Information, Coding, and Biological Function: The Dynamics of Life

Julyan H. E. Cartwright
CSIC-Universidad de Granada Instituto Andaluz de Ciencias de la Tierra
Department of Chemical Engineering
Instituto Carlos I de Física Teórica y Computacional

Jitka Čejková*
University of Chemistry and Technology Prague
cejkovaj@vscht.cz

Elena Fimmel
Mannheim University of Applied Sciences
Institute of Mathematical Biology

Simone Giannerini
University of Bologna
Department of Statistical Science

Diego Luis Gonzalez
University of Bologna
Department of Statistical Science
CNR, Area della Ricerca di Bologna

Greta Goracci
Free University of Bozen-Bolzano
Institute of Mathematical Biology

Clara Grácio
Universidade de Évora
CIMA
Faculty of Economics and Management

Jeanine Houwing-Duistermaat
Radboud University
Department of Mathematics IMAPP

Dragan Matić
University of Banja Luka
Faculty of Natural Science and Mathematics

Nataša Mišić
Research and Development
Institute Lola Ltd.

Frans A. A. Mulder
Johannes Kepler University
Institute of Biochemistry

Oreste Piro
Universitat de les Illes Balears
Department of Physics
Mediterranean Institute for Advanced Studies

Abstract In the mid-20th century, two new scientific disciplines emerged forcefully: molecular biology and information-communication theory. At the beginning, cross-fertilization was so deep that the term *genetic code* was universally accepted for describing the meaning of triplets of mRNA (codons) as amino acids. However, today, such synergy has not taken advantage of the vertiginous advances in the two disciplines and presents more challenges than answers. These challenges not only are of great theoretical relevance but also represent unavoidable milestones for next-generation biology: from personalized genetic therapy and diagnosis to Artificial Life to the production of biologically active proteins. Moreover, the matter is intimately connected to a paradigm shift needed in theoretical biology, pioneered a long time ago, that requires combined contributions from disciplines well beyond the biological realm. The use of information as a conceptual metaphor needs to be turned into quantitative and predictive models that can be tested empirically and integrated in a unified view. Successfully achieving these tasks requires a wide multidisciplinary approach, including Artificial Life researchers, to address such an endeavour.

Keywords
Complex biological systems, nonlinear dynamics, classical and quantum information, biological codes, theoretical biology

* Corresponding author.

1 Introduction

This article is a contribution to a paradigm shift in theoretical biology that can be subsumed in the concept of informational biology and that can be achieved only through a strong commitment from very different fields coming from a panoply of hard science disciplines. As such, there is not a comprehensive state of the art of informational biology; rather, there is the superposition of different states of the art of the involved fields. Thus, to establish a framework, one should necessarily begin with works relating information with biology, an endeavour that starts with the seminal paper of Schrödinger (1944) (see Figure 1) and that was strongly reinforced by the birth of modern information theory (Dimitrov et al., 2016; Gatenby & Frieden, 2007; Hamming, 1950; Shannon, 1948) (see Figure 2) and by the important steps of the elucidation of the molecular structure of DNA (Watson & Crick, 1953) (see Figure 3) and the informational structure of the genetic code (Nirenberg et al., 1965; Nirenberg & Matthaei, 1961).

Owing also to some promising results at the time of writing, this first phase was optimistically reviewed by Johnson (1970) in a *Science* article. However, as Gatenby and Frieden (2007) pointed out in their analysis of information theory (IT) in living systems, "it seems clear that, in the 35 years since Johnson's original article, IT using traditional Shannon methods has not become, as predicted, the 'general calculus' of biology" (p. 644). Looking for the causes of this failure, the author points out that (a) Shannon methods that quantify the content of information in a biological structure say nothing about its meaning, cost, or function; (b) traditional IT models typically do not address the precise mechanisms by which information is used to perform work and maintain cellular stability; and (c) IT tends to view information as a quantity that is simply exchanged between individuals. It has become increasingly clear that biological information flows along complex pathways, with positive and negative feedback loops and substantial temporal and spatial plasticity. This last argument implies that, to tackle complex biological problems in informational terms, we need to take into account the dynamics of information flux. And this leads to the other main field involved: dynamical systems theory. This physical-mathematical theory has been one of the three physical revolutions of the 20th century, alongside relativity and quantum mechanics.

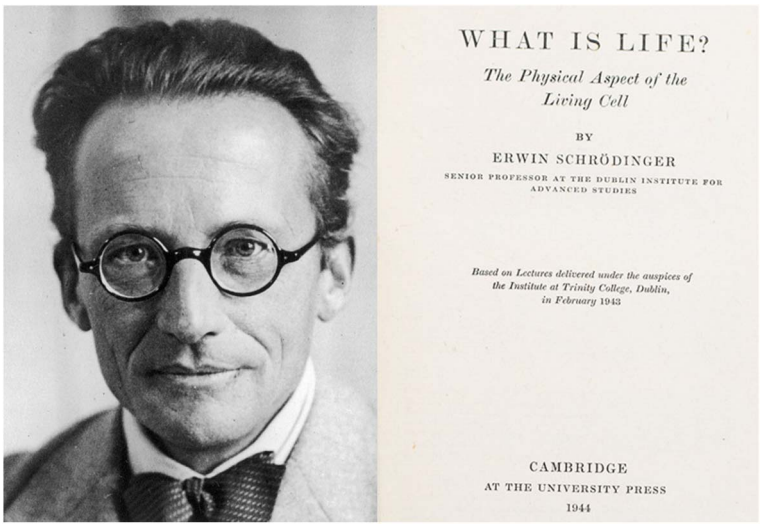

Figure 1. Erwin Schrödinger (August 12, 1887–January 4, 1961) and his book *What Is Life?* From Wikimedia Commons.

