EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

Brussels, 27 May 2022

COST 083/22

DECISION

Subject:	Memorandum of Understanding for the implementation of the COST Action “Information, Coding, and Biological Function: the Dynamics of Life” (DYNALIFE) CA21169

The COST Member Countries will find attached the Memorandum of Understanding for the COST Action Information, Coding, and Biological Function: the Dynamics of Life approved by the Committee of Senior Officials through written procedure on 27 May 2022.

Funded by
the European Union

MEMORANDUM OF UNDERSTANDING

For the implementation of a COST Action designated as

COST Action CA21169

INFORMATION, CODING, AND BIOLOGICAL FUNCTION: THE DYNAMICS OF LIFE (DYNALIFE)

The COST Members through the present Memorandum of Understanding (MoU) wish to undertake joint activities of mutual interest and declare their common intention to participate in the COST Action, referred to above and described in the Technical Annex of this MoU.

The Action will be carried out in accordance with the set of COST Implementation Rules approved by the Committee of Senior Officials (CSO), or any document amending or replacing them.

The main aim and objective of the Action is to to turn the conceptual, qualitative metaphor of information in the
biological sciences into a quantitative holistic view that
encompasses the different kinds of biological information and their
dynamical exchange mechanisms. The ultimate challenge is to produce
a sound rationale for the development of a quantitative theoretical
biology. This will be achieved through the specific objectives detailed in the Technical Annex.

The present MoU enters into force on the date of the approval of the COST Action by the CSO.

OVERVIEW

Summary

In the mid-twentieth 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 take advantage of the vertiginous advances in the two disciplines and presents more challenges than answers. These challenges are not only 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 long time ago in Europe, and that requires combined contributions from disciplines well outside 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. The successful achievement of these tasks requires a wide multidisciplinary approach, and Europe is uniquely placed to construct a world leading network to address such an endeavour. The aim of this Action is to connect involved research groups throughout Europe into a strong network that promotes innovative and high-impact multi and inter-disciplinary research and, at the same time, to develop a strong dissemination activity aimed at breaking the communication barriers between disciplines, at forming young researchers, and at bringing the field closer to a broad general audience.

Areas of Expertise Relevant for the Action	**Keywords**
● Physical Sciences: Non-linear physics ● Chemical sciences: Applied mathematics, statistics, non-computational modeling ● Biological sciences: Biological systems analysis, modelling and simulation ● Mathematics: Discrete mathematics and combinatorics ● Medical biotechnology: Applied mathematics, statistics, non-computational modelling	● Complex Biological Systems ● Non-linear Dynamics ● Classical and Quantum Information ● Biological Codes ● Theoretical Biology

Specific Objectives

To achieve the main objective described in this MoU, the following specific objectives shall be accomplished:

Research Coordination

- To develop and test the different models and theories about genetic coding with state-of-the-art biological data and modern methods from statistics and data science at large.
- To integrate in a holistic view the different models and theories developed and tested in Objective 1, contributing to the development of quantitative theoretical biology.
- To extend modelling strategies to the study of information flow in non-coding DNA.
- To extend modelling strategies to biological codes other than the genetic code.
- To explore potential applications of theories and models, in particular to medicine and synthetic biology.
- To create a panel with representatives of all the disciplines involved in the Action to coordinate the key multidisciplinary tasks regarding objective 1 and described in WP3.
- To create a panel on peer-to-peer dissemination to promote a collaborative multidisciplinary dialogue between the different research groups.
- To promote communication, dissemination, and science-art crossover activities related to informational biology at large, remarking the importance of basic research for the evolution and the welfare of human society.

Capacity Building

- To improve training of Young Researcher and Innovators of ITC countries: stimulating short-term scientific missions to research groups working in different disciplines, promoting and organizing courses and summer schools, stimulating and facilitating participation to Conferences and Workshops, assigning responsibility positions to young researchers and ensuring gender balance.
- To develop a prototypical National Platform for Communication and Dissemination of informational biology at large in an ITC country.
- To promote the creation of national-level groups of interest for researchers and teachers in the field of informational biology.
- To create a forum for discussing with stakeholders the usability and portability to other countries of the prototypical National Platform described in the objective 2 and for consolidating at large the transnational reference character of the present Action.
- To bridge barely communicating scientific fields for attaining, maintaining and developing a critical mass of researchers and consolidate a renewed scientific field that we can define as a modern vision of informational biology.
- To fit into the wider global research community in this area, to engage with other major centers both inside and outside Europe.

TECHNICAL ANNEX

1. S&T EXCELLENCE

1.1. SOUNDNESS OF THE CHALLENGE

1.1.1. DESCRIPTION OF THE STATE OF THE ART

The COST Action DYNALIFE 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 between 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, in order to establish a framework, one should necessarily begin with works relating information with biology, an endeavour that starts with the seminal paper of Schrödinger in 1944 and that was strongly reinforced by the important steps of the elucidation of the molecular structure of DNA and the informational structure of the genetic code. Due also to some promising results at the time of writing, this first phase was optimistically reviewed by Johnson in a Science article in 1970. However, as pointed out by Frieder in his analysis of information theory in living systems, “it seems clear that, in the 35 years since Johnson’s original article, IT (Information Theory) using traditional Shannon methods has not become, as predicted, the “general calculus” of biology." Looking for the causes of this failure, the author points out that i) Shannon methods that quantify the content of information in a biological structure say nothing about its meaning, cost, or function; ii) traditional IT models typically do not address the precise mechanisms by which information is used to perform work and maintain cellular stability; iii) 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, in order 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 in the Action: dynamical systems theory. This physical-mathematical theory has been one of the three physical revolutions of the 20th century, alongside relativity and quantum mechanics. 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., *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". Chen et al. 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” remarking the importance of the modelling approach, however intended as computational, not theoretical as in the physical sciences. In any case, systems biology is a flourishing discipline that includes fields where the dynamical aspects are paramount, for example, metabolic networks or cell signalling 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

Funded by
the European Union

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 system 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 modelling complex systems based upon two notable facts: i) simple equations can produce complex behaviour, something generically denoted as deterministic chaos, and ii) real world phenomena are intrinsically non-linear, implying a great qualitative richness. Thus, aspects of complex systems, such as quasi-periodicity, dynamical attractors, chaos control theory, symbolic and topologic dynamics, etc. 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 modelling of biological systems. 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), 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, 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 reported by Ball 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”. 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 of the National Academies, “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, multi-scale in organization, and international in collaborations, sharing, and relevance.” The NRC 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. 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.” 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 abovementioned 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, and one of the main aims of the present COST Action is to contribute to the construction of the grammar of modern biology.

