model Organism

Experimental Planning

▶ Designing Experiments for Sound Statistical Inference

Experimental Standard Conditions of Enzyme Characterizations

Carsten Kettner
Beilstein-Institut zur Förderung der Chemischen Wissenschaften, Frankfurt am Main, Germany

Synonyms

ESCEC

Definition

The Experimental Standard Conditions of Enzyme Characterizations (ESCEC) is a conference series biennially hosted by the Beilstein-Institut in Rüdesheim, Germany (Hicks and Kettner 2010). The ESCEC symposia serve as a platform for the exchange of ideas and link to the scientific community. Experts from all fields of experimental, theoretical, and bioinformatics enzymology and metabolic network investigation present and discuss

new results, approaches, and methodologies as well as pitfalls and problems of data generation and reproduction. Suggestions from the STRENDA Commission form the basis of subsequent discussions in following symposia where they were are improved, rejected, or replaced by alternatives.

Cross-References

► STRENDA

References

Hicks MG, Kettner C (2010) Experimental standard conditions of enzyme characterizations. Logos, Berlin

Experimental Unit

Olga Vitek
Department of Statistics, Department of Computer Science, Purdue University, West Lafayette, IN, USA

Definition

Experimental unit is the basic unit of experimental design and is the subject, sample, or object that carries the condition or the treatment.

Cross-References

► Designing Experiments for Sound Statistical Inference

Expert Elicitation

► Prior Elicitation

Expert System

► Knowledge-based System

Explanation in Biology

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

Definition

An explanation of a characteristic of a system is an account of why that system has that characteristic.

Characteristics

Explanations in biology, unlike those in physics, rarely proceed by filling out the parameters in mathematical equations (laws) that describe the general characteristics of the relevant form of organization. With the notable exception of population genetics and theoretical ecology, biologists typically employ mechanistic, functional, and historical strategies of explanation, rather than mathematical ones. To some scientists, this is a sign of the immaturity of biology, perhaps resulting from the difficulty to take into account the extraordinary high number of parameters and variables relevant to the phenomena that interest biologists. Structuralist biologists such as Brian Goodwin (e.g., 1994) have pointed to the dominance of gene-centered and historical approaches as the main source of immaturity. They hope that in the future, more research, more data, more computational power, better mathematics, and above all, more insight into the proper way of explaining in the natural sciences and more willingness to pursue the structuralist strategy will enable what is thought to be a more scientific approach.

Many others reject the view that the minor role of general laws in biological explanation is a sign of its immaturity. They argue that the objects of study in biology have characteristics that justify approaches different from those suitable in physics. One such view points to the process by which life gets its shape: evolution by natural selection. For example, the evolutionary biologist Ernst Mayr (1904–2005), one of the founding fathers of the modern synthesis, argues that it makes sense to ask why questions in biology but not in physics and chemistry (e.g., Mayr 1997, Chap. 6). The reason is that organisms owe their characteristics to their history,

whereas history is, in general, not important to the characteristics of physical and chemical systems. So whereas physics and chemistry must limit themselves to how questions, biologists should address why questions in addition to how questions. Within biology, how questions and why questions are, according to Mayr, the subject of two "largely separate fields": functional biology and evolutionary biology, corresponding with two modes of explanation: functional explanation and evolutionary explanation. Functional explanations answer how questions by describing the operation of mechanisms. Evolutionary explanations answer why questions by revealing the evolutionary history of those mechanisms. Both kinds of explanation are legitimate and needed to understand the living world, but evolutionary explanations provide the distinctive biological perspective.

Whereas Mayr seeks the justification of the biologist's concern with why questions merely in the importance of history for understanding the characteristics of organisms, many philosophers (e.g., Dennett 1995) refer to the specific character of that history as being a selection history. Unlike physical objects, organisms have the character they have because ancestral variants with those characteristics were favored over variants with other characteristics. For that reason in biology, but not in physics and chemistry, the answer to the question how a certain characteristic is produced must involve an answer to the question why (e.g., for what effects) that characteristic was favored in the selection process.

In my view, attempts to justify types of explanation in biology that differ from explanations in physics and chemistry by appeal to the importance of history for understanding organisms or by appeal to the special character of that history are unsatisfactory. Such attempts take the importance of history for granted, whereas they should explain it. Moreover, they fail to address the many differences between *functional* biology and physics and chemistry (see Fox Keller 2002 to get a feeling for the weirdness of mechanistic explanation in biology in the eyes of a theoretical physicist). The most notable differences are a preference for mechanistic models that visualize interactions over abstract mathematical system descriptions (i.e., equations that do not bear an obvious relation to the physical properties of the parts of the system), appeal to role functions to explain how mechanisms work and the appeal to advantages and requirements to explain why a certain mechanism has the characteristics it has.

A better justification starts with the observation that organisms are highly organized (see also the essay on ▶ Organization): their capacities and characteristics critically depend not only on the characteristics of their parts but also on the spatial arrangement of those parts and on the order and timing of their activities. Explanations in physics and chemistry are typically aggregative: they derive the behavior of a system (such as the behavior of a volume of gas as described by the Boyle-Charles' law) by aggregating the behavior of the parts of that system (the molecules of which the gas is composed as described by Newton's laws of motion). As philosopher William Wimsatt (e.g., 2007, Chap. 12) has argued, very few system properties are aggregative under all possible decompositions of a system into parts. The list is pretty much exhausted by the quantities that appear in conservation laws: mass, energy, momentum, and net charge. All other system properties are more or less organized in the sense that they break down under at least one of the four conditions for aggregativity identified by Wimsatt (invariance under rearrangement and substitution, size scaling, decomposition and reaggregation, and linearity). By carefully choosing a decomposition, by limiting explanations to certain forms of organization and by introducing idealizations and approximations (in the description of the system and in the derivation of the system's behavior), the power of aggregative explanation can be expanded immensely. However, if a system is very highly organized, there comes a point where aggregative explanation is no longer possible and where the system's organization must be taken into account.

In biology, this is done by viewing organisms as mechanisms (▶ Mechanism) for being alive (Wouters 2005). For the purpose of this essay, a mechanism for a certain behavior can be defined as a complex system that produces that behavior by the organized interaction of its parts (cp. Glennan 1996; Machamer et al. 2000). Because the behavior of a mechanism results from its organization, a mechanism can be seen as a solution to the problem of how to organize the parts and their interaction in such way that this behavior is generated. Note, that this problem need not be *experienced* by the mechanism or some other system. The difficulty (i.e., the amount of organization needed) to produce or maintain that behavior in the circumstances

in which it occurs suffices to talk of problems. As George Cuvier (1769–1832), the founding father of functional zoology, already noted, the very existence of organisms means that they have solved the problem of how to stay alive (cf. Reiss 2009). Organisms exist far from thermodynamic equilibrium and can exist only by actively maintaining their organization (Prigogine and Stengers 1984). To understand how this form of existence is possible, a combination of explanations of four kinds must be employed, as many biologists have recognized (e.g., Tinbergen 1963): mechanistic explanation, functional explanation, developmental explanation, and evolutionary explanation.

Mechanistic explanations explain how the behavior of a mechanism arises from the properties of the parts, their interaction, and the way in which this interaction is organized. Biology's functional perspective glues explanations of different mechanisms together by viewing those mechanisms in terms of their role in maintaining the state of being alive (i.e., in terms of their biological role). The different organ systems have specific roles in bringing about the living state. Each organ system consists of organs with specific roles in bringing about the properties of the organ system that enable that organ system to perform its role in the maintenance of the living state. Each organ in turn is divided into subsystems, each with a specific role in bringing about the properties relevant to that organ's biological role. And so on, until a level is reached in which the relevant system properties arise out of unorganized components. Attributions of biological roles (often called "function ascriptions") (see ▶ Function, Biological) situate the different mechanisms in this encompassing hierarchical organization, making it possible to see different mechanistic explanations as part of the larger project to understand how organisms are able to stay alive. Attributions of biological roles are, hence, the key to explanation in biology.

The very fact that the behavior of mechanisms (including organisms) results from the *organization* of their parts (in addition to their composition) means that in order to understand a mechanism, it does not suffice to explain how it works. The mechanism's solution to a certain problem need not be the only possible solution to that problem. On the other hand, by definition, not any form of organization will solve any problem. So in order to understand a mechanism, one must not only explain how its organization results in a certain behavior (as is done in mechanistic explanations) but also why that mechanism's organization solves the problem, whereas other forms of organization (often called "designs") do not. This opens up the possibility and the need to explain why a mechanism has the characteristics it has on the basis of the requirements it has to satisfy: the constraints imposed on it by the problems it must solve, the other characteristics of that mechanism, and the conditions under which it works. This is what functional explanations (▶ Explanation, Functional) do.

Furthermore, because, by definition, one does not get a mechanism by throwing its parts together, the question of how the organization that solves the problems arises in the course of time needs consideration too. As became clear in the wake of Charles Darwin's theory of evolution, in the case of living mechanisms, the answer requires a two-staged explanation. Developmental explanations (▶ Explanation, Developmental) explain how the organization arises in the course of individual development. Evolutionary explanations (▶ Explanation, Evolutionary) explain how the machinery to develop the required organization arose in the course of evolutionary history.

So, the view of organisms as solutions to the problem of how to maintain the living state provides a unifying perspective in biology that explains and justifies the importance of the four kinds of explanation employed in biology, the relation between those explanations, and the main differences between biology on the one hand and physics and chemistry on the other.

Cross-References

- ▶ Explanation, Developmental
- ▶ Explanation, Evolutionary
- ▶ Explanation, Functional
- ▶ Function, Biological
- ▶ Mechanism
- ▶ Organization

References

Dennett DC (1995) Darwin's dangerous idea. Simon & Schuster, New York

Fox Keller E (2002) Making sense of life. Harvard University Press, Cambridge

Glennan SS (1996) Mechanisms and the nature of causation. Erkenntnis 44(1):49–71

Goodwin BC (1994) How the leopard changed its spots. Phoenix Books, London
Machamer PK, Darden L, Craver CF (2000) Thinking about mechanisms. Philos Sci 67:1–25
Mayr E (1997) This is biology. Harvard University Press, Cambridge
Prigogine I, Stengers I (1984) Order out of chaos. Heinemann, London
Reiss JO (2009) Not by design: retiring Darwin's watchmaker. University of California Press, Berkeley
Tinbergen N (1963) On aims and methods of ethology. Z Tierpsychol 20:410–433
Wimsatt WC (2007) Re-engineering philosophy for limited beings. Harvard University Press, Cambridge, MA
Wouters AG (2005) The functional perspective of organismal biology. In: Reydon TAC, Hemerik L (eds) Current themes in theoretical biology. Springer, Dordrecht, pp 33–69

Explanation, Developmental

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

Synonyms

Embryological explanation

Definition

A developmental explanation explains how a trait came into being through a process starting at the single-celled stage of the existence of the organism. Since developmentally explaining a trait is an explanation of its development, it requires explaining some detail of the developmental processes which were involved; "developmental explanation" secondarily means the explanation of a particular feature, moment, or process *of* development itself, for example, cell division, apoptosis, and specific pattern formations.

