**Bioinformatics in drug discovery & Development
not being updated
Mary Chitty** mchitty@healthtech.com **781 972 5416**
**Overviews  & introductions** [**Bioinformatics**](http://www.genomicglossaries.com/content/chapterinfosourcestext.asp#bioinfo)[**cheminformatics**](http://www.genomicglossaries.com/content/chapterchemogenomics.asp#chem info jrnls)[**Molecular Medicine informatics**](http://www.genomicglossaries.com/content/chapterinfosourcestext.asp#med info)

**Information resources** [**Bioinformatics**](http://www.genomicglossaries.com/content/chapterinfosourcestext.asp)   [**Cheminformatics**](http://www.genomicglossaries.com/content/chapterchemogenomics.asp#chem info ov)[**Drug discovery & development**](http://www.genomicglossaries.com/content/chapterddv.asp)[**Molecular Medicine**](http://www.genomicglossaries.com/content/chaptermolmed.asp)  [**Business**](http://www.genomicglossaries.com/content/chapterbusiness.asp)

**Bioinformatics** is inextricably intertwined with the biological, chemical and medical resources in all the other sections.

**What is Bioinformatics?** Many definitions, difficult to reach agreement on.

The field of science in which biology, computer science, and information technology merge into a single discipline. The ultimate goal of the field is to enable the discovery of new biological insights as well as to create a global perspective from which unifying principles in biology can be discerned. There are three important sub- disciplines within bioinformatics: the development of **new algorithms and statistics** with which to assess relationships among members of large data sets; the **analysis and** [**interpretation**](http://www.ncbi.nlm.nih.gov/Education/Bioinformatics/dataanal.html)of various [**types of data**](http://www.ncbi.nlm.nih.gov/Education/Bioinformatics/datatypes.html) including nucleotide and amino acid sequences, protein domains, and protein structures; and the development and implementation of tools that enable efficient access and management of different types of information. "Education" NCBI, 2003 http://www.ncbi.nlm.nih.gov/Education/index.html

The definition of bioinformatics is not universally agreed upon. Generally speaking, we define it as the creation and development of advanced information and computational technologies for problems in biology, most commonly molecular biology (but increasingly in other areas of biology). As such, it deals with methods for storing, retrieving and analyzing biological data, such as nucleic acid (DNA/ RNA) and protein sequences, structures, functions, pathways and genetic interactions.  Some people construe bioinformatics more narrowly, and include only those issues dealing with the management of genome project sequencing data. Others construe bioinformatics more broadly and include all areas of **computational biology**, including population modeling and numerical simulations. Russ Altman "What is bioinformatics?" Stanford Univ. 2002 http://smi-web.stanford.edu/people/altman/bioinformatics.html

Roughly, bioinformatics describes any use of computers to handle biological information. In practice the definition used by most people is narrower; bioinformatics to them is a synonym for "computational molecular biology" - the use of computers to characterise the molecular components of living things. Damian Counsell, bioinformatics.org FAQ <http://bioinformatics.org/faq/#whatIsBioinformatics>

Conceptualizing biology in terms of molecules (in the sense of physical- chemistry) and then applying "informatics" techniques (derived from disciplines such as applied math, CS [computer science] and statistics to understand and organize the information associated with these molecules on a large- scale. Mark Gerstein "What is Bioinformatics?" MB&B 474b3, 2001
<http://bioinfo.mbb.yale.edu/what-is-it.html>

Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data.  NIH, BISTIC Biomedical Information Science and Technology Initiative, 2005   <http://www.bisti.nih.gov/>

[**More bioinformatics definitions**](http://www.genomicglossaries.com/content/chapternotes.asp#bioinfo defs)  **More** [**bioinformatics**](http://www.genomicglossaries.com/content/chapterinfosourcestext.asp#informatics terminology) **terminology**

**Computational biology**A field of biology concerned with the development of techniques for the collection and manipulation of biological data, and the use of such data to make biological discoveries or predictions. This field encompasses all computational methods and theories applicable to MOLECULAR BIOLOGY and areas of computer-based techniques for solving biological problems including manipulation of models and datasets. [MeSH, 1997] Computational biology maps to bioinformatics in PubMed.

***Computational biology FAQ***, Robert D. Phair, US, 2000 <http://www.bioinformaticsservices.com/bis/resources/faq/faq.html>

I find that people use "computational biology" when discussing that subset of bioinformatics (in the broadest sense) closest to the field of classical general biology.  Computational biologists interest themselves more with evolutionary, population and theoretical biology rather than cell and molecular biomedicine. It is inevitable that molecular biology is profoundly important in computational biology, but it is certainly not what computational biology is all about ... Richard Durbin, Head of Informatics at the Wellcome Trust Sanger Institute, expressed an interesting opinion on this distinction in an interview on this distinction:  "I do not think all biological computing is bioinformatics, e.g. mathematical modelling is not bioinformatics, even when connected with biology- related problems. In my opinion, bioinformatics has to do with management and the subsequent use of biological information, particular genetic information."  [Damian Counsell, bioinformatics.org FAQ, 2001] [https://bioinformatics.org/faq/#definitionOfCompbiol](http://bioinformatics.org/faq/#definitionOfCompbiol)

