B

Belief Revision

The stability and malleability of belief is the concern of belief revision. As with most theoretical language, belief and its revision have multiple meanings depending on purpose, context, and logic, so any case must define its terms and their scope. This entry focuses on belief revision within the context of scientific reasoning.

Stakeholders are divided by whether they orient belief in the realm of cognitive (knowledge and reasoning) or affective (emotions, attitudes, values, and spiritual belief). The topic of belief revision need not address the stability or malleability of the world one is believing in but simply the dynamics that influence belief persistence, adjustment, or abandonment. Although individuals may revise beliefs pragmatically, beliefs in the sense of scientific theory or a computer software's logic would require a deliberate strategy reflecting a philosophy of science. For instance, Karl Popper (1902–1994) would reject a prototype model if any part diverged, but Willard Van Orman Quine (1908–2000) would revise or *fine-tune* the prototype until all parts became congruent.

Defining Belief and Its Revision

The revision function depends upon the defined scope of the belief, the logic used to determine doubt or confidence, and the epistemic aims and values regarding the belief. Belief may represent *propositions of truth*, justified by rational argument to produce knowledge. Belief may also suggest the degree of confidence in the proposition (faith) or it may provide the justification for other beliefs (theory). Belief may denote a pattern of interpretation (habits of mind) or judgment (ethics and morals). A belief system is one in which truths form a coherent framework of understanding (ideology); that is, an integrated, coherent protocol for making decisions (priorities), predictions (hypotheses), or solving problems (algorithms).

Theories Explaining Belief Revision

The first definition above is the standard used in *scientific reasoning* and especially in computer programming. Each coded statement contributes a belief to the mechanical logic. These statements reduce the complexity of a scenario to very small increments, forming a cohesive set of responses to new information. Advanced programs are called *artificial intelligence* if they *learn*, that is, if they can change their own coding. Belief revision thus refers to advanced coding logic for designing computational programs. Apart from the technicalities involved, this discourse informs analytic philosophy, which will symbolize propositions in transformational language akin to algebra.

If each belief serves to scaffold further beliefs, there is a foundation required to reach complex heights. The *belief set* is a closed set of propositions under examination, while *belief base* can examine separate propositions that are elements of the base and can tell the difference between those that are explicitly stated and those that are

a logical extension of the explicit beliefs. The primary goal of such a *foundation theory* is to discover truth, and thus a faulty premise casts doubt on the entire code sequence. This coding logic will identify an erroneous proposition and then either accept it (expansion), remove it (contraction), or both add something and take something else away (revision). It is a syntactic approach, akin to using syllogism for deductive reasoning. By contrast, a *coherence theory* is a synthetic approach, accepting new propositions for their logic in maintaining the meaningfulness of the whole framework; in other words, *verisimilitude*. Stability of the network of decisions is more important than semantic integrity, so the existing belief structure is not revised so much as its applications. A computer program example would be the apparently intuitive helpfulness of online search engines to provide advertisements consistent with topics related to your own history of links. The purpose of the algorithm is to generate all links consistent with what appears to be your consumer interest; no ideology is at risk for the computer by adding more search items. As long as the additional information is consistent with your own belief about the purpose of being online, you do not revise your belief; however, once the cost of these additions is too great, such as when the ads are distractingly animated and intrusive, you may revise your belief. According to foundation theory, you might abandon that browser or online shopping if you conclude the practice is wrong, but according to coherence theory, you might simply work around the ads if they are tolerable, or the stability of your habit takes on greater value than the dissonance of incompatible behavior.

Experimental studies of belief revision may be structured in series with multivariate analyses used to identify effect sizes for a range of possible influences. The formulas used in the statistical packages are themselves the source of inquiry now that computation can occur so quickly. Repeated measures of cognitive learning outcomes in schools are used to monitor cognitive change interpreted according to pedagogical theory.

Belief Persistence

One reason a belief is held in the face of contrary evidence is simply habit for individuals and the momentum of a complex network for organizations. Habits are automatic, whereas belief revision requires deliberation. There is also an emotional dimension of enjoying the predictability of customary beliefs and not trusting that the work and discomfort of extinguishing the old ones will be adequately compensated by a better future. Yet another reason beliefs do not change is that memories are faulty at best, it being impossible to retrieve all details of arguments that lead to confidence in the incumbent decision. Also, individual belief systems are fraught with inconsistencies, and people often will tolerate beliefs that do not seem to be aligned. For instance, one can wholeheartedly agree with the need to buy from independent local businesses but will nonetheless search out a bargain at a distant warehouse store.

Many beliefs about the world are formed intuitively at a young age. Teachers must diagnose their students' knowledge base, including misconceptions, in order to design instruction that prompts a more enlightened understanding and assessment that measures the degree of confidence in the new perspective. In this way, learning can be construed as a revision of belief. A particular challenge is posed when the academic objective is dispositional; that is, when students are expected to become more tolerant, frugal, democratic, generous, or healthy. Meanwhile, the aforementioned factors of habit, memory, comfort, and purpose conspire to maintain false beliefs or disbelief.

Belief profiles have been found to correlate with patterns of behavior and conceptual change, and thus belief revision is of interest to those hoping to influence change as well as those simply curious to understand the phenomenon. Given their individual conventions regarding the concept of belief, there are different fields of inquiry concerned with belief revision, for example, education (nature of teaching and learning), psychology (existential crises; self-efficacy), religion (conversions; hermeneutics), business (market analysis; organization management), mathematics (proofs; inferential statistics), science (empirical justification), and computer programming (artificial intelligence).

Many theories have emerged out of interest in belief revision. Thomas Kuhn (1922–1996) described a *paradigm shift* after which a revolutionary idea is

widely assumed to be true, for example, evolution or climate change, each of which has substantial empirical evidence to support it. The correlation for personal beliefs would be the loss of faith after some traumatic event, for instance, one's belief in God or one's trust in banks or one's commitment to marriage. The result is an existential crisis, a conversion to a new belief system, or perhaps agnosticism; that is, not knowing what is true, or perhaps an antagonistic stance such as atheism. Holocaust survivor Viktor Frankl (1905–1997) developed *logotherapy* to explain the sudden deaths of prisoners even when circumstances had not changed. He concluded that all people search for meaning in their lives, a useful construct in mental health.

Cybernetics is a system theory concerned with the dynamics of interdependence. While associated with mechanical or nonhuman systems, the principles of interaction apply to social contexts, especially regarding feedback, whether cognitive or affective. Conflict resolution techniques are a result, making the case for interrupting cycles of violence in any scale, such as within families or between countries. In such matters of diplomacy and legality, terms of formal agreements between parties will specify behaviors associated with target beliefs about their identities and status, given that in any social system there is a continuously challenging landscape of problems with sharing resources and power. Both what is believed and how it is believed will influence the stability of the relationships which are maintained on the basis of continual interaction with trusted sources of information. Open communication is blurred by tactics of silencing voices and withholding or distorting information and of limiting sources to an *echo chamber* that accepts only that which conforms to one's belief.

Naomi Jeffery Petersen

See also Fact Versus Theory; Generalization; Mental Models; Theory Change

Further Readings

Godden, D. M. (2012). Rethinking the debriefing paradigm: The rationality of belief perseverance. *Logos & Episteme: An International Journal of Epistemology*, *3*(1), 51–74

Hansson, S. O. (2014). Logic of belief revision. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2014 ed.). http://plato.stanford.edu/archives/spr2014/entries/logic-belief-revision

Lane, J. D., & Paul, L. H. (2014). Confronting, representing, and believing counterintuitive concepts: Navigating the natural and the supernatural. *Perspectives on Psychological Science*, *9*(2), 144–160.

Lee, B. D. (1998). The paradox of belief instability and a revision theory of belief. *Pacific Philosophical Quarterly*, *79(*4), 314–328.

Mason, L. (2010). Beliefs about knowledge and revision of knowledge: On the importance of epistemic beliefs for intentional conceptual change in elementary and middle school students. In L. D. Bendixen & F. C. Feucht (Eds.), *Personal epistemology in the classroom: Theory, research, and implications for practice* (pp. 258–291). Cambridge University Press.

Schlottmann, A., & Norman, H. A. (1995). Belief revision in children: Serial judgment in social cognition and decision-making domains. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *21*, 1349–1364.

Urban, T. (2019). The thinking ladder. *Wait But Why.* https://waitbutwhy.com/2019/09/thinking-ladder.html

Big Data

Our current age is often referred to as the era of *big data*. While the term has become increasingly popular in the last 10 years, the concept is now used in a wide variety of ways, often without referring to an acknowledged technical definition of the concept. In fact, an agreed-upon and technical definition can be argued not to exist, and the term *big data* is mainly used to refer to broad and general tendencies related to a rapid increase in the data collected in various domains and the accompanying need to develop and employ new methods of analysis in order to make sense of such data. These methods of analysis connect big data intimately to data mining, artificial intelligence (AI), and machine learning.

Big data has become a ubiquitous phenomenon, and it is applied to domains such as research, business intelligence, social networks, health care, crime prevention, and anti-terrorism. Data is routinely described as the "new oil," and in parallel with the historical discovery of other precious

resources, the quest to drill, mine, and refine it is both intense and characterized by competition, rapid development, and private sector innovation. Big data carries a number of actual and potential benefits, but it is also related to a number of ethical challenges. Many are related to the nature of the data employed, some relate to a purported blind and dangerous faith in quantitative data and algorithms as the solution to and source of insight in *all* aspects of human lives, and some are mainly related to the use of algorithmic decision-making in general. The next section of this entry presents an expanded definition of big data, followed by an exploration of the philosophical basis of the concept and its roots in the quantitative and empiricist approach to knowledge. Before concluding, the entry offers a discussion of the ethical considerations surrounding the growing applications of big data.

Definition

Big data refers to a development toward generating and gathering increasing amounts of data at increasing speeds, while the data is structured in heterogenous data sets. The data in question can, for example, be social and personal, or related to industrial processes or scientific measurements, and several sources and types of data are often connected in large data sets.

Although *data* is of course a key component of big data, the tools and techniques used to store, manage, and analyze the data are also integral parts of the concept. The problem with the data sets handled in big data is that traditional methods of storage and analysis are insufficient. Distributed storage systems and cloud computing are used for storing and managing the data. Data mining and machine-learning-based AI have become terms intimately associated with the analysis of the large data sets. The entire chain of generating, gathering, storing, and analyzing data is included as part of what is referred to as *big data*.

Data in isolation—gathered and stored—has limited potential for positive or negative consequences. It is through the quest to derive value from the data that, for example, benefits in business intelligence and concerns over discriminatory practices arise. The term *big data* is now used by both data and computer scientists, social scientists, and the general public, and the latter two groups often refer to a general and nonspecified concept without referring to particular forms of database management or specific machine learning techniques.

The techniques used to analyze the data are not entirely new, and these stem from earlier academic work in AI and statistics. However, the volume of today's data sets has mainly become a reality through private corporations focused on gathering and using data to provide various services, and this has led to a situation in which much development is done outside of academia. Although private companies are important for developing big data techniques, government and other public institutions are important secondary beneficiaries. There is now a growing tendency for government agencies to use big data in order to, among many other things, prevent crime and terrorism, find effective pandemic responses, and both discover fraud and automate decision-making in social services.

Philosophical Foundation

As data is analyzed in order to answer certain questions, or achieve a set of goals, some philosophical assumptions are inevitably involved. These *can* be made explicit, but far more often, as big data is applied to solve issues of importance for businesses and society, the assumptions remain implicit. One reason is that a lot of the techniques are developed and applied in nonacademic settings, but another important reason is that proponents of big data have repeatedly argued that big data is *not* based on theory or ideology. One of the major causes of disagreement over the implications of big data revolves around the possibility of achieving neutral and objective science through analyses of raw data and partially autonomous AI.

Scientific and other applications of big data are characterized by a multitude of approaches, and in the following a *typical* (but not the only possible) approach is described. By removing the human scientist from the equation, it is argued, true objective science has finally been realized. Such claims disregard the nature and origin of the data, and that data in itself is not neutral. The

various choices related to how data is gathered, which variables are collected, how data is coded, and so forth, by necessity involve a set of choices resulting in data that is a *model* of reality and not an accurate reproduction of it. Also, human beings are heavily involved in the creation of the algorithms used in the analysis of data, decisions about when and how to use big data solutions, and also in choosing how the results are interpreted and utilized.

Big data is based on a quantitative and empiricist approach to knowledge, in which observable phenomena are seen as the basis of knowledge. As machine learning methods have become increasingly advanced, some even argue that human interpretation is no longer an integral part of deriving insight from these observations. Furthermore, *more* data is seen as conducive to more knowledge, whereby observations both strengthen the results of the analysis *and* improve the tools of analysis as data is used to train and further develop machine learning algorithms.

As data is gathered and fed to the computer for analysis, the intent is that AI tools sift through it and uncover interesting patterns in the data which are not identifiable directly by humans or through traditional statistical methods. This process is referred to as *data mining*. Big data is thus usually characterized by inductivism and the idea that we blindly approach data in order to identify patterns and develop theories about dependencies and causation. In practice, however, human beings are necessarily involved in a dynamic process in which initial results from the analysis are examined and used to guide and shape further analysis. Such a process would be more accurately described as a combination of the deductive and inductive approaches, at times referred to as an *abductive* or *retroductive* approach.

Big data as applied to human phenomena is often based on a behaviorist approach, in which observable phenomena and actions are emphasized over, for example, cognitive phenomena and interpretation. When an individual's subjective experiences and evaluations constitute the data analyzed, the goal is likely to be how this translates into actions, and how various variables are connected, rather than a deeper understanding and explanation of what goes on inside people's heads.

In short, big data is philosophically closer to the natural sciences than the social sciences and humanities. The knowledge and insight produced by big data can thus be argued to constitute a part of the answer when human phenomena are examined, and not a complete picture. In order to explain and understand *why* the patterns discovered occur, and how the phenomena and correlations described are actually experienced, human scientists, other methods of analysis, and other forms of data are still vital parts of making sense of the world as it appears through the results produced through big data.

Ethical Considerations

Big data is a major part of much contemporary debate about surveillance, privacy, social networks, and liberty. Although not directly related to big data as *large data sets*, such debate is clearly related to some of the more prominent applications of the general phenomenon of big data understood as large data sets combined with AI for analysis. In this section, the discussion narrows in on *social* and *personal* big data, since data related to individuals and their social relations is the most obvious cause of concern.

A key concern related to big data is a purported lack of transparency and difficulties of explaining the results of AI-based analysis of data sets. Many of the purported benefits of big data are incontrovertible, but an opaque process of analysis often likened to a "black box" creates a situation in which unintended consequences often go unnoticed. For this reason, much research on AI is now focused on creating *explainable AI* because this is required for securing transparency and ensuring that big data and AI are developed and deployed in an accountable and responsible manner.

Because big data is often based on historical data, its use can perpetuate and even obfuscate structural racism, sexism, and other forms of inequality. For example, if a bank employs an algorithm for determining who should receive a loan based on historical data, the models trained on this data will perpetuate historical bias related to, for example, race and gender. Furthermore, although we know that human beings are biased, some argue that machines are *not* biased; therefore, unless the algorithms are audited and

developed in order to foster explanation, discriminatory practices may go unnoticed and even become legitimized.

Privacy and surveillance are other concepts often related to big data, as the generation and gathering of data can be perceived as a form of surveillance. Concerns relate to a violation of rights to privacy in general, and to more insidious phenomena, such as the combination of large numbers of data points from different databases in order to create highly detailed personality profiles. This allows actors to influence individuals more effectively than they could otherwise have done and could thus be construed as inimical to liberty and independence. Such influence relates to all sorts of behavior, including political behavior and government efforts to exert social control.

Whereas the aforementioned use of personal data often involves efforts to market and sell products and services more effectively, the societal consequences of big data are also a cause for concern. Particularly in relation to social media, some authors have shown how big-data-based personalization might allow actors to influence individuals' perceptions of the world (trap them in "filter bubbles" where people are increasingly being exposed to what they *want* to see), influence how people associate and relate to each other (and create "echo chambers" in which they associate only with like-minded people), and thus contribute to societal polarization, which is a growing concern.

In contrast to the preceding challenges, the benefits derivable from big data can potentially help improve individuals' lives and alleviate and solve important social problems. For example, preventing crime and terrorism is clearly a good thing and can easily be argued to warrant the sacrifice on some privacy. The same applies to using big data in health care, where it is used to improve the understanding of diseases and interventions and for more effective diagnosis. The purpose of highlighting the causes for concern mentioned above is simply to show that while big data is highly effective and provides a plethora of new opportunities in all domains of society, it is important to include perspectives from ethics and the social sciences in the development and deployment of big data projects.

Conclusion

While big data is arguably a continuation of a long quest for more information and better methods of analysis, recent developments related to the massive amounts of data gathered from various devices such as household appliances, smart electricity meters, global-positioning-system-enabled devices and other devices with a variety of sensors, and social media suggest that it is reasonable to describe Big Data as a new and important phenomenon. The amount of data has increased rapidly, and even if the machine learning techniques used are not entirely new, they are now trained on far larger data sets than before, which leads to increasingly useful results and a renewed interest in research and development.

Big data is used in all domains of modern societies, and the benefits it provides are substantial. These include, among other things, benefits in terms of scientific advancements, improved business intelligence and analytics, new tools and methods for law enforcement, innovation in health care, and new ways for people and societies to connect, communicate, and share information.

At the same time, there are important ethical concerns raised by big data, which include invasions of privacy, the non-neutrality of data and discrimination, societal effects of social networks, and the use of big data to exercise political influence and control. Those in possession of large data sets gain power and the means to influence others more effectively than ever before.

Henrik Skaug Sætra

See also Artificial Intelligence; Data and Phenomena; Data Models; Philosophy of Science; Prediction

Further Readings

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, *19*(2), 171–209. https://doi.org/10.1007/s11036-013-0489-0

Dean, J. (2014). *Big data, data mining, and machine learning: Value creation for business leaders and practitioners*. John Wiley & Sons.

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1). https://doi.org/10.1177/2053951714528481

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

Sætra, H. S. (2018). Science as a vocation in the era of big data: The philosophy of science behind big data and humanity's continued part in science. *Integrative Psychological and Behavioral Science*, *52*(4), 508–522. https://doi.org/10.1007/s12124-018-9447-5

Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2013). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, *26*(1), 97–107. https://doi.org/10.1109/TKDE.2013.109

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for the future at the new frontier of power*. Profile Books.

Biochemistry, 19th Century

The field of biochemistry has a long-standing history of immense contributions to the progress of medicine. Notwithstanding, biochemistry has attracted less attention than the other biomedical sciences. Much of this can be attributed to the fact that as a separate entity it is rather young when compared with the likes of anatomy or physiology. Before 1750, biochemistry was not a full-fledged theory in its own right. It did gain a measure of respectability as its own discipline, however, in the later part of the 19th century and would come to fruition thereafter.

Although a strong association can be traced between biochemistry and organic chemistry, the history of biochemistry has received scant coverage within the history of chemistry generally. This is perhaps because biochemistry was once regarded by chemists as a lesser subdiscipline or as a part of medicine. In its broader sense, biochemistry can be regarded as the study of living matter and its chemical composition as well as of the biochemical processes essential for the maintenance of growth and vital life activities. This entry aims to examine the various developments within the sphere of biochemistry in the 19th century that were important to its subsequent emergence in the 20th century. Significant contributions by eminent scientists, such as the discovery of the activity of yeasts in alcoholic fermentation, discovery of the first enzyme, and the classification and nomenclature of the enzymes, were all crucial leads for the future generations of biochemists. In this phase, biochemistry was mainly concerned with the study of chemical reactions taking place within living cells (both in microorganisms and in humans).

Roots of Biochemistry: An Overview

The roots of biochemistry can be found in 19th-century studies of organic chemistry, animal chemistry, and physiological chemistry. It was already found by then that the chemistry of living and nonliving materials was markedly different. In the 1830s, it was thought that the jellylike homogenous substance, protoplasm, present within the organisms was responsible for carrying out most of the vital activities of biosynthesis, respiration, and intracellular food breakdown (metabolism). However, this general belief did not find support from such eminent chemists as Justus von Liebig (1803–1873) or Ernst Hoppe-Seyler (1825–1895). Although the hydrolytic enzymes such as pepsin, maltase, and amylase were known to the scientific world at that time, they were not considered to be acting within the cells. Eduard Buchner in 1897 perhaps performed the single most significant experiment that commenced the study of biochemistry. He prepared a cell-free yeast extract which he named zymase and observed its efficiency in fermenting glucose into ethanol and carbon dioxide. He later went on to receive the Nobel Prize (in 1907) for his discovery of cell-free fermentation and contribution to biochemical research. Zymase was regarded by Buchner as a single enzyme; however, this was soon disproved by others and was found to contain several other enzymes. Nonetheless, Buchner's work discredited the protoplasm theory and confirmed fermentation as a chemical process. Also, the presumed differences between the catalytic activities of the intracellular enzymes and the extracellular hydrolytic enzymes were found not to exist.

The French chemist Louis Pasteur (1822–1895) started studying the alcoholic fermentation of sugar by yeasts. He went on to conclude that the fermentation of sugar to alcohol is carried out by a vital force called *ferments*, present within the yeast cells, that can only function inside the living

cells. He further went on to clarify that fermentation was not associated with the putrefaction or death of the cells but rather was correlated with the organization and life of the yeast cells. In 1833, the first enzyme, called *diastase*, was discovered by the French chemist Anselme Payen, and in 1878 the German physiologist Wilhelm Kühne coined the term *enzyme* to describe this process. Nonliving substances such as pepsin was later referred to as *enzymes*, while the term *ferment* was used to describe the chemical processes of the living organisms.

Anselme Payen and Diastase

Payen, the son of an entrepreneur who had started a number of chemical production factories, rose to fame with his discoveries of the carbohydrate cellulose and the enzyme diastase. He had received knowledge of chemistry from his father and later from the likes of Michel-Eugène Chevreul and Nicolas Louis Vauquelin. Payen was promoted to become the head of his father's borax production plant in 1815. Here, he was able to use boric acid (available cheaply from Italy) to prepare borax. With the production costs going down significantly, Payen succeeded in besting his competitors in selling borax. Payen then shifted his attention in studying the sugar production from sugar beets and thereby established a method of sugar decolorization using charcoal. Subsequently, charcoal filters began to be used in gas masks to absorb dangerous organic gases. In 1833, Payen came up with yet another discovery: a chemical derived from malt extract that could catalyze the conversion of starch to sugar. The catalyst was named diastase, the first enzyme to be produced in concentrated form. The pattern of naming the enzyme diastase, as initiated by Payen, is still existent today with enzymes being named with the suffix *-ase*. He also concentrated his research on wood and extracted from it a substance resembling starch. This substance he named *cellulose*, which he could find in abundance in the plant cell walls he studied. Here also Payen became the pioneer in starting a nomenclature of the carbohydrates that would end with the suffix *-ose*. Although Payen was the first to isolate cellulose, it was the American botanist Wanda K. Farr (1895–1983) who ultimately discovered the mechanism of its production by plants. Many other subsequent inventions used cellulose as a building block. It became the main component in a number of products, including cellophane, celluloid, rayon, collodion, guncotton, nitrocellulose, and other related products.