Characteristics

Development has narrow and large meanings: very narrowly, it is the embryogenesis, that is, the process leading from conception to hatching or birth; less narrowly, it means the process leading from the zygote to the adult form. The larger meaning covers the whole life of the organism, which includes episodes such as reproduction and senescence (West Eberhardt 2003; Gilbert 2003). Developmental explanations pinpoint events in the lifetime accounting for an individual carrying a trait, whereas evolutionary explanations (▶ Explanation, Evolutionary) often stand at the population level.

What Development Explains

The general explanandum of developmental explanations is twofold: to explain why and how a zygote develops into an adult of its species, with its species-typical traits; and to understand why and how systematic deviations from the normal type occur, which usually concerns "teratology." Often, a back and forth movement between teratology and normal development characterizes developmental explanation, because knowing which alterations (e.g., silencing genes; injecting proteins, etc.) produce a specific abnormal development allows understanding what is necessary to develop the corresponding normal trait; therefore, many experimental developmental theorists proceed by such controlled perturbations of development, at any hierarchical level (genes, cell, tissues).

Developmental explanations target ▶ model organisms such as nematodes, sea urchins, frogs, chicken; through those models they aim at understanding what is common across many species and clades. Development has many varieties: some species hatch, some have larval stages, some develop from a unicellular zygote as a sort of bottleneck between generations – whereas others do not. Lowest level mechanisms involving genes and gene transcription are likely to be very general across multicellular organisms; but some higher level traits typical to a clade (i.e., eyes, hearts, etc.) require targeting an appropriate model organism.

In principle, any trait in an *individual* organism can receive a developmental explanation. However, the focus is often put on species(or clades)-typical traits: heart or central nervous system in vertebrates, tetrapod's limb, butterflies' wings striped, etc. Some theoreticians view the scope of developmental explanation as in principle limited to *species-typical* traits and the genealogy of a species type, then of a family type, etc., as in early descriptive embryology. This conflicts with the core Darwinian idea that *variation* is proper to species.

Inversely, variants in a population can often be explained by a same core mechanism, whose various outputs depend on the initial conditions, and which will yield both statistically typical and also eccentrically varying forms (e.g., stripes on echinoderms). But this idea of a developmental matrix of variation seems not enough to solve the issue of the individual versus typical scope of developmental explanations.

Developmental explanations explain the generation of organic form and therefore, given that individuals of a same species are typically like their parents, contribute to explain the conservation of form across time. The transmission of traits across generation is "inheritance"; for biologists of the nineteenth century, explaining development was supposed to explain inheritance, whereas the Modern Synthesis (Darwinian) dissociated inheritance (explained by Mendelian transmission of characters from parents to offspring) from development, as the process of growing individual forms. The integration of Mendelian genetics with Darwinism from the 1930s underpinned the split between developmental and evolutionary theories, and the current call for a new synthesis between development and evolution (Gilbert et al. 1996). In a radical developmentalist perspective, the constancy of the main developmental processes (and not the transmission of traits) explains the reproduction of form across generation and then contributes to explain the conservation of form.

Besides *physiology* – the science of the functioning of adult organisms – *embryology* has been constituted as the science of study of structure and function of *developing* organisms, including the function of developmental processes. In this sense, and given that *functionality* as well as persistent *form* metaphysically characterize organisms, their difference exemplified the two poles of biology, identified in 1911 by E.S. Russell in *Form and function*: the science of *function* and the science of *form*. Developmental explanations typically instantiate the explanatory style proper to questions about biological form. Many of the current controversies about the purported absence of development within evolutionary biology express this conflict between these two stances on biology (Amundson 2005). Especially, ▶ adaptationism in evolutionary explanation neglects the coherence of organismal form (in favor of explaining isolated traits as optima), whereas developmental explanation aims at accounting for it.

Alternative Epistemological Options in Developmental Studies

Embryology emerged in a context where debates opposing ▶ preformation and epigenesis were salient. The current idea that genes and ▶ epigenetics both contribute to development sounds like a synthesis of preformationism and epigeneticism. More generally, developmental explanations can be divided into two styles (often combined): those which pinpoint processes occurring step after step within development, and those which specify some informational content or specified dispositions present along development.

Two other general stances in biology have been conflicting within embryology: vitalism and mechanism. Vitalists hold that development presents us with the clearest expression of vital forces, what Driesch called "entelechies," supposed to make sense of the classical sea urchin experiments (▶ Developmental Biology, Classical Sea Urchin Experiments). Mechanists claim that physical and chemical forces are alone at work in development, a process on a par with other chemical processes like fermentation or corruption. Famously, the school of *Entwicklungsmechanik* (Roux, His), in late 1800s, studied development by experimental controlled perturbation, in order to characterize the conditions under which physical processes result in an embryogenesis. Nowadays, nobody endorses vitalism, yet there is an informal consensus about the fact that some "▶ emergence" occurs in development, even if what this means is still debated. But what shaped modern developmental explanations, in its difference with initial embryology initiated by Wolff (1764) and culminating with von Baer (1828), is cell theory and Mendelian genetics, cytology, and molecular biology.

Cells, Genes, and Epigenetics

Elaborated by Schwann, Schleiden, and then Virchow in the late 1800s, cell theory rooted the concepts of embryologists – layers, tissues, organs – within their substrate, namely, cells (Mazzarello 1999). Cell proliferation proved to be the basic operation which underpinned embryogenetic development as well as reproduction (the zygote results from cells from the mother and the father). Genes, postulated by Mendelian theory as the carriers of inheritance, turned out to be localized on chromosomes and finally identified to their material substrate, DNA in 1953. (However, the *assumed* identity between genes of molecular biology and the genes of classical transmission genetics is

highly controversial (Moss 2003)). Genes have then been central both in evolutionary (▶ Explanation, Evolutionary) and in developmental explanations because they have been seen both as substrate for inheritance (which pertains to evolution) and as causes of development (as they code for traits). The extreme view on the role of genes in development consists in claiming that they are a program (▶ Evolution Programming) for development (or a "recipe," as Dawkins says), which is undoubtedly the modern form of preformationism, whereas this has been challenged by recent advances in our knowledge of ▶ epigenetics, as well as by the ongoing understanding that "genes" are much more complex than a mere linear continuous strain of coding DNA.

Each cell in a eukaryotic multicellular organism has the same DNA in its nucleus, which conditions proteins and then organismal traits. The genesis of the organisms is precisely the process through which the zygote multiplies and differentiates into various cell types which in their turn combine into various tissues and organs, stemming from the various layers, which emerged in the first stage (blastula) of the process, and then brings about organogenesis. ▶ Mitosis explains cell multiplication; however, there remains the question of why and how it is the case that, for example, cells in the brain develop into neurons whereas cells in the skin develop into epithelial cells; hence developmental explanations first aim at uncovering the reasons why each cell expresses distinct genetic resources. Morphogenesis, or pattern formation, assuming cell differentiation, is the second issue to be solved.

Genes are either expressed (active) or repressed. Their expression is mediated by signals from external environment factors, which allow a cell to express the "right" genes in the place it is, and therefore, produce proteins, change this environment and contribute to the extant signals around it, so that it will trigger other events of gene expression; reciprocally, this cellular environment contains products like transcription factors already synthesized by some genes.

Essential in developmental explanation is the identification of episodes of ▶ embryonic induction, where a cell is instructed to a specific cell fate. Identifying morphogenetic fields as set of cells likely to undergo a specific fate (limb, eye, etc.) is a crucial step in such explanation, even if morphogenetic fields can be transformed.

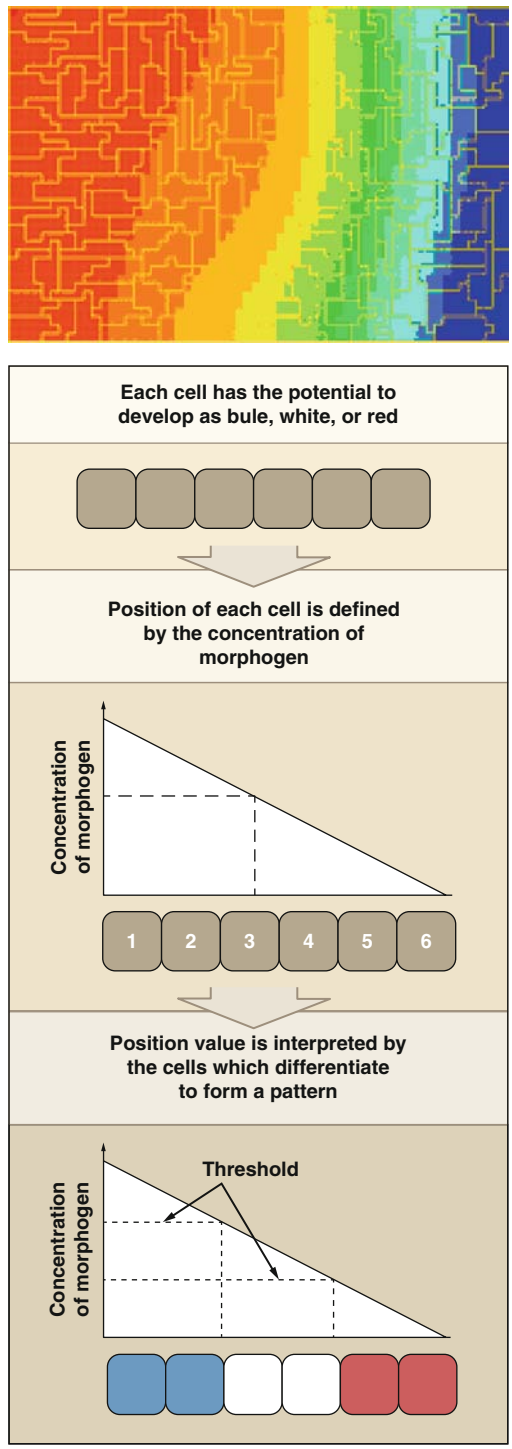
The (study of the) set of nongenetic factors, intra- or extracellular, which conditions gene expression, is often called ▶ epigenetics. Hence, epigenetic factors are sought, which are able to switch off and on genes. While some genes are precisely acting upon other genes (Regulatory Genes), others are expressing proteins required for building up the body. Developmental explanations are thereby in principle multiscale explanations because causal factors intervene at each level (signals in the tissues, e.g., gradients of a substance, signals within the cell and genetic instructions, signaling inter-tissues).

A specific set of genes, developmental genes, condition across many eukaryote phyla the anteroposterior and dorsoventral axes, as well as many other feature which seemed previously to be analogous (such as eyes). Understanding the specific timing of the activation of those genes is crucial to understand the process of development. Some biologists (e.g., Carroll 2005) would reconcile evolutionary and developmental perspectives by seeing developmental theory as the genetics and epigenetics of developmental genes, yet other Evo-Devo scientists (e.g., Müller and Newman 2005) view this option as still too much gene-centered.