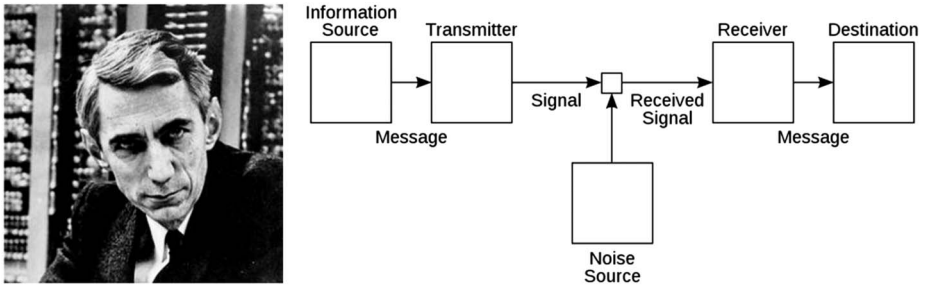

Figure 2. Claude Elwood Shannon (April 30, 1916–February 24, 2001) and his diagram of a general communications system, showing the process by which a message sent becomes the message received (possibly corrupted by noise). From Wikimedia Commons. Author: Konrad Jacobs, Oberwolfach Research Institute for Mathematics.

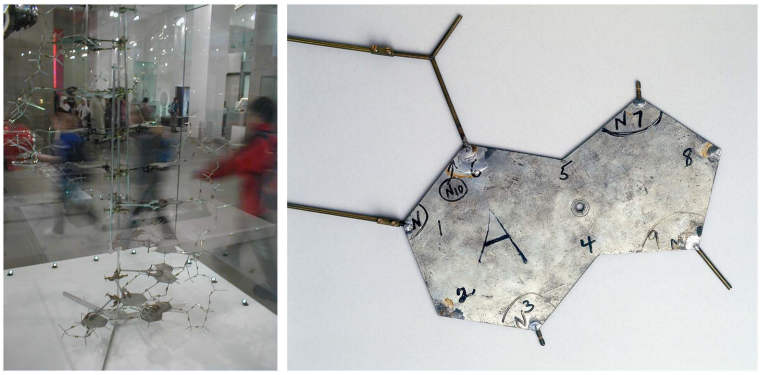

Figure 3. (left) Original DNA model built by Crick and Watson in 1953, on display in the Science Museum, London. Public domain. (right) Adenine template from Crick and Watson's DNA molecular model. Wikimedia Commons, licensed under a Creative Commons Attribution-Share Alike 2.0 Generic license.

The intersection of dynamical systems theory and biological sciences has blossomed in a new discipline: systems biology. As stated in the book chapter by Papakonstantinou et al. (2019), "Systems Biology Consolidating State of the Art Genetics and Bioinformatics,"

> the definition of systems biology has not yet been clarified; In general, systems biology is an interdisciplinary field of biology that involves the computational and mathematical modelling of complex biological systems. The purpose of this field is to understand the complex interactions and functions at the organism, tissue or cell level with direct application in biomedical research. (para. 1)

Chen et al. (2014), in "Current State-of-the-Art and Future Directions in Systems Biology," add that "Systems Biology offers the promise of decoding genetic information, optimizing pharmaceutical design, and aiding in the development of precision medicine. These advances require the bimodal approach of deriving information from experimental data and integrating this information through computational modelling" (p. 12), remarking on the importance of the modeling approach, however, intended as computational, not theoretical, as in the physical sciences.

In any case, systems biology is a flourishing discipline that includes fields in which the dynamical aspects are paramount, for example, metabolic networks or cell signaling networks. Moreover, it has an impact on many "-omics" fields by contributing to identifying patterns that point to an emerging order (and thus information) and by providing a rational systematization of dynamic biological processes. Examples of this kind abound and are at the root of many modern biological breakthroughs, such as the Human Genome Project. Beyond genomics, we can mention a few examples, such as

phenomics, transcriptomics, glycomics, or lipidomics, that represent some of the present challenges for systems biology. However, not all the corpus of the rich theory of dynamical systems has yet entered systems biology.

The most recent results provide a comprehensive framework for modeling complex systems based on two notable facts: (a) simple equations can produce complex behavior, something generically denoted as deterministic chaos, and (b) real-world phenomena are intrinsically nonlinear, implying a great qualitative richness. Thus aspects of complex systems, such as quasi-periodicity, dynamical attractors, chaos control theory, or symbolic and topologic dynamics, represent a guide for understanding the dynamics of biological information. As mentioned, information theory has not fully entered the field, despite the enormous developments related to the dynamics and modeling of biological systems (Yockey, 1992, 2002; Yockey et al., 1958).