Louis Pasteur and Alcoholic Fermentation

The first of Pasteur's contribution to the study on microbial activities was in relation to lactic acid fermentation. Pasteur was consulted by a local alcohol producer, a Monsieur Bigo, who was facing serious problems with fermentation. After careful examination of the fermenting liquor under a microscope, Pasteur found that when the fermentation was satisfactory, the globules present were rounded in shape, but when they deteriorated and became elongated, the end product was lactic acid fermentation. There are four general requirements for such work, as suggested by the papers published by Pasteur on lactic acid fermentation. The conditions to be fulfilled include the following: (1) Optimum conditions must be prevalent in order to study the process of fermentation, (2) the substances used should be from the simplest possible source, (3) the appearance of the organism during the process of fermentation should remain constant, which should be confirmed by careful examination under the microscope, and (4) the characteristic fermentation can be produced by even a minute trace of the presumptive cause. In 1857, Pasteur published the first paper on alcoholic fermentation. In accordance with the catalytic theory of von Liebig and Berzelius, the ferment takes nothing from the fermentable material and gives up nothing during the fermentation process. However, when the ingredients were weighed prior to the commencement of, and following, fermentation, it was clear that the yeast cells were taking something from the sugar. The sugar's breakdown into carbon dioxide and alcohol was associated by Pasteur to the living processes, wherein the part of the material for the yeast was provided by sugar. The role of yeast in alcoholic fermentation was affirmed by Pasteur in 1860. In sharp contrast to von Liebig's assumptions, ethanol and carbon dioxide constituted only 95% of the products of fermentation of invert

sugar (mixture of glucose and fructose). The rest of the 5% included cellulose, succinic acid, and glycerol. Thus, Pasteur proved that yeast indeed took something from the sugar which was not returned; hence he claimed that fermentation is a physiological process. Also, Pasteur was able to produce a yeast type from a defined medium containing inorganic phosphate and ammonium tartrate along with sugar. There was no existence of any substance that can be putrefied by oxygen and make sugar unstable, as had been proposed by von Liebig. Thus, the theory that the origin of yeast is dependent on the action of oxygen on the fermentable liquid was refuted by Pasteur. Here, Pasteur has also encountered the problem in differentiating between fermentation by intact cells and by enzymic action. There was a long-standing difficulty around the confusion regarding the difference between enzymic action and fermentation, and Pasteur gave special attention to this issue. He had a particular difference of opinion with Marcellin Berthelot (1827–1907), an eminent figure in the French scientific community. In 1860, around the same time when Pasteur published his alcoholic fermentation paper, Berthelot provided an important account of how sucrose is broken down into fructose and glucose by beer yeast. Earlier, Eilhard Mitscherlich had given accounts on the activity of yeast extract to produce a levorotatory sugar from cane sugar. Augustin-Pierre Dubrunfaut further went on to show that glucose and fructose constituted the levorotatory sugar. In order to refute the views of Pasteur, Berthelot in 1860 described the method of invertase (β-fructofuranosidase) isolation that can break down sucrose. Pasteur was of the opinion that it was the succinic acid produced during the fermentation process that was responsible for the breakdown of sucrose. Succinic acid, however, could not invert sucrose at all in conditions similar to that which is prevalent during the process of fermentation, and inversion was only possible in an alkaline medium. During the interval from 1860 to 1880, chemists had a definite change in stance concerning the activities of the enzymes and the process of fermentation. The findings of Berthelot and Pasteur are thought to have provoked these changes in attitudes of the scientific community toward the biological processes and helped immensely in building up and shaping the field of biochemistry during its periods of infancy. In 1878, Wilhelm F. Kühne, in order to remove the confusion surrounding the double meaning of the term *ferment*, introduced the term *enzyme* to designate the soluble ferments.

Contribution of Eduard Buchner

Eduard Buchner started his work with Adolf von Baeyer in chemistry in 1884 and with C. von Naegeli in botany. His brother Hans was his special supervisor at the Botanic Institute, Munich. In 1885, his first publication explained the effect of oxygen on fermentation. He received special stimulation and assistance for his research in organic chemistry from the likes of H. von Pechmann and T. Curtius. It was possible for him to continue his studies with the aid of the Lamont Scholarship. With grant aids from von Baeyer, Buchner established a small laboratory of chemical fermentation, where he performed his experiments on chemical fermentations. The rupture of the yeast cells was experimentally studied by him in the year 1893, although the board members of the laboratory believed that this would accomplish nothing. In 1907, Buchner went on to receive the Nobel Prize for his discovery of cell-free fermentation and contributions to biochemical research. In 1897, the experimental design was set up with the production of a cell-free extract from yeast cells. Buchner went on to show that the cell-free extract could ferment sugar. The theory of vitalism received a severe blow when it was shown that the process of fermentation did not require the presence of living yeast cells. He prepared the cell-free yeast extract by a combination of kieselguhr (diatomaceous earth) and quartz with dry yeast cells after pulverizing the dry cells with mortar and pestle. With the cell contents coming out of the yeast cells, the mixture became moistened. Following this step, a press was used for the mixture to be passed through that would result in the production of the *press juice* that contained fructose, glucose, or added maltose. Carbon dioxide sometimes was found to develop. It was found through microscopic analysis that the extract contained no living yeast cells. Buchner believed that the proteins secreted by the yeast cells in their environment were responsible for the fermentation of the

sugars, but later it was confirmed that it was within the yeast cells that the fermentation took place.

Sugar Metabolism and Enzyme Specificity

Toward the end of the 19th century, Emil Fischer emerged as a leading organic chemist who today is regarded as the founder of carbohydrate chemistry. The first ideas about enzyme specificity and its molecular mechanisms were drawn from his studies on sugar metabolism in yeast. A new hexose, mannose, was prepared in 1888 by Fischer and Hirschberger, which at room temperature could be fermented avidly by beer yeast. Yeast was earlier encountered by Fischer when his father had invested in a Dortmund brewery. He used sugars such as galactose, mannose, and glucose to test for the fermenting ability of beer yeast. He found that fermentation was possible for the D-sugars only. Thus, he could separate this form from the racemic mixtures with L-forms. Similar approach was undertaken by Pasteur for the tartaric acids. Subsequently, the osazones and hydrazones of the corresponding L-sugars were characterized by Fischer. Prior to this, Fischer was able to discover the reaction between the sugars and phenylhydrazine, which was later used in deriving the configurations of the sugars. The sugars produce characteristic osazone crystals; this was an important finding since previously the major handicap of sugar chemistry was the problem in obtaining sugar crystals. In line with the shift of Pasteur from chemistry to biology, Fischer, after establishing the sugar configurations and monosaccharide classifications, thought of applying his findings to biological research. To save material, he made use of a very small fermentation tube, since the sugar preparations were often laborious and the experiments needed to be repeated frequently. Until exploited by Fischer, Pasteur's finding about the potential ability of the microbes to discriminate between the L- and D-substrates received little attention. Accordingly, Fischer observed that only the D-forms of the sugars like galactose, mannose, and glucose were fermented by the microbes and not their L-counterparts. Additionally, he found that the different sugars were fermented by different yeasts and their fermentation ability also varied considerably, which can be largely attributed to their structural characteristics. Fischer in 1894 developed the image of lock and key that formed the basis of enzyme-substrate complex formation, as developed later on by Leonor Michaelis, Maud Menten, and Victor Henri. It is also implicated in the idea of substrate destabilization or transition state stabilization and in the role of selective binding energy—the basis for substrate activation theory by J. B. S. Haldane, enzyme transition state complementarity by Linus Pauling, and induced fit theory by Daniel Koshland. All these eminent scientists have acknowledged the findings of Fischer as the basis for their works.

Kühne and Physiological Chemistry

Wilhelm Kühne joined Heidelberg in 1871 to replace Helmholtz. Here in Heidelberg, Kühne went on to produce some of the greatest discoveries of his life, including the digestion of protein and the chemistry of vision. He continued to work selflessly and enthusiastically in Heidelberg and left a rich legacy in the history of biochemistry and physiology. Besides his mainstream work, his research included a series of miscellaneous studies. Kühne published papers on the origin of hippuric acid, artificial diabetes in frogs, resuscitation of carbon monoxide, poisoned dogs, the cause of jaundice, blood ozone content, occurrence of ammonia in blood, cholera transfer and treatment, and on many other aspects of physiological chemistry, thereby demonstrating his breadth of talents and versatility of interests. Living muscle fiber sarcoplasm was considered by Kühne as a protein-rich fluid whereby clotting of the proteins takes place during heat or death rigor. Kühne observed sarcoplasm fluidity with the accidental finding that inside a single frog muscle fiber, a living nematode could pass through the cross striations by performing active movements. This was when the concept of myosin came to the fore. (Albert Szent-Györgyi, a hundred years later, would study the phenomenon of muscle contraction based on Kühne's myosin discovery.) Kühne then turned his focus to the study of digestive enzymes and ultimately went on to discover trypsin, which can be regarded as a landmark discovery in the field of proteomics research. Besides

discovering trypsin, his efforts also identified a number of protein digestion products. Using a mixture of dog pancreas and beef fibrin, he observed that fibrin solubilized with the concomitant appearance of peptone, leucine, and tyrosine. In a piece of dog small intestine which still had in place the pancreatic duct, similar digestion products could be obtained. The digestive function of pancreatic juice was confirmed, and Kühne found that this activity was different from the role, earlier discovered, that gastric juice performs. At the Naturhistorisch-Medizinischen Verein, Heidelberg, Kühne delivered a lecture in 1876. Here, he went on to introduce the term *enzyme* to designate unorganized ferments such as ptyalin and pepsin. The actions of these substances could already be distinguished from that of the living cells known to contain ferments in an organized specialized form. The pancreatic powerful proteolytic enzyme was termed *trypsin* (the cleaver). Kühne is also credited with the extraction of trypsin and examination of some of its properties. This was followed by a series of other observations in relation to trypsin, such as the fistula-mediated collection of pancreatic juice; protein clotting by this juice; observation of the pancreatic cells that secrete this juice (which undergo a change in form during secretion); and the microscopic observation during rest and digestion of the rabbit pancreas. With Ewald, Kühne tested the activity of trypsin on different animal tissues, being aware of the potential usefulness of trypsin in histology. His outstanding understanding of physiology and chemistry allowed him to bring insights from both to solidify the basis of biochemistry. Trypsin was unsuccessful in digesting some of the tissue elements, and these he was able to isolate as keratinized and collagenous structures. This reignited Kühne's earlier interests in protein chemistry and protein digestion, and with Russell Henry Chittenden he resumed his works on these aspects of protein biology. He had earlier established that proteins and carbohydrates were distinct in their composition by employing alkaline and acid hydrolysis to obtain some of the products of protein hydrolysis that were entirely distinct from that of the carbohydrate hydrolysates. The research was extended by Kühne to make it more physiological and relevant from the biological perspective by using pancreatic and gastric juice enzymes. Two of the general groups of substances found as the early products of digestion, namely peptones and albumoses, had previously been found by different investigators as products of alkaline and acid hydrolysis of proteins. Eventually, a number of amino acids could be detected. It was quite obvious that Kühne and Chittenden were successful in identifying the sequential nature of intestinal (pancreatic) and gastric digestion of proteins, since the breakdown of proteins could proceed to a certain extent in an acid-trypsin environment and required the alkaline-trypsin environment to proceed further. Knowledge regarding the difference in action of trypsin and pepsin on the peptides has been much extended to the present day and terms such as *antipeptones*, *antialbumoses*, *hemipeptones*, and *hemialbumoses* that were used frequently by Kühne and Chittenden are no longer existent. However, it cannot be denied that these pioneering works are the cornerstone for the field of proteolysis, which has immensely enriched the sphere of biochemistry.

Syed Feroj Ahmed

See also Biochemistry, 20th Century

Further Readings

Dubos, J. (1951). Louis Pasteur: Free Lance of Science, Gollancz. Quoted in Manchester K. L. (1995): Louis Pasteur (1822–1895): Chance and the prepared mind. *Trends in Biotechnology, 13*(12).

Fischer, E. (1894). Synthesen in der Zuckergruppe II. *Berichte der Deutschen Chemischen Gesellschaft*, *27*.

Kühne, W. (1877). Über das Verhalten verschiedener organisirter und sog. ungeformter Fermente. *Verhandlungen des naturhistorisch-medicinischen Vereins zu Heidelberg. Neue Folge, 1*.

Mitscherlich, E. (1842). Ueber die chemische Zersetzung und Verbindung mittelst Contactsubstanzen. *Annales des Chimie et des Physique, 55*.

Pasteur, L. (1858). Memoire sur la fermentation appele´e lactique. *Annales des Chimie et des Physique, 52*.

Payen, A. & Persoz, J. F. (1833). Mémoire sur la diastase, les principaux produits de ses réactions et leurs applications aux arts industriels (Memoir on diastase, the principal products of its reactions, and their applications to the industrial arts), *Annales de Chimie et de Physique, 53*(2).

Ukrow, R. (2004). Nobelpreisträger Eduard Buchner (1860–1917): Ein Leben für die Chemie der Gärungen und - fast vergessen - für die organische Chemie.

Biochemistry, 20th Century

Biochemistry may be defined as the study of different chemical processes in all living organisms. These processes typically govern all life functions, as for example by controlling information flow through biochemical signaling networks and the intrinsic flow of chemical energy by way of metabolism.

The history of biochemistry spans more than eight centuries. Its origins can be traced to the ancient and medieval societies of China and India, with subsequent advancements made in the Islamic world during the centuries that followed. With the introduction of clinical trials and clinical pharmacology in the modern era, biochemistry had continued to make significant breakthroughs, although it was only in the year 1903 that the term *biochemistry* was proposed by Carl Neuberg, a German chemist, to denote a single interdisciplinary subject combining elements of what hitherto had been two separate disciplines, biology and chemistry. Over the course of the 20th century, biochemistry gradually developed into an essential discipline within the life sciences.

Much research deals with the basic structures and effective functions of cellular components such as carbohydrates, proteins, nucleic acids, lipids, and many other biomolecules, although the study of various natural processes involving close-linked networks rather than those of individual molecules is quickly becoming the focus of biochemistry. During the last half century, biochemistry has succeeded in explaining a variety of life processes, and as a result many disciplines, from medicine to botany, are seriously engaged in biochemical research. Today, it has become the main focus of pure biochemistry to understand how biological molecules actually give rise to the processes that typically occur within living cells, which in turn significantly contributes to the understanding of whole organisms.

Among the vast numbers of many different biomolecules that exist in the living system, most are large, highly complex and molecules (called *biopolymers*), which are composed of many similar repeating subunits (called *monomers*). Each of these classes of polymeric biomolecules has a very different set of various subunit types—for example, a *protein*, which is essentially a polymer whose different subunits are selected from a typical set of 20 or more amino acids. *Carbohydrates*, in the form of polysaccharides, are formed by linking monosaccharides or oligosaccharides; *lipids* are typically formed from combining fatty acids and glycerols; and the nucleic acids are formed from the simple nucleotides. Biochemistry also studies the basic chemical properties of the very important biological molecules, which includes proteins with catalytic activities, in particular the hard-core chemistry of the enzyme-catalyzed reactions. Furthermore, the metabolism of specific cell types and the working of the endocrine system have also been quite extensively described. Some other areas of focus in biochemistry include the understanding of the genetic code as embodied in the nucleic acids deoxyribonucleic acid (DNA) and ribonucleic acid (RNA); cell membrane transport; protein synthesis; and the phenomenon of signal transduction, by which a chemical or physical signal is transmitted through a cell as a series of molecular events. Thus, biochemistry as a field of study is highly dynamic and has the potential to continue to develop with time. The following sections discuss some of the significant discoveries and achievements in biochemistry during that period and identify the key subfields and related disciplines that have emerged since the turn of the 20th century.

Major Developments in the 20th Century

Before the turn of the 20th century, the notion was widely held among scientists that living cells were occupied by a substance called *protoplasm*, in whose huge molecules obscure and perhaps incomprehensible chemical changes continuously took place. This view would become discredited when it was discovered that complex cell constituents like the proteins and enzymes could be effectively isolated and carefully crystallized. In 1901, for example, the English biochemist Frederick Gowland Hopkins, with his graduate student Sydney W. Cole, in 1901, discovered the amino acid tryptophan to be a typical constituent of many proteins.

Enzymes

In 1897, the German chemist Eduard Buchner began to study the specific ability of yeast extracts to efficiently ferment sugar despite the absence of any living yeast cells. In a series of critical experiments at his laboratory in Berlin, he found that fermentation of sugar occurred without the presence of living yeast cells owing to the action of some chemical compound in the extract or mixture. He determined this to be an *enzyme* (i.e., a catalytic protein that brought about the fermentation of sucrose), which he named *zymase*. In 1907, Buchner received the prestigious Nobel Prize in Chemistry for his outstanding biochemical research that led to the discovery of typical cell-free fermentation. Following Buchner's example, researchers started naming enzymes according to the kind of reaction they perform in the system. For example, *lactase* is the name of the enzyme that cleaves lactose; *polymerase* is the enzyme that catalyzes the formation of DNA polymers. Many early workers noted that enzymatic activity was typically associated with proteins, but many scientists at that time, including the 1915 Nobel laureate Richard Willstätter, vigorously argued that the proteins were merely the carriers for true enzymes and that these proteins were virtually incapable of performing catalysis. However, in 1926, James B. Sumner clearly showed that urease, a common enzyme, was basically a pure protein and efficiently crystallized it; he again did likewise for another enzyme, catalase, in 1937. Further, the distinct conclusion that many pure proteins can act as enzymes was definitively proved by John Howard Northrop and Wendell Meredith Stanley, who crystallized the three digestive enzymes trypsin, pepsin, and chymotrypsin. For their pioneering efforts Sumner, Northrop, and Stanley were jointly awarded the Nobel Prize in Chemistry in 1946. The critical discovery that enzymes could be effectively crystallized meant that scientists could eventually solve their molecular structures by the technique of X-ray crystallography. This was for the first time done for the enzyme *lysozyme*, which is found significantly in saliva, tears, and egg white, and which is capable of breaking down the cell wall of many gram-positive bacteria. This structure was actually solved by a group of researchers led by David Chilton Phillips, and the work was published in the year 1965. This crystallized high-resolution structure of lysozyme marked the beginning of the field of structural biology and also led to the effort to understand how these enzymes actually work at the atomic level.

Metabolism

One of the most significant and prolific of 20th-century investigators was the biochemist and physician Hans Adolf Krebs (1900–1981), who made major contributions to the study of metabolism. He himself discovered the urea cycle and later, while working with Hans Kornberg (1928–2019), discovered the glyoxylate cycle and the TCA or citric acid cycle. These significant discoveries led to Krebs being awarded the Nobel Prize in Physiology in 1953, which he shared with another well-known biochemist, Fritz Albert Lipmann, who was in addition the codiscoverer of the essential factor *coenzyme A*. In 1960, the biochemist Robert K. Crane revealed his extraordinary discovery of the phenomenon of sodium-glucose cotransporters as the major mechanism for glucose absorption in the intestine. This was the first significant proposal of a typical coupling between the basic fluxes of a substrate and an ion, and it has been viewed as having efficiently sparked a new revolution in biology. This discovery would not have been possible, however, without the work of Emil Fischer, who had actually discovered the structure and chemical properties of the molecule glucose, for which he was awarded the 1902 Nobel Prize in Chemistry, nearly 60 years before. Because the process of metabolism entails the breakdown of larger molecules into simpler ones (catabolic processes) and the synthesis of larger complex molecules from these simple particles (anabolic processes), the importance of glucose and its involvement in the formation of adenosine triphosphate (ATP) are fundamental to our understanding of metabolism. Glycolysis, also known as the *Embden–Meyerhof–Parnas (EMP) pathway*, named after the three scientists who discovered it, is one of the main pathways by which breakdown of glucose occurs under both aerobic and anaerobic conditions. These three researchers discovered that the glycolysis is actually a strongly determinant process, which is essential for the functional efficiency of the human body.

Understanding this pathway led to the opening of a new field of clinical biochemistry that participates in the study of metabolic disorders. Subsequently, it was also possible to delineate the difference between the glycolytic pathways in the mammalian and microbial systems.

Major Instrumental Advances

Since the mid-20th century, biochemistry has significantly advanced with the rapid development of more and more new techniques such as electron microscopy, chromatography, radioisotopic labeling, protein nuclear magnetic resonance spectroscopy, X-ray diffraction, and studies of molecular dynamics simulations. These extraordinary techniques have enabled the discovery and detailed analysis of many significant cellular functions, such as the citric acid cycle and glycolysis.

Modern biochemistry was revolutionized after the invention of the primary gene amplification technique called the *polymerase chain reaction* (PCR), which was developed by Kary Mullis in the 1980s. This technique basically allows copying a single gene, which is then gradually amplified into millions of copies and has therefore become a cornerstone in the standard protocol for any biochemist who wishes to work with gene expression. PCR is used not only for basic gene expression research but also in aiding research and clinical laboratories in diagnosing leukemia, lymphomas, and other malignant diseases. Without PCR development, many important advances in the field of protein expression study and bacterial study would not have come to fruition. Along with the development of the basic theory and the process of PCR, the invention of the instrument called the *thermal cycler* proved to be critically important, since the use of PCR would not have been possible without it. This is a clear instance of the combined value of the advancement of technology and the painstaking research that resulted in the development of key theoretical concepts in biochemistry and their subsequent application.

Vitamins

In 1912, the Polish biochemist Casimir Funk put forward the arresting theory that scurvy, beriberi, rickets, and perhaps pellagra were actually caused by the lack or deficiency of some special substances in the diet, which he named *vitamins*. With the typical exception of vitamin B_{12} and folic acid, by 1940 all of these substances had been mostly identified, purposely isolated, and chemically characterized.

Before the onset of World War II, not only the structures of almost all the vitamins had been discovered but their actual modes of action had also been largely recognized. Vitamin B_1 (thiamine) deficiency was found to cause polyneuritis in the kidney due to the faulty utilization of pyruvate. It was actually the cofactor thiamine pyrophosphate which proved to be extensively needed for the typical processes of natural or biological decarboxylation of molecules like pyruvate and others. During this era, nicotinic acid was found to be an important growth factor. In addition, its presence is required for the effective biosynthesis of nicotinamide coenzymes, which play an important and fundamental role as typical hydrogen carriers, in both aerobic and anaerobic organisms. This latter fact makes it quite possible that nicotinamide adenine dinucleotide (NAD) was among the first substances ever formed in living cells, perhaps millions of years ago when life on earth greatly flourished in an anaerobic atmosphere. Subsequent observations pointed out the efficacy of pyridoxine, or vitamin B_6, as another important cofactor, shown later to play an important role, in the form of its pyrophosphate, in the amino acid transaminations and decarboxylation.

By midcentury, most of the now well-known vitamins had been discovered and their typical modes of action explained. Pantothenic acid was actually found to be an integral part of the coenzyme A molecule. At this time, biotin had been discovered, and its structure had been totally elucidated, then its mode of action in the path of causing major carbon dioxide assimilation in animal cell metabolism was largely clarified by Feodor Lynen and colleagues.

Molecular Genetics

With its roots in biochemistry and closely allied with molecular biology, another discipline of life sciences with vast potential is molecular genetics. A prime target of its investigations is the genetics of enzyme-deficiency diseases and their

basic clinical aspects. Molecular genetics offers tools for anticipating major health risks and for the prevention of diseases. Yet another related field of importance is molecular pharmacology, which aims to alleviate and cure our aches, pains, and diseases.