Cascades, Signals and Networks

Models of development involve various kinds of information: instructions coded in genome, and signals exchanged between cell environments and receptors in cells (Gilbert 2003). The role of the latter has been modeled by the "French flag model" proposed by Lewis Wolpert, which provides a general view on pattern formation (e.g., veins in insect wings, cartilages in vertebrate limbs) (Wolpert 1994). Cell types develop according to the gradient of a "morphogen" produced by the cells, a threshold in the gradient triggering a specific gene expression in some cells, and then a cell type; the continuity of the diffusion of the morphogen through the gradient translates into the discontinuity of cell types, the concentration of the product due to its position informs on the cell fates (Fig. 1). This dialectics between genetic instructions and "positional" information provides a scheme of genes-cells-tissues interaction ruling morphogenesis. Philosophers should note the pervasiveness of a vocabulary such as: cells *interpret*, *read* their position, etc. Even if those words are rather theoretical reformulations than metaphors, one can still question whether they are reformulated according to a unique information-theoretical framework.

A developmental explanation therefore can appear as a *cascade* of events involving signaling, production of morphogens, expression and repression of gene products. Modeling those cascades and identifying



Explanation, Developmental, Fig. 1 (**a**) The French flag model pattern. (**b**) The causal underpinning of the flag pattern (After Kerszberg and Wolpert 2007)

the morphogens as well as their propagation dynamics and the cells' receptivity mechanisms yields an explanation of morphogenesis, exemplified by the emergence of asymmetry in the heart (▶ Asymmetry of the Heart, Development).

The nonlinear, complex relationships between thousands of regulatory genes in a cell, transcription factors, and finally a cell lineage involved in morphogenesis are likely to be mathematically modeled by *networks* (Davidson 1986). The *topology* of the network represents the signaling, inducing, repressing, and expressing pathways, whereas its *dynamics* captures the variation in quantitative variables (amount of genes products, etc.) according to some equations (e.g., Michaelis-Menten, etc.). ▶ Gene Regulatory Networks (GRN), since the initial investigation of the GRN proper to Endo 16 in sea urchins by Davidson, have proved therefore to be crucially involved in the formation of patterns during development (Davidson et al. 2003). The network's view is powerful but, while it avoids a crude idea of a genetic program ran by development, it still gives genes and transcripts the leading role in developmental processes. An alternative view emphasizes the fundamental role of a limited set of molecules (e.g., cadherin) involved in many developmental episodes that is pervasive across many clades, finally compacted into developmental building blocks (DPM, "Dynamical patterning modules," see Forgacs and Newman 2005; Gilbert et al. 1996), irrespective of which genes will be recruited to supervise development. GRN are clearly developmental modules. At a higher level, endoderm and mesoderm are also developmental modules, as well as Newman's DPN. Plausibly, alternative interpretations of developmental explanation – for example, more or less gene-oriented – are also distinguishable through the way they identify developmental modules. This is because they do not use the same criteria to pinpoint them.

Nevertheless, GRN are also likely to account for many features of developmental processes, especially ▶ canalization (because they offer multiple pathways between a given output and input, so that the failure of some genes does not often affect the final product) and ▶ plasticity, two key issues for Evo-Devo.

Articulating Explanatory Regimes

"Developmental explanation" in general means explaining a trait by unraveling a developmental process.

When and how are developmental explanations needed? According to Ernst Mayr (1961), developmental explanations pertain to proximate causes of phenomena, and evolutionary explanations (▶ Explanation, Evolutionary) to their ultimate causes. However, given that natural selection requires variation and that some variation relies upon possibilities given by developmental processes (and not mere mutations and recombinations), developmental explanations may be implied by evolutionary questions. In general "developmental ▶ constraints" condition the generation of variation, through which selection shapes traits and organisms. Therefore, evolutionary and developmental explanations seem complementary, the former specifying which variations are available and the second explaining which actual variants will emerge on the basis of these variations.

However, it may be that sometimes a developmental explanation is sufficient because it unravels a physicochemical structure involved in a large set of phyla, and therefore explains the pervasiveness of some traits without a need to appeal to selective pressures. The self-organizing gene networks investigated by Kauffmann's *Origins of Order* (1993) provide developmental explanations of this type. An example would be the modularity of cell metabolic networks: even if it seems that modularity confers a selective advantage (through ▶ robustness), so that natural selection seems responsible for the fact that all cell types in many clades exhibit modular networks, the mere topology of networks will indeed promote mostly modular networks without the need for natural selection to prune a set of randomly varying networks (Solé and Valverde 2008). In cases like this, the developmental explanation seems enough to explain the generality of a trait, without hypothesizing a fixation through natural selection. The final status of Evo-Devo seems is tied to a decision about whether such cases are exceptions, or not.

Cross-References

- ▶ Canalization
- ▶ Emergence
- ▶ Epigenetics
- ▶ Explanation, Evolutionary
- ▶ Gene Regulatory Networks
- ▶ Model Organism
- ▶ Plasticity
- ▶ Preformation and Epigenesis

References

Amundson R (2005) The changing role of embryo in evolutionary theory. Cambridge University Press, Cambridge
Carroll S (2005) Endless forms most beautiful: the new science of evo devo and the making of the animal kingdom. Norton, New-York
Davidson EH (1986) Gene activity in early development. Academic, Orlando
Davidson E, McClay D, Hood L (2003) Regulatory gene networks and the properties of the developmental process. PNAS 100(4):1475–1480
Forgacs G, Newman S (2005) Biological physics of the developing embryo. Cambridge University Press, Cambridge
Gilbert S, Opitz G, Raff R (1996) Resynthesizing evolutionary and developmental biology. Development and evolution 173:357–372
Gilbert S (2003) Developmental biology. Sinauer, Sunderland
Kerszberg M, Wolpert L (2007) Specifying positional information in the embryo: looking beyond morphogens. Cell 130:205–210
Mayr E (1961) Cause and effect in biology. Science 134(3489):1501–1506
Mazzarello P (1999) A unifying concept: the history of cell theory. Nat Cell Biol 1:E13–E15
Moss L (2003) What genes can't do. MIT Press, Cambridge, MA
Müller G, Newman S (2005) The innovation triad: an evo-devo agenda. J Exp Zool 304B(6):487–503
Solé R, Valverde S (2008) Spontaneous emergence of modularity in cellular networks. J R Soc Interface 5:129–133
von Baer KE (1828) Entwickelungsgeschichte der Thiere: Beobachtung und Reflexion, Bornträger, Königsberg
West Eberhard MJ (2003) Developmental plasticity and evolution. Oxford University Press, Oxford
Wolff CF (1764) Theorie von der generation. Birnstiel, Berlin
Wolpert L (1994) Positional information and pattern formation in development. Dev Genet 15:485–490

Explanation, Evolutionary

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

Definition

Evolutionary theory is the general framework for modern biology, in the sense that all living phenomena have

an evolutionary history which somehow accounts for them being the way they are. Ernst Mayr usefully distinguished two sets of inquiries in biology: the "functional biology," looking for "proximate causes" of a trait in an organism, that is, causes pertaining to the lifetime of the individual, and the "evolutionary biology," looking for "ultimate causes," namely, causes at the level of the history of the species to which belongs the individual. The former includes physiology, molecular biology, developmental biology, etc., whereas the latter includes paleontology, population genetics, behavioral ecology, systematics, etc. Evolutionary explanation is the set of explanatory styles to be met in this field (Ridley 2001).

Characteristics

Explananda and Explanantia

Evolutionary theory, as it was formulated by Darwin, holds two key ideas: all living forms share a common history and (as the naturalists hypothesized) a common origin; natural selection is a process responsible for many of the features of this Tree, and of the species and organisms it includes. In the usual theory, called the "Modern Synthesis," namely, a synthesis of Darwinism and Mendelism on the basis of population genetics in the 1930s, evolution is related to the dynamics of gene frequencies.

Let us first specify what evolutionary explanations are intended to explain. First, they should explain *evolution*; then, they should explain the *diversity* of living beings; the *adaptedness* of organisms to their environments – on the basis of the fact that they display features which suit them to their environment, for example, the webbed feet of ducks are suited to swimming. So they explain the traits themselves, which are adaptations. Given that the Modern Synthesis has defined evolution as the "change in allelic frequencies," a first explanandum is the *change in the relative frequency* of genes. The existence of a trait in this context is explained when the fixation of a gene underlying it is explained.

The set of explanantia mostly used to explain those explananda are natural selection, genetic drift, and "*constraints*." The last of these is loosely defined. It includes "phylogenetics inertia," that is, the fact that a same trait in a population is indeed a character state of an ancestor and did not happen through natural selection even if it is adaptive. It also means "developmental constraints" (see ▶ Explanation, Developmental). *Natural selection* is the process of differential reproduction of individuals carrying traits which give them different chances of having their traits represented in the next generation. As soon as a population of individuals has traits which are varying and heritable, and if those traits causally influence their chances of reproduction, then we have natural selection. This third property means that those traits have, or contribute to, ▶ fitness. Finally, *drift* is the statistical effect due to the fact that generation after generation a stochastic sampling of the population takes place. The smaller the population, the higher the expected amount of drift; whereas in an infinite population, selection would be the sole causal factor, hence the fittest traits or alleles would in general go to fixation.

Thus, evolutionary explanation is always a population level explanation. All factors are differential, so the whole process needs a population of varying individual; it contrasts with an explanation which would consider an individual and explain its traits by investigating how they developed. To this extent, population level and individual level explanations are rather complementary than competing: "why does individual A have trait X?" can be answered by considering its history, whereas population level explanations answer questions such as: "why does each individual in the population have trait X?," or "why do many of them (or: a minority) have trait X?" (e.g., why are some people left-handed?)

Two Evolutionary Questions

Evolutionary explanations often build models, analytic or simulated, which on plausible assumptions describe the evolutionary dynamics at stake or the way it may have produced the explanandum; model-building and comparing with data are the two stages of the explanation. ▶ Fitness and population size influence evolutionary dynamics (as well as migration and mutation rates). Yet ▶ fitness is a probabilistic term, and has causes: all the interactions within which an organism can engage, and into which the trait of interest is involved, determine in the end the fitness value of the trait or of the organism. This entails a difference between two kinds of questions:

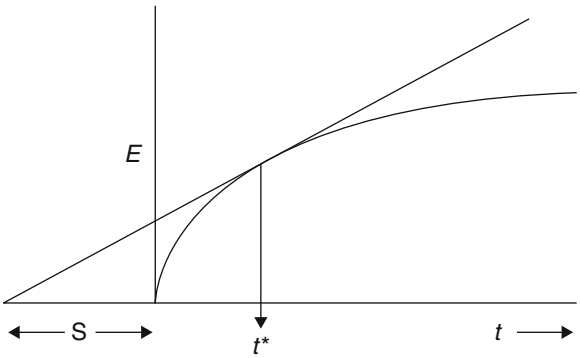
(a) *How does selection proceed?* With fitness as a variable, one can build a model describing the dynamics of a population of alleles with specific genetic structure (epistasis, pleiotropy, dominance, etc.) and fitness values. This is what population genetics does.