Beyond genomics and molecular biology (with evolutionary biology, biological data compression, and bio-inspired software), a few examples are represented by networks (properties of large graphs and reliable data storage), control theory (the capacity of feedback channels and communication control), and statistics and machine learning (computational efficiency and complexity together with multivariate statistics). At a recent workshop on statistical physics, information processing, and biology organized at the Santa Fe Institute in New Mexico, a mecca for the science of complex systems and the Artificial Life field, a fundamental question in theoretical biology was addressed: Does the relationship between statistical physics and the need of biological systems to process information underpin some of their deepest features? As Ball (2017) reported in the article "Information Theory: How Life (and Death) Spring From Disorder,"

> it's hardly surprising that there was no consensus. But one message that emerged very clearly was that, if there's a kind of physics behind biological teleology and agency, it has something to do with the same concept that seems to have become installed at the heart of fundamental physics itself: information. (para. 9)

As stated in 2006 in *Globalization, Biosecurity, and the Future of the Life Sciences*, a report of the Institute of Medicine and National Research Council (2006),

> the kinds and levels of expertise needed to address the most challenging problems of contemporary biology stretch the current state of knowledge of the field. A new level of sophistication in computing and informatics is required for interpretation of much of the data generated today in the life sciences. These data are highly heterogeneous in content and format, multimodal in collection, multidimensional, multidisciplinary in creation and analysis, multiscale in organization, and international in collaborations, sharing, and relevance. (p. 3)

This report notes also that these data are windows into structures of immense complexity. Moreover, it emphasizes that the life sciences of the future will be information sciences and will

> use computing and information technology as a language and a medium in which to manage the discrete, non-symmetric, largely non-reducible, unique nature of biological systems and observations. (p. 32) … In some ways, computing and information will have a relationship to the language of 21st century biology that is similar to the relationship of calculus to the language of the physical sciences. (p. 3)

It can be noted that there is a missing piece in this analysis. Computing and information are indeed a part of the language of 21st-century biology: They represent the orthographic level, just as in the afore mentioned example, calculus is the orthographic level of the physical sciences. But the success of physical sciences has been determined mainly by the achievements at the level of

grammar; within this metaphor, the great theories, such as relativity or quantum mechanics, represent such a grammatical level. To contribute to the construction of the grammar of modern biology is one of the main aims of the European Cooperation in Science and Technology (COST) Action "Information, Coding, and Biological Function: The Dynamics of Life" (Dynalife, 2022–2026).

Dynalife is a new network, founded by us and funded by COST, that is scheduled to run from 2022 to 2026. Our objective is to build fluid and viable pathways between the various communities and research groups to understand the role of information and information management in biological systems, together with its connection with spatial organization and temporal dynamical evolution. This integrated fundamental knowledge will be shared with research groups focusing on specific applications, as well as with groups involved in dissemination activities at large.

2 Challenges

Seventy years ago, two groundbreaking scientific fields emerged: molecular biology and communication theory. Initially, their close connection led to the widespread adoption of the term *genetic code* (Kay, 2000) to explain mRNA's role in amino acid formation. However, today, this collaboration poses more challenges than solutions, despite the rapid progress in both disciplines. These challenges are not only theoretically significant but also pivotal for future biology, including personalized genetic therapy, Artificial Life creation, and the production of biologically active proteins for diverse applications.

Last but not least, the question is intimately connected to the development of a multidisciplinary and integrated scientific view of theoretical biology, introduced more than 70 years ago (Schrödinger, 1944) and formally stated two decades later (Waddington, 1968) but that has remained underdeveloped and very incomplete to date. The key aspect to attain these ambitious milestones is to understand the role of information and information flux in biological systems and promote a paradigm change in informational biology. The use of information as a conceptual metaphor for a qualitative description of the development of dynamics, and evolution of life, needs to be turned into quantitative and predictive models that can be tested against the huge amount of data produced by molecular biology (Nurse, 2021). Successfully completing these tasks requires a strong multidisciplinary approach that incorporates fields well outside the biological realm (Cartwright et al., 2016): classical and quantum information theory, the mathematics of codes, the physics of non-linear dynamical systems, and the engineering of reliable communication devices are a few significant examples.

Informational biology develops at the intersection between molecular biology and information sciences (which broadly cover information, coding, and communication theory) (E. E. May, 2006). It is not a new science (Hamming, 1950; Shannon, 1948). Its starting point coincides with the dawn of molecular genetics, that is, the discovery that the structure of DNA is a double helix. Incidentally, the result had been anticipated by Schrödinger's visionary work. Despite the initial expectations that involved the work of great scientists such as Watson and Crick (1953), Gamow (1954), and Golomb et al. (1958), both sciences followed weakly interacting, and even divergent, pathways. In the late 1960s, Pattee et al. (2012) asked the question about the connection between matter and symbols:

> I want to know how to distinguish communication between molecules from the normal physical interactions or forces between molecules which we believe account for all their motions. Furthermore, I need to make this distinction at the simplest possible level, since it does not answer the origin question to look at highly evolved organisms in which communication processes are reasonably clear and distinct. Therefore I need to know how messages originated. (p. 55)

Moreover, the practical consequences of the interaction remained at a metaphorical or philosophical level, e.g., "The Concept of Information in Biology" (Smith, 2000).