**Bioinformatics Overviews & introductions**
NCBI, NLM, NIH: **Science Primer** <http://www.ncbi.nlm.nih.gov/About/primer/index.html> Bioinformatics and molecular modeling

**Very very very short introduction to protein bioinformatics**, Patricia Babbitt et. al., 57 pages <http://baygenomics.ucsf.edu/education/workshop1/lectures/w1.color2.pdf>

**What is Informatics?
Informatics according to the OED**[translation Russian *informatika* from information SEE –ICS.] (See quotation 1967) Cf. *information science* 1967 *FID News Bull. XVii 73/2 Informatics* is the discipline of science which investigates the structure and properties (not specific content) of scientific information, as well as the regularities of scientific information activity, its theory, history, methodology and organization. Oxford English Dictionary Oxford English Dictionary, 2nd edition.

**According to NIH's Office of Rare Diseases**
The study of the application of computer and statistical techniques to the management of information. In genome projects, informatics includes the development of methods to search databases quickly, to analyse DNA sequence information, and to predict protein sequence and structure from DNA sequence data. ORD Office of Rare Diseases, NIH glossary.  [http://ord.aspensys.com/asp/resources/glossary\_f-m.asp#I](http://rarediseases.info.nih.gov/RareDiseaseList.aspx)

It is interesting that the OED definition specified domain independent information, while the ORD NIH definition is very domain specific. While "ontologies" offer the hope of cross domain interoperability, much effort is still being devoted to facilitating communication within domains. While the pharmaceutical research is increasingly interdisciplinary and NIH has come out with new initiatives such as the NIH Road map http://nihroadmap.nih.gov/index.asp  there are still many obstacles to truly interdisciplinary research.

**from the Dept. of Biopharmaceutical Sciences, UCSF Bioinformatic** and experimental analysis of protein superfamilies for understanding protein structure- function relationships and developing strategies for protein engineering. Using superfamily analysis to understand how protein sequence and structure determine protein function. Our computational approach begins with identifying the sets of divergently related proteins that comprise enzyme superfamilies and then attempts to correlate their conserved and variable structural features to similarities and differences in their functions.

This work also requires the development of new tools in protein bioinformatics to identify and evaluate distant relationships and to distinguish those elements of structure that provide common function from those that determine specificity. Designed to take advantage of the huge volumes of data coming out of the genome projects, this approach provides a much more contextual picture of the structure- function paradigm than can be achieved by studying a single protein at a time. This work has been successfully applied to such problems as the prediction of function for unknown reading frames and elucidation of enzyme mechanisms. Patricia Babbitt, Dept. of Biopharmaceutical Sciences, Univ. of California San Francisco, US http://www.ucsf.edu/dbps/faculty/pages/babbitt.html

**Introductions to protein bioinformatics**

**Protein bioinformatics**
Some/ many? of the above definitions specifically include proteins.

**[Protein] Structural bioinformatics**Involves the process of determining a protein's three- dimensional structure using comparative primary sequence **alignment**, secondary and tertiary structure prediction methods, **homology modeling**, and crystallographic diffraction pattern analyses. Currently, there is no reliable *de novo* predictive method for protein 3D- structure determination. Over the past half- century, protein structure has been determined by purifying a protein, crystallizing it, then bombarding it with X-rays. The **X-ray diffraction** pattern from the bombardment is recorded electronically and analyzed using software that creates a rough draft of the 3D structure. Biological scientists and crystallographers then tweak and manipulate the rough draft considerably. The resulting spatial coordinate file can be examined using modeling- structure software to study the gross and subtle features of the protein's structure. Christopher Smith "Bioinformatics, Genomics, and Proteomics"  Scientist 14[23]:26, Nov. 27, 2000
<http://the-scientist.com/yr2000/nov/profile_001127.html>

**See also** [**systems bioinformatics**](http://www.genomicglossaries.com/content/chapterinfosourcestext.asp#sysbio)

**What are cheminformatics, chemoinformatics, chemi- informatics?** The terminology is even less standardized here.

Google hits for:
cheminformatics about 16,300 hits Dec. 11, 2003 about 168,000 Oct. 14, 2005
chemoinformatics about 8,670 Dec. 11, 2003, about 85,300 Oct. 14, 2005
"chemical informatics" about  3,300 Dec. 11, 2003, about  2,100,000 Oct. 14, 2005
chemiinformatics  about 35 Dec. 11, 2003; about 168 Oct. 14, 2005