(b) *Why is/was there selection?* One can investigate the causes of the fitness values, which implies considering the physiology of individuals and the ecology of the population. Behavioral ecology, asking about the ▶ function of various traits – that is, the effect which led them to spread into the population and allows for its maintenance – exemplifies such research program, as well as paleontology.

In (a), population genetics explicates the process of natural selection and shows which combinations of fitness values, size, and mutation rates will plausibly lead a trait to fixation. In this context, an evolutionary explanation of the frequency of traits in a population would start from equilibrium considerations: the ▶ Hardy-Weinberg law for two alleles with initial frequencies p and q in a sexual panmictic population defines the equilibrium frequencies that each allele would have with no selection in a panmictic infinite isolated population. This is a null hypothesis for the system, like the inertia principle in mechanics (Sober 1984). Then, one compares the actual population to the model; if there is deviation from this equilibrium, one will hypothesize drift and selection as causes. Factoring in the fitness values as they are measured in nature in the equations, will give predictions about the frequencies; differences with such frequencies in the actual population are therefore caused by drift (Gillespie 2004). In several cases, an explanation can prove that a trait is here because of selection, while not being able to show what selection was for – because the causes of fitness or selection are unknown, ecological context being insufficiently understood.

With question (b) one is interested in knowing why a trait is there, whether it is by selection and for which selective pressures. The pervasiveness of altruism (i.e., behaviors increasing other's fitness at the cost of one's fitness), whereas selfishness by definition seems privileged by natural selection, compelled biologists to consider selection at the level of genes, where altruism toward kin appear evolutionarily favored since the ▶ relatedness of the beneficiary of the act mitigates the cost of altruism. Thus "kin selection" (West et al. 2007), together with ▶ sexual selection, is another form of natural selection to which evolutionary explanations must appeal.

Several kinds of explanations are therefore possible, which need not take the genetic makeup into account, and which involve basically the idea that selection, being the over-reproduction of the fittest, optimized the traits. Comparisons between species are also used in addition, or independently.



Explanation, Evolutionary, Fig. 1 Optimality model for time t spent in patches when foraging (E is calorific intake, s is the mean time spent to move between patches); t^* is optimal because of the law of diminishing returns, which makes increasing t less and less beneficial (After Charnov 1976)

Optimality Modeling

If one can show how a trait makes the highest contribution to fitness among all the actual or possible variants, then it is highly plausible that it results from natural selection. Often, one picks up a proxy for fitness, such as energy intake, chances of survival at some age, etc., not necessarily correlated to fitness in a linear way. An evolutionary explanation would therefore model one of these values as a function of the trait, given some parameters representing the selective pressures in the environment (Fig. 1): for example, the scarcity of resources, the chances of meeting a predator, and so on. If the extant trait values in various environments are the maxima of the function, this is evidence for the trait being selected for meeting those environmental demands. The optimal value for one selective pressure may not be the same as the optimum for another; hence the result of natural selection should be a trade-off between those demands, for example, foraging and looking for mates. As a methodology, this ▶ adaptationism is extensively used in behavioral ecology. If reality does not match an optimality model, one can say either that the selective pressures have not been correctly identified, or that genetic structure, amount of genetic variation, drift, or other factors affect the evolutionary process, so that the models only show what the trait would be if selection were acting alone.

Game-Theoretic Modeling

Often, the fitness of traits depends on the frequency of traits-carriers in a population: for example, a camouflage is all the more efficient when it is not so frequent. Frequency dependence is pervasive when it comes to social interactions, where the payoffs of the possible traits depend upon what the others will do, so that the evolutionary dynamics at one moment in turn depends upon the chances of meeting an interactor of a given type (e.g., altruist, selfish, aggressive, etc.). Evolutionary explanation of social behavior therefore relies on Evolutionary Stable Strategy models (Maynard-Smith 1982), an ESS being a strategy such that if all the population adopts it, it is not vulnerable to invasions. Often, ESS is a mixed strategy (e.g., cooperate at a chance 0.7, defect at a chance 0.3). The distribution of traits in a population matching what an ESS model would predict is clear evidence that the maintenance of those traits results from natural selection. Often, ESS models are devised in the frameworks of ▶ replicator dynamics. Recently, "adaptive dynamics" (Metz 2008) has provided a specific way of modeling evolutionary dynamics on the basis of "invasion fitness," that is, the probability of each mutant (individual or species) to invade a population, therefore synthesizing population genetics, behavioral, and community ecology.

Those explanations concern the *maintenance* of traits, rather than their *origin*; ESS models are built on the assumption of a strategy set but the evolutionary origins of strategies do not matter. Explaining origins of traits is another evolutionary question, which one undertakes, for example, in paleontology, where the environment of the organisms is not known. In such context, an available method is *reverse engineering*: given the traits and their mechanisms, reconstruct the environmental demands to which they may have responded.

Comparisons

Comparing a trait in an environment with one in a sister or ancestor species, and whose origin is already known, allows one to check whether it is simply inherited, or whether it is an adaptation for analogous selective pressures (Endler 1986). Those comparisons may justify a claim that some trait results from selection, but also supplement other evolutionary models, for example, in order to exclude rival nonselective hypotheses. Comparative methods are therefore pervasive.

Experimental tests may be used, for instance, changing the trait value enables one to check whether selective pressures exist which will reestablish the original value of the trait – a strong argument for its being maintained by selection (Williams 1966). A complete explanation of a trait therefore relies on ecological knowledge (of the environment), phylogenetic knowledge (for comparisons), and genetics (the underlying genetic makeup, and the heritability of traits) (Brandon 1990).

Formal Justifications

Evolutionary biology is a historical science – hence acknowledging an important role for contingency – but heavily relies on mathematical modeling (e.g., ▶ Hamilton's rule for the evolution of altruism). Adaptive dynamics provide for those models a "canonical equation" of change of fitnesses, yet the equation is often not solvable (Metz 2008). Many equations of the population genetics models can be derived from the ▶ Price equation. Other models of natural selection are stated in terms of ▶ replicator dynamics (especially with ESS). Many models are indeed simulated rather than analytically solved.

Optimization as an assumption is supposedly derived from Fisher's fundamental theorem of natural selection, which states that the mean fitness of a population increases equally to the genetic additive variance (by nature always positive). Fisher's result has been discussed and is indeed very restricted, concerning only the evolutionary change in fitness due to the direct action of natural selection (Frank and Slatkin 1992). It can be derived from the Price equation, when the value z considered is the fitness itself. Optimality assumptions of behavioral ecologists seem at odds with results of population genetics emphasizing the large restrictions on optimization (especially, in cases of frequency-dependence selection, e.g., social interactions). However, there is now (Grafen 2007) a formal proof of an isomorphism between population genetics models of natural selection dynamics and optimization, which provides a mathematical basis to the implicit pervasive intuition that natural selection is an optimizing process. The maximand of such a process is inclusive ▶ fitness. Therefore, optimality modeling and ESS receive a formal justification from the mathematics of allele frequencies.

Cross-References

- ▶ Adaptation
- ▶ Fitness
- ▶ Function
- ▶ Price Equation
- ▶ Relatedness
- ▶ Replicator Dynamics
- ▶ Sexual Selection

References

Brandon R (1990) Adaptation and environment. MIT Press, Cambridge
Charnov EL (1976) Optimal foraging: the marginal value theorem. Theor Popul Biol 9:129–136
Endler J (1986) Natural selection in the wild. Princeton University Press, Princeton
Frank S, Slatkin M (1992) Fisher's fundamental theorem of natural selection. TREE 7(3):92–95
Gillespie J (2004) Population genetics – a concise guide. The John Hopkins University Press, Baltimore
Grafen A (2007) The formal Darwinism project: a mid-term report. J Evol Biol 20:1243–1254
Maynard-Smith J (1982) Evolution and the theory of games. Cambridge University Press, Cambridge
Metz JAJ (2008) Fitness. In: Jørgensen S, Fath B (eds) Evolutionary ecology. Encyclopedia of ecology II. Elsevier, Oxford, pp 1599–1612
Ridley M (2001) Evolution. Oxford University Press, Oxford
Sober E (1984) The nature of selection. MIT Press, Cambridge
West S, Griffin A, Gardner A (2007) Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. J Evol Biol 20:415–432
Williams GC (1966) Adaptation and natural selection. Princeton University Press, Princeton

Explanation, Functional

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

Synonyms

Design explanation; Ecological explanation; Viability explanation

Definition

Functional explanations explain why certain organisms have certain traits rather than some conceivable alternatives by appealing to the advantages for those organisms of having those traits rather than the alternatives.

Characteristics

The term "functional explanation" is sometimes used in a very wide sense, meaning any explanation in functional biology or even any explanation that refers to functions in any sense of "function" (▶ Function, Biological). This essay discusses functional explanations in a narrower sense, namely, reasonings that purport to explain why certain organisms have a certain trait by elucidating why that trait is advantageous to or needed by those organisms. An example is the well-known explanation of why many organisms have a circulatory system by Nobel Prize winner August Krogh in 1919 (see Krogh 1941). Applying Fick's law of diffusion, Krogh calculated that diffusion (passive transport) is too slow to provide the inner organs with oxygen at the rate needed to stay alive if the distance between the inner cells and the outside is more than 0.5 mm. The presence of a circulatory system solves this problem by providing an additional and faster means of transport. To avoid terminological confusion, philosopher Arno Wouters introduced the terms "viability explanation" (Wouters 1995) and "design explanation" (e.g., Wouters 2007) to refer to functional explanations in this sense. Behavioral biologists and evolutionary biologists often call them "ecological explanations."

Functional explanations are part and parcel of biology, but they are not considered legitimate in other natural sciences such as physics and chemistry. Reductionist and structuralist schools in biology tend to reject functional approaches in biology because they would rest on illegitimate teleological assumptions. Another worry concerns the possibility to provide evidence for counterfactual claims concerning what would happen if an organism lacked the trait to be explained.

Functional explanations are concerned with the way in which the different traits of a living system (organism) functionally depend on each other. Functional explanations point out that because an organism has certain traits (in the example of Krogh: a certain size and a certain level of activity), its ability to maintain

the living state would diminish if the trait to be explained (the presence of a circulatory system) was replaced by a specific alternative (no active transport). The explaining traits are functionally dependent on the trait to be explained in the sense that the ability to maintain the living state of an organism with the explaining traits would diminish if the trait to be explained was replaced by the alternative, whereas replacing the trait to be explained would not make much difference if the organism lacked the explaining traits. For example, having a certain size and activity is functionally dependent on the presence of active transport because removal of active transport would diminish the life chances of organisms with that size and activity, whereas the replacement would have little effect on that organism's capacity to maintain itself if it were small enough and not very active.