Neurochemistry

Another important domain, which emerged in the latter half of the 20th century, is a neurochemistry, which studies the chemistry of the nerves and the brain. Here, the cell surfaces play a fundamental role, as do the *synapses*, the gaps between the terminal dendrites of adjacent neurons where electrochemical events occur between communicating cells. At each synapse there is a ceaseless and rapid surge of potassium, sodium, calcium ions as well as some other ions and amino acids, various types of organic substances, and at specific gates or pores; that is, points having special permeability. A neuron can effectively pass an electrochemical signal to another by a chemical messenger through the synapse and a very long, quite sustained, chained effect is accomplished whereby a typical stimulus at a particular sensory organ is finally communicated to the brain and a response by the brain is generated. The vesicles that exist in the synaptic cleft of the neuron possess a high concentration of chemical messengers. This organized phenomenon of typical exocytosis is the foundation of our understanding of perception and of life functions ranging from simple reflexes to conscious awareness and abstract thought.

Conclusion

Biochemistry interfaces with both biology and chemistry and is concerned with the various chemical processes that occur within living cells. Contemporary biochemistry emerged from what in the late 19th and early 20th century was called *physiological chemistry*, which dealt chiefly with extracellular bodily fluids and the processes of digestion. Biochemistry as such is thus largely, though not quite exclusively, a 20th-century discipline. Molecular biology, which came into existence as an offshoot of biochemistry itself, has come to mean a study of the three-dimensional structure and function of such biologically relevant and important biomolecules as nucleic acids and proteins. Molecular biology is basically as much a typical interface of biology with chemistry as of general biology with physics. In many such respects biochemistry and molecular biology represent the basic realization of the dream of the mechanistic biologists of the early 20th century, who were convinced that the most fundamental biological processes could of course ultimately be well understood in terms of the fundamental laws of both physics and chemistry.

Syed Feroj Ahmed

See also Biochemistry, 19th Century

Further Readings

Bartlett, J. M. S., & Stirling, D. (2003). A short history of the polymerase chain reaction. *Methods in Molecular Biology, 226*.

Florkin, M., & Stotz, E. H. (1970–1995). *Comprehensive biochemistry, section VI (Volumes 30–39): A history of biochemistry*. Elsevier.

Fruton, J. S. (1999). *Proteins, enzymes, genes: The interplay of chemistry and biology*. Yale University Press.

Funk, C. (1912). The etiology of the deficiency diseases. Beri-beri, polyneuritis in birds, epidemic dropsy, scurvy, experimental scurvy in animals, infantile scurvy, ship beri-beri, pellagra. *Journal of State Medicine, 20*.

Hopkins, F. G. (1913). The dynamic side of biochemistry. *Report of the British Association for the Advancement of Science, 652*.

Kohler, R. E. (1973). The enzyme theory and the origin of biochemistry. *Isis,* 64(222).

Quastel, J. H. (1985). The development of biochemistry in the 20th century. *Molecular and Cellular Biochemistry, 69*(1).

Simoni, R. D. (2002). The discovery of Glutathione by F. Gowland Hopkins and the beginning of biochemistry at Cambridge University. *The Journal of Biological Chemistry, 277*.

Sumner, J. B. (1926). The isolation and crystallization of the enzyme urease. Preliminary paper. *Journal of Biological Chemistry, 69*.

Templeton, N. S. (1992). The polymerase chain reaction. History, methods, and applications. *Diagnostic Molecular Pathology: The American Journal of Surgical Pathology, Part B, 1*(1).

Weckbecker, A., Gröger, H., & Hummel, W. (2010). Regeneration of nicotinamide coenzymes: Principles and applications for the synthesis of chiral compounds. *Advances in Biochemical Engineering/ Biotechnology, 120*.

Wilson, B. A., Schisler, J. C., & Willis, M. S. (2000). Sir Hans Adolf Krebs: Architect of metabolic cycles. *Laboratory Medicine, 41*(6).

Biochemistry, Contemporary

Biochemistry is a discipline that examines and implements new knowledge of the physical, biological, and chemical processes within organisms, how these constitute interrelated systems, and how their structures determine their physiological functions. The study of biochemistry describes the structures and mechanisms of life in molecular terms and the evolution of complex physiological functions. The chemical nature of simple organic molecules yields a world of its own—one in which 20 standard amino acids assemble to generate essential functions of complex proteins and structures. This examination of contemporary biochemistry aims to discuss the physical and chemical processes involved within organisms and how larger systems ultimately develop. After a presentation of some basic concepts, the entry provides an explanation of the features of cells, followed by a discussion of the role of nucleic acids in living systems. Next, a detailed review of protein structure and functions precedes a discussion of drug development and technology, which identifies the primary targets of drug research. The entry continues with an account of cellular metabolism and disease and concludes with an exploration of the biochemistry of the brain.

Fundamental Concepts

Life consists of three major properties: (1) the ability to transform energy and matter, (2) the ability to grow and reproduce, and (3) the ability to respond to and modify the environment. Traditional undergraduate biochemistry curriculums introduce the chemistry of life sequentially, separating general chemistry from biochemistry by a year's study of organic chemistry. While governed by the physical and chemical laws of the nonliving universe, the distinction between the subjects lies in differentiating living organisms from nonliving matter. For example, carbon dioxide is an organic molecule and is also a biochemical substrate for the enzyme carbonic anhydrase, which catalyzes the production of bicarbonate and protons. The resulting Bohr effect is a biological phenomenon in which carbon dioxide reduces hemoglobin's affinity for oxygen. Carbonic anhydrase is itself a biomolecule, yet more complex and animated in its function than carbon dioxide.

Biomolecules are a collection of organic molecules, often complex chemical structures, which, in an organized and hierarchical manner, maintain and perpetuate life governed by the many physicochemical principles associated with the movement of energy through matter. Yet, when trying to understand the chemical processes within and between organisms, what is already known often needs refinement. Acquiring new knowledge of how these molecules interact allows biochemists to formulate potentially alternative models on established concepts. Contemporary biochemistry builds upon traditional knowledge to accumulate a new understanding of perceived biological phenomena that often lead to novel approaches in biotechnology.

Additionally, *biomedicine* is the application of biochemical theories toward insightful discoveries in developmental biology, anatomy, and physiology. Contemporary scientists use fundamental core concepts to find key ways to tackle new problems, so as to develop an increasingly comprehensive understanding of human development and disease, with a focus on case studies and clinical application. Concepts that constitute contemporary biochemistry and distinguish them from traditional biochemistry aim to answer many questions about science and life, and how the two together can broaden perspectives about cellular interactions.

The Cell

All cells share similar chemical features. They contain approximately 70 percent water by mass. They have lipid membranes that separate and protect the intracellular components from the external environment. The organelles and other significant

cell components are composed of four major macromolecules: *proteins*, *lipids*, *carbohydrates*, and *nucleic acids*. Simple organic molecules are used as building blocks for the larger complex macromolecules. The origins of the cell are not well understood, yet there is striking consistency in the biochemistry of life from the simplest prokaryote to complex organisms. Norman Pace presented an insightful perspective on biochemical unity in which he reasons that the phylogenetic tree of life, based on small ribosomal ribonucleic acid (RNA) sequences, indicates that the three domains of simple cellular life—archaea, eukaryotes, and bacteria—have a common evolutionary origin, and that inorganic molecules may have been used to support early metabolic processes. The relative ordering of evolutionary events between the earlier form and the three domains is the subject of an ongoing debate among scientists.

Understanding the nature of cells through exploration, discovery, inquiry, and consideration has led to accelerated advancements in stem cell research. Stem cells are a unique type of cell, which can either remain unspecialized or differentiate into specific cell lineages. Discovering the active biochemical agents that guide decisions made by stem cells provides insight into tissue repair and wound healing treatments.

Nucleic Acids

Genomic information of the cell is stored and expressed by nucleic acids. Nucleic acids are long, linear polymers formed from four types of monomers. Each monomer consists of a sugar, phosphate, and nucleobase. The sequence of the bases yields the information content of nucleic acids. Information storage and active use through metabolic processes ensure fidelity. Optimizing information transfer is provided at the level of genetic replication, which must access and evolve the stored information to be faithful to the environment. Deoxyribonucleic acid (DNA) is the biomolecule responsible for storing information, and as a memory molecule it possesses the collective history of all the environmental factors that led to the survival of our species. The importance of genetic information was noticed after polymeric DNA was inferred by Oswald Avery in 1944 to contain genetic information. In 1953, Rosalind Franklin made further contributions when she collected X-ray diffraction data on purified DNA, which led to the development of a structural model of the DNA helix by James Watson and Francis Crick. From this era, experimental discoveries from studies of bacteria and bacterial viruses provided evidence that DNA alone functions to store and transmit information.

The structural and biochemical knowledge of DNA gained since the 1970s is now being used in the emerging fields of nanotechnology and biotechnology. The 1970s development of recombinant DNA technology sparked a revolutionary shift in the world of science. Contemporary scientists can now modify DNA sequences in their endogenous contexts and observe the functional organization of the genome at a systems level. Continued study of prokaryotic organisms such as bacteria and an increased understanding of their adaptive strategies against viral infection led to the discovery of the clustered regularly interspaced short palindromic repeats (CRISPR) system by Yoshizumi Ishino in 1987. In 2020, Jennifer Doudna and Emmanuelle Charpentier received the Nobel Prize in Chemistry for developing the system into a genome editing tool in which the CRISPR nuclease cas9 is targeted by a short guide RNA that recognizes the target DNA of choice and integrates exogenous DNA into endogenous sites. Precise modification of genetic building blocks could further enhance drug development processes and medical therapeutics by introducing new drug targets or correcting harmful mutations. Knowledge of DNA modification is imperative for understanding, teaching, or experimenting in contemporary biochemistry and impacts research in bacterial genetics, prokaryotic genetics, prokaryotic gene regulation, developmental biology, and cancer.

Another area of interest is messenger RNA (mRNA) vaccines, which were released under emergency authorization in response to the COVID-19 pandemic in 2020. As noted by science writer Elie Dolgin, in the journal *Nature*, the history of mRNA vaccines is not straightforward. Multiple threads of interwoven discoveries from many scientists contributed to the mosaic of knowledge that led to the repurposing and delivery of mRNA as a vaccine. As it stands, a small segment of mRNA corresponding to an

encoded area of the viral protein is introduced into the cell. Once translated into protein, an immune response is generated in which antibodies are produced and can recognize and mark invading viruses for destruction. Lipid-based nanoparticles are an essential strategy for mRNA delivery, since mRNA is unstable and unable to cross the lipid membrane owing to its negatively charged backbone. The lipid nanoparticles enclose the negatively charged RNA strand, which enables mRNA transport across the lipid membrane and evades detection by RNases once inside the cell. Many details of this process are still not well understood and are an area of interest for biochemists.

Protein Structure and Function

The unpacking of encoded information into proficient proteins occurs at the ribosome in a process called *translation*. Proteins are composed of amino acids linked together by a peptide bond. The distinct ordering of amino acids determines the primary structure of proteins. The diversity of proteins arises from the unique primary sequences of amino acids and their side chains. Proteins fold to create three-dimensional structures with multiple levels of organization. Secondary structures are held together by hydrogen bonding within the backbone of a single polypeptide chain, which collapses into a foldable tertiary structure driven into self-organization and stabilization by the hydrophobic effect. Noncovalent interactions such as hydrogen bonding, ionic bonds, and dispersion forces occur between side chains and help to further stabilize the tertiary structure of a single polypeptide chain. Additional stabilization is provided when the side chains of the amino acid cysteine are oxidized to form disulfide bridges, which are covalent bonds and stronger than noncovalent interactions. Quaternary structures arise from the same folding principles of tertiary structures but occur between multiple polypeptide chains.

Protein folding does not appear to be a random process. A protein with four disulfide bonds can fold in a matter of seconds, even though there is a 1/105 probability of reforming the bonds correctly. The bacteria *Escherichia coli* can make a functional protein in 5 seconds, which includes translation and folding, but if each possible conformation was sampled, it would take 10^{77} years to sample a reduced number of conformations for a protein containing 100 amino acids (single vibration is 10^{-13} seconds per conformation; 10 different conformations per amino acid). In the hierarchical model of protein folding, hydrogen bonding helps to stabilize local areas of secondary structure, which may narrow the number of possible rearrangements, with distant areas folding together through the hydrophobic effect. This differs from the molten globule model in which the polypeptide chain collapses into a compact globular structure driven primarily by the *hydrophobic effect*, whereby a gain in the stability of the folded protein is due to the release of ordered water molecules from nonpolar side chains of the polypeptide. The brief time that it takes to fold shows that all the information necessary for protein folding is present in the primary sequence. Computational biochemistry builds upon this knowledge by providing insightful discoveries into the mechanism of protein folding with computational software that predicts protein structures based on sequence and structure homology.

Proteins exhibit a multifunctional nature. At a physiological level, proteins are the primary drivers of metabolic processes and the expression of phenotypic function. At the cellular level, proteins serve as enzymes, hormones, or antibodies; additionally, there are structural and membrane-bound proteins. The geometric and chemical complementarity between proteins and their substrates or ligands ensures selectivity, and at a grander scale it enriches the structural and functional diversity of cellular proteins. *Enzymes* are proteins that catalyze reactions in the human body. They cannot change the direction of the reaction but can lower the activation energy needed to speed up a reaction—in other words, how fast the reaction reaches equilibrium. While most enzymes are proteins, in some cases, RNAs are also capable of catalyzing reactions. As biochemical catalysts, enzymes are often regulated by effector molecules and inhibitors. They are often dependent on cofactors that assist in functional group transformations and redox reactions.

Drug Development and Technology

Academics, clinical researchers, and the private sector continue to collaborate in developing specific compounds against targets. G-protein receptors

(GPCRs), the ribosome, and enzymes are known to be amenable targets for drug discovery. The *ribosome* is the site of protein synthesis, translating genetic information into a linear polymer through the condensation of amino acids. Ribosome inhibitors are among the most successful antimicrobial drugs to treat infections and are a central focus in the development of drug design and mechanisms of inhibition.

Nearly all therapeutic drugs are enzyme inhibitors that perform target-specific functions, and so they continue to play a great role in developing commercial antibiotics and instrumentations. The technological discovery of biosensors, motivated by the need to measure blood glucose levels for patients with diabetes, revolutionized gaining control and analysis of specific conditions. The pioneering efforts of researchers in mutating and inhibiting enzymes have been a significant focus in contemporary science, as they allow for a genome-wide study of enzyme evolution. Through genome sequencing, a recent class of enzymes called *pseudoenzymes* were identified. Pseudoenzymes are proteins that have lost enzyme activity owing to mutations and are now areas of target or anti-target in drug development.

Antibodies are also areas of drug target and development. Monoclonal antibodies—molecules produced in vitro that serve as substitute antibodies—are currently used to fight against infection, since the immune system protects against foreign substances by producing antibodies. Antibodies are specific proteins that target certain antigens and mark them for destruction. They bind to their target cells and kick-start a cascade system to eliminate specified foreign substances. Monoclonal antibodies have high specificity for an antigen or epitope. To effectively design target drugs, scientists are searching to identify mRNA/protein levels to determine gene expression abnormalities and whether the presence of specific proteins correlates with disease progression.

Imitating the body's natural mechanisms for evading foreign pathogens has challenges. In the pathogenesis of a disease, several complex pathways are often involved and driven by proteins. Promiscuous proteins, so called, can sometimes bind to many ligands within living cells, becoming problematic in immunotherapy.

Cellular Metabolism and Disease

Proteins and enzymes affect systems at various levels—from genomic sequences to physiological states of an organism or cell. The dynamic nature of cellular metabolism and disease is a growing concern among contemporary scientists. *Metabolism* is broadly defined as a series of reactions within cells that produce or consume energy for vital processes. It can be subdivided into three classes: anabolism, catabolism, and waste disposal. Traditional biochemists have identified core metabolic networks within these classes as responsible for nutrient utilization and energy production in humans and organisms. These include fundamental processes such as glycolysis, respiration, tricarboxylic acid (TCA) and urea cycles (Krebs), and oxidative phosphorylation. These fundamental processes take food such as carbohydrates or fats and oxidize them into carbon dioxide and water. Energy is extracted and converted into adenosine triphosphate (ATP), the universal currency in biological systems, which pays the price for biological functions such as motion, active transport of molecules across the cell membrane, biosynthesis, and signal amplification.

Research in metabolism is beginning to shift the focus on cellular metabolism toward examining metabolic processes that accompany human diseases. For example, Facundo Fernández and colleagues showed that metabolomic profiling of serous ovarian cancer in mouse models allowed visualization of tissue heterogeneity and lipid alterations in tumor cells, thus providing a tool for understanding disease origin and progression. The regulatory state of a cell or tissue is necessary to realize, as it is often driven by signaling pathways and specific transcription factors, which can impose themselves upon the dynamics of its metabolic state. Contemporary scientists also ask whether the reciprocal idea can also be possible: whether the metabolic state imposes itself on the regulatory state of the cell. Understanding the relationship between cellular-level interactions and biological systems will help discover key proteins and how they function.

Biochemistry of the Brain

The brain is the most complex organ in the human body, composed of an intricate web of synaptic

connections and neurochemicals necessary for a normal functioning central nervous system. The brain accounts for more than 20% of total oxygen metabolism and represents the largest source of energy consumption in a human. Oxidative DNA damage is considered a significant cause of neurological diseases in humans. Mitochondria, a type of organelle within cells, produce reactive oxygen species (ROS) as a by-product of normal cell activity. Michelle Watts and associates have shown that overproduction of ROS and disruption of mitochondria function are consistent causes of many neurodegenerative diseases, such as dementia. The regulation of tissue metabolite supply and cellular energy metabolism is essential to maintaining healthy cellular and systemic function, although prognostic and diagnostic approaches need refinement. Daniel Silverman and colleagues acknowledge that assessing correlations between brain metabolism with neurodegenerative diseases is difficult, since there are only a few cases in which patients are monitored over long-term clinical follow-up. Prolonged studies on the effects of brain metabolism on inner-neuronal connections among a larger patient pool will give rise to answering fundamental questions about cognitive health. Such studies eventually may provide insights into the relationship between memory and biological information, revealing how the retention and active use of information passing through the nervous system and mental processes impacts the way life is experienced and defined.

Systems neuroscience aims to understand how complex interactions between networks of neurons give rise to perception, behavior, and neurodegenerative diseases. The catecholamine neurotransmitter dopamine (DA) plays an essential role in behavioral and cognitive functions relating to memory, motivation, learning, and movement. The progressive degeneration of neurons of the substantia nigra causes a depletion of DA in areas of the brain and can lead to disorders such as Parkinson's disease. Richard Meade and associates report that the etiology of Parkinson's disease remains unknown, but the pathogenic hallmark of the disease is associated with protein misfolding and the presence of alpha-synuclein, which is itself poorly defined both in structure and function. According to Marek Cieślak and colleagues, dopaminergic agonists, in conjunction with the dopamine precursor L-DOPA, are used to slow the dopaminergic neurons' progressive death. The complex nature of the brain's chemistry and structure continues to perplex scientists, as poor diagnostic tools and the lack of reliable biomarkers complicate the selection of drug targets and candidate molecules. In addition to stem cell treatments, contemporary biochemists are working to produce efficient drug targets that surpass the blood–brain barrier, a highly selective border of endothelial cells.

Many avenues have been explored with drugs on the study of the brain and the biochemical basis for protein misfolding in neurodegenerative diseases. Yet, there remains a vast landscape of unknowns for contemporary biochemists to explore, particularly in how to unify neurochemical processes with recent machine technologies. Furthermore, abstract notions of human existence and reality are fundamentally created by neurochemical processes, yet the reconciliation of these two concepts is far from complete. Their fundamental claims ask the same questions scientists have asked for centuries: How is the external knowledge of the world possible in science? Questions concerning the meaning of life must be accounted for at a molecular level as well as at the universal level. Although yet to be convincing, Freeman Dyson once offered an explanation to reconcile the higher centers of human consciousness and the lower level of atoms: namely, that consciousness is not just a byproduct of neurochemistry but an active agent, forcing atoms and subatomic particles to make choices between one quantum state and another.

The more options, or choices, the higher the probability. This is one way to interpret entropy. Macroscopic properties emerge from microscopic interactions, which are often driven to rearrange molecules in space by entropy. Choice continues to be an interesting theme to consider, whether it concerns the mind, the differentiation of stem cells, or the quantum states of particles. Contemporary scientists have yet to develop a way to explore the notion of intellectual intuition in human life with neurochemical processes. The reality of things in how they are perceived and how this is coextensive with the mind thinking those things may help scientists to understand whether or how intuition is a form of cognition,

and how organisms can receive sensory data and form understanding. Emerging subgenres of contemporary biochemistry—drug-target kinetics, neuroscience, brain–machine interfaces, metabolic disorders, and nutrition—dive into such ideas surrounding the human condition.

Sami Dhungana and Tammy Y. Campbell

See also Biochemistry, 19th Century; Biochemistry, 20th Century; Biophysics, Contemporary; Cell Theory

Further Readings

Cieślak, M., Komoszyński, M., & Wojtczak, A. (2008). Adenosine A(2A) receptors in Parkinson's disease treatment. *Purinergic Signalling*, *4*(4), 305–312.

Doglin, E. (2021). The tangled history of mRNA vaccines. *Nature*, *597*, 318–325.

Dyson, F. (2004). *Infinite in all directions: Gifford lectures given at Aberdeen, Scotland April–November 1985* (1st Perennial ed.). HarperCollins.

Grosjean, H. (2009). *Nucleic acids are not boring long polymers of only four types of nucleotides: A guided tour*. Landes Bioscience, 2000–2013. Available at Madame Curie Bioscience Database. https://www.ncbi.nlm.nih.gov/books/NBK6489

Hughes, J. P., Rees, S., Kalindjian, S. B., & Philpott, K. L. (2011). Principles of early drug discovery. *British Journal of Pharmacology*, *162*, 1239–1249.

Meade, R. M., Fairlie, D. P., & Mason, J. M. (2019). Alpha-synuclein structure and Parkinson's disease: Lessons and emerging principles. *Molecular Neurodegeneration*, *14*, 1–14.

Nelson, D. L., & Cox, M. M. (2017). *Lehninger principles of biochemistry* (8th ed.). W. H. Freeman.

Pace, N. R. (1997). A molecular view of microbial diversity and the biosphere. *Science*, *276*, 734–740.

Sah, S., Ma, X., Botros, A., Gaul, D. A., Yun, S. R., Park, E. Y., Kim, O., Moore, S. G., Kim, J., & Fernández, F. (2022). Space- and time-resolved metabolomics of a high-grade serous ovarian cancer mouse model. *Cancers*, *14*(9), 2262. Published online 30 April 2022. https://doi.org/10.3390/cancers14092262

Silverman, D. H. S., Small, G. W., Chang, C. Y., et al. (2001). Positron emission tomography in evaluation of dementia: Regional brain metabolism and long-term outcome. *JAMA*, *286*(17), 2120–2127.