**Cheminformatics definitions
Cheminformatics: Going by the literature**
Mixing of information technology and management to transform data into information and information into knowledge for the intended purpose of making better decisions faster in the arena of drug lead identification and optimization. . In Chemoinformatics there are really only two [primary] questions: 1.) what to test next and 2.) what to make next. The main processes within drug discovery are lead identification, where a lead is something that has activity in the low micromolar range, and lead optimization, which is the process of transforming a lead into a drug candidate. Frank Brown, "Chemoinformatics: What is it and How does it Impact Drug Discovery" Annual Reports in Medicinal Chemistry 33: 375-384, 1998

Increasingly incorporates "compound registration into databases, including library enumeration; access to primary and secondary scientific literature; QSAR Quantitative Structure Activity Relationships) and similar tools for relating activity to structure; physical and chemical property calculations; chemical structure and property databases, chemical library design and analysis; structure- based design and statistical methods. Because these techniques have traditionally been considered the realms of scientists from different disciplines, differences in computer systems and terminology provide a barrier to effective communication. This is probably the single most challenging problem that chemoinformatics must solve. M Hann and R Green "Chemoinformatics  a new name for an old problem?" Current Opinion in Chemical Biology 3:379- 383, 1999

Many people view chemoinformatics as an extension of chemical information, which is a well established concept covering many areas that employ chemical structures, data storage and computational methods, such as compound registration databases, on- line chemical literature, SAR analysis and molecule- property calculation. Timothy Ritchie "Chemoinformatics; manipulating chemical information to facilitate decision- making in drug discovery" Drug Discovery Today 6(16) : 813- 814, Aug. 2001

**Chemical informatics**Variously known as chemoinformatics, cheminformatics, or even chemiinformatics, chemical informatics is the application of computer technology to chemistry in all of its manifestations. Much of the current use of cheminformatics techniques is in the drug industry. Indeed, one definition of chemical informatics is "the mixing of information resources to transform data into information and information into knowledge, for the intended purpose of making decisions faster in the arena of drug lead identification and optimization." Now chemical informatics is being applied to problems across the full range of chemistry.  Gary D. Wiggins, "What is Chemical Informatics?"  Indiana Univ., US, 2006  http://www.chembiogrid.org/resources/whatis.html

**Cheminformatics overviews & introductions**
**25 Years of Research in Cheminformatics:** A Portrait of the Research Group of Prof. Johann Gasteiger, Computer Chemie Centrum and Institute of Organic Chemistry, Univ. of Erlangen- Nurnberg, 2001 <http://www2.chemie.uni-erlangen.de/presentations/symposium/torvs_e.pdf>

**Cheminformatics and beyond**Drug discovery and development is in the midst of a critical transition, from a discipline dominated by empirical tests and brute force to one in which biological and chemical structural knowledge are exploited intelligently, using computational assistance. Cheminformatics, the combination of chemical synthesis, biological screening, and data mining approaches used to guide drug discovery and development, cheminformatic tools that allow for the rational selection of designed compounds with drug- like properties from an almost infinite number of synthetic possibilities, building smarter focused libraries for virtual and high- throughput screening and the exploitation of previously obtained discovery data to guide lead optimization efforts are all important.

There are many sources of chemical data; registered chemical structures with stereochemistry, synthesis records, spectral data including NMR Nuclear Magnetic Resonance, purity determinations, not to mention the volume of data generated by HTS High Throughput Screening, SAR Structure Activity Relationship studies and the calculation of physiochemical  properties. Accessibility, manipulation, and data mining of chemical information translates to knowledge for smarter drug development. Chemoinformatic tools for storage, design and mining of chemical databases/ information have had success in lead identification and optimization. Chemoinformatics is about presenting and integrating a vast and complex array of information so that people who make the decisions in drug discovery can make better choices (relatively) quickly and easily.

**Molecular modeling and systems biology**Many people include these concepts under chemical informatics.

**Molecular modeling:**
A technique for the investigation of molecular structures and properties using computational chemistry and graphical visualization techniques in order to provide a plausible three- dimensional representation under a given set of circumstances. IUPAC Medicinal Chemistry

***in silico*:** Literally "in the computer" (as contrasted with *"in vitro"* (in glass) or *"in vivo"* (in life). Can be used to screen out compounds which are not **druggable**.

In a white paper I wrote for the European Commission in 1988 I advocated the funding of genome programs, and in particular the use of computers. In this endeavour I coined "in silico" following "*in vitro*" and "*in vivo*" I think that the first public use of the word is in the following paper: A. Danchin, C. Médigue, O. Gascuel, H. Soldano, A. Hénaut, [From data banks to data bases.](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=1784830&dopt=Abstract) Res. Microbiol. (1991) 142: 913- 916.  You can find a developed account of this story in my book [The Delphic Boat](http://www.hup.harvard.edu/catalog/DANDEL.html), Harvard University Press, 2003, personal communication Antoine Danchin, Institute Pasteur, 2003

**Mapping and modeling networks and pathways**The experimental task of mapping genetic **regulatory networks** using genetic footprinting and [**yeast] two- hybrid** techniques is well underway, and the kinetics of these networks is being generated at an astounding rate. ... If the promise of the genome projects and the structural genomics effort is to be fully realized, then predictive simulation methods must be developed to make sense of this emerging experimental data.