Functional dependence is a synchronic relation at the level of the individual (the size and activity of an organism are dependent on *that* organism having a system of active transport). The relation is not of a causal nature (see ▶ Causality), but rather a ▶ constraint on what can be alive: our universe is such that living systems of a certain size and activity cannot exist without mechanisms for active transport. There may, of course, exist causal relations between the trait to be explained and its dependents (perhaps the transport system was maintained in the lineage because variants with a less well-developed transport system were less active and for that reason eliminated by natural selection or perhaps the activity of the developing organism influences the development of a transport system in the course of the ontogeny), but the functional dependence exists independent of the history of the organism and hence independent of these causal relations (it would exist also if the same organism arose out of different ontogenetic and evolutionary processes).

Functional explanations not merely identify the traits that are functionally dependent on the trait to be explained, but they also account for the dependency itself. This is done by relating the dependency to invariant relations that result from the way the organism is wired and from more general invariant relations that scientists call "laws." Such an explanation often introduces a structure of functional dependencies intermediate between the trait to be explained and the dependent traits. For example, the structure of functional dependencies in Krogh's explanation is as follows: (a) the size and the rate of oxygen delivery functionally depend on active transport (follows from Fick's law of diffusion, a well-known law of physical chemistry), (b) the rate of oxygen consumption functionally depends on the rate of oxygen delivery (follows from the obvious invariance that the rate of consumption cannot be higher than the rate of delivery), and (c) the rate of activity of the organism functionally depends on the rate of oxygen consumption (follows from a well-established invariant relation (very roughly: the more active the organism, the more oxygen it consumes) that results from the way in which those organisms generate the energy needed for their activities).

Several types of evidence support or undermine claims about the existence of a relation of functional dependence. *Comparisons* provide an important clue. It is necessary that all organisms with the dependent traits have the trait to be explained or a functional equivalent, but the support provided by comparisons for the conclusion that a certain trait is dependent on others is rather weak. *Experimental manipulation* provides better evidence. By producing organisms that are in relevant aspects similar to the real organism, except that the trait to be explained is replaced by an alternative, one gets good evidence for the advantageousness of that trait. In order to provide evidence for or against the thesis that the need is due to the presence of the dependent traits, the dependent traits should be modified too. Techniques of genetic manipulation have greatly increased the possibilities to provide this kind of evidence. The *explanation* of a functional dependency on the basis of invariances is strong evidence for the existence of that dependency, provided that both the invariances and the assumptions on which the explanation rests are well established. Finally, the development of *simulation models* in systems biology offers a new and powerful way to provide evidence for or against functional claims because they allow for precise modifications of large numbers of relevant and potentially relevant variables *in silico*.

We can now see that functional explanations do not assume ▶ teleology. Functional explanations are concerned with what is needed or advantageous for staying alive, but they do not tell us how the required traits are brought about and, in fact, make no assumptions at all about how or why living systems and their traits come into being. For that very reason, they do not assume that traits are brought about because of their advantages, in order to fit the requirements or because something or someone had a certain goal.

Functional explanations are, nevertheless, crucial to understand life because they show us how the characteristics of an organism fit into the requirements for being alive. Living systems exist far from thermodynamic equilibrium and can exist only because they are able to maintain themselves actively (i.e., by using energy). Although there are many ways to stay alive, not all combinations of matter will do. The ability of a system to maintain the living state is critically dependent on its organization, that is, on the composition, character, and arrangement of its parts and the order and timing of its activities. Causal explanations can tell us how a certain form of organization came into being and how that form of organization brings about the ability to stay alive, but not why certain forms of organization are able to stay alive and others not nor why certain forms of organization are better in staying alive than others. This is the kind of understanding provided by functional explanations, and this is why a combination of functional and causal explanations is needed to understand life (see ▶ Explanation in Biology).

The failure to understand the noncausal nature of functional explanations and the failure to understand how noncausal explanations can contribute to scientific understanding have instigated or strengthened many misconceptions. If it is, erroneously, assumed that all explanations purport to explain how the phenomenon to be explained was brought about, one might easily, but erroneously, take functional explanations as resting on the teleological assumption that traits of organisms are brought about because they perform a certain function. In response to this misconception, one might reject functional explanations, erroneously, as illegitimately teleological, as both structuralist and reductionist biologists tend to do, or look for a place for teleology within the Darwinian theory of evolution, as many naturalistic philosophers tend to do. The idea that explanations should explain how the trait to be explained is brought about might also lie behind the tendency to present the conclusion of functional explanations in evolutionary terms. Several books advertise a functional approach as an evolutionary one, and an increasing number of articles present functional explanations as showing that a certain trait "evolved as an adaptation for" or "was selected for" some advantage, need, or dependent trait. This is unfortunate not only because of the teleological odor of this kind of conclusion but also (and more importantly) because it suggests much more than the evidence allows (claims about selection require evidence about the variants that occur in the population, the occurrence of selection, the heritability of the trait, the structure of the population, and phylogenetic polarity, in addition to evidence about the advantageousness of the trait) (see ▶ Explanation, Evolutionary). It is also unfortunate that this way of presenting functional explanations has led some critics to reject functional explanations as mere speculation, thus ignoring the valid insights provided by functional explanations together with the unsubstantiated evolutionary conclusions drawn from them.

Cross-References

- ▶ Causality
- ▶ Constraint
- ▶ Explanation in Biology
- ▶ Explanation, Evolutionary
- ▶ Function, Biological
- ▶ Teleology

References

Krogh A (1941) The comparative physiology of respiratory mechanisms. University of Pennsylvania Press, Philadelphia

Wouters AG (1995) Viability explanation. Biol Philos 10(4):435–457

Wouters AG (2007) Design explanation: determining the constraints on what can be alive. Erkenntnis 67(1):65–80

Explanation, Reductive

Marie I. Kaiser
Department of Philosophy, University of Cologne, Cologne, Germany

Synonyms

Explanatory reduction

Definition

Reductive explanations (or explanatory ▶ reduction) are a special kind of scientific ▶ explanation. It has

been argued that three central features distinguish reductive explanations in biology from non-reductive ones (Kaiser 2011; for other analyses see Sarkar 1998; Hüttemann and Love 2011). The first and third are necessary conditions for a biological explanation to be reductive, whereas the second is only a feature that is typical for most (but not for all) reductive explanations in biology.

In reductive explanations, the behavior of a biological system S is explained

1. by referring only to factors that are located on a *lower level* of organization than S,
2. by focusing on factors that are *internal* to S (i.e., that are genuine parts of S),
3. by describing parts of S only as being *parts in isolation* (i.e., by appealing only to those relational properties of and interactions between parts that can be studied in other contexts than in situ).

For instance, the contraction of a muscle fiber is explained by appealing to certain molecules (e.g., actin, myosin, calcium ions), cell organelles (e.g., sarcoplasmic reticulum), and by describing the interactions between them. This explanation possesses a reductive character because all factors mentioned in the explanation are internal to the muscle fiber (condition 2). External factors like the incoming neuronal signal are, if mentioned at all, simplified as being mere input- or background conditions. Furthermore, all explanatorily relevant factors (i.e., actin and myosin molecules, calcium ions, sarcoplasmic reticulum) are located on a lower level of organization than the muscle fiber itself (condition 1). Finally, only those interactions between the muscle fiber's parts are represented in the explanation that can be investigated by taking the parts out of the original system (i.e., the muscle fiber) and studying their properties in other contexts than *in situ* (condition 3).

Cross-References

▶ Explanation in Biology
▶ Reduction

References

Hüttemann A, Love AC (2011) Aspects of reductive explanation in biological science: intrinsicality, fundamentality, and temporality. Br J Philos Sci 62:519–549
Kaiser MI (2011) Limits of reductionism in the life sciences. Hist Philos Life Sci 33:389–396
Sarkar S (1998) Genetics and reductionism. Cambridge University Press, Cambridge

Explanatory Reduction

▶ Explanation, Reductive

Exploratory Experimentation

Richard M. Burian
Department of Philosophy, Virginia Tech, Blacksburg, VA, USA

Definition

Experiments count as exploratory when the concepts or categories in terms of which results should be understood are not obvious, the experimental methods and instruments for answering the questions are uncertain, or it is necessary first to establish relevant factual correlations in order to characterize the phenomena of a domain and the regularities that require (perhaps causal) explanation. Elliot (2007, 322 ff.) offers a useful taxonomy of exploratory experimentation. Rather than testing hypotheses; it varies parameters or circumstances to see what will happen; it utilizes background knowledge (including theories and experimental lore) to establish novel correlations, follow anomalies, seek improvements in instrumentation and experimental protocols, and the like; and it employs a variety of systematic strategies to govern appropriate variation of parameters and appropriate orientation to the primary questions in the background. Thus, some of the investigations well suited for exploratory experimentation

- Aim at developing new experimental technologies for use in data-intensive research
- Seek to track causal interactions that cross hierarchical levels
- Deal with cases in which the state space of relevant variables is not well delimited

- Aim at developing new concepts for "fixing phenomena" calling for explanation by establishing relevant regularities or classifying relevant entities (e.g., regulatory RNAs, compounds that act as endocrine disruptors, and protein networks) that have not yet been well characterized
- Seek to resolving anomalies, especially those arising across disciplinary boundaries in the handling of complex systems

Exploratory experimentation is often a key component in investigations of the importance of potentially relevant variables, parameters, and sources of error. It also facilitates the discovery of unrecognized patterns and regularities (▶ Pattern Mining) and the design of instrumentation for investigating open questions and for analyzing complex phenomena, particularly when new or revised concepts and classifications are needed to help explain novel phenomena. It may help to establish system boundaries and resolve anomalies that have turned up in previous investigations (Elliott 2007).

Characteristics

Philosophical Background

Until about 1980, mainstream philosophers of science focused mainly on *hypotheses*, *theories*, and their *justification* rather than *discovery* and *hypothesis generation*, holding that the latter can neither be analyzed philosophically nor made subject to philosophical norms. They paid far more attention to the hypothetico-deductive method as opposed to inductive methods for inference from data to hypotheses: if justification depends on deduction from true or justified premises, inductive inferences (which are deductively invalid!) are suspect and, arguably, even illegitimate (▶ Deduction; ▶ Induction). This centrality of theory and of hypothetico-deductive methodology is nicely exemplified in the philosophy of Sir Karl Popper, who held that science is composed of conjectures from which testable hypotheses are deduced with the aid of further hypotheses (e.g., about boundary conditions, experimental setups, and the behavior of experimental instruments):

> The theoretician puts certain definite questions to the experimenter, and the latter, by his experiments tries to elicit a decisive answer to these questions and to no others. . . (Popper 1968)

Of course, studies of experimental methods and practices always recognized that experiments may serve other purposes – e.g., to help design or calibrate instruments, standardize reagents, refine key measurements, establish correlations between changes in experimentally controlled parameters and dependent variables (▶ Experiment). Ian Hacking (1983), helped spark a philosophical reexamination of experimental practice in its own right and reinforced recognition within philosophy of science that experiment "has a life of its own" and is not simply a handmaiden of theory. Philosophers supporting the "new experimentalism" generally agree that experimental protocols can be freed from pernicious dependence on the hypotheses or theories under investigation (Arabatzis 2008; Hacking 1992; Weber 2004) and that experiments serving exploratory and inductive purposes serve different purposes and should meet different evaluative criteria than those designed for hypothesis testing.