Watts, M. E., Pocock, R., & Claudianos, C. (2018, June 22). Brain energy and oxygen metabolism: Emerging role in normal function and disease. *Frontiers in Molecular Neuroscience*. https://doi.org/10.3389/fnmol.2018.00216

Bioinformatics

Paulien Hogeweg and Ben Hesper coined the term *bioinformatics* in 1978 to refer to the study of informatics processes in biotic systems. Bioinformatics is the application of information technology to the field of molecular biology and more generally asking biological questions with a computer. There was a time when biology happened mostly in dissection labs and test tubes and under microscopes. Owing to the development of genomic technologies, biology has been transformed from a science in which the human effort was mainly oriented toward data gathering to a science that generates a huge volume of data. Many scientists today refer to the next wave in bioinformatics as systems biology, a new approach to tackling new and complex biological questions. The range of possible applications of bioinformatics is enormous: molecular biology, clinical medicine, pharmacology, biotechnology, forensic science, anthropology, and many other disciplines.

Use of Sophisticated Analytic Tools

Bioinformatics is about searching biological databases, comparing sequences, and looking at protein structures. Systems biology involves the integration of genomics, proteomics, and bioinformatics information to create a whole-system view of a biological entity.

The origins of bioinformatics derive from the existence of biological databases. The first bioinformatics or biological databases were constructed a few years after the first protein sequences of amino acids became available, resulting from work on insulin in 1956. After the formation of the databases, tools became available to search sequence databases. Since these early efforts, significant advances have been made in automating the collection of sequence information. Rapid innovation in biochemistry and instrumentation has brought us to the point where the entire genomic sequence of several organisms is known. Projects to elucidate more than 100 prokaryotic and eukaryotic genomes are currently under way. The Internet is the virtual laboratory in which genomic research is now conducted.

In the early 1990s, scientists at the European Organization for Nuclear Energy (CERN) invented the World Wide Web (WWW) technology on the Internet (the now ubiquitous computer network developed earlier in the United States). The Web was the platform that solved many problems of maintenance, update, access, and integration of databases in molecular biology. In a way, without the WWW technology, the Human Genome Project would not have been possible.

The information archive within each organism is its genetic material (DNA and RNA). The human genome is only one of the many complete genome sequences known. The ENCODE Project (ENCyclopedia Of DNA Elements) has the ultimate goal of developing methods for comprehensive identification of functional regions of the human genome, including coding and regulatory regions.

Roderic Guigó, one of the main characters in the race for the genome project culmination, said that life begins when the nucleotides are arranged in the sequence of the genome. It is the particular order of nucleotides in this sequence, rather than its physical and chemical properties, that dictate the biological characteristics of living beings.

The human genome sequence is now complete, and it is joined with 18 archaea, 155 bacteria, and over 30 Eukarya, as well as many other organelle and viral sequences that are now known.

Key databases containing this information include the archive of nucleic acid sequences known as the International Nucleotide Sequence Database Collection, maintained by GenBank, based at the U.S. National Center for Biotechnology Information, in Bethesda, MD; the EMBL Nucleotide Sequence Database, or EMBL-Bank, based at the European Bioinformatics Institute, in Hinxton, U.K.; and the Center for Information Biology and DNA Databank of Japan, located at the National Institute of Genetics in Mishima, Japan.

The archive of amino acid sequences of proteins is maintained by The United Protein Database, a merger of the databases SWISS-PROT, The Protein Identification Resource, and Translated EMBL. Three systems use Web technology to facilitate access to genomic information via distributed databases: ENSEMBL in Europe and NCBI and Genome Browser, both in the United States.

The implementation of genetic information occurs through the synthesis of RNA and proteins. However, not all DNA is expressed in proteins or structural RNA. Most genes contain internal untranslated regions called *introns*. Some regions of the DNA sequence are devoted to control mechanisms, and for a substantial number of regions, we do not yet understand the function. Proteins, in contrast, show a great variety of three-dimensional configurations with diverse structural and functional roles. Originally, bioinformatics concentrated on the study of the genome, but today it also extends to the study of the proteome, involving the patterns of gene expression and the complex networks of regulatory interactions associated with protein functions.

Alignment: Toward Structure and Function

Aligned sequences of nucleotide or amino acid residues are typically represented as rows within a matrix. In the precomputer era, sequences of DNA, RNA nucleotides, or protein amino acids were assembled, analyzed, and compared manually. Later, these problems were solved using algorithms instead. An algorithm is a complete and precise specification of a method for solving a problem. As soon as computers became available, the computational biologist started to enter these manual algorithms into the machines.

In general, this new discipline called *bioinformatics* could be summarized in the term *alignment*. In bioinformatics, a sequence alignment is a way of arranging the sequences of DNA, RNA nucleotides, or protein amino acids to identify regions of similarity that may be a consequence of functional, structural, or evolutionary relationships among the sequences. Sequence—structure—function: This is now a central concept of both molecular biology and bioinformatics.

Protein sequence alignment has become an essential task in modern molecular biology research. A number of alignment techniques have been documented in the literature, and their corresponding tools are made available as both freeware and commercial software. The two most commonly used programs are the Basic Local Alignment Search Tool (BLAST) and FASTA. BLAST finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches.

BLAST can be used to infer functional and evolutionary relationships between sequences and to help identify members of gene families. BLAST was developed to provide a faster alternative to the earlier FASTA without sacrificing much accuracy.

These programs are an ideal starting point to determine whether a related sequence, or a family of sequences, already exists in a database. The results from these programs will provide evidence of function, utility, and completeness of the gene product. In the 21st century, the sensitivity of sequence searching techniques has been improved by profile-based or motif-based analysis, which uses information derived from multiple sequence alignments to construct and search for sequence patterns.

Núria Pérez-Pérez

See also Informatics

Author's Note: Adapted from Pérez-Pérez, N. (2010). Bioinformatics. In S. H. Priest (Ed.), *Encyclopedia of science and technology communication*. SAGE.

Further Readings

Claverie, J. M., & Notredame, C. (2003). *Bioinformatics for dummies* (2nd ed.). Wiley.

Lesk, A. M. (2019). *Introduction to bioinformatics* (5th ed.). Oxford University Press.

Serra, R. G. (2007). Bioinformática [Bioinformatics]. *Les biotecnologies: Treballs de la Societat Catalana de Biologia* [*The Biotechnologies: Works of the Catalan Society of Biology*], *58*, 11–24.

Biology, Evolutionary

Evolution is the study of the history of life on earth and how it has changed over time. One of the most famous statements on the subject is Theodosius Dobzhansky's 1973 dictum, "Nothing in biology makes sense except in the light of evolution" (*The American Biology Teacher*, p. 125). As we shall see, it is equally true to note that nothing in evolution makes sense except in the light of biology. Over almost three centuries of modern thought on the topic, theories of evolution have been shaped by the current status of knowledge about biological processes, most especially the nature of heredity and hereditary variation. The major key to understanding why evolution theories have changed and continue to do so today is to be found in new technical and conceptual advances in the science of inheritance. We are in a revolutionary period in study of that subject because we can now directly read and modify genomic DNA sequences. Consequently, the topic of evolution currently has more complexity and intellectual ferment than the general public realizes.

Setting the Scientific Stage for Evolutionary Thinking

Although in antiquity there were some precursors to ideas about evolution, the dominance of religious thought about divine creation of all living organisms dominated the Christian and Islamic worlds through the Middle Ages and even survives today among fundamentalist groups of both faiths. The scientific basis for modern ideas about evolution originated in the 18th and early 19th centuries around two related subjects: taxonomy and paleontology.

Although he believed in the fixity of species, the main figure in the origins of scientific classification and the idea of biological relatedness was the Swedish scholar Carl Linnaeus (1707–1778). In 1751, Linnaeus published *Philosophia Botanica*, setting out the bases for taxonomic classification into species and classes of species by defining relationships through particular inherited characteristics plants did or did not share.

The systematic study of fossils and their connection to taxonomy originated with the Frenchman Georges Cuvier (1769–1832). Known as the founding father of paleontology, Cuvier studied vertebrate paleontology in the context of comparative anatomy. He extended Linnaean taxonomy by recognizing that multiple classes could be grouped into higher level classifications called *phyla* and by incorporating both fossils and living species into the classification scheme. The inclusion of fossils of no longer living organisms established the complementary realities of species extinction and survival by distinct but recognizably similar species. While the extinction/survival duality has become critical to evolutionary thought, Cuvier too accepted the fixity of species. He was the first scientific catastrophist and interpreted the fossil record as evidence for recurring

cycles of species destructions in global mass extinction events followed by creation of functionally similar species without an intermediate process of biological transformation.

Descent With Modification, Natural Selection, and *the Origin of Species*

The transition to thinking about the diversification of life as a biological process of acquiring and modifying inherited characteristics, known as Descent with Modification, had its origins in the late 18th and early 19th centuries. The idea that living organisms could undergo hereditary variation and transform themselves to new life forms arose in Britain and France. In Britain, the major figure was Erasmus Darwin (1731–1802), Charles's grandfather. He had translated Linnaeus from Latin into English with two colleagues. In E. Darwin's most important scientific work, *Zoonomia* from 1794, he anticipated the modern theory of evolution:

> Would it be too bold to imagine, that in the great length of time, since the earth began to exist, perhaps millions of ages before the commencement of the history of mankind, would it be too bold to imagine, that all warm-blooded animals have arisen from one living filament, which THE GREAT FIRST CAUSE endued with animality, with the power of acquiring new parts, attended with new propensities, directed by irritations, sensations, volitions, and associations; and thus possessing the faculty of continuing to improve by its own inherent activity, and of delivering down those improvements by generation to its posterity, world without end! (p. 453)

Erasmus Darwin also anticipated survival of the fittest in *Zoonomia* mainly when writing about the "three great objects of desire" for every organism: "lust, hunger, and security…the strongest and most active animal should propagate the species, which should thence become improved" (p. 452).

In France, the first comprehensive statement of evolution by "transmutation of species" came from the French self-described student of "biology," Jean-Baptiste Lamarck (1744–1829), in his 1809 *Philosophie Zoologique*. Lamarck was the first scientist to state explicitly that all species arise from other species and illustrate a branching "tree of life" for animals (p. 649). Lamarck's solution to explaining the apparent direction of evolution was to postulate an inherent drive toward greater organismal complexity ("le pouvoir de la vie," the life force) and environmental inputs by change linked to the use and disuse of characters that distinguish one species from another. The life force postulates an inherent goal-oriented drive or teleology to the evolutionary process, different from Erasmus Darwin's "lust, hunger, and security," and has been denounced by modern evolutionary biologists as unscientific. The second adaptive factor is the well-known "inheritance of acquired characteristics." That process has been singled out by Darwinists to distinguish Lamarck from Charles Darwin as an evolutionary theorist, although Darwin himself explicitly subscribed to the same idea in *Origin of Species* (chapters 5, 6, etc.).

The turn to what is considered by most people to be the definitive theory of evolution came in 1858 with the publication of papers by Alfred R. Wallace and Charles Darwin entitled jointly "On the Tendency of Species to Form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection" (Vol. 3 of *Zoological Journal of the Linnean Society*). The major advances proposed by Darwin and Wallace were (a) discarding any appeals to unexplained internal forces or drives, (b) the assumption of random variation in inherited organismal properties, and (c) the concept of evolutionary advances resulting from differential reproductive success of organisms with improved adaptive characters in a Malthusian struggle for survival and reproduction. In the Darwin–Wallace theory, the purely objective process of *natural selection* for reproductive advantage determined the course of evolutionary change. Darwin explained this idea in the subtitle to his 1859 book, *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*.

Since its introduction, evolution "by Means of Natural Selection" has received broad acceptance as a foundational explanation of the evolutionary process. However, there were basic missing elements in Darwinism, as even Wallace came to call it (A. R. Wallace, 1889, *Darwinism*). Understanding "Descent with Modification" requires a theory of hereditary transmission and of how hereditary

variants altering adaptive traits arise. Darwin and Wallace had neither, because the fundamental biology was missing. Moreover, Darwin made some *ad hoc* assumptions about variation, apparently influenced by the uniformitarian views of his Edinburgh geology professor, Charles Lyell (1767–1849). Darwin quoted Linnaeus's statement "*Natura non facit saltus*" (Nature does not make jumps) and insisted on evolutionary advances by gradual change over long periods of time in his first edition: "If it could be demonstrated that any complex organ existed, which could not possibly have been formed by numerous, successive, slight modifications, my theory would absolutely break down. But I can find out no such case" (*Origin of Species*, 1859, chap. 6, p. 189). Only later, in his 1872 6th edition, did Darwin recognize "sports" as discontinuous "variations which seem to us in our ignorance to arise spontaneously. It appears that I formerly underrated the frequency and value of these latter forms of variation, as leading to permanent modifications of structure independently of natural selection" (chap. 15, p. 395). As we shall see later, the issue of punctuated versus continuous change has come to be one of the key issues in debates over evolution theory for the last 80 years.

The Science of Heredity: Mendelian Genetics, the Modern Synthesis and the Central Dogma of Molecular Biology

While Darwin and Wallace were proclaiming the idea of evolution by natural selection, the unknown Czech monk Gregor Mendel (1822–1884) took the first steps toward filling in our knowledge about the biological basis of heredity. Mendel documented the regular inheritance patterns of discrete dominant and recessive factors determining coat color and seed shape in breeding (hybridization) experiments with peas in 1866. His results contradicted the prevailing idea that somehow the properties of the parents were blended in their offspring. Mendel must have appreciated the relevance of his findings to evolution because he sent his paper to Darwin, but Darwin ignored it, as did the rest of the scientific world for over 30 years. In 1900, at least two botanists searching for discontinuous inheritance, Hugo de Vries and Carl Correns, rediscovered Mendel's paper and confirmed his rules governing specific trait inheritance.

The widespread dissemination of Mendel's rules at the very start of the 20th century led to an explosion of research and the development of the science of Mendelian genetics, which initiated the scientific examination of biological heredity. The Danish botantist Wilhelm Johannsen (1857–1927) introduced the term *gen* (gene) in 1903 to denote any kind of inherited factor that follows Mendel's rules and in 1909 coined the terms *genotype* (an individual's complete collection of inherited factors) and *phenotype* (the individual's visible properties influenced by the inherited genes). Genetic and microscopic analysis of model organisms, such as maize (corn) and the fruit fly *Drosophila melanogaster*, studied by Thomas Hunt Morgan and his colleagues, established the chromosomes in the eukaryotic cell nucleus as the physical structures carrying Johannsen's genes from one generation to the next. Microscopic and genetic analysis dissected the chromosomal basis for sexual reproduction by haploid gamete (sperm/pollen, egg/ovule) formation in meiotic cell divisions. These studies defined homologous chromosome pairing as a critical step in meiosis that has two major evolutionary consequences:

1. Meiotic chromosome pairing facilitated recombinational exchanges between homologues to generate novel combinations of gene variants along each chromosome and thereby produced additional hereditary variation. This process was quickly recognized and incorporated into mainstream evolution theory.
2. Pairing also restricted successful mating to organisms which shared the same chromosome and thus contributed to reproductive separation of closely related species that had different "karyotypes" or structural nuclear constitutions. Mainstream evolutionists resisted highlighting this feature of species differentiation.

In the 1920s to 1930s, the mathematically gifted pioneers J. B. S. Haldane, Ronald A. Fisher, and Sewall Wright developed the quantitative discipline of population genetics, which adapted the discontinuous, gene-based approach of Mendelian genetics to the idea of continuous evolutionary change. This new, highly statistical approach treated the distribution of the ensemble of individual gene variants within populations as a

major source of evolutionary variation: Natural selection of individuals with favorable combinations of alleles would change the "gene pool" and provide significant adaptive advantages. The assertion that "evolution is a change in gene frequencies" was made on more than one occasion. Together with geneticists, naturalists, and less quantitative theorists, the population geneticists promulgated the modern synthesis (ModSyn) of Darwinism and Mendelian genetics in the early 1940s as a comprehensive and, they believed, definitive scientific statement of evolutionary change. This quickly became the mainstream view of evolutionary processes in the 1940s and has remained so in the public awareness over the past 80 years.

The ModSyn retained Darwin's view of random mutations making small differences in phenotypic traits that could be selected over time to produce major adaptive novelties. To these random mutations, the ModSyn added genotypic and phenotypic variability resulting from meiotic recombination, patterns of chromosome transmission, and population-level events. Exceptional occurrences of major genotypic change associated with formation of new species or taxonomic groups were explicitly excluded. In the words of Ernst Mayr, a major figure in the ModSyn: "The proponents of the synthetic theory maintain that all evolution is due to the accumulation of small genetic changes, guided by natural selection, and that transpecific evolution is nothing but an extrapolation and magnification of the events that take place within populations and species" (1970, p. 351).

The early years of biochemical genetics and molecular biology appeared to reinforce the ModSyn. One major unanswered question was, how does genetic information function to specify phenotypic characters? The answer came in 1941 from genetic experiments with the fungus *Neurospora crassa* when George Beadle and Edward Tatum reported that mutants deficient in biosynthetic capacity lacked specific enzymes. Their "one gene–one enzyme" hypothesis was rapidly generalized to "one gene–one protein," and the information content of the genotype was interpreted to consist of the instructions for the synthesis of an organism's complete repertoire of proteins, the molecules with the biochemical abilities needed to construct a phenotype.

The 1944 discovery by Oswald Avery and colleagues that DNA carries genetic information in bacteria and by Alfred Hershey and Martha Chase in 1952 that bacterial viruses inject DNA into their hosts when they reproduce quickly led to working out the DNA double helix structure in 1953 by Rosalind Franklin and Ray Gosling, James Watson, and Francis Crick. The structure of DNA immediately answered two questions about this central molecule of heredity: How it is able to duplicate accurately, as required for genotype maintenance during cell division and growth? and How it could specify the sequence of amino acids in protein molecules by a nucleotide sequence code? Working out the details of protein synthesis and the genetic code for amino acids in the products inspired Crick to articulate the "central dogma" of molecular biology, which states that sequence information flows unidirectionally from DNA to RNA to proteins, sometimes from RNA to DNA, but never from proteins to DNA. This tidy division of labor in cellular informatics also fit very well with the ModSyn view of the genotype as the controlling blueprint for organismal characteristics. It also provided a molecular explanation of random genetic change as DNA replication errors.

Diverging Views Prior to the Modern Synthesis

The most arbitrary feature of the ModSyn—and of Darwin's original theory in *Origin of Species*—was the *ad hoc* assumption that the mutational events generating hereditary variation in the first place must be random accidents of small phenotypic consequence. There was no evidence that this was actually correct. The gradualist assumption was made to give predominance to natural selection as a directing force and it effectively assigned a purely passive role to the organism in the evolutionary process. In discussions of evolutionary change over time, there had always been an ongoing debate between gradualists like Darwin and saltationists, such as Hugo de Vries, who believed that nature does make discrete mutational leaps, which can explain the origins of many diverse life forms. A very different disagreement with evolutionary gradualism came from Russian and American scientists who studied cell biology and

proposed that cell fusions between distinct organisms had generated completely novel hybrids by an abrupt process called *symbiogenesis*, whereby one cell becomes an internalized organelle of the new hybrid. This argument was revived in the 1970s by Lynn Margulis (1938–2011) to explain the origins of eukaryotic cells. Both saltatory and symbiogenetic views were treated as far-fetched fantasies by ModSyn followers.

The most serious contemporaneous intellectual challenge to the ModSyn was posed by the developmental geneticist Richard Goldschmidt in his 1940 book, *The Material Basis of Evolution*. Goldschmidt argued that a fundamental difference exists between the accumulation of localized mutations in particular genes that account for *microevolution* improvements that remain within each species and the occurrence of structural changes to the chromosomes that produce *macroevolution* changes forming new species. Goldschmidt also argued that mutational alterations in embryonic development could produce "hopeful monsters," organisms with major morphogenetic changes (i.e., "monsters") that occasionally would provide adaptive benefits (i.e., "hopeful"), as a way to explain major changes in organismal body plans. The mainstream ModSyn evolutionist community dealt with Goldschmidt's well-documented dissent by a form of scientific excommunication, either totally ignoring him or relegating his views on evolution to what they saw as the clearly preposterous hypothesis of "hopeful monsters." (Today, of course, the hopeful monster idea is seen as the beginning of a major field studying the evolution of multicellular development, widely known as *evo-devo*.)

Problems Arising After the Modern Synthesis

One major problem came out of the research of a plant geneticist credited with being one of the ModSyn founders, G. Ledyard Stebbins. Stebbins studied the role of hybridization between different plant species to generate new organisms for agriculture. He documented the rapid formation of new plant species with new chromosome constitutions following interspecific hybridization and called this process *cataclysmic evolution* (*Scientific American*, April 1951, p. 54). While Stebbins himself considered his observations a special exception to gradual speciation, they exposed a major evidentiary problem for ModSyn arguments. No amount of gradual selection for changes in individual traits ever produced a novel species, but a very different process has been generating them in agriculture for thousands of years (at least since the origins of flour wheat in the Fertile Crescent about 10,000 years ago in Stebbins's 1951 narrative).

The most transformational results challenging the assumptions of the ModSyn came from two markedly different sources in the middle of the 20th century. The first source was Barbara McClintock's studies of chromosome breakage and repair in maize. McClintock found in the 1930s that what were thought to be X-ray-induced "gene mutations" were actually the products of chromosome breakage repaired by the maize plants. She then set out to study chromosome breakage and repair. She documented processes of major chromosome restructuring and unexpectedly discovered genetically mobile "controlling elements" that could insert at new genome locations, where they reprogrammed patterns of expression by adjacent genes. McClintock's results received a hostile reception when she first presented them to a scientific meeting in 1952, but she was ultimately vindicated and received a Nobel Prize in 1983, recognizing that her controlling elements are universal in biology.

The second transformational episode was humanity's worldwide evolution experiment combating bacterial pathogens with antibiotics after World War II. When the ModSyn was formulated, it was still debated whether prokaryotic cells without a nucleus had any genetics at all. But postwar experiments pioneered by medical microbiologists and physicists-turned-molecular-biologists established the science of bacterial genetics. However, the rules were strikingly different from those of eukaryotic Mendelian genetics. Bacteria had so-called *infectious heredity* and could readily transfer DNA from cell to cell, often between unrelated types of cells. When antibiotic resistance emerged as a serious clinical problem in the 1950s and 1960s, it did not originate by localized mutations as had been predicted—and experimentally confirmed in laboratory experiments. The resistance determinants

were carried by transmissible DNA molecules, and these "R-factors" accumulated resistances to multiple antibiotics by acquiring mobile DNA elements analogous to those discovered by McClintock. Bacterial versatility in engineering and spreading novel DNA structures is one of the reasons that multidrug-resistant "superbugs" now threaten our health.

Bacterial genetics was also central to pioneering studies of how cells control genome expression. Such studies began to dissect the "gene" and identified two features of genomes absent from the ModSyn: (1) the presence of regulatory DNA sequence elements needed to control the expression of every genetic locus and (2) the shared use of repeated regulatory elements at multiple loci to establish control networks across the genome. In bacteria, but even more so in complex eukaryotes, these multi-locus networks regulate the expression of protein-coding genes under changing conditions. They are the reason so many phenotypes have complex multifactorial genetic determination. Explaining how multisite networks evolved is a serious challenge for ModSyn gradualism but is a straightforward application of the genome-wide distribution of mobile controlling elements.

What Does Genomics Tell Us?