There are three bottlenecks in the numerical analysis of biochemical reaction networks. The first is the **multiple time scales** involved. Since the time between biochemical reactions decreases exponentially with the total probability of a reaction per unit time, the number of computational steps to simulate a unit of biological time increases roughly exponentially as reactions are added to the system or rate constants are increased. The second bottleneck derives from the **necessity to collect sufficient statistics** from many runs of the Monte- Carlo simulation to predict the phenomenon of interest. The third bottleneck is a practical one of **model building and testing**: hypothesis exploration, sensitivity analyses, and back calculations, will also be computationally intensive. Lawrence Berkeley Lab "Advanced Computational Structural Genomics" Glossary, c. 1999

**more on** [**Networks and pathways**](http://www.genomicglossaries.com/content/chapternotes.asp#networks and pathways)**Systems biology**There are two opinions on what *systems biology* is supposed to be. One group sees *systems biology* as another level of combining data from different levels (like DNA, RNA and protein level) (see [Leroy] HOOD). Another group wants to combine classical molecular and cell biology with systems theory and focus on the new forms of behavior that emerge when systems of genes and proteins are studied in a wholistic way. For this they need data from all those different levels as well, of course. That is why they see *systems biology* as a cooperative effort, with systems theory providing a theoretical framework and a new view on things for biologists, along with lots of experience with complex systems, and biology providing in-depth knowledge of the field of application as well as practical handling experience. This data is the basis for developing the kind of detailed models that are necessary for such studies of systemic properties and behavior. For both groups, the goal is to reach a new level of understanding of biological systems often referred to as 'systems level' understanding. A glossary for Systems Biology, Systems Biology Group, Stuttgart [http://www.sysbio.de/projects/glossary/SYSTEMS\_BIOLOGY.shtml#systems\_biology](http://www.sysbio.de/projects/glossary/SYSTEMS_BIOLOGY.shtml/ohttp%3A/www.sysbio.de/projects/glossary/SYSTEMS_BIOLOGY.shtml#systems_biology)

**Institute for Systems Biology,** Seattle WA <http://www.systemsbiology.org/> Lee Hood's group.

**Systems bioinformatics**With the completion of the Human Genome Project, the scientific community is now faced with the even greater challenge of analyzing the resulting data from this and other large- scale genome projects to better understand the networks underlying biological function. Second International Computational Systems Bioinformatics Conference To be Held August 11-14, 2003 at Stanford University, IEEE CS Bioinformatics Technical Chair via BizWire <http://quickstart.clari.net/qs_se/webnews/wed/bx/Bca-ieee-cs_csb2003.RMsB_DuP.html>

**Drug discovery and development informatics**
**Pharmainformatics**
The multidisciplinary informatics needs of the pharmaceutical industry (HTS High Throughput Screening data, Computational Chemistry, Combinatorial Chemistry, ADME Informatics, Cheminformatics, Toxicology, Metabolic Modeling, Bioinformatics in Drug Discovery and Metabolism etc. information access and communication between various departments like the development and discovery.  Yahoo Groups Pharmainformatics <http://health.groups.yahoo.com/group/pharmainformatics/>

**Pharmaceutical bioinformatics**Bioinformatics and structure- aided drug design are really part of the same continuum. Bioinformatics offers a means to get to a structure through sequence; while structure- aided drug design offers a means to get to a drug through structure. We plan to combine innovative computational techniques with biochemical and structural expertise to bring bioinformatics and structure- aided drug design even closer together. In particular, we intend to blend computational chemistry with computational biology to create software that will aid protein chemists in understanding, evaluating and predicting the structure, function and activity of medically and industrially important proteins. My laboratory is currently involved in three "bioinformatics" projects. These include: (1) the development of novel methods to identify remote sequence/ structure relationships; (2) the creation of a compact, relational database with advanced bioinformatics functionality; and (3) the development of novel methods for predicting and evaluating protein secondary and tertiary structure. David Wishart, Wishart Pharmaceutical Research Group, Univ. of Alberta, Canada  <http://redpoll.pharmacy.ualberta.ca/projects/bioinfo.html>

**Research informatics**The explosion of genomic information, from sequences and gene expression to **SNPs** and protein structures, is of limited value for pharmaceutical researchers without powerful software capable of interpretation and comparisons. Data mining, multiple location data sharing, and computational enhancements of biological and chemistry projects, as well as integration of these efforts need various approaches for overcoming the problems of legacy information systems, the very different language and perspectives of chemists and biologists, and the organizational issues of compartmentalization and **information silos**.