Conditions that Exploratory Experiments Should Meet

Several major steps are required for success in exploratory experiments. These include:

- Establishing conditions in which experimental findings are stable (Hacking 1992) and/or system behavior robust to various perturbations (Ferrell 2009)
- Establishing which downstream effects follow on small changes in the location, timing, or order of events
- Iteratively adjusting experimental conditions and technologies to reduce noise in experimental findings
- Rethinking the identities of the "players" (boundary conditions, entities, mechanisms, parameters, processes, and unknowns) relevant to the findings
- Iteratively deploying revised or new simulations and computational and theoretical models of the experimental systems and situations

These steps are intensively experimental. Satisfactory results are dependent on

- Careful tuning of experimental techniques and skills
- Integration of appropriate calculational tools and theoretical commitments
- Reworking of the experimental conditions, protocols, instrumentation, and analytical categories in terms of which the experiments are described

- Iterative reexamination of the scope of tentative findings and the applicability of the techniques and concepts to additional research problems and systems
- Interactive stabilization of the protocols, descriptive terminology, and experimental findings, with iterative procedures for refining the experimental apparatus, the calculational tools, and of the analyses of patterns in the phenomena and causal pathways at issue (Kelder et al. 2010)

Exploratory Experimentation in Systems Biology

Because systems biology is concerned with establishing the strength and scope of interactions across many scales in complex systems that often have partially specified architectures and boundaries, exploratory experimentation typically serves different purposes and employs different modalities than conventional hypothesis testing (e.g., high-throughput technologies and data-intensive research with high levels of noise) (Franklin 2005; O'Malley 2007). Recent debates about the structure and value of "data-driven research" and "discovery science" are therefore relevant (▶ Data-Intensive Research; ▶ Data Mining; ▶ De Novo Computational Discovery of Motifs; ▶ Functional/Signature Network Module for Target Pathway/Gene Discovery; ▶ Rule Discovery). Furthermore, much research in systems biology is comparative and classificatory, e.g., comparing the distinctive functions of genes or proteins in different systems, different developmental stages of a developing system, or the networks and pathways of different systems. Such comparisons often require use of data from multiple databases and disciplines. Ensuring robustness of findings and comparability of data from diverse sources is often problematic, fraught with the difficulties of developing or meeting the constraints of curated databases and/or standardized "ontologies" such as the Gene Ontology (▶ Bio-Ontologies; ▶ Disease Ontology; ▶ Gene Ontology; ▶ Ontology Analysis of Biological Networks; ▶ Ontology Structure).

Again, the relationship between high-throughput exploratory experiments and well-controlled hypothesis-testing experiments is crucial to evaluation of systems biological experiments. Recently, protocols have been developed that combine the two with means of calibrating findings (e.g., Pfeifer et al. 2010); the resulting mutual reinforcement is potentially very powerful. Often (as illustrated by Pfeifer et al.), data from several sources need to be reconfigured to make the data comparable and to cope with the high dimensionality of the interrelationships involved. Such work often requires innovative statistical and computational tools as well as significant adjustment of terminologies and conceptual apparatus from different disciplines and organisms. A common issue in analyzing pathway and networks is, therefore, managing and reducing the semantic difficulties arising from differing disciplinary commitments.

An important type of investigation drawing on exploratory experimentation is the interactive revision of classifications and analyses to discover empirically meaningful patterns that can facilitate the production of coherent and reliably reproducible results, transportable from one problem or database to another. Thus the recognition of novel patterns exhibited in data and/or turned up while developing novel explanatory concepts is a major purpose of exploratory experimentation. Pattern recognition of this sort encounters many obstacles in the interdisciplinary approaches to complex problems now prevalent in systems biology and beyond. An extreme example illustrates the difficulty: consider recent debates about the (partly anthropogenic) causes and effects of climate change in relation to ecological change and the spread of various pathological organisms. The resolution of the issues about which patterns are salient and involve anthropogenic causes must reconcile perspectives from biogeography, biological systematics, ecology, economics, epidemiology, genomics, meteorology, molecular biology, paleoclimatology, pathology, and many other disciplines. It is hardly a surprise that the integration and interpretation of data to attempt a systematic approach to some of the issue involved is, by itself, an enormously difficult problem. Problems of this sort, though not always as dramatic, arise often enough when exploratory experimentation is employed in systems biology.

Conclusion

The most important points of this entry are the distinctiveness of exploratory experimentation and the inadequacy of the characterizations it has received in standard accounts of experimental methods. Philosophers and methodologists are only beginning to develop models of exploratory experimentation. There has been little appreciation of the relatively indirect role of hypothesis testing in exploratory experimentation. For example, standard reviews of experiment in systems biology (e.g., Kreutz and Timmer 2009) retain strong rhetoric about the need to specify exact hypotheses even when they concentrate mainly on issues described here as

pertinent to exploratory experiments. Unlike hypothesis testing, exploratory experiments do not begin with a null model and are not able to specify in detail the expected outcome of key experiments. They must focus, instead, on reducing noise in the data, on how best to perturb systems to uncover relevant parameters or the triggers of threshold transitions, and on ensuring robustness of findings. Exploratory experiments do, however, have a strong basis in (theoretical) background knowledge of relevant mechanisms and comparative knowledge of related systems. One analytical approach of great interest compares the ways in which exploratory experiments utilize complex data sets and revision of concepts and classifications with the methods utilized by natural history in dealing with complex data, revision of concepts, and revision of classifications (Strasser 2010). The point is partly that the highly experimental work discussed above seeks to understanding of complex interactions in natural conditions among "natural assemblages" of molecules, genomes, organisms, and a welter of interacting mechanisms, "players" and processes – and this is quite like what is done in natural history except that it is in carried out in highly experimental contexts.

Natural historical research is often characterized as "descriptive." Fair enough. But it should be recognized that descriptive findings are equally essential to intensively experimental laboratory research required to characterize the systems and the processes of interest in systems biology. This is why broadly inductive uses of exploratory experimentation and natural-historical methods employing high-throughput technologies are so central in systems biology. And it is why iterative interaction between the methods of exploratory experimentation and those of hypothesis testing is crucial to the ongoing development of systems biology (Kelder et al. 2010).

Cross-References

- ▶ Bio-Ontologies
- ▶ Data Mining
- ▶ Data-Intensive Research
- ▶ De Novo Computational Discovery of Motifs
- ▶ Deduction
- ▶ Disease Ontology
- ▶ Experiment
- ▶ Functional/Signature Network Module for Target Pathway/Gene Discovery
- ▶ Gene Ontology
- ▶ Induction
- ▶ Ontology Analysis of Biological Networks
- ▶ Ontology Structure
- ▶ Pattern Mining
- ▶ Rule Discovery

References

Arabatzis T (2008) Experiment. In: Psillos S, Curd M (eds) The Routledge companion to the philosophy of science. Routledge, London/New York, pp 159–170

Elliott KC (2007) Varieties of exploratory experimentation in nanotoxicology. Hist Philos Life Sci 29(3):313–336

Ferrell JE Jr (2009) What is systems biology? J Biol 8:2

Franklin LR (2005) Exploratory experiments. Philos Sci 72:888–899

Hacking I (1983) Representing and intervening: introductory topics in the philosophy of natural science. Cambridge University Press, Cambridge

Hacking I (1992) The self-vindication of the laboratory sciences. In: Pickering A (ed) Science as practice and culture. University of Chicago Press, Chicago, pp 29–64

Kelder T, Conklin BR, Evelo CT, Pico AR (2010) Finding the right questions: exploratory pathway analysis to enhance biological discovery in large datasets. PLoS Biol 8(8):e1000472

Kreutz C, Timmer J (2009) Systems biology: experimental design. FEBS J 276:923–942

O'Malley MA (2007) Exploratory experimentation and scientific practice: metagenomics and the proteorhodopsin case. Hist Philos Life Sci 29(3):337–360

Pfeifer AC, Kaschek D, Bachmann J, Klingmüller U, Timmer J (2010) Model-based extension of high-throughput to high-content data. BMC Syst Biol 4:106

Popper KR (1968) The logic of scientific discovery. Harper and Row, New York

Strasser BJ (2010) Laboratories, museums, and the comparative perspective: Alan A. Boyden's serological taxonomy, 1925–1962. Hist Stud Nat Sci 40(2):149–182

Weber M (2004) Philosophy of experimental biology. Cambridge University Press, Cambridge and New York

Exposure

Catherine M. Lloyd
Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand

Definition

An exposure in the CellML model repository refers to the published view of a specific revision of a CellML

model. It can be regarded as the "display page" of that particular version of a model. It is important to note that a model file can exist in a workspace without being exposed. That model is still accessible if it has been made publicly available; in such a case, it has to be downloaded directly from the workspace.

Cross-References

▶ CellML Model Repository

Expression Profiling

▶ Cell Cycle Analysis, Expression Profiling

Expression Signature

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

Extended Gaussian Image

Virginio Cantoni, Alessandro Gaggia and Luca Lombardi
Department of Computer Engineering and Systems Science, University of Pavia, Pavia, Italy

Synonyms

Orientation histogram

Definition

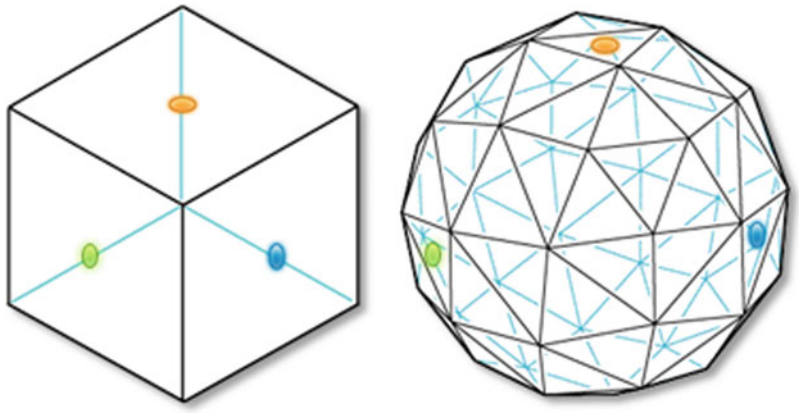
A Gaussian Image is the mapping of all the normals of an object on a sphere of unitary radius (Gaussian Image): all the tails of the vectors are on the center of the sphere while the heads lie on the surface. The Extended Gaussian Images (EGI) (Horn 1984) include the associate area: each point on the sphere has a mass proportional to the area (Fig. 1). The EGI is exploited for characterizing the objects shape. Herman Minkowski in 1910 demonstrated that a convex polyhedron is fully described by the area and orientation of its faces, that is, if two different objects are convex, their EGI representations are necessarily different.

The EGI can be easily built from needle or depth maps generated by range or stereo devices and, most importantly, they can also be determined from the 3D model of an object.