Now that we are able to read DNA and RNA sequences, we need to formulate a more genomics-based view of evolution. One of the first results of sequencing was the discovery by Carl Woese and colleagues as recently as 1977 that bacteria are not the only prokaryotic organisms on earth. There is an equally ancient, abundant and diverse, but evolutionarily distinct kingdom named *Archaea*. This unexpected discovery ultimately led to molecular confirmation of the symbiogenetic origin about 1.8 billion years ago of the ancestor of all eukaryotes by fusion of a *Proteobacterium* ancestor to the mitochondrion and an *Asgard* archaeon related to the eukaryotic cytoplasm. The eukaryotic nucleus contains sequences of both bacterial and archaeal origin, but how it formed as a separate organ of the cell remains unknown. Genome sequencing has also confirmed the symbiogenetic origins of algae, plants, and other photosynthetic eukaryotes. Thus, the deepest evolutionary divisions in biology that we can explain arose symbiogenetically, not by Darwinian gradualism.

Other features of eukaryotic genome evolution require the formulation of novel theoretical perspectives. The first of these is the discovery of *horizontal DNA transfer* across all phylogenetic boundaries. Examples include bacterial and fungal DNA sequences encoding digestive enzymes in herbivorous invertebrates and sequences encoding eukaryotic regulatory proteins in bacterial pathogens, where they serve to establish the bacterium as an intracellular parasite. Acquiring horizontally transferred proteins that execute specialized functions means that the evolution of key adaptive traits need not be a slow gradual process. The patterns of horizontal transfers are numerous and complex and indicate multiple modes of DNA acquisition. For some researchers in this area, the Tree of Life has transformed into a branching Web of Life.

One of the most important features of complex eukaryotic genomes, the presence of large amounts of repeated DNA sequences, was discovered pre-sequencing in 1968 by a simple biophysical experiment. The presence of abundant DNA repeats in the genome did not accord with the idea of chromosomes carrying unique protein-coding genes like "beads on a string" and demanded an explanation. The response was to call this repetitive DNA "junk," which accumulated because it had "selfish" properties of replicating faster than unique coding DNA in the genes. Richard Dawkins even elaborated a whole philosophy on the basis of *The Selfish Gene* (1976). Such ideas notwithstanding, sequencing the human and other genomes plus functional studies have revealed important adaptive features of repetitive DNA elements (which outnumber protein-coding sequences in our genome by a factor greater than 30:1):

1. The majority constituents of repetitive DNA are mobile genetic elements distributed throughout the genome.
2. Many of these dispersed mobile elements have been co-opted in evolution as regulatory sites for proper expression of coordinated genome networks encoding complex traits such as pregnancy in mammals.

3. Other dispersed mobile elements serve as part of the DNA templates for the formation of noncoding "ncRNAs," which encode no proteins but instead fulfill a rapidly growing list of cell regulatory functions, some of which depend upon the repetitive nature of the sequences they contain.
4. A minor fraction of repetitive DNAs consists of tandem repeat arrays that format the chromosomes for proper replication and distribution at cell division. Some arrays also play roles in formatting the genome for epigenetic control of chromatin structure.

In summary, the repetitive DNA story tells us that genomes do more than encode proteins and that we will always find unexpected features, like regulatory ncRNAs, as we delve deeper into the biological basis for evolution. Repetitive DNA elements are typically the most labile genome components in evolution because the majority of the protein-coding sequences are essential for shared functions and conserved by selection, but the regulatory framework involving DNA repeats can change more freely as species diverge.

Mobile DNA elements have been the subject of comprehensive genetic and biochemical analysis. Although these elements come in many forms (e.g., viruses that integrate into cell genomes, *transposons* that migrate within genomes through DNA intermediates, *retrotransposons* that migrate through RNA intermediates), the proteins and nucleic acids involved in genetic mobility have been thoroughly characterized. In particular, the proteins that execute mobile DNA movements in the genome are nonrandom in their actions on the mobile element itself, and many have been found to be targetable for new insertions to particular sites or classes of sites in the genome. Additionally, various ecological triggers and stresses can activate the mobilization of these DNA elements. Thus, for a significant portion of evolutionary mutability, the changes are nonrandom with respect to both genome location and life history events.

While mobile DNA elements sometimes cause large-scale genome rearrangements, it has unexpectedly been the study of somatic evolution in cancer that has significantly expanded our understanding of major genome restructuring under stress in the 21st century. Whole genome sequencing of various tumor cells revealed highly nonrandom patterns of multiple DNA rearrangements that had occurred in a single chromosome or in a small subset of the 24 human chromosomes. We know from direct observation that these rearrangements often occurred in a single cell division. The concerted rearrangements are triggered by cell division errors and have received colorful names, such as *chromothripsis* ("chromosome shattering") and *chromoplexy* ("chromosome weaving"). Although cancer might be considered an aberrant condition and irrelevant to organismal evolution, the same nonrandom multisite rearrangement patterns were quickly found to occur in the human germ line and in other animals and in plants. The detailed biology of chromothripsis involves the ability of eukaryotic cells to encase "lagging" chromosomes or fragments left off the mitotic spindle in special subnuclear compartments and to recruit hyper-mutagenic DNA repair enzymes to catalyze the complex multisite rearrangements and generate a level of genome variability that would have been totally inconceivable at the time of the ModSyn.

The Beginnings of a Genomics-Based Theory of Evolution

The study of phenomena like noncoding DNA and ncRNA regulation, the evolutionary utilization of mobile DNA elements to form genomic regulatory networks, and the potential for rapid multisite genome restructuring combine to provide a dramatically different picture of what a genome is and how it can change in evolution. Rather than merely a string of protein-coding genes, the genome has become an intricately formatted and structured database encoding proteins and RNAs, which feed back onto the genome to ensure the correct expression of specific organismal characters. Rather than possessing a read-only ROM memory system that changes gradually by random copying errors, we know that evolving organisms have the biochemical and cellular tools to rapidly restructure their read–write genome databases in macroevolutionary transformations into new species with novel

adaptive traits. Unlike the ModSyn and the central dogma, the genomics-based theory envisions the organism playing an active role in changing its DNA in a manner responsive to its life history and ecology. This means that punctuated equilibrium should be the default mode of macroevolution. The main questions for the future will probably have less to do with biochemical details than with cybernetic questions about how complex DNA changes successfully produce adaptively useful biological novelties.

James A. Shapiro

See also Genetic Drift; Germ Theory; Infectious Disease Studies; Life Sciences, Contemporary; Punctuated Equilibrium

Further Readings

Bukhari, A. I., Shapiro, J. A., & Adhya, S. L. (Eds.) (1977). *DNA insertion elements, plasmids and episomes*. Cold Spring Harbor Press.

Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. John Russel.

Darwin, C. (1872). *The origin of species by means of natural selection, or the preservation of favoured races in the struggle for life* (6th ed.). John Murray.

Darwin, E. (1794). *Zoonomia; or, the laws of organic life*. J. Johnson.

Goldschmidt, R. (1982). *The material basis of evolution, reissued (the Silliman Memorial lectures series)*. Yale University Press. (Original work published 1940)

Gould, S. J. (1983). Punctuated equilibrium and the fossil record. *Science*, *219*(4584), 439–440. https://doi.org/10.1126/science.219.4584.439

Gould, S. J. (2002). *The structure of evolutionary theory*. Harvard University Press.

Heng, H. H. (2019). *Genome chaos: Rethinking genetics, evolution, and molecular medicine*. Academic Press.

Huxley, J. (1942). *Evolution: The modern synthesis*. Allen & Unwin.

Margulis, L. (1970). *Origin of eukaryotic cells*. Yale University Press.

Mayr, E. (1970). *Populations, species and evolution*. Harvard University Press.

McClintock, B. (1987). *Discovery and characterization of transposable elements: The collected papers of Barbara McClintock*. Garland.

Sapp, J. (2009). *The new foundations of evolution: On the tree of life*. Oxford University Press.

Shapiro, J. A. (2011). *Evolution: A view from the 21st century*. FT Press Science.

Shapiro, J. A. (2021). What can evolutionary biology learn from cancer biology? *Progress in Biophysics and Molecular Biology*, *165*, 19–28. https://doi.org/10.1016/j.pbiomolbio.2021.03.005

Biophysics, 19th Century

While the discoveries of inorganic chemistry were quickly advancing during the end of the Enlightenment period and the Chemical Revolution, many new avenues of scientific inquiry were being born. In the areas of physics and chemistry in particular, new methodologies such as mechanics (engineering and optics) and chemical production (chemical factories) were quickly laying the foundations for invention and chemical knowledge. In fact, the drive to develop industries contributed to the feverish pursuit of scientific knowledge that could be applied in the industrial arena. Among the subspecialties that began to emerge toward the mid-to-late 1800s was biophysics, the application of physics to investigate biological phenomena. Although the actual origin of biophysics is uncertain, the invention of the battery; increased understanding of animal electrophysiology; and developments in electrochemistry, physical chemistry, thermodynamics, X-ray crystallography, and electromagnetic theory all contributed to the field we now call biophysics.

Origins

Scientific inquiry gradually matured through the ages, documented as far back as the ancient Egyptian and Babylonians, though less organized before Aristotle (384–322 B.C.E.) defined the scientific method. In science, perhaps more so than any other school of thought, knowledge builds upon knowledge. A seemingly inconsequential discovery could later inform a different investigation, bringing forth a clarification that would not have been possible otherwise. Biophysics was born from such disjointed discoveries in early experiments that had a narrow focus but later informed or provided a template for a methodology that could be modified to test a new theory. Examples of

these types of experiments were the discovery of bioluminescence jellyfish and fireflies in studies by Athanasius Kircher, Isaac Newton's foray into the investigation of the nature of muscle movement (1687), and John Walsh's discovery, with Benjamin Franklin, of the electrical discharge from a torpedo fish or electric ray (1773).

Later in the 18th and 19th centuries, scientists became more interested in elucidating the force that produced movement in animals. The scientific method and experimental work concerning the nature of electricity and its role in animal life allowed scientists to ask questions that had not been possible to evaluate before. Most of the experiments conducted during this time were performed by professors of physics who were interested in describing biological phenomena. One such physicist, Abbé Giovanni Beccaria, professor of physics in Turin, performed experiments in which animal muscles received electrical stimulation. An English surgeon and professor of anatomy and physiology in Göttingen, Albrecht von Haller, hypothesized that electrical matter and "animal spirits" were the same.

Electrochemistry

Understanding the electrical component of muscle stimulation was facilitated by the experiments of Italian physiologist Luigi Galvani in the late 1700s. During an experiment, he had skinned a frog leg and was preparing to rub the leg to produce static electricity. His assistant probed the frog leg with a scalpel that unknowingly had a static charge. When the scalpel touched the leg, a spark of static electricity was produced, and the frog leg jumped. At the time, it was believed that electrical current within the animal produced a contraction and that the source was concentrated in the animal's pelvis.

In an effort to clarify the cause of the frog leg's response to the application of static electricity, Alessandra Volta, an Italian contemporary of Galvani, performed several experiments to reproduce the effects witnessed by Galvani. Volta believed that the frog leg only served as an electrical detector to the external electrical source. Meanwhile, those who sided with Galvani's explanation that the electrical source causing contraction was in the animal tissue, several experiments were conducted that appeared to confirm Galvani's hypothesis. The animal muscle contracted when placed in contact with charged metals. However, Volta performed experiments using his invention of the voltaic pile, so called, a series of alternating silver or copper and zinc discs that had cloth or cardboard soaked in brine placed between the discs and arranged in a circuit, connected to the frog leg. He demonstrated that the contraction was produced by the contact with dissimilar metals, and the frog leg was only behaving as a detector.

Volta's experiments using the voltaic pile resulted in the invention of the early electrical battery. This invention was immediately recognized by the scientific and industrial community for potential uses, which seemed more important than understanding the electrical potential in animals. Study in the area of electrophysiology did not regain popularity until 1827 when the galvanometer, or an instrument for detecting electric current, was built by German electrophysiologist Emil Du Bois-Reymond. This instrument proved able to detect small potential differences across nervous membranes. A new branch of science, *neurophysiology,* developed from the data that were produced by the galvanometer. It is not unusual for physiology and biophysics experimentation to overlap yet remain distinct fields of study.

In addition to neurophysiology, biophysical experiments investigating diffusion gradients and osmotic pressure in living organisms became an important avenue of study. *Osmotic pressure* is the pressure that develops in a solution of salts when separated from pure solvent by a semipermeable membrane. This phenomenon was first described by Abbé J. A. Nollet, a French physicist. The semipermeable membrane was discovered by French scientist René Dutrochet in 1828. The understanding of diffusion in biological systems was furthered by botanist W. F. P. Pfeffer, who obtained the first quantitative measurements. Adolf Fick then took the experimental data generated by Pfeffer and merged it with his understanding of the laws governing the flow of heat. He published the first biophysics textbook, *Die mediziniche Physik* (Medical Physics) in 1856.

Additional applications of electrophysiology to biophysical experimentation were devised by the German physiologist Hermann von

Helmholtz (1821–1894), who studied both nerve conduction and the physiology of hearing. In 1849, he studied the speed at which an electrical signal was carried along a nerve fiber. Before this measurement, scientists were under the impression that nerve signals traveled faster than anything their available equipment could measure. He utilized the galvanometer developed by Du Bois-Reymond to act as a timing device. Helmholtz attached a mirror to a needle connected to a sciatic nerve of a frog leg, which reflected a light beam across the room to a scale. He was able to detect transmission speeds between 24.6 and 38.4 meters per second. This information enabled scientists to begin to develop better technologies that could capture transmission speeds of nerve conduction.

Developments in Physical Chemistry

Important to the later applications of molecular biophysics (most commonly practiced today) are the atomic theory, the first and second laws of thermodynamics, and advancement in reaction kinetics, which allowed scientists to describe the structural and functional properties of enzymes and proteins. In the early 1800s, the English physicist John Dalton proposed his atomic theory, which stated that matter was composed of discrete, indivisible atoms. It was later corroborated by Amedeo Avogadro in 1811, by making the distinction between atoms and molecules of a specific substance. At the same time, advancements in understanding about heat, energy, work, and temperature were made. The first law of thermodynamics, which states that heat and work are interchangeable, was proposed by the German physicist Julius Robert von Mayer in 1842. Almost 10 years later, the second law of thermodynamics was presented by the German mathematical physicist Rudolf Julius Emanuel Clausius and William Thomson, saying that when energy is transferred or changes from one form to another in a closed system, disorder (entropy) increases.

Chemical kinetics studies were bolstered by the developments made in the area of gas properties and laws. Between 1860 and 1875, Austrian physicist Ludwig Boltzmann and British physicist James Clerk Maxwell demonstrated that the ideal gas law could be explained by the kinetic theory of matter. Subsequent analysis in chemical kinetics and the formulation of the laws of chemical equilibrium (e.g., Le Chatelier's principle) were based on the kinetic theory of matter. It was determined that the main factors that influence a reaction rate are the concentration, physical state, and temperature of the reactants, and whether there is a catalyst present in the reaction.

In 1864, Peter Waage and Cato Guldberg formulated the law of mass action, which states that the speed of a chemical reaction is proportional to the amount of the reactants. Also, it was found that in consecutive reactions, the kinetics are derived by the rate-determining step. This step represents the slowest step in the series of reactions and can be identified by the rate equation. Knowing the rate-determining step helps to optimize chemical reactions such as catalysis and combustion. This information was important to being able to test the thermodynamics of enzymes and proteins, which enable biophysicists and biochemists to characterize the function of both.

Advancements in optics were also important to the development and characterization of macromolecules, such as proteins that are crystallized (though this application was not practiced until the 20th century). In 1815, a French polymath named Jean-Baptiste Biot studied plane-polarized light (polarized light results from an electromagnetic wave being filtered so that it oscillates in just one plane of vibration rather than in several, unlike nonpolarized light such as light from the sun or from a candle flame); he focused polarized light on organic substances and noticed that the light could be rotated clockwise or counterclockwise, depending on the optical axis of the molecule. This knowledge contributed to determining the structure of proteins and enzymes by X-ray crystallography, which is a very important role that biophysics has played in the 20th and 21st centuries.

Late 19th Century

An increase in the understanding of thermodynamics and molecular structure helped to pave the way for future experiments in structural biology, which draws upon many physical and biological

disciplines to determine the structure of proteins. Scientists were particularly interested in studying the life-sustaining chemical reactions that happened in living cells. Among the important developments that helped to better understand molecular structures were stereochemistry and thermodynamics, as well as advancements in chemical theories.

In 1874, Dutch physicist Jacobus Hendricus van't Hoff established the foundation of stereochemistry with his experiments on optically active carbon compounds and was able to describe three-dimensional and asymmetrical molecular structures. *Stereochemistry* is the study of the spatial arrangement of atoms in a molecule and detection of a change in that arrangement when the molecule is manipulated (e.g. heated). A short time later, he began to apply the thermodynamics of chemical reactions, leading to defining the order of reactions. *Thermodynamics* is a component of physical chemistry that seeks to describe the physical properties of macroscopic systems—for example, enzymes and proteins. These physical properties consist of sensitivities to temperature, pressure, and volume, which are the classic parameters that are measured. In addition to these parameters, variables such as density, specific heat, compressibility, and the coefficient of thermal expansion can be evaluated to determine the full spectrum of physical characteristics and how a substance behaves in its environment.

Other scientists, such as Comte Claude Louis Berthollet, studied the rate and reversibility of reactions, and Benjamin Thompson, an American physicist, tried to elucidate the mechanical equivalent of heat. In 1875, Josiah Willard Gibbs presented the phase rule, which applies thermodynamics to heterogeneous substances. Swedish chemist Svante August Arrhenius produced the theory of electrolyte dissociation, later called *Arrhenius's theory*, in 1889. In 1906, the German physical chemist Walther Hermann Nernst produced Nernst's theorem, subsequently referred to as *the third law of thermodynamics*, which states that the entropy of a system at absolute zero is a well-defined constant. He also contributed to the advancement of physical chemistry by describing physical properties, molecular structures, and reaction rates of macroscopic molecules.

X-Ray Crystallography

One of the most important scientific modalities for biophysics of the 19th century (and beyond) was X-ray crystallography. Crystallography was invented to analyze crystals and their shapes in the 18th century. In 1784, René Just Haüy discovered that a crystal's face could be described in terms of stacking patterns of blocks of similar shapes and sizes. In 1839, William Hallowes Miller was able to assign each facet of the crystal a unique three-integer label. Eventually, it was proven that all crystals are a regular three-dimensional array of atoms and molecules. By the 19th century, a complete catalog of possible symmetries of a crystal were constructed by Johan Hessel, Auguste Bravais, Evgraf Fedorov, Aruthur Schönflies, and Willam Barlow. In the 1880s, a few crystal structures had been elucidated, though the structural information was not completely accepted due to the lack of techniques available to verify it (X-ray).

In 1895, Wilhelm Conrad Röntgen discovered the X-ray, though little was known about this form of electromagnetic radiation. Scientists were uncertain of the characteristics of X-rays but were able to reconcile some of the electromagnetic properties of X-rays according to the wave model of light (Maxwell theory of electromagnetic radiation). The visible wavelength of X-rays was determined by Arnold Sommerfield to be approximately 1 angstrom. By 1905, Albert Einstein had discovered that X-rays were not only waves of electromagnetic radiation but also photons and exhibited particle-like properties. It was observed that when an X-ray is directed toward a crystal, the electron density in the crystal will cause a scattering of light that is characteristic of the arrangement of the electrons within the structure. Based on these scatter patterns, a structure can be determined mathematically. However, this approach to solving the structure of a molecule did not become prevalent until 1912.

Electromagnetic Theory

Electromagnetic theory is important to the field of biophysics because it provides another avenue to observe crystal structures. Up until the late 1800s, electricity and magnetism were considered separate forces. However, when James Clerk

Maxwell published his 1873 *Treatise on Electricity and Magnetism*, he presented the observation that interactions of positive and negative charges seemed to be moderated by a single force. There were four points to his argument: (1) Electric charges attract or repel each other with an inversely proportional force multiplied by the square of the distance between them (opposite charges attract each other); (2) magnetic poles attract or repel based on their charge and occur in pairs; (3) an electric current in a wire generates a circular magnetic field around the wire, and its direction followed that of the current; and (4) a current is created in a loop of a wire when it is pulled away or toward a magnetic field, and follows the direction of the current. This observation helped to advance our understanding of electromagnetic forces, which led to the development of nuclear magnetic resonance (NMR). NMR allows for determination of the structure of solid materials at high resolution. Additionally, elucidation of intermolecular forces generated by the momentum of an electron's movement was made possible. Scientists in the 20th century were able to employ technological advances such as NMR in determining the structure of organic molecules or substances difficult to crystallize, compared with that observed by X-ray crystallography.

Mandy M. McBroom

See also Biochemistry, 19th century; Chemistry; Medicine, Enlightenment; Physics, 19th century; Physics, Quantum Theory.

Further Readings

Bragg, W. H. (1908). The nature of X-rays. *Nature*, *78*(2035), 665.

Brouwer, D. H. (2008). NMR crystallography of zeolites: Refinement of an NMR-solved, crystal structure using ab initio calculations of ^{29}Si chemical shift tensors. *Journal of the American Chemical Society*, *130*(20), 6306–6307. https://doi.org/10.1021/ja800227f

Glaser, R. (2012). *Biophysics: An introduction* (2nd ed). Springer.

Glynn, I. (2010). *Elegance in science*. Oxford University Press.

Lewis, G. N., & Randall, M. (1923). *Thermodynamics and the free energy of chemical substances*. McGraw-Hill.

Macholl S., Borner, F., & Buntkowsky, G. (2004). Revealing the configuration and crystal packing of organic compounds by solid-state NMR spectroscopy: Methoxycarbonylurea, a case study. *Chemistry*, *10*(19), 4808–4816. https://doi.org/10.1002/chem.200400191

Patterson, E. C. (1970). *John Dalton and the atomic theory*. Doubleday.

Biophysics, 20th Century

Biophysics is a branch of science that integrates several disciplines (physics, chemistry, and bioengineering) to study biological systems. It is applied to the molecular scale, in determining structures of nucleic acids and proteins; to whole organisms, and even ecosystems. One advantage of biophysics is the ability to set up experiments that elicit exact hypotheses and interpretations. The biophysicist is likely to be the first to employ new technologies (e.g., NMR, nuclear magnetic resonance) to apply complex physical theory to the problems posed in biology.

Another advantage is that the application potential of biophysics to the natural world seems endless. A challenge for the student of biophysics, however, is the disparate opinions on what exactly biophysics is. There is scientific equipoise on the actual definition of biophysics, as it spans so many different applications.

However, despite the disagreement among biophysicists on what biophysics really means, there are areas of application that can be used to categorize the contributions that biophysics has made in the 20th century, leading into the 21st century. Astrobiophysics studies the influence of astrophysical phenomena upon life on planet Earth. Medical biophysics uses the advances in biophysics methodologies for diagnostic and therapeutic purposes. Molecular biophysics is the most rapidly evolving version of biophysics; it characterizes biosystems, molecular structures, dynamic behavior of proteins (single molecules up to supramolecular structures), and viruses. Detlev Wulf Bronk, a renowned American physicist and administrator, established

biophysics as a distinct scientific discipline toward the middle of the 20th century.

Over the decades that followed, as technological advances were made, so too did the selection of instrumentation and equipment help to overcome a difficulty in experimentation. Some of the biophysical methods used to study biological systems include X-ray diffraction, chromatography, macromolecular crystallography, proteolysis, NMR spectroscopy, electron paramagnetic resonance, cryo-electron microscopy, multi- and small-angle light scattering, ultrafast laser spectroscopy, dual polarization interferometry, and circular dichroism.