**Laboratory informatics**The specialized application of information technology to maximize laboratory operations. Laboratory informatics encompasses data acquisition, data processing, laboratory information management system (LIMS), laboratory automation, scientific data management (including data analysis and long- term archiving), and electronic laboratory notebooks. Focus is on the application of this technology in analytical, production, and R&D laboratories.  Graduate Programs: Laboratory Informatics, Indiana Univ. School of Informatics, US  http://www.informatics.iupui.edu/Academics/graduate/laboratory\_informatics/index.php

**Toxicoinformatics**[**Toxicogenomics**](http://www.genomicglossaries.com/content/chaptergenomictext.asp#tox gen)

**Toxicoinformatics**An emerging scientific discipline that integrates approaches from multidisciplinary fields of bioinformatics, chemoinformatics, computational toxicology, informatics technologies and physiologically- based pharmacokinetic modeling with the objective of knowledge discovery and the elucidation of mechanisms of toxicity. NCTR's Center for Toxicoinformatics, National Center for Toxicological Research, FDA, 2003  <http://www.fda.gov/nctr/science/centers/toxicoinformatics/>

In the end, successfully moving research from the laboratory into the clinic is the ultimate **target validation**. While new technologies may be helpful and/or necessary, the challenges of scaling, automating (both for cost effectiveness and reproducibility) standardizing and simplifying are equally, if not more important.

**Medical Bioinformatics**Covers **haplotyping, genotyping** and **population genomics**, gene expression profiling, particularly for use in diagnosis, prognosis and therapeutic stratification of patients. Most of this work is being done first in oncology.

***Medical informatics***
Medical informatics has many different contexts.

The field of information science concerned with the analysis and dissemination of medical data through the application of computers to various aspects of health care and medicine. [MeSH 1987]

Medical informatics has to do with all aspects of understanding and promoting the effective organization, analysis, management, and use of information in health care. While the field of medical informatics shares the general scope of these interests with some other health care specialties and disciplines, medical informatics has developed its own areas of emphasis and approaches that have set it apart from other disciplines and specialties. For one, a common thread through medical informatics has been the emphasis on technology as an integral tool to help organize, analyze, manage, and use information. In addition, as professionals involved at the intersection of information and technology and health care, those in medical informatics have historically tended to be engaged in the research, development, and evaluation side of things, and in studying and teaching the theoretical and methodological underpinnings of data applications in health care. However, today medical informatics also counts among its profession many whose activities are focussed on dimensions that include the administration and everyday collection and use of information in health care.  FAQ, American Medical Informatics Association, 2003 http://www.amia.org/about/faqs/f7.html

**Health information data**Includes Clinical data captured during the process of diagnosis and treatment. Epidemiological databases , that aggregate data about a population. Demographic data used to identify and communicate with and about an individual. Financial data derived from the care process or aggregated for an organization or population. Research data gathered as a part of care and used for research or gathered for specific research purposes in **clinical trials**. Reference data that interacts with the care of the individual or with the healthcare deliver systems, like a formulary, protocol, care plan, clinical alerts or reminders, etc. Coded data that is translated into a standard nomenclature or classification so that it may be aggregated, analyzed, and compared.  Health Information Management; Professional definitions, Committees on Professional Development, American Health Information Management Association, 1999, 2000 http://www.ahima.org/infocenter/definitions/HIMprofessionaldefinition.htm

**Public health informatics**The systematic application of information and computer sciences to public health practice, research, and learning. It is the discipline that integrates public health with information technology. The development of this field and dissemination of informatics knowledge and expertise to public health professionals is the key to unlocking the potential of information systems to improve the health of the nation. [www.nlm.nih.gov/pubs/cbm/phi2001.html](http://www.nlm.nih.gov/pubs/cbm/phi2001.html)  [MeSH 2003]

[Emerging medical informatics specialties](http://www.genomicglossaries.com/content/chapternotes.asp#emerging)

**Social informatics**An important and often ignored piece of the puzzle. A serviceable working conception of "social informatics" is that it identifies a body of research that examines the social aspects of computerization. A more formal definition is "the interdisciplinary study of the design, uses and consequences of information technologies that takes into account their interaction with institutional and cultural contexts." ... Social informatics has been a subject of systematic analytical and critical research for the last 25 years. Unfortunately, social informatics studies are scattered in the journals of several different fields, including computer science, information systems, information science and some social sciences. Each of these fields uses somewhat different nomenclature. This diversity of communication outlets and specialized terminologies makes it hard for many non- specialists (and even specialists) to locate important studies. Rob Kling, What is social informatics and why does  it matter? D-Lib 5(1): Jan. 1999 <http://www.dlib.org/dlib/january99/kling/01kling.html>

**Deuterostomes** (taxonomic term: **Deuterostomia**; from the Greek: "second mouth") are a superphylum of [animals](http://en.wikipedia.org/wiki/Animal). They are a [subtaxon](http://en.wikipedia.org/wiki/Taxon) of the [Bilateria](http://en.wikipedia.org/wiki/Bilateria) branch of the subregnum [Eumetazoa](http://en.wikipedia.org/wiki/Eumetazoa), and are opposed to the [protostomes](http://en.wikipedia.org/wiki/Protostome). Deuterostomes are distinguished by their [embryonic development](http://en.wikipedia.org/wiki/Embryonic_development); in deuterostomes, the first opening (the [blastopore](http://en.wikipedia.org/wiki/Blastopore)) becomes the [anus](http://en.wikipedia.org/wiki/Anus), while in protostomes it becomes the [mouth](http://en.wikipedia.org/wiki/Mouth). Deuterostomes are also known as **enterocoelomates** because their [coelom](http://en.wikipedia.org/wiki/Coelom) develops through [enterocoely](http://en.wikipedia.org/wiki/Enterocoely).