1.1.2. DESCRIPTION OF THE CHALLENGE (MAIN AIM)

In the mid-twentieth century two new scientific disciplines emerged forcefully: molecular biology and communication theory. At the beginning, the 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, seventy years later, such a synergy presents more challenges than answers and has not taken advantage of the vertiginous advances in the two disciplines. These challenges not only are of great theoretical relevance but also are unavoidable milestones for next-generation biology: from personalized genetic therapy, to artificial life, to the characterization and production of biologically active proteins needed in a wide spectrum of applications. Last but not least, the question is intimately connected to the development of a multidisciplinary and integrated scientific view of theoretical biology, introduced precisely in Europe more than seventy years ago (Schrödinger, 1944) and formally stated two decades later (Waddington, Towards a Theoretical Biology, Nature, 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. The successful completion of these tasks requires a strong multidisciplinary approach that incorporates fields well outside the biological realm: 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 but significant examples. Europe is uniquely placed to develop a world leading network focused on informational biology, but progress is limited due to an inefficient combination of the talents of researchers involved in it. A main aim of this COST Action is to link such research groups throughout Europe to stimulate new, innovative and high-impact interdisciplinary scientific research in this area. At the same time, we aim to develop a strong dissemination activity to break the communication barriers between disciplines and to bring the field closer to a broad general audience. Moreover, special attention will be dedicated to training young researchers and developing a prototypical national communication and dissemination platform to promote this field and to amplify the impact of the expected results. 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.

Relevance and timeliness: Informational biology develops at the intersection between molecular biology and information sciences (that broadly cover information, coding, and communication theory). It is not a new science. Its starting point coincides with the dawn of molecular genetics, i.e., the discovery that the structure of DNA is a double helix. Incidentally, the result had been anticipated by the visionary work of Schrödinger. Despite the initial expectations that involved the work of great scientists such as Crick, Watson, Gamow, and Golomb, both sciences followed weakly interacting, and, even divergent, pathways. Moreover, the practical consequences of the interaction remained at a metaphorical or philosophical level (e.g., *The Concept of Information in Biology*, Maynard Smith). 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 far away, fields, such as, for example, quantum computing. We can say that these two sciences, in less than sixty 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 due to the fact that 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. However, 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 Boltzmann demons to the nervous system at the macroscale, learning and forgetting are essential operations that any biological organism needs to be able to perform in order 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 modelling challenge. From continuous and discrete low-dimensional dynamics, to the dynamic of complex networks, e.g., 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 completely new idea, it was proposed by Wiener, 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-twentieth century along with the advent of computers. This would allow dealing with the general problem of the interaction of biological information with dynamical behaviour. 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 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 explain some unexpected behaviours 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 due to some essential theoretical gaps, precisely in informational biology. This challenging situation begins 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 US National Science Foundation (NSF) that launched four new centres to bring mathematical perspectives to the biological search for the rules of life. Collectively, these centres are expected to produce a new generation of scientists equipped to explore questions such as, for example, how the information encoded in DNA results in complex organisms with diverse forms, functions and behaviours. Another important example is represented by the 12 centres established by the US 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 on 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 Genome Project, but it now seems hard to imagine how linear sequences of bases in nucleic acids could determine cell behaviour without the support of a sound theoretical background that is still lacking. 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. 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, that contributed to create the big data approach, arise from genome mapping and proteomics (sequence and structure).

The aforementioned international initiatives not only prove the relevance and the timeliness of this COST action, but also point to the fact that a strong interdisciplinary effort is needed in order to achieve significant advances in the comprehension of how information works and is managed at a biological level. On the one hand biology, chemistry and also physics tell us how the process of translating the 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. In this respect, Europe is particularly well placed to address this challenge; different recent publications such as theme issues, specialized conferences and workshops, demonstrate that this is the case. A significant problem is represented, however, by the lack of interaction between different groups and the absence of formative opportunities for young scientists in this promising area.