On one hand, molecular biology attained great achievements, supported by but almost independent of other intersecting fields; on the other hand, a similar independent evolution involved information science, not only in the explosive field of communication technology, which shaped irreversibly the information society, but also in other, apparently faraway fields, such as quantum computing. We can say that these two sciences, in fewer than 60 years, contributed indelibly, but almost independently, to shaping our modern world. This has produced a more or less fuzzy frontier that we can call informational biology, which is motivated by the fact that the informational paradigm for molecular biology is not at all completed. Today the concept of information is being extraordinarily enriched, and new theoretical and experimental advances motivate new applications in biology.

It is clear that the framework of information "a la Shannon" alone does not explain the key aspects of biology; this is mainly because biological information is essentially related both to meaning and to biological function, and these are missing concepts in Shannon's framework. Shannon's information relates to the "quantity" of different states that a given system can assume and is more directly related to the notion of entropy.

The discovery that information has a physical reality has a deep potential impact in biology: Information can be converted into energy, and conversely, we need a minimal quantity of energy for erasing a bit of information in any kind of practical memory. From molecular motors and Maxwell demons to the nervous system at the macroscale, learning and forgetting are essential operations that any biological organism needs to be able to perform to survive in a challenging environment.

Of course, the analysis of the primary structure of nucleic acids alone is not expected to explain biological meaning. There is now wide consensus about the essential complexity of biological systems and the need for new theoretical tools, such as dynamical systems, to address the modeling challenge. From continuous and discrete low-dimensional dynamics to the dynamics of complex networks (Walker & Mathis, 2018), for example, metabolic networks, or even the nervous system, the way in which genetic information interacts to determine and control the underlying dynamics is a major open challenge. It is probably the key for understanding the principles of biological control and homeorhesis: the dynamical version of the traditional homeostasis. This is not a new idea; it was proposed by Wiener (1948), who, with concepts of cybernetics, hypothesized that organisms can be considered as complex dynamical systems, implying optimal control and feedback. Wiener's program can be continued today thanks to the corpus of dynamical systems theory, developed from the mid-20th century along with the advent of computers. This would allow dealing with the general problem of the interaction of biological information with dynamical behavior. We can hypothesize the birth of the field of information flux dynamics.

Another relevant question is the extent to which the quantum reality of the atomic and molecular worlds is reflected in the realm of living systems. This also represents a great challenge related to the concept of information, and again, the recent advances in information theory could be key. The new science of quantum information is developing very rapidly and is empowered by its potential applications in quantum computing. Recent developments in the interpretation of quantum information (e.g., Basieva et al., 2021; Fimmel & Petoukhov, 2020; Matsuno & Paton, 2000) are fostering the application of quantum-like information processing to macroscopic biosystems, from genomes, proteins, and cells to the brain. New ideas and theoretical results in this area may find unexpected counterparts at the level of molecular biology and can contribute to explaining some unexpected behaviors and capacities of living organisms.

It is becoming evident that a paradigm change in biology can be ignited by the wealth of new ideas and approaches generated in the hard sciences. Moreover, some expected key advances are delayed or bogged down owing to some essential theoretical gaps, precisely in informational biology. This challenging situation has begun to be recognized by the international scientific community, and important initiatives are being taken worldwide. One important example is represented by the recent project of the U.S. National Science Foundation (NSF), which launched four new centers to bring mathematical perspectives to the biological search for the rules of life. Collectively, these centers are expected to produce a new generation of scientists equipped to explore questions such as how

the information encoded in DNA results in complex organisms with diverse forms, functions, and behaviors (Brenner, 2012). Another important example is represented by the 12 centers established by the U.S. National Cancer Institute, devoted to applying insights from the physical sciences to the problem of cancer. This initiative represents one of the reactions to the stagnation of cancer research despite the ambitious promises of the Human Genome Project. That project triggered strong expectations for the possibility of diagnosing and treating not only cancer but also many other serious diseases. However, these expectations have not yet been met.

Biologists are now divided over what, if anything, went wrong and what needs to happen next. On one hand, some researchers think that we do not have enough available information, so that more data are required. This line of thought has led to the 1000 Genomes Project (Auton et al., 2015), the 100,000 Genomes Project (Turnbull et al., 2018), the Tree of Life project (Darwin Tree of Life Project Consortium, 2022), and so on, but it now seems hard to imagine how linear sequences of bases in nucleic acids could determine cell behavior without the concurrence of a panoply of other intrinsic and environmental factors.

On the other hand, many scientists think that we need a shift to a new paradigm grounded in first principles from mathematics, chemistry, physics, and engineering (Kauffman & Roli, 2023). Therefore the shift implies a move from current approaches in biology toward a more fundamental and interdisciplinary approach, incorporating principles from various scientific disciplines to address the challenges and opportunities presented by modern biological research, particularly in dealing with the vast quantities of data generated. In fact, it seems that the real challenge in modern biology is more related to the overwhelming quantity of data than to their absence or scarcity. It is also clear that a great deal of such biological data, which contributed to creating the big data approach, arise from genome mapping and proteomics (sequence and structure).