There are four extant phyla of deuterostomes:

* Phylum [Chordata](http://en.wikipedia.org/wiki/Chordata) ([vertebrates](http://en.wikipedia.org/wiki/Vertebrate) and their kin)
* Phylum [Echinodermata](http://en.wikipedia.org/wiki/Echinodermata) ([sea stars](http://en.wikipedia.org/wiki/Sea_star), [sea urchins](http://en.wikipedia.org/wiki/Sea_urchin), [sea cucumbers](http://en.wikipedia.org/wiki/Holothuroidea), etc.)
* Phylum [Hemichordata](http://en.wikipedia.org/wiki/Hemichordata) ([acorn worms](http://en.wikipedia.org/wiki/Acorn_Worm) and possibly [graptolites](http://en.wikipedia.org/wiki/Graptolite))
* Phylum [Xenoturbellida](http://en.wikipedia.org/wiki/Xenoturbellida) (2 species of worm-like animals)

Superphylum Deuterostomia was redefined in 1995 based on molecular sequence analyses when the lophophorates were removed from it and combined with other protostome animals to form superphylum [Lophotrochozoa](http://en.wikipedia.org/wiki/Lophotrochozoa).[[1]](http://en.wikipedia.org/wiki/Deuterostome#cite_note-0) The phylum [Chaetognatha](http://en.wikipedia.org/wiki/Chaetognatha) (arrow worms) may also belong here. Extinct groups may include the phylum [Vetulicolia](http://en.wikipedia.org/wiki/Vetulicolia). Echinodermata, Hemichordata and Xenoturbellida form the clade [Ambulacraria](http://en.wikipedia.org/wiki/Ambulacraria).[[2]](http://en.wikipedia.org/wiki/Deuterostome#cite_note-1)

In both deuterostomes and protostomes, a zygote first develops into a hollow ball of cells, called a [blastula](http://en.wikipedia.org/wiki/Blastula). In deuterostomes, the early divisions occur parallel or perpendicular to the polar axis. This is called [radial cleavage](http://en.wikipedia.org/wiki/Radial_cleavage), and also occurs in certain protostomes, such as the [lophophorates](http://en.wikipedia.org/wiki/Lophophorate). Most deuterostomes display [indeterminate cleavage](http://en.wikipedia.org/wiki/Indeterminate_cleavage), in which the developmental fate of the cells in the developing embryo are not determined by the identity of the parent cell. Thus if the first four cells are separated, each cell is capable of forming a complete small larva, and if a cell is removed from the blastula the other cells will compensate.

In deuterostomes the [mesoderm](http://en.wikipedia.org/wiki/Germ_layer#Mesoderm) forms as evaginations of the developed gut that pinch off, forming the [coelom](http://en.wikipedia.org/wiki/Body_cavity). This is called [enterocoely](http://en.wikipedia.org/wiki/Enterocoely).

Both the Hemichordata and Chordata have [gill slits](http://en.wikipedia.org/wiki/Gill_slit), and primitive fossil echinoderms also show signs of gill slits. A hollow nerve cord is found in all chordates, including [tunicates](http://en.wikipedia.org/wiki/Tunicate) (in the larval stage). Some hemichordates also have a tubular nerve cord. In the early embryonic stage it looks like the hollow nerve cord of chordates. Because of the degenerated nervous system of echinoderms, it is not possible to discern much about their ancestors in this matter, but based on different facts it is quite possible that all the present deuterostomes evolved from a common ancestor that had gill slits, a hollow nerve cord and a segmented body. It could have resembled the small group of Cambrian deuterostomes named [Vetulicolia](http://en.wikipedia.org/wiki/Vetulicolia).

|  |
| --- |
| **Contents** [[hide](http://en.wikipedia.org/wiki/Deuterostome)] * [1 Formation of mouth and anus](http://en.wikipedia.org/wiki/Deuterostome#Formation_of_mouth_and_anus)
* [2 Origins](http://en.wikipedia.org/wiki/Deuterostome#Origins)
* [3 References](http://en.wikipedia.org/wiki/Deuterostome#References)
* [4 External links](http://en.wikipedia.org/wiki/Deuterostome#External_links)
 |

**[**[**edit**](http://en.wikipedia.org/w/index.php?title=Deuterostome&action=edit&section=1)**] Formation of mouth and anus**

All chordates, such as birds and mammals, are deuterostomes. Shown here is a [Barred Owl](http://en.wikipedia.org/wiki/Barred_Owl).