1.2. PROGRESS BEYOND THE STATE OF THE ART

1.2.1. APPROACH TO THE CHALLENGE AND 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 the management of 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 that are usually based upon sparsity and redundancy. The most studied apparatus within this framework is the genetic one. Different mathematical models, algebraic, geometric, and based on number theory, of the genetic code and genetic coding in general have been developed in Europe. 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 neighbouring areas; in particular, the new field of code biology shows that, beyond the genetic code, there are many other biological codes that could be implemented with analogous coding and decoding strategies. Beyond the mathematical modelling of the genetic code, different innovative and fundamental ideas about the interrelation between information, coding and function in biology have been proposed and developed at a European level. DYNALIFE intends to consolidate and empower the existing research groups in this area of theoretical research through the development of a strong network of active collaborations reaching a critical mass of expertise, and an intensive exchange of ideas and programmes. The necessary first step to consolidate the network and 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. The recent announcement of DeepMind, an initiative connected to Google, that, with the aid of artificial intelligence software AlphaFold2, collects all the human protein structures that have been elucidated and deposited in a public database, represents a very important step that allows testing fundamental theories and their predictions regarding the spatial structure of proteins. The availability of a proteome database 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 also open an incredible opportunity for theoretical biology and basic research. Neural networks and AI are strong statistical approaches that allow us to mimic nature very efficiently 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 testing and refining different theories about the genetic code: spatial interactions of sense/anti-sense 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 network will be formed in first place by theorists from different disciplines that will produce substantive mathematical and physical models. Moreover, it will also include statisticians and computer scientists that usually work on biological problems within a black-box-type unsupervised perspective. The implementation step of the Action, groups the theorists in one workgroup (WG1) whereas working group 2 (WG2) will include statisticians, bioinformaticians and computer scientists. The synergistic exchange between the two groups in terms of results and initiatives is the main engine of the whole action and will stimulate new high-impact research. Along with the testing and validation of models and theories against available experimental data the main aim is to produce an integrated view of informational biology. This task presents an inherent difficulty that, in many cases, has prevented its completion; as reported in *Information Retrieval and Management: Concepts, Methodologies, Tools, and Applications* (2018), “With the advancements in sequencing technologies, there is an exponential growth in the availability of the biological databases...These databases are widely spread across the globe and are being maintained by many organizations. A number of tools have been developed to retrieve the information from these databases. Most of these tools are available on the web but are scattered. So, finding a relevant information is a very difficult, and tedious task for the researchers. Moreover, many of these databases use disparate storage formats but are linked to each other. So, an important issue concerning present biological resources is their availability and integration at single platform.” Working group 3 will address this problem and focus on the cross-field interaction and validation. Experts from WG2 will deal with the availability of biological data to create a platform for testing and evaluation. The emergence of a unified interpretation will be fostered by a panel comprising both by theorists from WG1 and by data science experts from WG2.

The situation in Europe is due for a mandatory change that cannot be postponed any longer. The lack of significant progress in many important fields such as cancer research, genetic therapy or in the *de novo* prediction of biological properties of proteins, is motivating the quest for new methods and ideas grounded on exact sciences. May, a famous 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”. Of course, any progress in the general endeavour of developing such a “Newtonian” 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 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. Given the universality of the genetic apparatus, any step connecting information with biological function is crucial to understand how biological information determines the main features of an organism. Besides contributing to the basis of a new theoretical biology, this step is mandatory also for the development of 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 in order 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 in social issues related to public health and awareness, and also at an economic level through the involvement of the pharmaceutical and bio-technological industries. Moreover, even a little step ahead towards 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, etc, will be able to approach the challenge working together by removing barriers between researchers and produce major breakthroughs.

1.2.2. OBJECTIVES

1.2.2.1 Research Coordination Objectives

The main objective of this Action is to coordinate the expertise of a potentially critical mass of research groups throughout Europe in order 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 behaviour, 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 the understanding of coding and decoding strategies in other biological codes. A particular focus will be on the information contained in the so-called non-coding DNA, and the associated coding strategies. Also, the empirical testing of theoretical ideas on large-scale genomic data will be fostered by combining the probabilistic and statistical approach 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 primarily rooted in the fields of physics, mathematics, chemistry, statistics, and engineering, the Action 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 building new bridges among disciplines to enhance the potential of this field. The management of 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.

Specific	Measurable	Achievable	Relevant	Timely
To develop and test the different models and theories about genetic coding and the respective predictions (synonymous and non-synonymous mutations and evolution, protein	Publish a thematic volume with the results of the objective with an integrated comparative view of the different approaches.	Creation of a panel board with representatives of all the involved disciplines for coordinating the key multidisciplinary tasks regarding this objective (described in WP3). A strong peer-to-peer dissemination activity (WP4) for	The empirical testing of theoretical ideas on large-scale genomic and proteomic data are needed to shed light on the novel research area	The area of biological coding and decoding systems is not yet explored.

spatial structure, etc.)		promoting the collaborative multidisciplinary efforts will be established		
To extend modelling strategies to the study of information flow in noncoding DNA and to other biological codes.	Integrate in a holistic view the different approaches and related results published in peer-reviewed publications	State-of-the-art biological data and up-to-date methods from statistics and computer science will be used to achieve this objective	To increase available biological data, to design specific reference datasets, spread data standards and increase awareness of the biological information in the community	Study of information flow has already been proven to be effective in these and other fields, and now needs to be applied more extensively in the filed of biological coding
To explore potential applications of theories and models, in particular to medicine and synthetic biology.	Publication of best practices for experimental and computational results particularly oriented to application in medicine	Collaboration between multidisciplinary teams of experts will be performed to reach a consensus on theoretical findings and practical application	New methods and results in the field of information biology are needed in biotechnology, medicine, and synthetic biology	Dealing with coding information can be challenging due to the lack of efficient methodological solutions
To contribute to the development of a scientifically sound holistic view of theoretical biology, continuing the efforts that started in Europe more than 70 years ago.	Number of researchers that join the network and contribute to the key resources in the field	The community will be involved through gamification strategies, acknowledgments and joint publications	Providing a better understanding of genetic coding has noteworthy implications to increase knowledge on the biological functions, expressed in terms of quantifiable dynamical behaviour	Ensure high quality literature (including textbooks) available to the scientific community

LIST of specific Research Coordination Objectives:

1) To develop and test the different models and theories about genetic coding with state-of-the-art biological data and modern methods from statistics and data science at large. 2) To integrate in a holistic view the different models and theories developed and tested in Objective 1, contributing to the development of quantitative theoretical biology. 3) To extend modelling strategies to the study of information flow in non-coding DNA. 4) To extend modelling strategies to biological codes other than the genetic code. 5) To explore potential applications of theories and models, in particular to medicine and synthetic biology. 6) To create a panel with representatives of all the disciplines involved in the Action to coordinate the key multidisciplinary tasks regarding objective 1 and described in WP3. 7) To create a panel on peer-to-peer dissemination to promote a collaborative multidisciplinary dialogue between the different research groups. 8) To promote communication, dissemination, and science-art crossover activities related to informational biology at large, remarking the importance of basic research for the evolution and the welfare of human society.