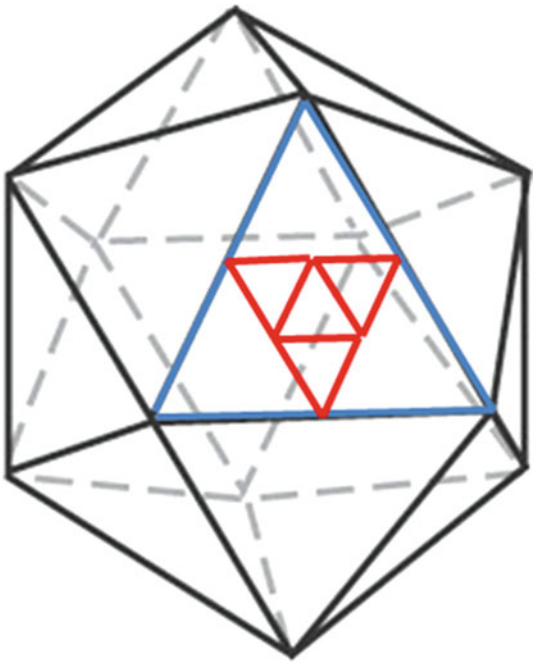
The EGIs can be used in machine vision also for determining the attitude in space of an object. The EGI is particularly suited to deal with the varying attitude of a 3D object in space: it is invariant to the object position and, in case of convex objects, there is a bijective correspondence with their EGI.

The basic idea is that each surface patch is mapped to a weighted point on the Gaussian sphere according to its surface normal and the value assigned to each orientation is the sum of area of all the surface patches having a common normal (Eq. 1, Fig. 1).

$$W_{\vec{d}} = \sum_{l=1}^{N_{\vec{d}}} A_{l,\vec{d}} \tag{1}$$



Extended Gaussian Image, Fig. 1 Extended Gaussian image



Extended Gaussian Image, Fig. 2 EGI with a triangular tessellation

where $\vec{d}$ is the direction associated with a point on the Gaussian sphere, $N\vec{d}$ the total number of surface patches with normal $\vec{d}$, and $A_{l,\vec{d}}$ the area of the l_{th} surface patch with normal $\vec{d}$.

For a useful computer representation the Gaussian sphere must be discretized, usually adopting a triangular tessellation, and this is called geodesic dome. Starting with a regular polyhedron (e.g., the Icosahedron), for a more detailed description each triangle is split (iteratively) into four smaller triangles (Fig. 2). This representation is usually called orientation histogram.

Cross-References

▶ Extended Gaussian Image for Pocket-Ligand Matching

References

Hermann M (1910) Geometrie der Zahlen, Leipzig and Berlin: R. G. Teubner, MR0249269

Horn BKP (1984) Extended Gaussian images. Proc IEEE 72:1671–1686

Extended Gaussian Image for Pocket-Ligand Matching

Virginio Cantoni[1,2], Alessandro Gaggia[1] and Luca Lombardi[1]
[1]Department of Computer Engineering and Systems Science, University of Pavia, Pavia, Italy
[2]Computational Biology, KTH Royal Institute of Technology, Stockholm, Sweden

Synonyms

Orientations histogram

Definition

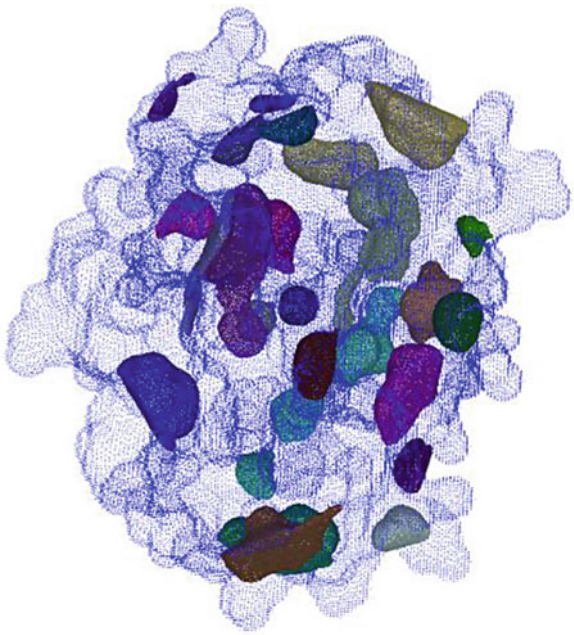
The ▶ Extended Gaussian Image (EGI) (Horn 1984) can be useful for a preliminary evaluation of pocket-ligand matching likelihood. This is supported by the effectiveness in representing object mainly convex (i.e., a pocket and a ligand). The analysis, which consists of a kind of necessary condition for matching, follows a sequence of three steps:

1. The solvent-excluded surface (SES) is analyzed and a number of pockets and tunnels sites are identified (Fig. 1).
2. The candidate binding sites are detected through a structural matching of pockets and ligand, both represented through a suitable EGI modality.
3. The loci of compatible positions of the ligand are identified through standard tools of ▶ mathematical morphology operators.

Characteristics

The aim, for docking applications, is the search for subregions that are complementary between different molecules. When we have a large molecule (receptor) and a small molecule (▶ Ligand), docking takes place in a protein cavity. In this connection the first subproblem to be solved, in protein-ligand interfaces, is to develop the representations and the data structures suitable to support the computational methods that allow a quantitative evaluation of the protein-ligand matching on the basis mainly of their 3D structure and

morphology. The EGI is a possible representation for the ligand molecule, with the correspondent data structure based on a first-order statistic of the surface orientations. After the segmentation of the protein SES (Cantoni et al. 2010), the interface regions, which potentially can be active sites, are represented by an EGI (Fig. 2).



Extended Gaussian Image for Pocket-Ligand Matching, Fig. 1 Pockets and tunnels extracted from a test protein

A given 3D molecule, modeled through its SES in a triangular mesh, is described by the set of triangles:

$$T = \{T_1, ..., T_m\}, T_l \subset R^3 \tag{1}$$

where each T_l consists of a set of three vertices (2):

$$T_l = \left\{P_{A,l}, P_{B,l}, P_{C,l}\right\} \tag{2}$$

Center, normal, and area of each triangle T_l, namely g_l, $\vec{d}_l$ and A_l, respectively, can be computed by (3), (4), and (5):

$$g_l = (P_{A,l} + P_{B,l} + P_{C,l})/3 \tag{3}$$

$$\vec{d}_l = (P_{C,l} - P_{A,l}) \times (P_{B,l} - P_{C,l}) \tag{4}$$

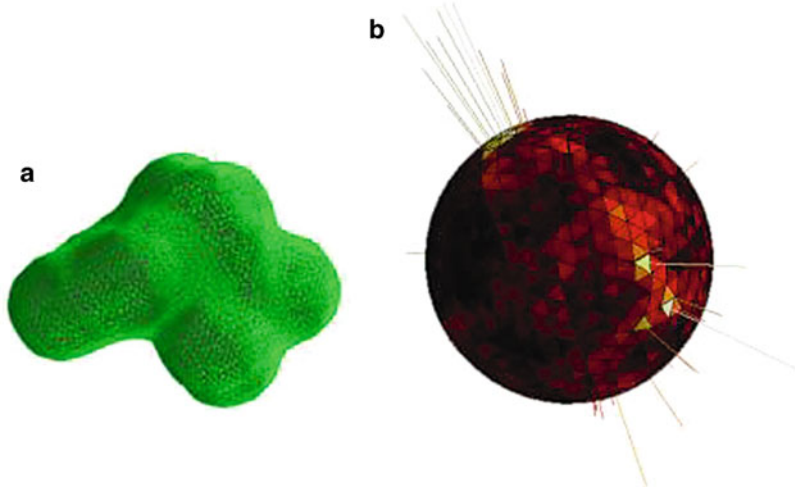
$$A_l = |(P_{C,l} - P_{A,l}) \times (P_{B,l} - P_{C,l})|/2 \tag{5}$$

The total area of the mesh A is given by cumulating the area of each single triangle (6):

$$A = \sum_{l=1}^{m} A_l \tag{6}$$

where the Gaussian sphere is partitioned into a number of cells m.

Then all the triangles T of the target molecule are mapped onto the corresponding cells on the basis of the discretized orientation $\vec{d}_l$. Dealing with pockets and external molecules that must match the cavity surface, the approximation that the small target molecule is mainly convex (at least for the segment involving the



Extended Gaussian Image for Pocket-Ligand Matching, Fig. 2 (**a**) Wireframe surface representation of a ligand. (**b**) The correspondent EGI representation

matching of the pocket concavity) is usually suitable. Nevertheless, if the concave parts of the target molecules become critical other approaches that reduce the known ambiguities of the EGI representation must be followed. In this connection new representations such as the complex EGI or the enriched complex EGI (▶ Complex EGI and Enriched Complex EGI) (Hu et al. 2010) even if do not eliminate the ambiguities certainly in practical cases results very effective. Here the ECEGI solution is considered. If the object is convex, the mass center of the three W_x, W_y, and W_z distributions on the Gaussian sphere coincides with the center of the sphere. In fact, this is true also for the EGI (and for the CEGI), as by (7):

$$\sum_d |W_{x,\vec{d}}|\vec{d} = \sum_d |W_{y,\vec{d}}|\vec{d} = \sum_d |W_{z,\vec{d}}|\vec{d} = \sum_d A_{\vec{d}}\vec{d} = 0 \quad (7)$$

Since for a convex object $|W_{x,d}| = -W_{y,d}| = -W_{z,d}| = A_d$, being A_d the area of the surface patch with normal $\vec{d}$. A tessellated sphere with uniform and isotropic subdivision is needed. These properties are obviously satisfied by the projection of a regular polyhedron onto the sphere. Adopting the highest order regular polyhedron, the icosahedron with 20 triangular cells as a basis (but it provides a too coarse sampling of the orientations), and proceeding further by dividing iteratively each triangular cells into four smaller triangles according to the well-known geodesic dome (Horn 1984) constructions, the required level of resolution can be achieved: being n the number of iterative subdivision steps, the cells number is $m = 10*2^{2n+1}$, and the area (solid angle) of the single cells is $\pi/10*2^{2n-1}$ respectively. The corresponding data structure is consequently a hierarchical one (in which each cell of one level contains, other than the specific orientation, the four pointers to cells of the subsequent level) and the searching strategy of the orientation histogram values becomes a hierarchical process.

Given two candidates molecules dual parts (i.e., a cavity and a ligand) the aim is to find if they are morphologically (or geometrically) compatible, in other words we look for the rigid motion that could bring the protrusion into the cavity. On this purpose, a preliminary coarse constraint is given by the mass of the ECEGI of the cavity and the ligand: $A_{cav} > A_{lig}$.