*Main article:* [*Embryological origins of the mouth and anus*](http://en.wikipedia.org/wiki/Embryological_origins_of_the_mouth_and_anus)

The defining characteristic of the dueterostome is the fact that the blastopore (the opening at the bottom of the forming gastrula) is turned into the anus, whereas in protosomes the blastopore becomes the mouth. The deuterostome mouth develops at the opposite end of the embryo from the blastopore and a digestive tract develops in the middle connecting the two.

**[**[**edit**](http://en.wikipedia.org/w/index.php?title=Deuterostome&action=edit&section=2)**] Origins**

The majority of animals more complex than [jellyfish](http://en.wikipedia.org/wiki/Jellyfish) and other [Cnidarians](http://en.wikipedia.org/wiki/Cnidarians) are split into two groups, the [protostomes](http://en.wikipedia.org/wiki/Protostome) and deuterostomes. Chordates (which include all the vertebrates) are deuterostomes.[[3]](http://en.wikipedia.org/wiki/Deuterostome#cite_note-Erwin20002-2) It seems very likely that [555](http://toolserver.org/~verisimilus/Timeline/Timeline.php?Ma=555) million years old [*Kimberella*](http://en.wikipedia.org/wiki/Kimberella) was a member of the protostomes.[[4]](http://en.wikipedia.org/wiki/Deuterostome#cite_note-Fedonkin2007-3)[[5]](http://en.wikipedia.org/wiki/Deuterostome#cite_note-Butterfield2006-4) If so, this means that the protostome and deuterostome lineages must have split some time before *Kimberella* appeared — at least [558](http://toolserver.org/~verisimilus/Timeline/Timeline.php?Ma=558) million years ago, and hence well before the start of the Cambrian [542](http://toolserver.org/~verisimilus/Timeline/Timeline.php?Ma=542) million years ago.[[3]](http://en.wikipedia.org/wiki/Deuterostome#cite_note-Erwin20002-2) The [Ediacaran](http://en.wikipedia.org/wiki/Ediacaran) fossil [*Ernietta*](http://en.wikipedia.org/wiki/Ernietta), from about [549 to 543](http://toolserver.org/~verisimilus/Timeline/Timeline.php?Ma=549–543) million years ago, may represent a deuterostome animal.[[6]](http://en.wikipedia.org/wiki/Deuterostome#cite_note-5)

Fossils of one major deuterostome group, the [echinoderms](http://en.wikipedia.org/wiki/Echinoderm) (whose modern members include [sea stars](http://en.wikipedia.org/wiki/Sea_star), [sea urchins](http://en.wikipedia.org/wiki/Sea_urchin) and [crinoids](http://en.wikipedia.org/wiki/Crinoid)) are quite common from the start of the Cambrian, [542](http://toolserver.org/~verisimilus/Timeline/Timeline.php?Ma=542) million years ago.[[7]](http://en.wikipedia.org/wiki/Deuterostome#cite_note-Bengtson2004-6) The Mid [Cambrian](http://en.wikipedia.org/wiki/Cambrian) fossil [*Rhabdotubus johanssoni*](http://en.wikipedia.org/wiki/Rhabdotubus_johanssoni) has been interpreted as a [pterobranch](http://en.wikipedia.org/wiki/Pterobranch) hemichordate.[[8]](http://en.wikipedia.org/wiki/Deuterostome#cite_note-7) Opinions differ about whether the [Chengjiang fauna](http://en.wikipedia.org/wiki/Chengjiang_fauna) fossil [*Yunnanozoon*](http://en.wikipedia.org/wiki/Yunnanozoon), from the earlier Cambrian, was a hemichordate or chordate.[[9]](http://en.wikipedia.org/wiki/Deuterostome#cite_note-ChenHangLi1996-8)[[10]](http://en.wikipedia.org/wiki/Deuterostome#cite_note-ChenHangLi1999-9) Another Chenjiang fossil, [*Haikouella lanceolata*](http://en.wikipedia.org/wiki/Haikouella_lanceolata), also from the Chengjiang fauna, is interpreted as a chordate and possibly a craniate, as it shows signs of a heart, arteries, gill filaments, a tail, a neural chord with a brain at the front end, and possibly eyes — although it also had short tentacles round its mouth.[[10]](http://en.wikipedia.org/wiki/Deuterostome#cite_note-ChenHangLi1999-9) [*Haikouichthys*](http://en.wikipedia.org/wiki/Haikouichthys) and [*Myllokunmingia*](http://en.wikipedia.org/wiki/Myllokunmingia), also from the Chenjiang fauna, are regarded as [fish](http://en.wikipedia.org/wiki/Fish).[[11]](http://en.wikipedia.org/wiki/Deuterostome#cite_note-10)[[12]](http://en.wikipedia.org/wiki/Deuterostome#cite_note-11) [*Pikaia*](http://en.wikipedia.org/wiki/Pikaia), discovered much earlier but from the Mid Cambrian [Burgess Shale](http://en.wikipedia.org/wiki/Burgess_Shale), is also regarded as a primitive chordate.[[13]](http://en.wikipedia.org/wiki/Deuterostome#cite_note-12) On the other hand fossils of early chordates are very rare, since non-vertebrate chordates have no bones or teeth, and none have been reported for the rest of the Cambrian.