1.2.2.2 Capacity-building Objectives

In order to overcome the relative isolation of researchers of different backgrounds – experimental, computational, theoretical or industrial – this Action will establish a joint research agenda in order to ensure the cooperation between heterogeneous research groups. In addition, the interchange between institutions will be fostered through short term scientific missions (STSMs) and workshops. This Action will prioritize the STSM Mobility and Collaboration activities, aimed at the training of young staff (especially PhD students and Young researchers and innovators) in areas of expertise different from those of the sending institution. A special focus is to include young scientists from non-research-intensive countries, as well as under-represented groups, such as female researchers and ethnic minorities.

Specific	Measurable	Achievable	Relevant
To facilitate joint research agenda	Regular meetings (WG and inter-WG meetings) to coordinate research organized; updated description of research activities on website and group collaboration platforms (like Slack)	Meetings will include roundtable discussions and sessions for ongoing research, coordination of activities with partner networks. Meetings will be organized either in-person and online	To void duplication of effort in parallel activities and foster collaborations
To develop a prototypical National Platform for Communication and Dissemination of Informational Biology at large in an ITC country	Included results of the present Action and social impact at different levels are presented on the prototype of the National Platform for Communication and Dissemination.	The will be coordinated preferentially by a member of an ITC country	To promote communication, dissemination, and science-art crossover activities related to informational biology at regional level and to consolidate at large the transnational reference character of the present Action.
To create national-level groups of interest for researchers and teachers in the field of Informational Biology.	Groups of researchers and teachers in countries involved in the Action. At least 30 scientific papers in total will be published by the researchers involved in such groups	Research communities on information biology will be created as the result of regular meetings, WG meetings and network-wide events.	Several objectives of this Action require an interdisciplinary approach at national and international levels
To improve the career prospects of PhD students and post-docs that participate to our network events	Contributed to the training of at least 20 PhD/post doc students that will attend our network-wide events and STSMs. Organize four 5-day schools, four 2-day workshops, four 1- day hackathons and at least 10 STSM calls	Develop an official research programme for PhD or postdoctoral internship. Organize STSMs, training schools and training events. Representatives in the core group meetings and in the management of STSMs, support organization of independent and satellite events, special sessions in regular meetings	Young researchers and innovators network essential to ensure interdisciplinarity and skills exchange as well as dissemination and communication
To promote geographical and demographic diversity, with special attention paid to supporting under-represented groups, such	We expect that at least 30% of the joint papers will have a female researcher as one of the leading authors (clearly indicated in the paper)	Joint projects and exchanges between groups will be performed through meetings and STSMs Conference Grants for Early Carrier	Support PhD students and Young researchers and innovators from participating ITCs to attend international conferences in the

as female researchers and ethnic minorities.	and at least 40% of the joint papers will have one of the leading authors from a target developing country.	Investigators from ITC. At least one of the key positions in the Action management will be reserved to a representative of ITC.	thematic domain of the Action

List of Capacity Building Objectives: 1) To improve training of early stage researchers of ITC countries by using different instruments: stimulating short-term scientific missions to research groups working in different disciplines, promoting and organizing courses and summer schools, stimulating and facilitating participation to Conferences and Workshops, assigning a percentage of responsibility positions to young researchers and ensuring also gender balance. 2) To develop a prototypical National Platform for Communication and Dissemination of Informational Biology at large in an ITC country (including results of the present Action and social impact at different levels). 3) To create national-level groups of interest for researchers and teachers in the field of Informational Biology. 4) To create a forum for discussing with stakeholders the usability and portability to other countries of the prototypical National Platform described in the objective 2 and for consolidating at large the transnational reference character of the present Action. 5) To bridge barely communicating scientific fields for attaining, maintaining and developing a critical mass of researchers and consolidate a renewed scientific field that we can define as a modern vision of informational biology. 6) To fit into the wider global community in this area, to engage with other major centres both inside and outside Europe.

2. NETWORKING EXCELLENCE

2.1. ADDED VALUE OF NETWORKING IN S&T EXCELLENCE

Networking excellence is determined by the effectiveness of the interaction between experts to achieve the objectives of the Action. Beyond the usual expression of collaborative research in the form of scientific publications, excellence in the DYNALIFE Action depends mainly on the degree of interaction between groups working in different disciplines. This is because the main objectives of the action should be attained only by means of a strong multidisciplinary effort. In order to obtain this, the proposed network has been determined as to ensure a balanced representation of all the involved.

2.1.1. ADDED VALUE IN RELATION TO EXISTING EFFORTS AT EUROPEAN AND/OR INTERNATIONAL LEVEL

This COST Action does not duplicate any other Action or Research Programmes supported by the European Commission through its various funding alternatives. This Action, however, is related with former activities that refer to partial aspects of this Action. Previous COST Actions are involved, for example, with mathematical models for interacting dynamics on networks, or systems chemistry, or also to the emergence of chemical systems, but none of these is focused on the connection between information, coding and dynamical systems as the present one. Moreover, some past initiatives in Europe have been related to the present main challenge of this COST Action. For example, an international workshop on mathematical and statistical models for genetic coding has been held in Germany in 2013; six international conferences on code biology, focused on the general problem of biological codes, have been organized in Europe in the last years; also, different thematic issues about the informational interpretation of the genetic code, the role of information in biology, and the study of different biological codes have been published in recent years. The present COST Action can be considered as having a natural continuity with these different activities and initiatives aimed to contribute to a change of paradigm in the field of informational biology. As mentioned above, different initiatives have been carried out also at an international level, mainly in the USA. The present COST Action, however, brings a particular emphasis on physical, mathematical and statistical models and theories that have been developed mainly at a European level; these have a strong potential that needs to be developed and validated against available biological data and conceptually integrated at a multidisciplinary level. In this respect, the present COST Action is complementary to former initiatives and can find important synergies through knowledge exchange and the identification of common challenges. Finally, we underline the added value of creating a network on informational biology in Europe. The integration with other activities at the European and International levels allows far greater returns than the simple superposition of different bilateral projects.