The aforementioned international initiatives not only prove the relevance and timeliness of the COST Action Dynalife but also point to the fact that a strong interdisciplinary effort is needed to achieve significant advances in comprehending how information works and is managed at a biological level. On one hand, biology, chemistry, and also physics tell us how the process of translating genetic information into life could possibly work. On the other hand, mathematics and statistics give us models and methods to describe such natural processes within a theoretical framework. Also, they provide us with hints and predictions that can be tested at the experimental level.

3 Progress Beyond the State of the Art

As biological systems are naturally very noisy systems, the problem of faithful transmission of information with a noisy background, typical of engineered communication systems, is of key importance for managing biological information at all scales, from molecules to ecosystems. Without error control, it is not possible to attain reliable and resilient biological systems. Moreover, the control of errors implies the presence of mathematical structures behind the coding strategies, which are usually based on sparsity and redundancy. The most studied apparatus within this framework is the genetic one.

There are still many urgent questions concerning the mechanisms of error detection and error correction in it: Does the quasi-universal genetic code carry more information in addition to that related to specific coding of the polypeptide chain? How can this information be used? Are the genetic code and other kinds of codes fundamental constituents of a complex decoding system that uses algebraic and group-theoretic properties for detecting and correcting informational errors? The questions that have risen in the molecular biological context have led to a new understanding of the role of the mathematical modeling: "Life is a partnership between genes and mathematics" (Stewart, 1999, p. xi). But what kind of mathematics is a partner of genes? And which mathematical methods are appropriate to model the genetic code? These questions are, among others, the topics of a special issue of *Journal BioSystems* published in 2021, titled "Foundations of Mathematics and

Theoretical Biology" (Fimmel & Rodin, 2021), to which many of the current participants in the COST Action Dynalife contributed as editors and authors.

Different mathematical models, algebraic, geometric, and based on number and graph theories, of the genetic code and genetic coding in general have been developed (Błażej et al., 2019; Cartwright et al., 2016; Fimmel & Strüngmann, 2018; Gonzalez et al., 2009; Gumbel et al., 2015; Gunawardena, 2014; Qi et al., 2011). Indeed, the genetic code can be used as a paradigm of biological coding, and the results obtained from studying its structure and the associated mechanisms may cross-fertilize other fields where the transmission of biological information has been ascertained. The COST Action Dynalife proposes to use knowledge about genetic coding as a paradigm to produce an integrated view of the mathematical understanding of regularities and to fertilize neighboring areas; in particular, the new field of code biology shows that, beyond the genetic code, many other biological codes could be implemented with analogous coding and decoding strategies (Barbieri, 2015).

Beyond the mathematical modeling of the genetic code, different innovative and fundamental ideas about the interrelation between information, coding, and function in biology have been proposed and developed. The necessary first step to produce a consistent scientific advancement will consist in testing existing models and theories and producing an effort to improve them on the basis of the results obtained. DeepMind, an initiative connected to Google, with the aid of artificial intelligence software AlphaFold2, allowed the determination of the 3D shapes of almost all human proteins with great accuracy (Jumper et al., 2021). Such data, deposited in a public database, including protein shapes of other organisms, open immediate opportunities for testing fundamental theories and their predictions regarding the spatial structure of proteins.

The availability of an almost complete proteome database (Callaway, 2022) that includes spatial structures opens immediate opportunities for applied research; many laboratories are already using these data, for instance, to develop life-saving cures for diseases that disproportionately affect the poorer parts of the world, to help engineer faster enzymes to recycle polluting single-use plastics, to study antibiotic resistance, to increase the understanding of SARS-CoV-2 biology, and so on. Furthermore, these results, considered almost unreachable only a few years ago, open an incredible opportunity also for theoretical biology and basic research that has not been fully appreciated.

Neural networks and other AI techniques allow us to efficiently mimic nature, but frequently, it is very difficult to put their results into a theoretical and conceptual framework. In our case, the existence of a public human proteome can allow for testing and refining different theories about the genetic code: spatial interactions of sense/antisense proteins; the effect of synonymous mutations on secondary and tertiary structures; the combined effect of mutations and spatial structure in protein expression; and, in more general terms, how the mathematical structure of the code relates to protein function. The research is done by theorists from different disciplines who produce substantive mathematical and physical models and by statisticians, bioinformaticians, and computer scientists who usually work on biological problems within a black box–type unsupervised perspective.