**Protostome**

From Wikipedia, the free encyclopedia

Jump to: [navigation](http://en.wikipedia.org/wiki/Protostome#mw-head), [search](http://en.wikipedia.org/wiki/Protostome#p-search)

**Protostomia** (from [Greek](http://en.wikipedia.org/wiki/Greek_language) meaning "mouth first") are a [clade](http://en.wikipedia.org/wiki/Clade) of [animals](http://en.wikipedia.org/wiki/Animal). Together with the [deuterostomes](http://en.wikipedia.org/wiki/Deuterostome) and a few smaller [phyla](http://en.wikipedia.org/wiki/Phylum), they make up the [Bilateria](http://en.wikipedia.org/wiki/Bilateria), mostly comprising animals with [bilateral symmetry](http://en.wikipedia.org/wiki/Symmetry_%28biology%29#Bilateral_symmetry) and [three](http://en.wikipedia.org/wiki/Triploblastic) [germ layers](http://en.wikipedia.org/wiki/Germ_layer). The major distinctions between [deuterostomes](http://en.wikipedia.org/wiki/Deuterostome) and protostomes are found in [embryonic development](http://en.wikipedia.org/wiki/Embryonic_development).

In animals at least as complex as [earthworms](http://en.wikipedia.org/wiki/Earthworms), the [embryo](http://en.wikipedia.org/wiki/Embryo) forms a dent on one side, the [blastopore](http://en.wikipedia.org/wiki/Blastopore), which deepens to become the [archenteron](http://en.wikipedia.org/wiki/Archenteron), the first phase in the growth of the [gut](http://en.wikipedia.org/wiki/Gut_%28zoology%29). In [deuterostomes](http://en.wikipedia.org/wiki/Deuterostome), the original dent becomes the anus while the gut eventually tunnels through to make another opening, which forms the mouth. The protostomes were so named because it used to be thought that in their embryos the dent formed the mouth while the anus was formed later, at the opening made by the other end of the gut. More recent research, however, shows that in protostomes the edges of the dent close up in the middle, leaving openings at the ends which become the mouth and anus.[[1]](http://en.wikipedia.org/wiki/Protostome#cite_note-0) However, this idea has been challenged, because the [Acoelomorpha](http://en.wikipedia.org/wiki/Acoelomorpha), a group which forms a sister group to the rest of the bilaterian animals, have a single mouth which leads into a blind gut (with no anus). The genes employed in the embryonic construction of this mouth are the same as those expressed around the protostome mouth, [[2]](http://en.wikipedia.org/wiki/Protostome#cite_note-Hejnol2008-1)

There are other significant differences between the protostome and deuterostome patterns of development:

* Most protostomes are schizocoelomates, meaning a solid mass of the embryonic [mesoderm](http://en.wikipedia.org/wiki/Mesoderm) splits to form a [coelom](http://en.wikipedia.org/wiki/Coelom). A few, such as [Priapulids](http://en.wikipedia.org/wiki/Priapulid), have no coelom, but they may have descended from schizocoelomate ancestors. On the other hand all known deuterostomes are [enterocoelous](http://en.wikipedia.org/wiki/Enterocoely), meaning that the coelom is formed from longitudinal pouches of the [archenteron](http://en.wikipedia.org/wiki/Archenteron) which then become separate cavities.
* Within the Protostomes a number of phyla undergo what is known as [spiral cleavage](http://en.wikipedia.org/wiki/Spiral_cleavage) which is determinate, meaning that the fate of the cells is determined as they are formed. This is in contrast to deuterostomes which have [radial cleavage](http://en.wikipedia.org/wiki/Radial_cleavage) that is indeterminate.

Current molecular data suggest that protostome animals can be divided into three major groups:

* [Ecdysozoa](http://en.wikipedia.org/wiki/Ecdysozoa), e.g. [arthropods](http://en.wikipedia.org/wiki/Arthropod), [nematodes](http://en.wikipedia.org/wiki/Nematode)
* [Platyzoa](http://en.wikipedia.org/wiki/Platyzoa), e.g. [platyhelminthes](http://en.wikipedia.org/wiki/Platyhelminthes), [rotifers](http://en.wikipedia.org/wiki/Rotifer)
* [Lophotrochozoa](http://en.wikipedia.org/wiki/Lophotrochozoa), e.g. [molluscs](http://en.wikipedia.org/wiki/Mollusc), [annelids](http://en.wikipedia.org/wiki/Annelid)

as well as a number of minor taxa of basal or ambiguous affinity.