2.2. ADDED VALUE OF NETWORKING IN IMPACT

2.2.1. SECURING THE CRITICAL MASS AND EXPERTISE

This COST Action, through Research Coordination and Capacity Building activities, will have a huge impact on the field of informational biology in the next years, both at the European and at the international level: first, it will contribute to a paradigm change both from the scientific and the philosophical points of view. Clearly, without a sound theoretical framework, the interpretation of the plethora of data produced by molecular biology is reaching a bottleneck. This expected contribution, depends on the consolidation of different theories through systematic testing on existing data and the simultaneous existence of a coherent critical mass of researchers that will succeed in demonstrating the validity and usefulness of the theoretical framework for the interpretation of biological data. Without the consolidation of a strong networking activity these conditions cannot be met. Moreover, the Research Coordination Activities related to the actual testing of the theoretical approaches will naturally lead to touching important challenges and/or shortcomings of current molecular biology (for example, in the above-mentioned cases related to medicine) and will generate new experimental tests and new data, as well as foreseen and even unexpected applications. A transversal workgroup will be dedicated to such key interactions between theory, experimental data, experiments and applications, ensuring cohesion and communication between research groups as well as reaching the needed critical mass of researchers. Due to the extreme cross-disciplinary character of this work-group, the activity within it will be intimately interrelated with that of the other transversal work-group dedicated to dissemination and scientific communication that will also contribute to the growth of the groups and to the contact with stakeholders, research groups from Neighbouring and International Partner Countries, International Organizations, and society in general. With this strong activity of dissemination, we expect to attract also other international research groups; in this respect, a first effort has already been produced by securing the participation of experts from International Partner Countries.

2.2.2. INVOLVEMENT OF STAKEHOLDERS

Throughout the preparation of the Action, leading researchers from different disciplines converging in informational biology have involved stakeholders and policy makers across Europe. A forum promoting dialogue with interested actors will be created and this activity will be associated to the development of a prototypical National Dissemination Platform in an ITC country with the aim to communicate and engage policy makers, industrial and health representatives, and the society in general. In particular, the platform will allow dissemination/development of knowledge in Informational biology, including exchanges and other networking activities related to medical procedures and biomedical applications of peptides, polypeptides and proteins. The creation of an attracting pole for health and industrial representatives will be promoted. Given the clear applications, we shall involve stakeholders at the level of medical and pharmaceutical companies. Moreover, the Action will coordinate the co-utilization of scientific resources (laboratories, relevant instruments and procedures, etc.), also in related fields such as biological engineering and green chemistry. Furthermore, the dissemination platform will take care of the new paradigms related to the genetic code manipulation that arose in recent years, e.g., de novo proteins, CRISPR based technologies, new diagnostic and therapy methods based on sense/anti-sense active proteins, etc. Bridging separate fields of science disciplines to achieve breakthroughs that require an interdisciplinary approach, will be of particular importance for the implementation of informational biology knowledge, and also in order to develop capacity at national, EU, and international levels.

2.2.3. MUTUAL BENEFITS OF THE INVOLVEMENT OF SECONDARY PROPOSERS FROM NEAR NEIGHBOUR OR INTERNATIONAL PARTNER COUNTRIES OR INTERNATIONAL ORGANISATIONS

The network of proposers involved three International Partner Countries: USA, Mexico, and Argentina. The internationalization of the Network offers many advantages, for example by increasing its resilience and by offering enlarged opportunities of knowledge exchange and ensuring a greater impact. In this respect, the involvement of International Partner Countries will be encouraged throughout the Action. For the IPC partners initially involved in the Action there are additional reasons that bring mutual benefit. In particular, the expertise of the secondary proposers is in Communication Engineering applied to the synthesis of proteins and to chemical and biological processes in general; there is not a critical mass of research in this field and such expertise complement very well the EU Network that has an important knowledge development in theoretical models. Such complementarity prompts for an important mutual benefit from the activities of the Action Network. The Action includes US researchers that gave a pioneering contribution in the application of communication engineering to informational biology; we can mention: the special session on communication and coding theory in biology, EMBS 2003, the edition of the IEEE EMB Magazine special issue on molecular biology and coding theory (2005), the ACM international conference on nanoscale computing and communication - 2015, 2016, 2019, and the Programmatic Oversight of the National Science Foundation that sponsored the workshop on biology through information, coding, and communication theory (BioTICC) in January 2020.

3. IMPACT

3.1. IMPACT TO SCIENCE, SOCIETY AND COMPETITIVENESS, AND POTENTIAL FOR INNOVATION/BREAKTHROUGHS

3.1.1. SCIENTIFIC, TECHNOLOGICAL, AND/OR SOCIOECONOMIC IMPACTS (INCLUDING POTENTIAL INNOVATIONS AND/OR BREAKTHROUGHS)

This COST Action promotes innovative and high-impact multi and inter-disciplinary research in a field that can represent a new paradigm in biology. The achievement of the objectives requires an extraordinary multidisciplinary effort that has not been attempted before. As such, we expect a great scientific impact implied by reaching the aims of the Action: first, the consolidation of a multi-disciplinary and holistic view of the role of information in biology. These aspects can have a great scientific impact, from the philosophical interpretation of life, also with consequences on theories on the origin of life, to a decisive boost for overcoming conceptual and practical bottlenecks that slow down the progress of different applied biological fields. As mentioned, specific examples come prevalently from medicine and biotechnology. Second, the multi-disciplinary synergy between research groups attained thanks to the Action can accelerate the advancement in the understanding of specific problems that are at present attacked in a more isolated fashion. It is envisaged that these problems come predominantly from the elucidation of coding strategies in newly discovered biological codes. Third, while in the short term the activities in the Action will deal mainly with fundamental science, the knowledge obtained and disseminated under this Action will have a long term strong socio-economic impact. In the middle term is expected to have returns on medicine and bio-technology; in the long term, the understanding of mechanisms that life optimized along millions of years of evolution, can found unexpected technological applications in a sort of reverse engineering process that can lead to developing biomimetic systems and devices, and information processing algorithms, that are essentially bio-inspired in the strategies of biological management of information. The aspect of competitiveness deserves the last word: as mentioned, different important initiatives to overcome the bottlenecks encountered in different applied biological problems are being taken around the world; these mainly consist in the involvement of the hard sciences and consistently introduce the ideas from information theory and dynamical systems. For this reason, the achievement of the goals in the COST Action DYNALIFE represent a necessary step for ensuring competitiveness of European science in the forthcoming years. Some of the long-term impact objectives imply interdisciplinary efforts that will be started through this action but will continue to have long-term impact on the disciplines involved leading to impacts at the societal level and to medicine a well as in more fundamental theory.