The lack of significant progress in many important fields, such as cancer research, genetic therapy, or the de novo prediction of biological properties of proteins, is motivating the quest for new methods and ideas grounded in exact sciences. R. M. May (2004), an Australian British physicist and biologist, put the question in these terms:

> A paradigmatic account of the uses of mathematics in the natural sciences comes, in deliberately oversimplified fashion, from the classic sequence of Brahe, Kepler, Newton: observed facts, patterns that give coherence to the observations, fundamental laws that explain the patterns .... The sequence information, however, represents only the Tycho Brahe stage. Current work on various genomes uses pattern-seeking programs to sort out coding sequences corresponding to individual genes from among the background .... We are only just beginning, if that, the Newtonian stage of addressing the deeper evolutionary questions posed by these patterns. (p. 791)

Of course, any progress in the general endeavor of developing such a "Keplerian" stage of theoretical biology will have enormous consequences, not only at the level of applied disciplines, mainly in medicine, but also at the fundamental and philosophical levels; such a contribution would represent a fundamental step in our understanding of life itself, with consequences also for present theories about the origin of life on Earth (Walker & Davies, 2013). Given the universality of the genetic apparatus, any step connecting information with biological function is crucial to understanding how biological information determines the main features of an organism (Cartwright et al., 2016).

Besides contributing to the basis of a new theoretical biology, this step is mandatory also for developing a synthetic biology and all the innovative applied fields deriving from it. A foreseen aspect of Dynalife is represented by the interaction with top-level experimental research institutions to explore novel applications. Moreover, the foreseen results should have a deep impact in constructing a rationale behind biologically active proteins and, as such, a profound impact on social issues related to public health and awareness, and also at an economic level through the involvement of the pharmaceutical and biotechnological industries. Moreover, even a little step ahead toward a new grammar for biological sciences represents an important advance beyond the present state of the art in this field. Researchers from various disciplines, computational biologists, data scientists, physicists, mathematicians, and so on, will be able to approach the challenge working together, by removing barriers between researchers, and produce major breakthroughs.

4 A Network to Coordinate Research

The main objective of the COST Action Dynalife is to coordinate the expertise of a potentially critical mass of research groups, not only throughout Europe, to produce collaborative high-impact interdisciplinary research on the comprehension of biological information; the logic and the strategies used by life for its management; and how and to what extent related biological functions, expressed in terms of quantifiable dynamical behavior, are determined and/or controlled by it. The knowledge about the flow of genetic information and the innovative ideas arising both from exact and computational sciences will be used as a paradigm for understanding coding and decoding strategies in other biological codes. A particular focus will be on the information contained in the so-called noncoding DNA and the associated coding strategies. Also, empirical testing of theoretical ideas on large-scale genomic data will be fostered by combining the probabilistic and statistical approaches with techniques from machine learning and computer science.

In turn, this is connected to an important objective: the exchange of knowledge and information to test ideas at the molecular biology level and the development of innovative applications in biotechnology, medicine, and synthetic biology. Being rooted primarily in the fields of physics, mathematics, chemistry, statistics, and engineering, the COST Action Dynalife reflects the wide range of biological coding and decoding systems to which the new ideas and methods can be applied. Moreover, the specific objectives identify possible key fields that can benefit from the present Action. The objectives will help in building new bridges among disciplines to enhance the potential of this field. Managing the coordinated actions regarding expertise, experimental testing, and applications will stimulate a faster growth of the role of information biology at the level of fundamental science, applications development, and general societal impact.

Particular research coordination objectives of Dynalife are as follows: (a) to develop and test the different models and theories about genetic coding and the respective predictions (synonymous and nonsynonymous mutations and evolution, protein spatial structure, etc.); (b) to extend modeling strategies to the study of information flow in noncoding DNA and to other biological codes; (c) to explore potential applications of theories and models, in particular, to medicine and synthetic biology; (d) to contribute to developing a scientifically sound, holistic view of theoretical biology, continuing the efforts that started more than 70 years ago; and (e) to promote a collaborative multidisciplinary dialogue between the different research groups and to promote communication,

Figure 4. The first Dynalife conference, titled "80, 70, 20 Conference: Towards Excellence and Convergence Research in Theoretical Biology," held in Venice, Italy, in May 2023. The conference celebrated three important anniversaries in the history of theoretical biology, including the 80th anniversary of Erwin Schrödinger's lectures that led to his book *What Is Life?* and the definition of the new field of epigenetics by Conrad Waddington in 1943, the 70th anniversary of the discovery of the double helix structure of DNA by Watson and Crick in 1953, and the 20th anniversary of the publication of the first complete sequencing of the human genome in 2003. Photo courtesy of Jitka Čejkova.

dissemination, and science–art crossover activities related to informational biology at large, remarking the importance of basic research for the evolution and welfare of human society.

COST Action Dynalife is an open convergence research project encompassing all the aspects of dynamics and information that are relevant for biological processes. Dynalife offers an open invitation to all interested minds to join its mission at any time. The first Dynalife conference was held in Venice, Italy, in May 2023 (see Figure 4). The dynamic nature of the COST network ensures that its doors remain wide open, welcoming new perspectives and expertise from across the scientific spectrum. As the journey of unraveling biological codes and their intricacies continues, the collective effort grows stronger with each addition, enriching the tapestry of knowledge and pushing the boundaries of research.

Acknowledgment

The authors acknowledge networking support by the COST Action CA21169, supported by COST (European Cooperation in Science and Technology).