Some matching indices normally adopted in EGI applications are:

- The Minkowski distance:

$$M = \sqrt[p]{\sum_{l=1}^{m} |A_{l,cav} - A_{l,lig}|^p} \quad (8)$$

in for p = 1 and p = 2 we obtain the Manhattan and the Euclidean distances respectively

- The Bray Curtis distance:

$$B = \frac{\sum_{l=1}^{m} |A_{l,cav} - A_{l,lig}|}{\sum_{l=1}^{m} |A_{l,cav} + A_{l,lig}|} \quad (9)$$

obviously $0 \leq B \leq 1$

- The Hausdorff distance:

$$H = \max\left(\left\|\max_{\forall l} A_{l,cav} - \min_{\forall l} A_{l,lig}\right\|, \left\|\max_{\forall l} A_{l,lig} - \min_{\forall l} A_{l,cav}\right\|\right) \quad (10)$$

- Another distance that can be defined for the geodesic dome is $E=n/m$ where n is the number of triangles satisfying this threshold criteria: (11) and m is the total number of triangles of the considered mesh.

$$\left[\left(\frac{|A_{l,cav} - A_{l,lig}|}{\max(A_{l,cav}, A_{l,lig})} \geq \theta\right) \bigcup \left(\max(A_{l,cav}, A_{l,lig}) = 0\right)\right]_{l=1}^{m} \quad (11)$$

the distance is given by $E = n/m$, i.e., the percentage of the triangles satisfying the threshold criteria.

Cross-References

- ▶ Complex EGI and Enriched Complex EGI
- ▶ Extended Gaussian Image
- ▶ Ligand
- ▶ Mathematical Morphology Operators

References

Cantoni V, Gatti R, Lombardi L (2010) Segmentation of SES for protein structure analysis. In: 1st international conference on bioinformatics. BIOSTEC 2010. Valencia (ES), pp 83–89

Horn BKP (1984) Extended gaussian images. In: Proceedings of the IEEE. IEEE Computer Society Press, Los Alamitos, vol 72, pp 1671–1686

Hu Z, Chung R, Fung KSM (2010) EC-EGI: enriched complex EGI for 3D shape registration. In: Machine vision and applications. Springer Berlin/Heidelberg, vol 2, pp 177–188

Extensible Markup Language

▶ XML

Extracellular Matrix

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

Synonyms

Basal lamina; Basement membrane; ECM; Ground substance

Definition

Extracellular matrix is the complex protein and carbohydrate material in which cells of multicellular organisms are embedded.

Characteristics

The extracellular matrix (ECM) is distinct according to the developmental stage, organ, and tissue type (Hynes and Naba 2012). The core "matrisome" comprises about 300 proteins, whose structure and function are affected by ECM-modifying enzymes, ECM-binding growth factors, and other ECM-associated proteins. These different categories of ECM and ECM-associated proteins cooperate to assemble and remodel extracellular matrices and bind to cells through ECM receptors. Together with receptors for ECM-bound growth factors, they provide multiple inputs into cells to control survival, proliferation, differentiation, shape, polarity, and motility of cells. There are two major ECMs: The basal lamina or basement membrane is a highly structured ECM layer on which epithelia rest. The interstitial ECM is that between tissues generated by the stroma.

ECM is composed of proteins, glycosaminoglycans (GAGs), and carbohydrates. The range of proteins is large as are the modifications, variations, and binding factors, making the insoluble constituents of a tissue important in not only structural integrity but also critical for proper tissue function (Lukashev and Werb 1998). ECM degradation and composition is central to many degenerative disease processes and cancer.

Collagen is the most abundant protein in animals and can be grouped according to types that include fibrillar, facit, short chain, and reticular (Orgel et al. 2011). There are at least 30 collagens in humans (Table 1). The three most abundant collagens form fibers that are important in tissues structure. Type I collagen forms that self-assemble in the extracellular space into triple-helical cross-linked fibers up to 300 nm long. Type II collagen constitutes 90% of the structural protein of cartilage. Type III collagen is important during embryogenesis and in wound healing. In contrast, type IV collagen organizes in a reticular manner present in basement membranes. Other collagens not as abundant are important in determining tissue architecture and cell behaviors.

Non-collagen ECM proteins are highly glycosylated (Table 2). A notable exception is elastin, which, as its name implies, is important for providing elasticity and forms sheets and fibers by cross-linking on a scaffold of fibrillin microfilaments. These are present around blood vessels and airways in particular. Adhesion is another function of ECM. There are 15 known laminin, a major protein of the basement membrane that is essential for epithelial cell adhesion by way of integrins, integral membrane ECM receptors. Laminins are heterotrimeric, cross-like structures with 3 short and 1 long arms. In the interstitial ECM, fibronectin is the major adhesion protein. Fibronectins are dimers of 2 similar peptides chain that are 60–70 nm long and 2–3 nm thick. A single fibronectin gene gives rise to at least 20 different fibronectin chains by alternative RNA splicing. The ECM of developing or fetal organisms contains specific fibronectin isoforms that enable and stimulate motility.

Proteoglycans are made up of more carbohydrate than protein. The GAGs chondroitin sulfate (CS), dermatan sulfate (DS), keratan sulfate (KS), and heparan sulfate (HS) are linear polysaccharides composed of an amino sugar and an uronic acid. These basic components are further varied by epimerization, sulfation, and deacetylation. The order of the carbohydrate chain and the other chemical modifications determine their

Extracellular Matrix, Table 1 Collagen types

Collagen type	Gene symbol	Characteristics	Primary source
I	COL1A1, COL1A2	300 nm, 67 nm banded fibrils	Skin, tendon, bone, etc.
II	COL2A1	300 nm, small 67 nm fibrils	Cartilage, vitreous humor
III	COL3A1	300 nm, small 67 nm fibrils	Skin, muscle, frequently with type I
IV	COL4A1 thru COL4A6	390 nm C-term globular domain, nonfibrillar	All basal lamina
V	COL5A1, COL5A2, COL5A3	390 nm N-term globular domain, small fibers	Most interstitial tissue, assoc. with type I
VI	COL6A1, COL6A2, COL6A3	150 nm, N + C term, globular domains, microfibrils, 100 nm banded fibrils	Most interstitial tissue, assoc. with type I
VII	COL7A1	450 nm, dimer	Epithelia
VIII	COL8A1, COL8A2		Some endothelial cells
IX	COL9A1, COL9A2, COL9A3	200 nm, N-term globular domain, bound proteoglycan	Cartilage, assoc. with type II
X	COL10A1	150 nm, C-term globular domain	Hypertrophic and mineralizing cartilage
XI	COL11A1, COL11A2	300 nm, small fibers	Cartilage

Extracellular Matrix, Table 2 Non-collagen ECM proteins

Family	Proteoglycans	Core protein kDA	GAG type	Tissue
Small interstitial proteoglycans	Decorin	36 kDa	CS	Secreted, connective tissue
	Biglycan	38 kDa	CS	Secreted, connective tissue
Aggrecan family of matrix proteoglycans	Aggrecan	208–220 kDa	CS, HA, KSII	Secreted, cartilage
	Brevican	96 kDa	CS	Secreted, brain
	Neurocan	145 kDa	CS	Secreted, brain
	Versican	265 kDa	CS	Secreted, fibroblasts
HS proteoglycans	Periecan	400 kDa	HS	Basement membrane
	Agrin	212 kDa	HS	Basement membrane
	Syndecans 1–4	31–45 kDa	HS, CS	Epithelial cells, fibroblasts
	Betaglycan	110 kDa	HS	Fibroblasts
	Glypicans 1–5	~60 kDa	HS	Epithelial cells, fibroblasts
	Serglycin	10–19 kDa	CS	Mast cells
KS proteoglycans	Lumican	37 kDa	KS 1	Broad
	Keratocan	37 kDa	KS 1	Broad
	Fibromodulin	59 kDa	KS 1	Broad
	Mimecan	25 kDa	KS 1	Broad
	SV2	80 kDa	KS 1	Synaptic vesicles
Claustrin	105 kDa	nr	CNS	

specificity and functionality. Hyaluronan is non-sulfated and is composed of an unmodified disaccharide repeat that is not covalently linked to any protein (Toole 2004).

The basement membrane is a permeable, thin ECM, evident as a three-layered basal lamina when using electron microscopy, placed between epithelial cells or endothelial cells and adjacent connective tissue. It surrounds smooth and striated muscle cells, fat cells, Schwann's cells, cells of adrenal medulla, and reticular epithelial cells of the thymus and the surface of brain and spinal cord. Basement membranes usually contain type IV collagen, heparan sulfate proteoglycan, chondroitin sulfate proteoglycan, entactin, hyaluronic acid, collagen VII, and laminin. Each of the components of the basal lamina is synthesized by the cells that rest upon it and are modified by adjacent stromal cells beneath. Basement membrane is both structural and instructional, providing key positional

and behavioral information to the attached cells. Epithelial functions mediated by the basement membrane include polarity, morphogenesis, cell cycle regulation, and differentiation. Epithelial stem cells are thought to reside in a specialized ECM that forms the niche.

ECM composition is context and process dependent in a manner that propagates fundamental information. The dynamic ECM compositional changes during embryogenesis and rapid ECM remodeling during wound healing are essential for developing and restructuring tissues. An important distinction of fetal ECM is that scars do not form around skin wounds, which is due to microenvironmental cues and specific cytokines (Shah et al. 1992). Scars contain disordered collagen I fibrils. Abnormal ECM composition or remodeling contributes to pathogenesis in many organs. Overproduction of interstitial collagens is characteristic of fibrosis, which may ultimately compromise vital organ function. Fetal and ▶ wound healing ECM proteins are also often reexpressed in cancers. An example is migration-stimulating factor (MSF), a truncated fibronectin monomer. MSF message and protein are expressed by fibroblasts, keratinocytes, and vascular endothelial cells in fetal skin but not in the majority of healthy adult skin (Schor and Schor 2009). MSF message and protein are also expressed by tumor cells and associated stromal ▶ fibroblasts and endothelial cells. Thus, ECM is both a response to and a regulator of homeostasis and pathology.

Cross-References

▶ Cancer
▶ Fibroblasts
▶ Wound Healing

References

Hynes RO, Naba A (2012) Overview of the matrisome—an inventory of extracellular matrix constituents and functions. Cold Spring Harb Perspect Biol 4(1):doi:10.1101/cshperspect.a004903

Lukashev ME, Werb Z (1998) ECM signalling: orchestrating cell behaviour and misbehaviour. Trends Cell Biol 8:437–441

Orgel JPRO, San Antonio JD, Antipova O (2011) Molecular and structural mapping of collagen fibril interactions. Connect Tissue Res 52(1):2–17. doi:10.3109/03008207.2010.511353

Schor AM, Schor SL (2009) Angiogenesis and tumour progression: migration-stimulating factor as a novel target for clinical intervention. Eye 24(3):450–458

Shah M, Foreman DM, Ferguson MWJ (1992) Control of scarring in adult wounds by neutralizing antibody to transforming growth factor b. Lancet 339:213–214

Toole BP (2004) Hyaluronan: from extracellular glue to pericellular cue. Nat Rev Cancer 4(7):528–539

Extrinsic DNA Damage Checkpoints

▶ Cell Cycle Arrest After DNA Damage