3.2. MEASURES TO MAXIMISE IMPACT

3.2.1. KNOWLEDGE CREATION, TRANSFER OF KNOWLEDGE AND CAREER DEVELOPMENT

Knowledge creation is a main aspect of research progress and technological innovation. The DYNALIFE Action will provide an optimal environment for knowledge creation by connecting highly skilled researchers working on complementary fields with a common endeavour. The strong effort dedicated to break communication frontiers between disciplines will facilitate collaborative research between experts in different disciplines catalysing the generation of new ideas and favour knowledge transfer. Indeed, at a multidisciplinary level, and to some extent also at the interdisciplinary one, communication barriers represent one of the main obstacles to collaborative research and knowledge transfer. Bottom-up generation of new knowledge will be also stimulated by using the tools of the Action. Short Term Scientific Missions will greatly contribute to knowledge creation and transfer. Knowledge transfer represents also a fundamental ingredient to achieve a unifying interpretation of models and theories on genetic and biological coding in general. The communication activities dedicated to a constructive dialogue with stakeholders and with the society in general, will allow, in addition, a top-down knowledge transfer at these different societal levels that will be put in practice through the implementation of an experimental Dissemination Platform on Informational Biology. As regards career development, young researchers will be encouraged to apply to STSM for visiting excellence research centres of the Network and to gain experience and expertise in informational biology. These training activities will be complemented with theoretical schools organized by the Action and aimed at completing the theoretical formation of Young researchers and innovators. The skills in autonomy and leadership also will be encouraged by means of the assignation of responsibilities and duties within the COST Action coordination and management activities. Also, the participation on dissemination activities will be encouraged.

3.2.2. PLAN FOR DISSEMINATION AND/OR EXPLOITATION AND DIALOGUE WITH THE GENERAL PUBLIC OR POLICY

Dissemination and communication represent key aspects of the DYNALIFE Action. Dissemination at the research level will help breaking barriers between disciplines, stimulate collaboration inside the Action, communicate important scientific results to peers and encourage new collaborations. This will enlarge the Network, consolidate the critical mass of researchers and the flux of knowledge, and contribute to the formation of young researchers in the field. The communication to a large public will inform researchers and professionals in other disciplines, communicate the importance of the Action to the public at large, ensure and amplify the societal impact, establish a dialogue with stakeholders potentially interested in the results of the Action at different levels: from government to health institutions, to industrial representatives. In order to implement such aims, an ambitious Plan for Dissemination and Communication is foreseen: a multidisciplinary panel will be created on WG4 for coordinating the dissemination activities across the activities of DYNALIFE. This panel will study and develop methodologies and tools to implement a strong action of dissemination in order to break barriers between disciplines. This action will be accompanied by the usual scientific dissemination tools: journal publications, collaborative special issues, Conferences, Workshops, and Schools directed to the training of young researchers. In particular, short-duration schools will be organized. The main themes are: Information Theory and Coding, Complex System Modelling, Mathematical Modelling of the Genetic Code, and Communication Engineering Applied to Coding in Biology. Regarding Communication at large, the ambitious program is mainly based on the development of a prototype of a National Dissemination Platform. This Platform will coordinate communication of knowledge, challenges and impact of the results at a National level (preferably in an ITC Cost Country). This activity goes well beyond the activities inside the Network and is foreseen to act as a contact point with the society in general, especially with stakeholders at all levels. This activity will be complemented by a Forum with a transnational reach with the same objectives. The National Platform is aimed at experimenting the best communicative methodologies and instruments for promoting the field of Informational Biology. The experience collected during the Action is expected to be useful in the implementation of similar initiatives in other participant countries. Beyond the implementation of the usual means (such a web site, or scientific social networks) particular attention will be dedicated also to the methodologies and instruments implemented in WG4; some participants have experience in dissemination and communication activities but the collaboration with communication professionals will be encouraged. Also, the use of video documentaries that could be used at different communication levels will be strongly supported. Such means will be used also for facilitating communication and collaboration with the pharmacologic and bio-technological industries. It is expected that the specific results regarding the theoretical prediction of protein properties on the basis of proteome data bases will be of great interest for industry stakeholders and a specific dissemination task will be foreseen with this aim. Adequate means for protection and exploitation of intellectual and industrial property rights will be also provided.