References

Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., & Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68. https://doi.org/10.1038/nature15393

Ball, P. (2017). How life (and death) spring from disorder. *Quanta Magazine*, *26*. https://doi.org/10.7551/mitpress/11909.003.0022

Barbieri, M. (2015). Code biology. In *Code biology: A new science of life* (pp. 171–189). Springer. https://doi.org/10.1007/978-3-319-14535-8_10

Basieva, I., Khrennikov, A., & Ozawa, M. (2021). Quantum-like modeling in biology with open quantum systems and instruments. *Biosystems*, *201*, 104328. https://doi.org/10.1016/j.biosystems.2020.104328, PubMed: 33347968

Błażej, P., Fimmel, E., & Gumbel, M. (2019). The quality of genetic code models in terms of their robustness against point mutations. *Bulletin of Mathematical Biology*, *81*(7), 2239–2257. https://doi.org/10.1007/s11538-019-00603-2, PubMed: 30953247

Brenner, S. (2012). Life's code script. *Nature*, *482*(7386), 461. https://doi.org/10.1038/482461a, PubMed: 22358811

Callaway, E. (2022). "The entire protein universe": AI predicts shape of nearly every known protein. *Nature*, *608*(7921), 15–16. https://doi.org/10.1038/d41586-022-02083-2, PubMed: 35902752

Cartwright, J. H. E., Giannerini, S., & González, D. L. (2016). DNA as information: At the crossroads between biology, mathematics, physics and chemistry. *Philosophical Transactions of the Royal Society of London, Series A*, *374*(2063), 20150071. https://doi.org/10.1098/rsta.2015.0071, PubMed: 26857674

Chen, S., Ansari, A., Sterrett, W., Hurley, K., Kemball, J., Weddell, J. C., & Imoukhuede, P. (2014). Current state-of-the-art and future directions in systems biology. *Progress and Communication in Sciences*, *1*(1), 12–26.

Darwin Tree of Life Project Consortium. (2022). Sequence locally, think globally: The Darwin tree of life project. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(4), e2115642118. https://doi.org/10.1073/pnas.2115642118, PubMed: 35042805

Dimitrov, A. G., Fekri, F., Lazar, A. A., Moser, S. M., & Thomas, P. J. (2016). Guest editorial biological applications of information theory in honor of Claude Shannon's centennial—Part II. *IEEE Transactions on Molecular, Biological and Multi-scale Communications*, *2*(2), 117–119. https://doi.org/10.1109/TMBMC.2017.2658058

Dynalife (2022–2026). *Information, Coding, and Biological Function: The Dynamics of Life*. COST Action. https://www.dynalife.eu

Fimmel, E., & Petoukhov, S. V. (2020). Genetic code modelling from the perspective of quantum informatics. In *Advances in Artificial Systems for Medicine and Education II* (pp. 117–125). Springer. https://doi.org/10.1007/978-3-030-12082-5_11

Fimmel, E., & Rodin, A. (Eds.). (2021). Foundations of mathematics and theoretical biology [Special issue]. *Biosystems*, *205*, 104416. https://doi.org/10.1016/j.biosystems.2021.104416, PubMed: 33785382

Fimmel, E., & Strüngmann, L. (2018). Mathematical fundamentals for the noise immunity of the genetic code. *Biosystems*, *164*, 186–198. https://doi.org/10.1016/j.biosystems.2017.09.007, PubMed: 28918301

Gamow, G. (1954). Possible relation between deoxyribonucleic acid and protein structures. *Nature*, *173*(4398), 318–318. https://doi.org/10.1038/173318a0

Gatenby, R. A., & Frieden, B. R. (2007). Information theory in living systems, methods, applications, and challenges. *Bulletin of Mathematical Biology*, *69*(2), 635–657. https://doi.org/10.1007/s11538-006-9141-5, PubMed: 17083004

Golomb, S. W., Gordon, B., & Welch, L. R. (1958). Comma-free codes. *Canadian Journal of Mathematics*, *10*, 202–209. https://doi.org/10.4153/CJM-1958-023-9

Gonzalez, D. L., Giannerini, S., & Rosa, R. (2009). The mathematical structure of the genetic code: A tool for inquiring on the origin of life. *Statistica*, *69*(3–4), 143–157. https://doi.org/10.6092/issn.1973-2201/3553

Gumbel, M., Fimmel, E., Danielli, A., & Strüngmann, L. (2015). On models of the genetic code generated by binary dichotomic algorithms. *Biosystems*, *128*, 9–18. https://doi.org/10.1016/j.biosystems.2014.12.001, PubMed: 25530514

Gunawardena, J. (2014). Models in biology: "Accurate descriptions of our pathetic thinking." *BMC Biology*, *12*(1), 1–11. https://doi.org/10.1186/1741-7007-12-29, PubMed: 24886484

Hamming, R. W. (1950). Error detecting and error correcting codes. *Bell System Technical Journal*, *29*(2), 147–160. https://doi.org/10.1002/j.1538-7305.1950.tb00463.x

Institute of Medicine & National Research Council. (2006). *Globalization, Biosecurity, and the Future of the Life Sciences*. National Academies Press.