Notable Inventions

One of the most important tools used to determine the structure of a macromolecule is *X-ray crystallography*. By the early 1900s, physicists were beginning to understand the possible scope this technique could have; most certainly, it was the best tool for gathering atomic data on crystals. These data could be mathematically plotted and analyzed for determining the shape of a substance, as long as the substance could diffract. Not just any substance can diffract light. Certain types of surfaces are better for diffraction than others (e.g., jagged edges diffract better than smooth edges). In 1924, while at Cambridge University, British physicist Jon Bernal solved the structure of graphite and worked on solving the structure of bronze. One of his more distinct and wide-reaching inventions was an *X-ray spectrogoniometer*. A goniometer is an instrument that allows an object to be rotated and measured at different angles. For X-ray crystallography, goniometers are used for measuring angles between crystal facets as well as rotating the sample.

NMR spectroscopy is a technique that measures the energy that resonates in a specific frequency within a magnetic field. It was first developed in 1938 by Isidor Rabi, who later won a Nobel Prize for this invention. It is used to collect physical, chemical, electrical, and structural specifics about molecules. It can provide information about topology, dynamics, and the three-dimensional structure of a molecule in solution or in a solid state. One of its strengths is that it is selective in regard to types of atoms studied within a molecule.

Molecular modeling is the array of methods that are used to investigate the dynamics, surface properties, and structure of biomolecules. Beginning with the modeling of the double helix of DNA (deoxyribonucleic acid, which carries genetic instructions for the development of all known organisms and many viruses), many other molecules and their properties have been modeled such as protein folding, enzyme catalysis, protein stability (thermostability assays), and producing conformational variants (open, intermediate, or closed states).

Patch clamp and single-channel recordings enabled both electrophysiologists and biophysicists to study multiple channels or single-ion channels and determine the effects of these channels on different stimuli. The stimuli could be anything from protons to chemicals, or ligands that are introduced via a special apparatus that controls the rate of flow over a cell. The electrical potential is then measured in amperes and assessed for length of the channel being in open or steady-state conditions. Using these data, dose–response curves can be generated for a particular drug that is introduced to the receptor on the cell's surface.

The *Ramachandran plot* was created in 1963 by biophysicist G. N. Ramachandran and colleagues in order to visualize backbone dihedral angles against amino acids in a protein structure. Today, the plot is generated using X-ray diffraction data for a sample protein and compares the distribution of data points from the sample to that of a known protein structure (structure validation). This task is more easily accomplished if there is a known structure within a family of proteins.

In the late 20th century, *condensed phase single-molecule fluorescence spectroscopy* was devised by W. E. Moerner and Lothar Kador. It uses the fluorescence of a molecule to derive measurements about its structure, position, and its environment. Its strength is that it allows the investigator to obtain information without losing important values due to ensemble averaging. When performing recording techniques, such as patch clamp electrophysiology, the signal that is obtained from the recording of many molecules at the same time is considered the average property of the molecule's dynamics. By focusing on one molecule, it is possible to obtain two-state trajectories.

Nerve Impulse

Following the animal experiments conducted by late-19th-century biophysicists, a prominent English biophysicist, Archibald Hill, performed significant work in the area of nerve conduction and enzyme kinetics. He is known for expanding Michaelis–Menten kinetics and authoring the Hill coefficient; the latter characterizes the binding cooperativity for enzymes and molecules. His work in heat generation by contracting muscles was conducted in 1920. He focused on studying the effects of electrical stimulation to muscle, the mechanical efficiency of muscle, muscle recovery through energy processes, interaction between oxygen and hemoglobin; and produced drug response curves from muscle. He was able to show that electrical activity was most robust in the presence of oxygen, and that during recovery the amount of oxygen required over resting was equivalent to 20% of the lactic acid produced during exercise. This amount of oxygen was needed to resynthesize lingering lactic acid into glycogen.

After World War II, the development of electronics vastly expanded, including radar technologies and the use of nuclear reactors in peacetime to produce large quantities of radioactive isotopes. Alan Hodgkin and Andrew Huxley utilized these advances and data from earlier studies in animal physiology (e.g., Julius Bernstein's membrane theory of nervous conduction), to produce a mathematical model that explains how action potentials are initiated and propagated in neurons (1952). They received the Nobel Prize in Physiology or Medicine in 1963 for such a groundbreaking and important work. Another post–World War II invention, the *electron microscope*, allowed the description of a muscular contraction at the structural level. Many advances in X-ray technology and electron microscopy have made it possible to observe and characterize many molecules that are involved in muscular contraction.

Sensory communication is another area that biophysics has impacted by technological advances that allow scientists to be able to measure stimulus, namely computerized equipment that is far more sensitive than earlier instruments. In general, quantitative analyses of sensory responses are difficult due to the multiple cell types involved. Before the advent of the computer in the 1970s, Georg von Bekesy, a Hungarian physician, performed experiments with cadavers on the human ear. He dissected the inner ear of the cadaver while leaving the cochlea partially intact. Then he used strobe photography and silver flakes as a marker (which the black-and-white film easily detected) and was able to witness the surface wave of the basilar membrane when stimulated with sound vibrations.

Molecular Biophysics

American scientist Alexander Hollaender founded the study of radiation biology and produced early evidence of nucleic acid as the genetic material. In 1939, Hollaender published data showing that the absorption spectra of nucleic acids of a mutated ringworm fungus was the same as isolated nucleic acids, and therefore concluded that nucleic acids must be the building blocks of genes. However, his discovery was not immediately accepted until other experiments corroborated his hypothesis. It is with this information that a spectrophotometer can be used at a certain wavelength (260 nm) to detect and quantify the amount of DNA in a sample. This type of technology is used routinely in biological laboratories today.

Robert Brainey Corey, American biochemist, is mostly known for his role in the discovery of the α-helix and the β-pleated sheet with Linus Pauling. In 1950, Pauling and Corey discovered that secondary conformations of peptides were arranged in such a way that their hydrogen-bonding capacity of the backbone NH (amine) and CO (carboxyl) groups was facilitated. First described was the α-helix, a rod-like shape that was composed of a tightly coiled backbone in the inner part with the side-chains extended outward forming a helical pattern. The coil is stabilized by the NH and CO groups of the primary structure. All of the main-chain NH and CO groups are hydrogen bonded. The angles were described in terms of angstroms (distance from one atom to another) and pitch (screw-shaped pattern). Later, the duo discovered another facet of secondary structure, the β-pleated sheet. This pattern differs markedly from the α-helix, taking an accordion-like appearance. In the βeta sheets, βeta strands, which are polypeptide chains of about three to 10

amino acids long, connect laterally by two to three hydrogen bonds in the backbone.

In 1953, biophysicist Francis Crick and biochemist James Watson determined the structure of DNA, one of the most important discoveries of modern science. Advancements in the use of X-ray diffraction for the determination of crystal structure offered the opportunity to investigate the molecular structure of DNA and RNA. Together the two utilized Chargaff's rule (which stated that for every adenine (A) there is a thymine (T) and for every cytosine (C) there is a guanine (G)) in conjunction with the X-ray crystallization data from Rosalind Franklin and Maurice Wilkins, to determine that the DNA structure was a double helix. There have been some minor changes in the model since it was published, but four of the major features still hold true: (1) DNA is a double-stranded helix, (2) most DNA helices are right-handed, (3) the double helix is antiparallel, and (4) the DNA base pairs are connected through hydrogen bonding, as well as the outer edges of the exposed nitrogen-containing bases. The methods used by Watson and Crick to determine the DNA structure are now used for model building in biophysics.

Donald L. Caspar, an American scientist, made significant contributions in structural biology, X-ray, neutron and electron diffraction, and protein elasticity. In 1962, Caspar and Aaron Klug introduced the concept of quasi-equivalence to account for the arrangement of proteins on the surface of icosahedron virus particles. Caspar–Klug theory has played an important part in shaping the subsequent study of viruses and other macromolecular assemblies. The original concept was based on electron microscopic studies and has now been refined to take into account the atomic resolution structure of viruses, and other details of protein–protein interactions that crystallography has elucidated. Quasi-equivalence continues to be an important component of the philosophical basis for how we think about macromolecular assemblies.

In the 1980s, Adriaan Bax, a molecular biophysicist, developed the field of biomolecular NMR spectroscopy, and he is responsible for many of the standard methods in the field. He pioneered the development of triple resonance experiments and technology for resonance assignment of isotopically enriched proteins. He was also heavily involved in the development of using residual dipolar couplings (dipole–dipoles) and chemical shifts for determining RNA and protein structures. Recent work focuses on the roles of protein in membranes. He is one of the most cited and important scientists in the field of biomolecular NMR research. Direct dipole–dipole coupling is useful for molecular structural studies. The estimation of the coupling allows for a direct spectroscopic translation of the distance between nuclei and the geometrical form of a molecule. However, in isotropic solution, dipole–dipole couplings are absent as a result of rotational diffusion.

Gary Ackers, Emeritus Professor of Biochemistry and Molecular Biophysics at Washington University, focused on the thermodynamic linkage analysis of biological macromolecules, addressing the molecular mechanism of cooperative O_2 binding in human hemoglobin, beginning in the early 1970s. Briefly, cooperative binding occurs when a substrate binds to an enzymatic subunit (often protein molecules are a series of homogenous or heterogenous subunits that connect on the cell membrane to form one unit) and then activates the remaining subunits. Cooperativity can be positive, negative, or absent. Positive cooperativity results when oxygen binds to hemoglobin. The ratio of oxygen to ferrous iron in a hemoglobin molecule is one-to-one. Deoxy-hemoglobin has a low affinity for oxygen, but when one molecule of deoxy-hemoglobin binds to a heme site, the molecule's overall affinity for oxygen increases. Once this occurs, more oxygen molecules can bind to the hemoglobin molecule. Conversely, in negative cooperativity, when a ligand binds to a protein, the affinity for the ligand will decrease. The affinity for a ligand can be measured by plotting the amount of binding over the course of time.

In standard gel electrophoresis, which separates DNA molecules, there was an issue in that large DNA molecules (larger than 20 kb) would migrate together and it was difficult to obtain adequate resolution or detect the correct size DNA bands. Charles Cantor, an American molecular geneticist, along with David Schwartz, developed pulse-field gel electrophoresis for very large DNA molecules. This technique works by using an alternating

voltage gradient to resolve larger DNA molecules. Opposed to standard gel electrophoresis that runs unidirectional, the pulse-field gel's voltage is pulsed in three directions. The process takes longer than standard gel electrophoresis. Other contributions by Cantor and his lab at Boston University include methods for separating DNA molecules, increasing sensitivity for detection of proteins, and enabling the study of complex protein and nucleic acid relationships.

Pamela J. Bjorkman, an American biochemist, focuses her study on the three-dimensional structures of proteins related to Class I MHC, or major histocompatibility complex, which are proteins of the immune system. Bjorkman and her students study protein interactions involved with immune recognition, using techniques such as X-ray crystallography and confocal or electron microscopy.

Biological Membranes

Membrane biophysics is study of biological membrane structure and function by mathematical, physical, computational, and biophysical methods. In tandem, the data from these methods can be used to generate phase diagrams of various membrane types, describing their thermodynamic characteristics. In contrast to membrane biology, membrane biophysics provides quantitative information and accurate modeling of a variety of membrane topological presentations, such as protein–lipid coupling, lipid and cholesterol flip-flop rates, and lipid raft formations. Also, elasticity of the biological membrane and its intracellular effects can be studied.

One of the methods for visualizing a membrane surface is by attaching a radioactive compound that can either be introduced by insertion by recombinant techniques or added to a solution to bind to a specific receptor. In the early 20th century, the Danish physiologist August Krogh set the foundation for using radioisotopes in molecular biology. His understudy, Hans Ussing, described the pathway that allows the transport of ions across membranes to be explained. He is responsible for the discovery of the active transport of water and ions across a cell membrane. However, the complete molecular mechanism for how this is achieved is still poorly understood.

Despite the great progress in understanding mechanisms of how molecules are assembled into larger biomolecules, there remains a deficit in understanding of exactly how molecules are assembled into biological membranes. The way in which a membrane is constructed is extremely complex and requires better technology to gain useful insight.

Structural Biology

Structural biology is a branch of molecular biology that combines the disciplines biology, biochemistry, and biophysics. The main objective is to obtain a molecular structure of macromolecules, such as nucleic acids and proteins. In addition to structural information, functional data are elucidated through experiments such as electrophysiology. The proteins are usually evaluated in their most complex structures (tertiary and quaternary) and native states because proteins are functional in these states. When a protein is not in these states, inactivation occurs. A great deal of structural biology is centered on determining the structure of receptor proteins that function within a specific disease pathway or has an important physiological role. Often these proteins are crystallized in a specific condition and ligands (or specific compounds that bind to the receptors) are added that will allow the structural biologist to observe the changes in conformation that take place while the ligand is bound. If available, the ligand is a known drug that has a high affinity to the receptor that will promote a physiological response or block it.

Between 1982 and 1985, the German biochemists Johann Deisenhofer, Hartmut Michel, and Robert Huber determined the first crystal structure of an integral membrane protein, the photosynthetic reaction center. It is a membrane-bound complex of proteins and cofactors that play an important role in photosynthesis. They achieved this through X-ray crystallography to determine the exact arrangement of more than 10,000 atoms that make up the protein complex. Not only did their research increased the general understanding of the mechanisms of photosynthesis, but seeing a structure solved for an integral protein was important to help other biophysicists piece together new protocols for their own projects.

Helen M. Berman, an American biophysicist, is the director of the RCB Protein Data Bank. Her biophysics work includes structural analyses of protein–nucleic acid complexes and the role of water in molecular interactions. The Protein Data Bank is an online repository for three-dimensional structural data of large biomolecules (proteins and nucleic acid). The database houses hundreds of protein structures. This online repository is the premier source of data for structural biologists. Most of the data commonly reported with the structure is X-ray diffraction or NMR data. Scientists from around the world are able to submit their structures and are able to download structural data for comparison with their own models. This enables scientists to adjust their models or raise questions about existing structural models.

Steven M. Block (Stanford University) pioneered the use of optical tweezers, a technique developed by Arthur Ashkin, to study biological enzymes and polymers at the single-molecule level. Optical tweezers are instruments that use a highly focused laser beam that attracts or repels according to the refractive index mismatch. This technique has been highly utilized in modern biophysics. The ability to control single molecules with nanometer precision and piconewton accuracy has allowed scientists to directly test the effects of DNA binding on DNA properties and to test the interactions involved in these systems. Understanding the complex mechanisms involved in biological processes is very important and has the potential to improve therapeutics and other molecular biology methodologies.

Related to optical tweezers and single-molecule fluorescence, visualization techniques such as scanning-force microscopy have allowed biophysicists to study the structure and function of nucleoproteins. Carlos Jose Bustamante is an American molecular biologist who pioneered use of optical tweezers and scanning-force microscopy. His laboratory focuses on developing methods of using single molecules to characterize the elasticity of DNA, to induce unfolding of individual protein molecules, and to probe the mechanical behavior of molecular motors.

Axel T. Brünger is known for developing a computer program used for solving structures based on X-ray diffraction data or solution NMR data. The program is widely used because of its various features. The program is a major extension of the program his team developed called *PLOR*, which came out in 1987. The program utilized a method called *simulated annealing* to refine X-ray crystal structures. It was a manually intense process. When Brunger introduced simulated annealing crystallographic refinement in his 1987 paper, the time to refine crystal structures was significantly reduced and it had a tremendous impact in the crystallographic community.

In some families of proteins, such as the G-protein-coupled receptors (GPCRs), the majority of receptors' structures have been solved. These receptors sense molecules outside the cell and activate signal transduction pathways, which lead to cellular responses. They are difficult to crystallize because there are seven transmembrane receptors that pass through the cell membrane. There are six classes of GPCRs that are homologous and functionally similar. Much information has been obtained regarding their roles in signaling and ligand-binding.

Computational Biology

Predicting the three-dimensional structure of a protein from its amino acid sequence can be a daunting task, especially when there are no homologous structures to compare a new protein structure with. Structure prediction is diametrically opposite from protein design. Protein prediction is one of the most important goals of structural biology. By knowing the structure, it is possible to design drugs that will have greater specificity with the receptor target, thereby causing less side effects overall. Also, enzymes can be created that have biological or industrial uses. In order to help remove some of the obstacles in protein prediction, a number of bioinformatic programs have been created that allows the user to do more tasks at a time.

Some example of molecular design software are Pymol, AMBER, Ascalaph Designer, BOSS, DOCK, Firefly, Maestro, SCIGRESS, SPARTAN, and TINKER. Pymol is an open-source molecular visualization program that was created by Warren Lyford DeLano. It is most commonly used to produce high-quality three-dimensional images of an array of biomolecules. It is one of the only

open-source visualization programs available to structural biologists. It is programmed in the Python programming language.

David Baker is the principal investigator of the 60+ member Baker laboratory. He and his team developed the Rosetta algorithm for *ab initio* protein structure prediction, which is associated with the computing project called *Rosetta@home* and *Foldit*. Members of the Baker group also focus on protein design; they are the first group to have designed a protein, named Top7, with a unique fold. Baker received the 2008 Sackler International Prize in Biophysics for his work in protein folding.

Newer Applications of Biophysics

Astrobiophysics

Astrobiophysics is another emerging application of biophysics that utilizes data obtained from many diverse fields of study, such as biochemistry, paleontology, atmospheric science, and astronomy. Dr. Brian Thomas and his research team at Washburn University (Kansas) apply these principles to study how the Earth is affected by radiation from space. There are several projects underway, most notably the study of the impact on Earth of a nearby supernova. The project uses computational modeling to observe the combined effects of atmospheric ionizations and radiation impact on land and sea life. The presence of the isotope iron-60 in sediment cores bears witness to the fact that a near-Earth supernova occurred within the last 2 million years. The discharge from the event was close enough to change the atmospheric UV levels and possibly modified the Earth's climate. Their work continues in hopes of gaining insight into what sort of impact could a future near-Earth supernova pose to life on the planet.

Medical Biophysics

A fairly new (2013) application of the biophysics discipline is in the field of medicine. There are many classic biophysicists who use methods for determining structures of molecules, membrane surfaces, and proteins in hopes of treating diseases or better understanding a particular disease pathway. Medical biophysicists employ the findings of these studies in the treatment of patients, in addition to using scanning equipment (e.g., NMR). Individuals who are interested in this new area of study have only a few institutions (Canada and Europe) that offer coursework explicitly aimed at medical biophysics. In the Unites States, graduate-level programs usually offer PhDs or dual degrees (MD/PhD) in physiology and biophysics, but the interest is growing.

It is easy to see from the history of various disciplines that a consensus on the definition of what biophysics is difficult to achieve. At best, biophysics supplies molecular biology with highly precise quantitative measurements in biophysical processes that could otherwise not be determined. Although its early emphasis was on physiological problems such as nerve conduction and sensory communication (the ear), there was a clear movement toward molecular biology after Watson and Crick used X-ray diffusion data to determine the structure of DNA. It was understood that the future of science leaned heavily toward elucidating the mysteries of biology. It is now the case, however, that all molecular techniques are used in every biomolecular area of study.

The current emphasis for biophysics is in the area of structural biology. The last 20 years of structural biology has been replete with determining novel structures, often the only member of a protein family, and have facilitated the determination of structures in other members of a protein family. The other half of structural biology, the development of novel proteins and enzymes, has implications for cancer treatments or genetic therapies. These therapies and diagnostic methodologies have paved the way for newer fields applying biophysics such as medical biophysics and even astrobiophysics (ecological impacts). These applications seem to be where the future lies for biophysics and likely will be so in the next decade and beyond.

Mandy M. McBroom

See also Biophysics, 19th Century; Biophysics, Contemporary; Chemistry; Health Care Science; Physics, 20th Century; Physics, Thermodynamics

Further Readings

Ashrafuzzaman, M., & Tuszynski, J. A. (2012). *Membrane biophysics*. Springer-Verlag.

Brunger, A. T. (1992). Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, *6359*, 472–475. https://doi.org/10.1038/355472a0

Caspar, D. L., & Klug, A. (1962). Physical principles in the construction of regular viruses. *Cold Spring Harbor Symposium in Quantitative Biology*, *27*, 1–24. https://doi.org/10.1101/sqb.1962.027.001.005

D'A Heck, H. (1971). Statistical theory of cooperative binding to proteins. Hill equation and the binding potential. *Journal of the American Chemical Society*, *93*(1), 1A-288.

Diesenhofer, O., Epp, K., Miki, R., Huber, R., & Michel, H. (1985). Structure of the protein subunits in the photosynthetic reaction center of *Rhodopseudomonas viridis* at 3Å resolution. *Nature*, *318*(6047), 618–624. https://doi.org/10.1038/318618a0

Gilman, A. G. (1987). G proteins: Transducers of receptor-generated signals. *Annual Review of Biochemistry*, *56*, 615–649. https://doi.org/10.1146/annurev.bi.56.070187.003151

Glaser, R. (2012). *Biophysics: An introduction* (2nd ed.). Springer.

Katz, B. (1986). *Archibald Vivian Hill. Dictionary of national biography.* Oxford University Press.

Liu, Y., Zhang, X., Tan, Y. L., Bhabha, G., Ekiert, D. C., Kipnis, Y., Bjelic, S., Baker, D., & Kelly, J. W. De novo-designed enzymes as small-molecule-regulated fluorescence imaging tags and fluorescent reporters. *Journal of the American Chemical Society*, *136*(38), 13102–13105. https://dx.doi.org/10.1021%2Fja5056356

Rasmussen S. G., Choi H. J., Fung J. J., Pardon, E., Casarosa P., Chae, P. S., Devree B. T., Rosenbaum, D. M., Thian, F. S., Kobilka, T. S., Schnapp, A., Konetzki I., Sunahara, R. K., Gellman, S. H., Pautsch, A., Steyaert, J., Weis, W. I., & Kobilka, B. K. (2011). Structure of a nanobody-stabilized active state of the β(2) adrenoreceptor. *Nature, 469*(7329), 175–180. https://dx.doi.10.1038/nature09648

Watson, J. D., & Crick F. H. (1953, April). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, *4356*. https://doi.org/10.1038/248765a0

Wu, H., Canfield, A., Adihakari, J., & Huo, S. (2010). Quantum mechanical studies on model alpha-pleated sheets. *Journal of Computational Chemistry*, *30*(3), 1216–1223. https://dx.doi.10.1002/jcc.21408

Wu, H., Wacker, D., Mileni, M., Katritch, V., Han, G. W., Vardy, E., Liu, W., Thompson, A. A., Huang, X. P., Carroll, F. I., Mascarella, S. W., Westkaemper, R. B., Moiser, P. D., Roth, B. L., Cherezov, V., & Stevens, R. C. (2012). Structure of the human k-opioid receptor in complex with JDTic. *Nature*, *485*(7398), 327–332. https://doi.org/10.1038/nature10939

Biophysics, Contemporary

Biophysics is the application of physical concepts and methods toward understanding the living world. As in other branches of physics, experimental biophysical investigation is tightly coupled to the construction of theoretical models that form a mathematical narrative for the phenomena under study, and which help guide and interpret experiments. Whereas 20th-century biophysics largely followed a reductionistic approach, focused on understanding isolated biomolecules and biomolecular processes, contemporary biophysics frequently seeks to understand whole cells, organisms, and communities as self-organized systems, whose properties reflect the combined influence of selective evolutionary pressure and physical constraints.