4. IMPLEMENTATION

4.1. COHERENCE AND EFFECTIVENESS OF THE WORKPLAN

4.1.1. DESCRIPTION OF WORKING GROUPS, TASKS AND ACTIVITIES

All the potential members of the Action are internationally recognized experts in various scientific fields, such as mathematical and statistical modelling and the analysis and interpretation of biological systems. The Action already includes 24 research groups from 12 different European countries that carry out high level research but lack a proper European scientific network: 1. Bosnia Herzegovina: Combinatorial optimization and machine learning methods for solving NP hard and other computationally challenging problems in bioinformatics and computational biology. 2. Croatia: Biological Engineering, evolutionary genetics and Data Analysis, Proteomics. 3. France: Quantum computing applied to the structure of information in the genetic code, Computer Science, Genetic Coding. 4. Germany: Mathematical modelling of the genetic code, error-detecting codes; Bioinformatics. 5. Hungary: Evolution of cooperation, origin of life, RNA world, ribozymes, landscape ecology, Spatial ecology, clonal plants. 6. Italy: Statistical modelling, Dynamical systems and Chaos, Mathematical modelling of the genetic code, Molecular biology. 7. Poland: Statistics, mathematical modelling of the extensions of the genetic code, bioinformatics. 8. Portugal: Complex networks, graph theory and applications. 9. Serbia: p-adic analysis, mathematical modelling of the genetic code, fractals and chaos theory. 10. Spain: Theory and applications of dynamical systems, spatial and temporal patterns, chemical self-organization. 11. Sweden: Quantum Information models in biology. 12. The Netherlands: Complexity, in prebiotic evolution (origin of life) and in man-made synthetic, biological and chemical systems. 13. Slovenia: Enzyme networks, Information-theoretic measures. 14. Denmark: Mathematical and statistical modelling and theory of biological systems. It is our stated aim to encourage other European groups to join the Action. Many of the players have consolidated collaborations in information biology and dynamical systems with non-European experts from renowned institutes, including research groups from ICP countries. In particular, there are 3 International Partner Countries participating to the Action: EEUU, Mexico and Argentina; these contributions complement optimally the expertise of the European groups by providing expert know-how in Communication Engineering; this aspect is under-represented in Europe within the target field of Informational Biology. These non-European scientists will be invited to the Action Workshops. An essential part of the action is to strengthen and expand both the European network and an international scientific network in Informational Biology. The participants in the Action will form four Working Groups (WGs) to pool existing expertise and to promote scientific cooperation at all levels. The aim is to enable close integration of scientific research and a high level of knowledge transfer, including dissemination directed to break the communication barriers between different scientific disciplines, and communication at large to consolidate the international reach of the network and amplify the societal impact of the Action. The first Working Group is focused on theoretical aspects; the main aim is to develop models and exchange knowledge between research groups. The development of an integrated view of the different approaches including successes and failures of extant models, will be promoted. The group will also focus on identifying present and future challenges in molecular biology and envisaging the necessary steps towards a paradigm change in theoretical biology rooted on informational biology. The second Working Group includes expert researchers in modern statistics, probabilistic modelling, bioinformatics, machine learning and data science at large. Its main focus is the analysis of the big corpus of data already available from experiments in molecular biology. Also, it will provide the support and expertise to mathematical modelling both from the point of view of statistics, probability and machine learning, and from the bioinformatics and computer science side. Moreover, it will provide expert know-how aimed at testing models and theories. The third Working

Group represents one of the great challenges of the Action because it will be dedicated to the transversal integration of all the research groups. The main tasks are: cross-field integration, the testing of theoretical models (mainly on the basis of existing data in molecular biology), and the proposal of new experimental protocols together with the identification of potential applications. The fourth Working Group is dedicated to Dissemination and Communication. It is expected that WG4 will promote the close interaction of all the research groups. A particularly strong interaction will be developed with WG3 regarding dissemination. The communication between peers will be developed not only through usual open access scientific publications; a specific task force will be dedicated to break down the communication barriers between different scientific disciplines inside the Action. Overall, we expect both a strong collaborative effort within Work Groups, as they will face common challenges, and a strong interaction between different Work Groups since the success of the whole Action depends on the collaborative actions among them. The working groups will be coordinated by WG leaders who have exceptional expertise and experience. They will plan the activities of the Working Groups and will supervise the networking actions. This will ensure that the working groups do not define themselves as purely research projects but will promote widespread network actions involving the whole community. In order to ensure a better distribution of the workload and also to enable optimal cooperation within the network, the working group leaders will be preferentially selected among ITC Countries; To comply with the COST Excellence and Inclusiveness Policy, a plan will be developed for ensuring the full involvement of ITC's representatives in all aspects of the Action implementation, including Action leader positions. Moreover, a particular care will be devoted also to assign strategic functions to young researchers and to promote a gender equilibrium both in the Management Committee and in the leadership positions of the Action.

WG1. Theoretical modelling of information flow in genetic coding and other coding systems. WG1 includes expert researchers in physics, mathematics, chemistry, biology, statistics, computer science and engineering. **Objectives**: Develop a theoretical unifying interpretation of the regularities observed in the genetic code and in coding regions of DNA by comparing and integrating existing models and theories; Apply the different theoretical approaches developed for the genetic code and for coding regions of DNA to its noncoding part. Extend these theoretical approaches to the case of other biological codes.

Tasks: 1.1 Develop existing theoretical models for coding, information management, and the associated dynamical behaviour, in the encoding, transcription and translation of proteins. This task mainly involves models based on number theory, quantum and classical information, and dynamical systems theory that exploit to some extent the mathematical regularities observed in the genetic code and in protein coding sequences (e.g. symmetries, topological features). 1.2 Apply the theoretical approaches developed within task 1.1 for the study of the informational organization of non-coding regions of DNA. In human DNA, only about 2% of base pairs are related to protein coding and about 98% of them are non-coding. According to the international Encyclopaedia of DNA Elements (ENCODE) project, at least 80% of human DNA has some biochemical functions. Hence, understanding the roles that biological information and coding play in non-coding DNA represents a major scientific challenge. 1.3 Explore the possibility of extending the approaches of task 1.1 to other biological codes. Many biological codes, other than the genetic code, have been identified in recent years; for example: the splicing code and the histone code in the genetic apparatus, and the signal transduction codes in the cellular metabolism. The theoretical models developed for explaining the genetic code and protein synthesis represent an important scientific prototypical corpus for studying other biological codes. 1.4 Integrate in a unifying view the different theoretical approaches and build a strong interaction among different research groups within WG1 contributing also to a strong collaborative research with WG 2,3, and 4. To this aim, a specific task force will be created inside the group.