Johnson, H. A. (1970). Information theory in biology after 18 years: Information theory must be modified for the description of living things. *Science*, *168*(3939), 1545–1550. https://doi.org/10.1126/science.168.3939.1545, PubMed: 4911870

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2, PubMed: 34265844

Kauffman, S. A., & Roli, A. (2023). A third transition in science? *Interface Focus*, *13*(3), 20220063. https://doi.org/10.1098/rsfs.2022.0063, PubMed: 37065266

Kay, L. E. (2000). *Who wrote the book of life? A history of the genetic code*. Stanford University Press. https://doi.org/10.1515/9781503617575

Matsuno, K., & Paton, R. C. (2000). Is there a biology of quantum information? *BioSystems*, *55*(1–3), 39–46. https://doi.org/10.1016/S0303-2647(99)00081-7, PubMed: 10745107

May, E. E. (2006). Communication theory and molecular biology at the crossroads. *IEEE Engineering in Medicine and Biology Magazine*, *25*(1), 28–29. https://doi.org/10.1109/MEMB.2006.1578660, PubMed: 16485388

May, R. M. (2004). Uses and abuses of mathematics in biology. *Science*, *303*(5659), 790–793. https://doi.org/10.1126/science.1094442, PubMed: 14764866

Nirenberg, M., Leder, P., Bernfield, M., Brimacombe, R., Trupin, J., Rottman, F., & O'Neal, C. (1965). RNA codewords and protein synthesis, VII. On the general nature of the RNA code. *Proceedings of the National Academy of Sciences of the United States of America*, *53*(5), 1161–1168. https://doi.org/10.1073/pnas.53.5.1161, PubMed: 5330357

Nirenberg, M. W., & Matthaei, J. H. (1961). The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences of the United States of America*, *47*(10), 1588–1602. https://doi.org/10.1073/pnas.47.10.1588, PubMed: 14479932

Nurse, P. (2021). Biology must generate ideas as well as data. *Nature*, *597*(7876), 305. https://doi.org/10.1038/d41586-021-02480-z, PubMed: 34522015

Papakonstantinou, E., Pierouli, K., Eliopoulos, E., & Vlachakis, D. (2019). Introductory chapter: Systems biology consolidating state of the art genetics and bioinformatics. In *Systems Biology*. IntechOpen. https://doi.org/10.5772/intechopen.86264

Pattee, H. H., Rączaszek-Leonardi, J., & Pattee, H. H. (2012). How does a molecule become a message? In *LAWS, LANGUAGE and LIFE: Howard Pattee's classic papers on the physics of symbols with contemporary commentary* (pp. 55–67). Springer. https://doi.org/10.1007/978-94-007-5161-3_3

Qi, X., Wu, Q., Zhang, Y., Fuller, E., & Zhang, C.-Q. (2011). A novel model for DNA sequence similarity analysis based on graph theory. *Evolutionary Bioinformatics*, *7*, EBO–S7364. https://doi.org/10.4137/EBO.S7364, PubMed: 22065497

Schrödinger, E. (1944). *What is life? The physical aspect of the living cell and mind*. Cambridge University Press.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Smith, J. M. (2000). The concept of information in biology. *Philosophy of Science*, *67*(2), 177–194. https://doi.org/10.1086/392768

Stewart, I. (1999). *Life's other secret: The new mathematics of the living world*. John Wiley.

Turnbull, C., Scott, R. H., Thomas, E., Jones, L., Murugaesu, N., Pretty, F. B., Halai, D., Baple, E., Craig, C., Hamblin, A., Henderson, S., Patch, C., O'Neill, A., Devereau, A., Smith, K., Rueda Martin, A., Sosinsky, A., McDonagh, E. M., Sultana, R., & Caulfield, M. J. (2018). The 100 000 genomes project: Bringing whole genome sequencing to the NHS. *British Medical Journal*, *361*, k1687. https://doi.org/10.1136/bmj.k1687, PubMed: 29691228

Waddington, C. H. (1968). Towards a theoretical biology. *Nature*, *218*(5141), 525–527. https://doi.org/10.1038/218525a0, PubMed: 5650959

Walker, S. I., & Davies, P. C. (2013). The algorithmic origins of life. *Journal of the Royal Society Interface*, *10*(79), 20120869. https://doi.org/10.1098/rsif.2012.0869, PubMed: 23235265

Walker, S. I., & Mathis, C. (2018). Network theory in prebiotic evolution. In C. Menor-Salván (Ed.), *Prebiotic chemistry and chemical evolution of nucleic acids* (pp. 263–291). Springer. https://doi.org/10.1007/978-3-319-93584-3_10

Watson, J. D., & Crick, F. H. C. (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, *171*(4356), 737–738. https://doi.org/10.1038/171737a0, PubMed: 13054692

Wiener, N. (1948). *Cybernetics or control and communication in the animal and the machine*. MIT Press.

Yockey, H. P. (1992). *Information theory and molecular biology*. Cambridge University Press.

Yockey, H. P. (2002). Information theory, evolution and the origin of life. *Information Sciences*, *141*(3–4), 219–225. https://doi.org/10.1016/S0020-0255(02)00173-1

Yockey, H. P., Platzman, R. L., & Quastler, H. (Eds.). (1958). *Symposium on information theory in biology*. Pergamon Press.