Theory in Biophysics

Physics aims to form a narrative for all natural phenomena. This narrative takes the form of theoretical models, which are used as everyday tools in physicists' hands to formulate hypotheses and to guide and interpret experiments. Beyond the critical requirement of providing predictive power toward future experiments, a theoretical physical model must exhibit several other characteristics to be considered satisfactory: (a) It must be *quantitative*, expressed in mathematical equations rather than in words, pictures, and so on, which are acceptable formulations for theories in some other disciplines. (b) It must be *simple*. Following Occam's razor, a model is considered better if it has fewer equations and parameters. (c) At the same time, a physical model should also be *universal*, explaining multiple diverse phenomena.

An example of how these requirements can be simultaneously fulfilled is Isaac Newton's law of gravity, whereby the same mathematical equation can be used to calculate the trajectory of an apple falling from a tree, a rocket flying to Mars, or planets orbiting the sun. And in the centuries since Newton's time, theoretical models guided by these same criteria have been developed to describe inanimate matter across scales, from the subatomic to the cosmological. *Biophysics* seeks to extend these efforts to living matter, such that we can understand and predict the behavior of living systems using a quantitative, simple, and universal narrative. It must be noted at the outset that, compared to theories of nonliving systems, the biophysical effort is in an infinitely more primitive state—akin, perhaps, to where the rest of physics was in the 17th century, with (as Newton said) the great ocean of truth lying all undiscovered before us.

In the 20th century, biophysical investigation largely followed a reductionist approach, applying known physical theories to understand the properties of biomolecules, such as DNA, proteins, and lipids, and biomolecular processes, such as the folding of proteins into their functional three-dimensional conformations. In the early 21st century, these theoretical interrogations continue to gain power owing to rapid progress in enabling experimental techniques, such as cryogenic electron microscopy and fluorescence microscopy, and computational methods, such as molecular dynamics simulations and machine learning tools. Altogether, these advancements enable molecular biophysics to move from the study of isolated molecules in an aqueous solution to examining more elaborate scenarios, where multiple molecular species interact in a chemically complex environment, thus beginning to emulate what takes place in the living cell rather than in a laboratory test tube.

The Physics of Living Systems

A complementary approach to biophysics, which was always present in the field but gained prominence in the 21st century, focuses not on the isolated constituents that make up living matter, but rather on how these constituents come together to form *living systems*—cells, organisms, and communities. This effort is inspired by, and aims to emulate, the success of statistical physics to predict how macroscopic properties of inanimate matter emerge from the interactions between its microscopic constituents. This theoretical success was originally limited to simple materials that are homogenous and orderly, such as crystals, but gradually grew to include more complex systems that are heterogenous and disordered, such as glass, or out of equilibrium (i.e., undergoing change) such as weather phenomena. Extrapolating from this success, living systems are viewed as the ultimate emergent phenomena, and understanding them as the natural next step in the study of complex systems.

Theories in this area of biophysics (sometimes called *physics of living systems*) thus seek to explain how biomolecules self-assemble into machines, complexes, and networks that underlie the function of the living cell; how individual cells likewise organize to form a brain, an immune system, and the other multicellular structures that make up a higher organism; and how organisms again self-organize into communities and ecosystems. Moreover, beyond the understanding of living systems as a subject matter, developing such theories will lead (it is hoped) to the discovery of new physical principles—*laws*—governing the process of self-organization. Thus, this branch of contemporary biophysics goes beyond the traditional strand by anticipating new physics from biology, rather than only new biology through physics.

But applying the statistical-physics paradigm toward understanding biology faces considerable obstacles, reflecting the unique properties of living systems as compared with their inanimate counterparts. First of those is the absence of *scale separation*. In traditional physics, different spatiotemporal scales can often be considered separately. For example, when describing the flow of liquid at the macroscopic scale, one can ignore the atomistic nature of material at the microscopic level. This conscious ignorance is enabled by the theoretical practice of coarse graining, where a multitude of details on a lower scale (e.g., the complex interaction between nearby molecules of the liquid) is mapped to much fewer observables on the higher scale (the resulting viscosity of the liquid). Matters are very different in biological systems, where critical information is

typically present at all spatial and temporal scales, from the conformational changes of individual biomolecules at the nanometer and nanoseconds range to the resulting changes in characteristics (phenotype) of a multicellular organism, taking place over meters and hours. Consequently, there is no standard procedure for coarse graining over the molecular details of a living system when aiming to describe it at the cellular, or higher, level. This poses a significant obstacle on the way to constructing a simple theoretical narrative in the physical tradition.

To make matters worse, the detailed information required for constructing a biophysical model is simply unavailable for the vast majority of molecular processes in the cell. Even in the best studied organisms, which serve as *model systems* for the rest of the living world, what typically exists is a partial list of molecular players involved in a given process (e.g., the transcription of RNA molecules from their DNA template), but none of the biophysical parameters—binding affinities, reaction rates, and so on—that characterize the underlying molecular interactions. Even for the handful of select processes on which researchers have focused over decades—for example, how the activity of a specific sugar utilization gene in *Escherichia coli* bacteria is regulated—what is mostly available are measurements made on isolated molecules in a laboratory test tube. The relevance of these measurements to the complex intracellular environment, where physical conditions are very different, and where numerous additional players are present, is uncertain. The bottom line is that, even if building a detailed molecular model for everything in the cell were computationally feasible, the information required for constructing such models is still missing.

Minimal Mechanistic Models and Occam's Rug

To circumnavigate these obstacles, simple theoretical models for biological phenomena are often constructed by trying to identify a minimal set of features that are essential for the observed phenomenon—and are therefore explicitly included in the mathematical representation—while ignoring many other details. The choice of which details to ignore reflects a conscious decision by the biophysicist, but also, invariably, the fact that (as noted above) those details are largely unknown. In addition, identifying which features to include in the model is often done via trial and error by testing what choices improve the agreement between theoretical predictions and experimental observations. Owing to this targeted process of model construction, such models are more properly considered phenomenological (top-down) rather than truly reductionistic (*bottom-up*). Another hallmark of these simplified models is referred to as *Occam's rug*: In contrast to the physics of nonliving systems, where a simple theory is thought to bear the hallmark of truth (Occam's razor), in the case of biophysics, a model's simplicity instead implies that most of the details have been swept under the rug.

The minimal modeling approach has proven powerful when applied to biological phenomena that are relatively simple and well characterized, with the canonical examples coming from *E. coli*, the best studied of all model organisms. Mechanistic biophysical theories have been formulated to capture the way a bacterial cell senses changes in its chemical environment and modulates its behavior accordingly, by swimming toward nutrients and by producing the enzymes required for consuming these nutrients. Following the earlier success in bacteria, the first decades of the 21st century have seen an explosion of minimal mechanistic models directed at systems of higher biological complexity, including processes taking place in eukaryotic (nucleated) cells, during embryonic development of multicellular organisms, and during the onset of human disease, such as the emergence of cancer.

Unfortunately, evaluating the success of biophysical models is less straightforward than it is for models of inanimate matter. This is because, as noted above, the construction of biophysical models involves cherry-picking what features to include, and, consequently, the resulting models are rarely unique mathematically or biologically. In other words, the fact that a given theory recreates existing data does not necessarily mean that it correctly captured the essential biological features. This problem is made more severe by the fact that matching theory and experiment almost invariably involves the process of fitting unknown

mathematical parameters. These parameters represent biophysical properties of the system, such as molecular concentrations and diffusion and reaction rates, but since their values are unknown, they become ad hoc correction factors that artificially increase the model's chances of success. This weakness of mechanistic biophysical models often reveals itself in their limited predictive power outside the narrow premise for which they were calibrated, thus failing the premise of universality desired in physical theories.

Modeling Cellular Individuality

Both the success and limitations of minimal biophysical models are demonstrated by the question of single-cell behavior. It has long been observed that individual living cells, even if they are genetically identical to each other and placed in the same environment, nevertheless may exhibit large differences in the expression of their genes and, consequently, their phenotype. This cellular *individuality* has important consequences across biology, from the emergence of resistance to antibiotics among bacteria to the choice of cellular identity during early embryonic development. The individuality of cells places an additional challenge for biophysical models: beyond capturing the average behavior of a cell population, also predicting the cell-to-cell differences in this behavior.

Attuned, perhaps, to the contemporaneous zeitgeist emphasizing heterogeneity over uniformity, biophysical theory turned its attention to cellular individuality during the late 20th century. To do so, it called on the concept of *noise*, as used in the study of electronic communication, where it indicates random, unwanted fluctuations that accompany the transmitted signal and may obscure it. In the cellular context, noise—the deviation of single-cell properties from the expected average behavior—arises from the inherent randomness of biomolecular events such as diffusion and binding, which in turn results in cell-to-cell differences in the production of regulatory molecules and, consequently, in phenotype. The physical theory of stochastic processes provides the tools for calculating the expected degree of cellular heterogeneity, in the form of the statistical distributions of any quantity, for example, the numbers of proteins produced from a gene. This theoretical approach has proven successful in reproducing experimental measurements of gene expression in individual bacterial cells, and it has since been expanded to describe other facets of cellular individuality.

But utilizing the concept of *noise* to describe cellular heterogeneity also comes with a strong caveat. A key implication of the stochastic picture is that single-cell behavior is inherently unpredictable, and that a living cell's every choice is subject to significant randomness. However, by implying that single-cell observables are not merely unknown but also *unknowable* (in that they are unpredictable), the noise idea legitimizes our ignorance of these cellular properties, sweeping them too under Occam's rug. A stochastic description of cell behavior may thus create a facade of understanding, while in fact impeding the efforts to reveal deterministic drivers of that behavior. Conversely, forgoing the assumption of randomness and identifying those *hidden variables* (a term borrowed from quantum mechanics) will allow biophysical theories to regain predictive power at the single-cell level. The search for hidden variables that drive cellular individuality is an active area of study.

Beyond Mechanistic Models: Evolved Optimality Under Physical Constraints

A complementary approach to developing a theoretical description of living systems posits that these systems' overwhelming complexity, and our considerable ignorance of the underlying molecular details, should argue against the construction of mechanistic models. Instead, biophysical theories should leverage the one universal law of biology already known to us, namely, evolution. Specifically, the premise of theoretical analysis should be that the observed properties of any biological system reflect the outcome of selective pressure to optimize its function, as the system evolves under a set of physical constraints. This investigative approach, typically referred to as *systems biology*—although that term carries different meanings to different people—is highly interdisciplinary. In addition to physics, it leverages theoretical tools from fields as remote as information theory (originally used to describe the

communication of digital information), control theory (used to describe human-made devices), economics, and other areas where the question of optimal performance in the presence of external constraints is often of interest.

Systems biology has had multiple successes in identifying and explaining the properties of living systems. One example is the discovery of common topological features in biological networks. These network *motifs*—typical patterns of connectivity between network nodes—optimize the processing of information by the network. Strikingly, the same motifs are found across systems as disparate as the genetic circuits within a cell and the neuronal networks in the brain, as well as the (human created) World Wide Web, all of which share the requirement to transmit information efficiently to function properly. Other types of motifs endow networks with the property of *robustness*, defined as the ability to function in the presence of uncontrolled fluctuations of both the environment (e.g., changes in temperature) and of the system's internal components (e.g., varying numbers of cellular proteins). Applying principles from control theory led to the identification of robustness in the biochemical networks that underlie bacteria's ability to sense and swim toward nutrients, and in genetic circuits that drive the early stages of embryonic development. In another example of using a systems biology approach, researchers utilized the economic principle of *Pareto optimality*, which describes how trade-offs between different tasks can be best reconciled, to explain several quantitative relations observed across biology, from the shapes of Darwin's beak of the finch, and the sizes of bat wings, to the sets of genes bacteria express under changing growth conditions.

As exemplified by the cases above, the theoretical models of systems biology often constitute a satisfactory physical narrative, in being relatively simple but at the same time universal—that is, applicable across diverse biological instances. But this attractiveness does come with a price tag. For one, while a self-consistent explanation for the observed phenomenon is offered, the underlying mechanistic details (the *how*), which invariably differ between biological systems, often remain unknown. Another aspect sometimes missing from these theories is novel predictions regarding future experiments; for example, the expected response to specified perturbations. Such predictions are required as means to falsify the proposed theory, which, in their absence, risks becoming more of a *Just So* story.

Concluding Remarks

In evaluating contemporary biophysical theory, it must be noted that, even for the best characterized living systems, we are still very far from achieving a comprehensive *molecules to organisms* narrative that fulfills the simultaneous requirements of quantification, simplicity, and universality. An honest assessment of the field must question whether such models are even an achievable goal. Put differently, beyond important but limited *niches*, such as illuminating the function of biological molecules, what roles can physical theories play in understanding the living world? Of course, one may simply argue that physicists' interest in pursuing theoretical interrogation of living systems is sufficient justification for the endeavor. After all, we must remember that intellectual curiosity has always been the driver of discovery. Moreover, the biophysical endeavor has significant pedagogical value, as it involves the training of interdisciplinary scientists capable of applying rigorous quantitative analysis to biological and medical problems. On the other hand, if one considers that the role of theory is, above all, to provide predictive power—here, forecasting the future behavior of a living system under previously untested conditions—then, in the 21st century, it may be argued that a more promising avenue to achieving that power is offered by machine learning, whereby a computerized algorithm learns from past information to predict future data, without going through the *middleperson* of a human-curated model. But a more balanced view suggests that model-free prediction can provide benchmarking for the performance of human-created models, benchmarking which physicists never had before. Furthermore, machine learning methods can aid in the construction of theoretical models, by helping to delineate critical from ignorable features, and to identify model rules that predict experimental observations. Thus, rather than harbingering a future of *prediction*

without understanding, machine learning methods promise to become an invaluable tool in the theoretical arsenal of biophysicists.

Ido Golding

See also Biophysics, 20th Century; Cell Theory; Kinetic (Molecular) Theory; Modeling; Natural Selection; Physics, Contemporary; Systems Science

Author's Note: Ido Golding is supported by grant R35 GM140709 from the National Institutes of Health.

Further Readings

Alon, U. (2019). *An introduction to systems biology: Design principles of biological circuits* (2nd ed.). CRC Press.

Berg, H. C. (2004). *E. coli in motion*. Springer.

Bialek, W. S. (2012). *Biophysics: searching for principles*. Princeton University Press.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Dill, K. A., & Bromberg, S. (2011). *Molecular driving forces: Statistical thermodynamics in biology, chemistry, physics, and nanoscience* (2nd ed.). Garland Science.

Golding, I. (2011). Decision making in living cells: Lessons from a simple system. *Annual Review of Biophysics*, *40*, 63–80. https://doi.10.1146/annurev-biophys-042910-155227

Ingalls, B. P. (2013). *Mathematical modeling in systems biology: An introduction*. MIT Press.

Nelson, P. C., Bromberg, S., Hermundstad, A., & Prentice, J. (2015). *Physical models of living systems*. W.H. Freeman & Company.

Phillips, R., Kondev, J., & Theriot, J. (2009). *Physical biology of the cell*. Garland Science.

Swain, P. S. (2016). *Lecture notes on stochastic models in systems biology*. https://doi.org/10.48550/arXiv.1607.07806

Biopsychosocial Model

The biopsychosocial model requires that biological, psychological, and causal factors causally interact with each other. Hitherto this has been hard to theorize, but the current life and behavioral sciences show how within- and cross-domain causation works. The biopsychosocial model is a model for medicine proposed by the physician and psychoanalyst George Engel at Rochester, New York, in a series of papers in the late 1970s and early 1980s. The main contrast is with the biomedical model, which focuses on biological factors in disease, while the proposed extension broadens the scope to include psychosocial as well as biological factors. The biopsychosocial model and its comparisons and contrasts with the biomedical model raises many questions for theory across the range of the relevant sciences: the life sciences, the behavioral sciences, and the social sciences, as well as the environmental sciences.

Medicine is an applied science making extensive use of technology for detection and treatment, of many kinds, ranging from physics-based technologies to psychosocial. The biopsychosocial model can also be used as a frame for health professional–patient relationships, emphasizing the importance of taking account of the person as a whole. Although primarily a model for medicine, the biopsychosocial model arose at the time of revolutionary changes in the life and behavioral sciences. Centuries-old assumptions about physical nature, mind, and causation were being overturned, being replaced by foundational concepts such as information processing and systems science. A consequence of these changes is that what was hitherto a disunity among the sciences has given way to increasing recognition of the need for cross-disciplinary research programs.

Biomedical and Biopsychosocial Models

The Basic Contrast

Biomedicine is a set of basic biological and clinical science research programs with associated technologies for detection and treatments. It has made great advances in the understanding of many diseases and in the treatment of many, most spectacularly for infectious diseases, but many others besides. The core basic science is physiology,

increasingly enhanced by advances in the understanding of basic cellular processes and genetics. For practical purposes, biomedicine hardly needs a general model, but it does have the working, methodological assumption that diseases can be understood only or primarily in terms of biological factors. This assumption can be called the *biomedical model*. In 1977, George Engel proposed a new model for medicine, the *biopsychosocial model*, arguing that the biomedical model was too narrow in its focus on biological factors, and that, rather, medicine needed to take account of the full range of biological, psychological, and social factors.

Multiple Issues Involved: Etiology, Treatment, Boundaries

The issues involved in the biopsychosocial model are large, abstract, and complex. The broad question whether psychological and social factors as well as biological factors are involved in health and disease arises in many contexts: the etiology of disease, course of disease post-onset, complications by further health problems, and treatments. In addition, there is the question of how we draw the boundaries between what is disease and what is within the normal range of variation. In each of these cases the question arises, Are the main factors and considerations biological, or are they rather broader, biopsychosocial? All of these issues are of major importance in the health sciences and health care, and they are all in play in Engel's papers proposing the biopsychosocial model. This gives rise to significant ambiguity as to exactly what is being proposed, which is likely one reason for the model's mixed reception, as discussed in what follows.

Deep Theory Also Involved: The Nature of and Relation between Biological, Psychological, and Social Factors

Further, Engel's papers penetrate into deep theoretical issues, including whether biology is reducible to physics and chemistry; the need to abandon the dualism of mind and body; and the importance of causal interactions between biological, psychological, and social factors. These theoretical issues will be taken up later in this entry.

Mixed Reception of the Biopsychosocial Model

Widespread Acceptance

It would be reasonable to say that the biopsychosocial model is the dominant headline model of the contemporary practice of health care. It is endorsed as obvious by many health scientists, clinicians, and health educators, and as especially important for quality of clinical care and health care education. It is obvious enough that the patient as a whole has a biology and psychology and social life, and that this is important to bear in mind when treating.

Radical Criticisms of the Biopsychosocial Model

Alongside widespread assent to the biopsychosocial model, however, amid its apparent obviousness, another theme in the reception of the biopsychosocial model is that it is vague and useless. The criticism is that the model is mere handwaving, that it makes no specific scientific claims and gives no specific clinical guidance but can be used to justify anything or nothing. Thus, we have the intriguing picture as a whole that the biopsychosocial model is widely endorsed, but at the same time criticized for being too general and vague to be of any use.

Maybe the Biomedical Model Is Enough?

In the meantime, biomedicine continues to make substantial advances in the treatment of many diseases and continues to attract a high proportion of health research and clinical funding. At the time of writing, Fall 2020, promising interim results of the first to report Phase 3 trial of a vaccine for COVID-19 have just been released; the latest example of great success for biomedicine. More generally, the two most rapidly expanding basic health care sciences, genetics and neuroscience, both appear to be biological, insofar as they focus on biological factors—issues picked up in

the last section. In the context of the continuing success of biomedicine, it can be readily acknowledged as obvious that the person as a patient has a psychological and social life as well as a biological body, and that this has to be kept in mind when managing an illness, while noting that that this can be so even if the core disease process is primarily biological, requiring a biological intervention.

The Need for Evidence

Following the preceding line of thought, the biopsychosocial model only presents as a genuine alternative to, actually as an extension of, the biomedical model, inasmuch as it can be shown that psychosocial factors as well as biological factors are causally relevant to the onset and course of diseases. Such evidence has been gradually accumulating over recent decades as reviewed in the following section.

Evidence Base

New Group Research Methodology

The biopsychosocial model was proposed at a time, the late 1970s and early 1980s, when there still tended to be competing theories about the nature of illness in general—that it was caused by one or other of biological, psychological, or social factors. Research was confined mainly to case studies. It was not until the closing decades of the 20th century that randomized controlled treatment trials came into widespread use. Treatment trials are in a class of controlled research designs and associated statistical analytical modeling capable of estimating the proportion of variance in a health outcome of interest attributable to particular factors. The same approach has been applied in epidemiology to investigate risks for development of disease. These new research methodologies have been used to study empirically and quantifiably the role of biological, social, and psychological factors in the onset and treatment of diseases.

Evidence Base for Psychosocial Factors

Findings in epidemiology over the past few decades have included identifying the so-called social determinants of health. The novel finding has been that social disadvantages are implicated in the onset of many kinds of health problems, physical and mental. There are also many clinical trials showing the importance of psychological treatments in affecting (a) the course of mental health conditions and (b) the adjustment and mental health complications of long-term physical health conditions.

Evidence Base Is Typically Condition- and Stage-Specific (Not "General")

As these findings have accumulated, they have lent clear support to the biopsychosocial model. That said, the research evidence is specific rather than general; it distinguishes between conditions and stages of condition. In advanced stages of some physical illnesses, such as Ebola virus disease, or Grade 4 cancers, or advanced cardiovascular disease, the driving processes are typically biological only, and treatments, if there are any, are accordingly biomedical. On the other hand, if one looks for risks of disease, whether they be lack of environmental hygiene, or unhealthy diet, or chronic stress associated with economic disadvantage, the causal factors broaden out into biopsychosocial. This broader picture also applies to adjustment and quality of life in chronic health conditions. There is no biomedical cure for long-term conditions, though biomedical management may be critical, and questions of lifestyle and quality-of-life implicating psychological and psychosocial factors come to the fore.

Practical Implications of the Biopsychosocial Model

Public Health

The new findings in social epidemiology imply that reduction in population prevalence of disease would require policy to address chronic stress, with its associated biological damage, by alleviating socioeconomic disadvantages. Public health in this new context is closely linked to economic policy, and generally to policy to alleviate social exclusion from health-related resources such as clean air, healthy diet, and a living wage. Psychosocial factors are also increasingly emphasized in public health strategies to manage infectious disease epidemics, such as population cooperation

with infection control measures and high-enough vaccine uptake.

Clinical Treatments

Treatment trials for many health conditions have suggested efficacy of a range of psychosocial treatments. Such findings lend support to the biopsychosocial model. However, the biopsychosocial model is general and is no substitute for the detailed findings of treatment trials. We now have systematic reviews of treatment trials that summarize the findings for a particular condition or for a particular stage of a particular condition. These reviews discriminate between the scientific quality of the trials and make recommendations as to management and treatment. These recommendations, or treatment guidelines, always come with the qualification that each patient and presentation is unique, having a unique combination of factors that cannot all be taken into account when inferring an individual management plan from a range of group treatment studies; and that, accordingly, individual clinical judgment in deciding a clinical management plan is required. Treatment plans may be biological, psychological, or social (as in so-called social prescribing), alone or in some combination, typically requiring a multiskilled, multidisciplinary team. Good clinical care of the patient always involves respecting the whole person.

Theorizing Biological, Psychological, and Social Factors

Dualism and the Reduction Problem: Material Causes Only

At the theoretical core of the issues considered so far are the questions of what biological, psychological, and social factors are, and how, if at all, they interact. The life and human sciences have developed over the past or so against the background of mind/matter dualism that created deep theory prejudices: that nature consisted of physical matter and physical causes alone, and mind, the realm of cognition, was either noncausal or in some way identical to the material brain and behavior. As to social reality and causes, in this picture they are hardly conceivable. In the 1970s, it was commonly and still reasonably believed that biology was reducible to physics and chemistry. The stunning successes of biomedicine in identifying the biological causes of many diseases thus coincided with the deep theoretical or philosophical assumption that anyway all causation had to be biological, ultimately a matter of physics and chemistry. Engel proposed the biopsychosocial model in the 1970s against the background of the deeply antithetical assumptions that psychological and social causes were highly problematic ideas, whereas biological causation alone seemed legitimate. The proposal, however, was a signal of what was to come; around the very same time interrelated revolutions in biology and psychology were in the process of undoing the traditional assumptions.

What Is Biology? A New Science of Regulatory Control

A key theory contribution to revolutionizing biology was made by the physicist Ernst Schrödinger, famous for his work in quantum mechanics, in a lecture delivered in Dublin in 1943, in which he proposed that life could be understood as a local system in which the overall direction of the second law of thermodynamics ran, temporarily, in reverse: in biological systems, entropy is decreased; order, energy differences, increased. Schrödinger saw further that this extraordinary feat was accomplished by genetic coding—a new concept that remained to be worked out. Ten years later Watson and Crick published their work on the structure of DNA and the genetic code. From around that time, biology, physiology, including biomedicine, has developed as a balanced combination of two sciences: the biophysics and biochemistry of energy exchanges, for example underlying basic cellular metabolism, but this combined with a new science of regulatory control of those processes based on information exchange or communication. Current biomedicine is in large part a study of regulatory mechanisms regulating metabolic processes and other regulatory mechanisms. This new science has logical formulae for its mathematical representation: regulatory mechanisms are basically stop/start, represented by negation, and dependent on conditions, represented by "if . . . , then. . . ." Information processing science interweaves with computing.

The new framework of causal explanation in biology in terms of information- or communication-based regulatory control is also applicable in the behavioral and social sciences. This enables a more unified science in which there can be causal interactions between these previously disparate domains.

The Psychological as Embodied Agency and the Social as Resource

The new information-processing paradigm also revolutionized psychology at around the same time, in the so-called cognitive revolution in psychology. While early computational models of mind in the 1970s were computerlike, emphasizing only manipulation of symbols, increasingly, as the research program has evolved, the mind is seen as more biological, embodied, in close interaction with the environment. These theories of embodied cognition were first recognized in the tradition of philosophical phenomenology but are now at the cutting-edge of neuroscience. Important for the purpose of unifying the science, and implied by the common use of the information-processing paradigm, there are information-based regulatory pathways between the mind–brain and some biological systems below the neck, linking neuroscience with biomedicine. The mind–brain also regulates behavior in the environment; accordingly, agency as the capacity to act is an overriding function. Regulation is a core feature of social causes, as well as of biological and psychological causes. The environment for us includes the whole range of factors—physical, chemical, biological, and social—and our biological and psychological functioning depends on it. The biopsychosocial model needs to include the environmental sciences.

Causal Interactions

The core consequence of these theory shifts in the basic sciences are that causal interactions are possible between what were previously disparate domains, which now become linked in a unified ontological–causal space. A sign of this in health sciences are multifactorial models of causal interaction, such as between genes, lifestyle, and environmental exposure, in pathways to many health conditions. This development in the health sciences is part of a broader recognition of the need for problem-solving research programs to be multidisciplinary.

Derek Bolton

See also Health Care Science; Medicine, 20th Century; Medicine, Contemporary; Philosophy of Mind

Further Readings

Bolton, D., & Gillett, G. (2019). *The biopsychosocial model of health and disease. New philosophical and scientific developments*. Springer Palgrave Open Access. Available at https://www.palgrave.com/gp/book/9783030118983

Bolton, D. (2020). The biopsychosocial model and the new medical humanism. To appear in French translation in *Archives de Philosophie*, Special Issue, J. Ferry-Danini & E. Giroux (Ed.). English version at https://www.cairn-int.info

Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. *Science*, *196*, 129–136.

Engel, G. L. (1980). The clinical application of the biopsychosocial model. *American Journal of Psychiatry*, *137*, 535–544.

Frankel, R. M., Quill, T. E., & McDaniel, S. H. (Eds.). (2003). *The biopsychosocial approach: Past, present, future*. Rochester University Press.

Ghaemi, S. N. (2010). *The rise and fall of the biopsychosocial model: Reconciling art and science in psychiatry*. Johns Hopkins University Press.

Lehman, B. J., David, D. M., Gruber, J. A. (2017). Rethinking the biopsychosocial model of health: Understanding health as a dynamic system. *Social and Personality Psychology Compass*, *11*(8). https://doi.org/10.1111/spc3.12328

Marmot, M. G. (2006). Status syndrome: A challenge to medicine. *Journal of the American Medical Association*, *295*(11), 1304–1307. https://doi.org/10.1001/jama.295.11.1304

Schrödinger, E. (1944). *What is life?* Cambridge University Press.

White P. D. (Ed.) (2005). *Biopsychosocial medicine: An integrated approach to understanding illness*. Oxford University Press.

Biostatistics

Biostatistics is both a practice and a profession. As a practice, it involves addressing biological or medically related themes using quantitative reasoning. As a profession, biostatistics can be understood as a different type of practice, one that is widely institutionalized, primarily in North America and Europe, with its own societies, associations, journals, commonly taken-for-granted skills and credentials, and applications. After describing this basic terminology, this entry provides context for both the practice and profession of biostatistics as a series of three historical epochs: (1) its *gestation phase*, which largely began in the 19th-century Europe and the United States; (2) its *foundation phase*, whereby in the early 20th century, renewed debate over the nature of evolution led to one of the most historic accomplishments in science: the consensus across a broad array of disciplines in the process of Darwinian natural selection; (3) its *institutionalization* phase, in which the practice and profession of biostatistics became more widely institutionalized, particularly in the United States, along with a larger process of societal medicalization. The entry concludes with a discussion of some of the promises and perils for biostatistics in the 21st century.

Terminology: The Duality of Biostatistics

This section makes two main points. First, that the term *biostatistics* is semantically vacuous: most definitions of biostatistics have either been terse and limited to a one-sentence description, or definitions of biostatistics have been stretched to a dizzying array of domains. To clarify the first point: Biostatistics encompasses a wide variety of statistical applications to biological phenomena. Beyond this statement, further clarification is warranted.

Second, I claim that understanding biostatistics can be clarified further by distinguishing biostatistics as a practice, or *what some people do* and biostatistics as a profession, or *what some people are*. In other words, biostatistics as a concept can be purposefully widened beyond the tight definitional circularity in which it presently finds itself.

What Is Biostatistics?

The term *biostatistics* is vacuous because, as a concept, it is too broad to be very meaningful. Its sacrificial breadth is not unreasonable to compare it to Jorge Luis Borges's famous short story about a map that was made to such exacting detail that it became as large as the empire it was intended to represent. This *map-territory* problem is a conceptual problem in clarifying the concept of biostatistics: as a construct, it is too encompassing to be very meaningful without being circular semantically. The concatenation of "bio" with "statistics" is most accurately described as a "shotgun wedding at best" (by Morgan, 1986, p. 1105), primarily because it does little to clarify in specific terms or in specific contexts the practice of biostatistics. To define biostatistics as the application of statistics to biological phenomena does not clarify much if the reader knows the word root *bios* is Greek for *life* and that the word *statistics* refers broadly to the classification, collection, analysis, and inference of numerical facts or data. Several exchangeable terms predate the term *biostatistics*. For example, the first recorded use of the term *biostatistics* in the third edition of a dictionary of "medical lexicon," by notable English and American physician Robley Dunglison, refers the reader to the earlier term *medical statistics*, which was used in the volume's 1842 second edition (1865, pp. 138, 654). Even before the concept of biostatistics there were a plethora of other concepts that were used interchangeably with the term *biostatistics*, but which mean different things, including *biostatics*, *biometry*, *biometrics*, *biological statistics*, *medical statistics*, *biostatistical science*, and *biomedical sciences*. More recent terms include *environmental statistics*, *pharmaceutical statistics*, and *public health statistics*.

The Practice/Profession Distinction

To clarify the concept of biostatistics, we need to distinguish between the term as referring to a *practice* and as referring to a *profession*. To preview, biostatistics as a practice emphasizes what is meant when one *does biostatistics*, such as how a question is framed, what topic is addressed, and the procedure or method employed. In contrast, biostatistics as a profession focuses on how people *be biostatisticians*; that is, on how biostatistics has developed as discipline with its own societies, journals, and its own linkages with governmental institutions.

Biostatistics as a practice is relatively ambiguous; however, intuitively, the word itself references both the *statistics* (*statistics* broadly speaking, but more fundamentally a positivist approach toward knowledge construction) and *biology* (*bio*, referring to anything biological). The practice of biostatistics has varied over time and space in its application. About the only unifying methodological approach is its emphasis on quantification. While the practice of biostatistics is difficult to define, greater traction is possible by viewing it as a profession. In this context, professionalization refers to the social processes by which an occupation or a particular status is symbolically and socially constructed.

Beyond these two categories, biostatistics may refer to any number of data analytic techniques and may include any number of applications. The following are only a few topical examples, such as: the study of population prevalence of births and deaths (calculation of so-called *vital statistics*), the study of population health (*epidemiology*), the distributions and changes of genetic diversity of a population (*population genetics*), the sequence of *genomes* or genetic material of an organism or organisms (*bioinformatics*), the study of drug action (*pharmacology*), not to mention applications in agriculture, ecology, animal studies, health economics, and botany, among many other fields. The best way to understand biostatistics is to likely paraphrase famous statistician John Tukey's line: While statisticians "get to play in everyone's backyard," biostatisticians get to play in everyone's "garden," focusing not on the space as a play area, but rather on its flora and fauna.

Another point to be made is that, in large part because of the field's influence on statistics in general, while there are some statistical applications more commonly applied in biostatistics, such as sequence analysis, the majority of biostatistical methods could apply to almost any discipline, although the terminology may differ. For example, while biostatisticians refer to as *hazard* or *survival analysis* as a technique to understand the timing of an outcome (such as "number of years lived before death"), this same technique is sometimes referred to in economics as *duration analysis* and in history and sociology as *event history analysis*.

It is perhaps not surprising that most biostatisticians focus primarily on methods and on data applications. For example, in Wayne Daniel's accessible (and recommended) introduction, *Biostatistics: A Foundation for the Health Sciences*, nearly all of the textbook focuses exclusively on statistical applications, using examples in the health sciences, but with only a brief definition of the term on the third page. Accordingly, the major specialty journals such as *Biometrika*, *Biostatistics*, *Statistics in Medicine*, and *Pharmaceutical Statistics* tend to address new methodological innovations and findings from particular analyses, avoiding the ontological trap of attempting to define the subfield.

The Genealogy of Biostatistics in Three Acts

This section traces the *genealogy* (to borrow Michel Foucault's term) of biostatistics to understand its present incarnation as a historically situated phenomenon. By sticking to historical facts, one can begin to understand biostatistics *not* as I conceitfully introduced it—as a term so vacuous as to embody everything—but rather to understand its *specificity*.

Although biostatisticians are apt to identify pre–19th century, canonical figures, it is important to consider that biostatistics codeveloped with science more broadly and statistics and biology in particular. Biostatistics is best understood as a 19th-century phenomenon, closely aligned with progressivism and with eugenics in the early 20th century, but primarily an established discipline in

the post–World War II period in the United States. As a practice and as a profession, biostatistics has made great contributions to statistics, science, as well as population health. While biostatistics as a practice will undoubtedly continue, there are important questions regarding the future of the profession, in part due the inchoate boundaries of the discipline combined with rapid advancements in technology.

Gestation (1800–1900)

First, and perhaps most importantly, there was the rise of the bureaucratic nation-state, which created an avalanche of printed numbers that required bureaucratic thinkers to both consolidate and manage each state's *informational capital*. By the early 19th century, France, Germany, and the United Kingdom created governmental bodies devoted to collecting statistics on their citizens. A notable example in the United Kingdom was the transition from the church of vital records (e.g., births, deaths, marriages) to the government. In particular, in 1836, the United Kingdom established the General Register Office to manage to supply all vital records and assigned physician William Farr to manage the records. A centralized governmental agency took over the role of the Church of England in large part because of the increasing number of nonparticipants and Catholics in the United Kingdom.

Second, in large part because of the rise of the nation-state, European scientists and philosophers began rejecting the philosophy of causal determinism in favor of a theory of probabilistic thinking that provided an understanding of the world that made it seem less disorderly. Causal determinism was the view that the universe is established mechanistically, with every event occurring in a lawlike fashion in a predetermined causal chain of events. In contrast, probabilistic thinking provided thinkers to develop knowledge that focused on variation among populations and to think of social problems in terms of chance or risk. Bureaucrats armed with an *avalanche of numbers* used such resources to generate a philosophy of chance that guides much of statistical thinking today.

Third, the 19th century underwent political and social movements that promulgated *progressivism*, a broad philosophical idea that asserts that human advancement is made possible through technological economic development in conjunction with the practical application of scientific principles. Progressivism has its roots in 17th-century European Enlightenment ideals emphasizing reason, individualism over tradition, and with such famous French and English thinkers as Francis Bacon, John Locke, Voltaire, David Hume, and Isaac Newton. Although a review of such a grand concept is beyond the scope of this analysis, it is important to underscore that early thinkers were largely (and in many respects) are at least implicitly influenced by this moral worldview. Early 19th-century researchers such as John Snow, who provided evidence regarding the spread of cholera, sought to apply statistics in the service of social good; that is, of addressing the problems that came alongside the industrial revolution, particularly the spread of disease.

The fourth branch of influence can be traced to Charles Darwin's theory of natural selection, which he famously published in *Origin of the Species* in 1859. Despite Darwin's huge influence in biostatistics, it is important to note that biological applications of statistics had occurred previously for centuries, first as the Catholic Church documented births and deaths in its *registry of souls* and then in the early 19th century, particularly as British bureaucrats sought to determine the *quantum of sickness* through cataloging human existence in *life tables*.

Nevertheless, the late 19th and early 20th centuries underwent a series of social movements focused on manipulating population genetics for the betterment of humanity. This fourth branch can be traced to the ideas of Francis Galton, the developer of one of the first surveys and the creator of *regression* models, who also coined the term *eugenics*, combining the Greek roots for *good* and *genes* or *born*, based on the ideology that improving the genetic composition of society requires both the application of rational thinking and progressive ideas toward understanding and manipulating human population genetics.

In summary, it was a combination of events in the 19th century, predominantly resulting from the *avalanche of numbers* created in the early 19th century, combined with new social problems in during the industrialization of Europe, that led to changes in the way great minds thought about causality and about the role of science. These seeds gestated into what can be called the *golden era* of biostatistics, as it rose to great prominence in human society.

Foundation: 1901–1946

The foundation phase of biostatistics is marked by one of the most remarkable achievements in Western civilization: the adoption across a broad range of disciplines of Darwin's gradualist theory of natural selection. The early 20th-century contributions of members of the *Biometric school* helped resolve a decades-long debate regarding the process of evolution (i.e., whether it is gradual or characterized by a series of *jumps*). The debate was reignited in 1900 with the rediscovery of Gregor Mendel's pea plant experiments. Using quantitative evidence, scientists demonstrated that fundamental questions touching on religion and philosophy, in addition to biology, could be answered scientifically. By the 1940s, scientists reached consensus that Darwin's theory of evolution as a slow, gradual process was empirically valid. This *modern evolutionary synthesis*, as it has been famously called, underscored to both scientists and to the lay public the great possibilities with biostatistics.

The foundation of biostatistics cannot be appreciated without recognition of the discipline's trailblazers in the early 20th century. Members of the Biometric school such as Karl Pearson, R. A. Fisher, and W. F. Weldon achieved fame for viewing evolutionary debate as largely a *statistical problem*, as Weldon put it, and by developing foundations of modern statistical analyses. For example, Fisher developed the *analysis of variance*, the method of *maximum likelihood*, and Pearson developed the eponymous *R correlation coefficient* widely used today. Together, they developed standards for assessing statistical significance and the concept of a *Type I Error* (a false detection of a statistically significant result). The growing methodological development of statistics as applied to biological questions was marked by the establishment by Fisher of the journal *Biometrika* in 1901.

In addition to the remarkable achievements of the founders of biostatistics, it is important to underscore that they also viewed their developments within the political project of progressivism and of eugenics. Accordingly, Fisher participated in several eugenics societies and published his research in the journal *Eugenics Review*. Largely influenced by Darwin's theory of evolution, eugenicists sought to improve population health; not surprisingly, these ideas transmigrated to Europe, where they were carried into action by the Nazis. After World War II, eugenics as a social movement largely collapsed. Nevertheless, the rise of biostatistics, particularly in American society, maintains at least an implicit Progressivist approach toward societal health, leaving behind eugenics ideology.

Despite the early association of biostatistics with political controversy, its utility as a practice far outweighed any controversy; it is important to note that statistical methods were also developed during World War II, even if these did not have biological applications. Later biostatisticians would build upon the theories developed by Fisher and Pearson, focusing on broader progressive ideals addressing public health more generally.

Institutionalization: 1947 to the Present

Biostatistics did not become its own professionalized field until the mid-20th century. For example, biostatistics first became an independent discipline recognized by the National Institutes of Health in 1946 alongside the establishment of the Center for Disease Control. Professional societies for biostatistics also lagged behind those of statistics. For example, the American Statistical Association was established in 1839, and in England the Royal Statistical Society was established in 1834. The International Biometric Society, however, was not established until 1947, and the Biometrics Section of the American Statistical Association was not adopted until 1950. Other leading journals such as *Statistics in Medicine* were not established until much later, in 1982.

One way of understanding the development and professionalization of biostatistics in the postwar United States is to understand biostatistics

within the *medicalization of society* in the United States more broadly. The term *medicalization*, typically used by sociologists at least since the 1970s, refers to the increasing use of medical language and diagnoses to understand social phenomena. As the 20th century progressed, biostatisticians increasingly influenced the practice of care, as medical professionals used statistics to make important decisions regarding social policy and diagnoses.

A number of advancements occurred within the postwar period, building primarily from the early work of founders. First, physicians and statisticians worked in unison to define and to understand population prevalence of health risks with the development of large-scale surveys. New methods such as survival models were developed to assess mortality and other health risks. Second, with the aid of institutions such as the National Institutes of Health, biostatisticians began developing formal clinical trials to test the efficacy and safety of drugs and other health interventions, largely influenced by R. A. Fisher's theories regarding randomization, but also influenced by the notable statistician and epidemiologist Sir Austin Bradford Hill, who established a set of criteria (the so-called Bradford Hill criteria) for assessing causality in randomized-controlled trials.

In more recent decades, as sociologist Peter Conrad points out, *medicalization* has become increasingly privatized, with biotechnological revolutions, the rise of large pharmaceutical industries, and the rise of managed care. In contemporary Europe and the Untied States, biostatisticians are increasingly working in commercial fields rather than in academic domains, as private entities rely on biostatistical knowledge to define and simultaneously treat medical problems.

As higher education expanded in the United States and beyond, biostatistics departments began to develop in the postwar era, producing specific credential programs. Nevertheless, the disciplinary boundaries of biostatistics within the academy are still porous at best: the field is often subsumed under statistics departments or amalgamated within schools of medicine, public health, or epidemiology. Nevertheless, increasing numbers of advanced degrees in biostatistics are offered in major universities around the world, almost all of which are postgraduate degrees.

Conclusion: Are We All Biostatisticians Now?

This entry has distinguished biostatistics as a practice and as a profession. As a practice, biostatistics involves the marriage of quantitative methods to biological subject matter. As a profession, biostatistics has grown enormously over the 20th century and thereafter, and, like any profession, has its own institutions, educational credentials, professional groups and societies, and specialty journals. There are two new dynamics that affect the future of biostatistics. In what follows, we will revisit the practice/profession distinction and assess the implications of the big data revolution, so-called, on the behavioral, ecological, educational, and social dimensions of biostatistics.

The *big data revolution* is a term coined in recent years to refer to four aspects of quantitative data analysis. First, owing to increases in computing power, the digitization of records, and the incorporation of technology into everyday social life, there is greater availability of data than has ever been produced before in human history. Second, statistical software is increasingly accessible, with the availability of free, user-generated programs such as R or Python, the increased efficiency of computing power to produce faster results, and the greater availability of more user-friendly interfaces. Third, the big data revolution has led to the availability of online learning sources that have made esoteric statistical knowledge more accessible both within and outside the academy. Fourth, there is a recent cultural revolution: motivation to learn statistics has exploded, as statistical applications have found their way into every aspect of social and professional life. In short, technological advancements have made data more available, analyses easier to implement, skills more accessible, and motivation more widespread.

These implications are both positive and negative for biostatistics. From a practice perspective, biostatistics may far surpass the *golden age* of the 20th century. Soon, nearly everyone may become a biostatistician. From a professional orientation, however, biostatistics may be difficult to maintain as an exclusive profession as more nonprofessionals make substantial contributions to the field. Another concern iterated in biostatistical journals

is ensuring *good* statistical standards in light of greater numbers of participants.

Notwithstanding, if professionalism is the primary bulwark against expanding the contributions of biostatistics, one should be reminded of the pioneers *avant la lettre* who made major contributions even as amateurs. Take for example, John Graunt, often cited in biostatistics as a founding figure, who used public records in an attempt to quell the spread of bubonic plague in 17th-century London. Along the way, he was the first to calculate the population of the city and has been called the *father of demography*. Graunt was, however, a haberdasher by trade. How many future John Graunts are there today, working to contribute to our collective knowledge of the world around us?

Nathan E. Fosse

See also Bioinformatics; Biology, Evolutionary; Health Care Science; Statistics

Further Readings

Allen, G. E. (2001). Eugenics as an international movement. In N. Smelser & P. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (Vol. 16, pp. 4882–4889). Elsevier.

Borges, J. L. (1975). *A universal history of infamy*. Penguin.

Bourdieu, P., Wacquant, L., & Farage, S. (1994). Rethinking the state: Genesis and structure of the bureaucratic field. *Sociological Theory*, *12*(1), 1–18.

Daniel, W. W. (2005). *Biostatistics: A foundation for analysis in the health sciences*. Wiley.

Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life* (1st ed.). W. Clowes and Sons.

Dunglison, R. (1865). *Medical lexicon. A new dictionary of medical science, containing a concise account of the various subjects and terms, and formulae for preparations etc* (3rd ed.). Blanchard and Lea.

Huxley, J. (1943). *Evolution: The modern synthesis*. Harper & Brothers.

Lew, R. (2005). Experimental design. In P. Armitage & T. Colton (Eds.), *Encyclopedia of biostatistics* (3rd ed., pp. 1835–1844). Wiley.

Morgan, P. P. (1986). What are my chances of understanding biostatistics? *CMAJ: Canadian Medical Association Journal*, *134*(10), 1105–1106.