WG2. Statistical and probabilistic modelling and analysis – bioinformatics, data mining, machine learning. WG2 includes expert researchers in modern statistics, probabilistic modelling, bioinformatics, machine learning and data science at large. **Objectives**: Provide expertise in statistics, probability, bioinformatics, machine learning to address modern biological problems and for the analysis of all kind of (high-dimensional) genomic and proteomic data. Provide know-how for data mining, data management and testing theories and models. Provide the statistical and probabilistic support and expertise to mathematical modelling and experimental design.

Tasks: 2.1 Analysis of the big corpus of data coming from molecular biology in specific problems. Modern biological challenges are multifaceted and require a multi-disciplinary approach. Expert researchers in bioinformatics will deal with the more technical and technological aspects such as, for instance, the management of big databases or devising high performance computing solutions to complex combinatorial problems. The experts in statistics and machine learning will interact with them as to provide the expertise and rigour in statistical and probabilistic modelling by using state-of-the-art methods in hot areas such as high-dimensional classification and estimation problems, repeated testing, networks, just to name a few. 2.2 Provide the statistical and probabilistic support and expertise to mathematical modelling. Complement the theoretical modelling step with a statistical/probabilistic approach. For instance, the theory of dynamical systems is closely linked to time series analysis and the theory of stochastic processes. Second, in order to apply a theoretical model to real data and obtain estimates, a necessary step that involves mathematical statistics is needed, usually in the form of deriving the mathematical properties of the estimators. Third, the theory of design of the experiments will be possibly used to devise new laboratory experiments that will enable to test and validate theories and models. 2.3 Provide bioinformatic and machine learning expertise to WG1 to solve computational problems and for the analysis of all kind of (high-dimensional) genomic and proteomic data. The theoretical problems that are addressed by WG1 will pose new challenges that require modern technical expertise in the fields of bioinformatics, machine learning and data science. These include both supervised and unsupervised learning, protein structure prediction, solving combinatorial problems and many more.

WG3. Cross-field Interaction, Validation, and Potential Applications. **Objectives**: Create a multidisciplinary panel board to foster cross-field interaction, validation, and the exploration of potential applications; Stimulate and coordinate the validation and tests of models and theories; Encourage the design of new experiments and explore potential applications suggested by the different modelling frameworks.

Tasks: 3.1 Creation of a multi-disciplinary panel board with theoretical experts from WG1, statistical and computer science experts from WG2, and members of the dissemination panel. This panel will coordinate the collaborative research efforts between WG1 and WG2, and will promote a paradigm shift in theoretical biology, which can be identified as the informational interpretation of life. 3.2 Validation and testing of the different approaches in WG1 with data and tools from WG2 and the interdisciplinary coordination of the Cross-Field Interaction panel of 3.1. 3.3 Exploring the potential applications suggested by the Interdisciplinary results on the research, validation, and testing performed within task 3.2, stimulated and coordinated by the Cross-Field Interaction panel; proposal of new experimental tests on the basis of the results of task 3.2.

WG4. Dissemination and Communication. **Objectives**: Create a panel dedicated to dissemination to facilitate the interdisciplinary communication between different groups of the Action; Develop a prototypical “National Platform” for Dissemination and Scientific Communication in the field of Informational and Theoretical Biology in an ITC country; promotion of this prototypical National Platform as a model for other European countries involved in the Action and also at an international level; promote communication, dissemination, and science-art crossover activities related to informational biology at large remarking the importance of basic research for the evolution and the well-being of the human society.

Tasks: 4.1.- Creation of a multidisciplinary panel dedicated to dissemination and communication; the panel will be preferentially coordinated by a member of an ITC country and will deal with both the dissemination activities to promote multidisciplinary interaction and the communication activities at large. It will also devise tools and procedures to break the communication barriers between different disciplines. The involvement of experts in scientific communication and dissemination is foreseen, also through an expansion of the initial network. 4.2 – Development of a "National Platform" for Dissemination and Scientific Communication in Informational and Theoretical Biology in an ITC Country. 4.3 – Implementation of 5-day theoretical schools for training young researchers, four 3-day Workshops and four 1-day hackathons for training young PhD or Post-Doc researchers. 4.4 – Realization of two short videos, one directed to illustrate the scientific and historical aspects of informational biology, and the second for illustrating the results of the Action and the foreseen social impacts. These videos will be mainly used to promote communication with stakeholders, policy makers, and potential industrial partners. 4.5 – Implementation of a Forum for communication with potential industrial beneficiaries and stakeholders at different levels; invitation of key leaders in related applications and policy makers to seminars and Conferences organized by the Action. 4.6 – Implementation of the digital content, including interactive websites, blogs and wiki pages.

4.1.2. DESCRIPTION OF DELIVERABLES AND TIMEFRAME

WG1, Milestones: M1.1: A unified description of the organization and management of information in coding DNA. M1.2: Extension of the unified description in M1 for this application in non-coding DNA and eventually other biological coding systems. **Deliverables**: D1A: Publication of a special issue with and integrated interpretation of the regularities in the genetic code. D1B: Final report on theoretical results.

WG2, Milestones: M2.1: Integration of a set of statistical methods and bioinformatic tools for testing the theories and models developed in WG1. M2.2: Scientific publications on Statistical and Black-Box models of specific problems. **Deliverables**: D2A: Report on a universal platform for validation and testing, including recommended tools and procedures. D2B: Final report on results regarding WP2.

WG3, Milestones: M3.1: Holistic and unifying interpretation of the corpus of theoretical modelling on the basis of the results on validation and test with existing experimental data. M3.2: Scientific publications about the results on tests and validation of current models and theories. **Deliverables**: D3A: Publication of a position paper on the progress toward the paradigm of Informational Biology. D3B: Final report on the results attained in WG3 together with the identification of potential public health and industrial applications.

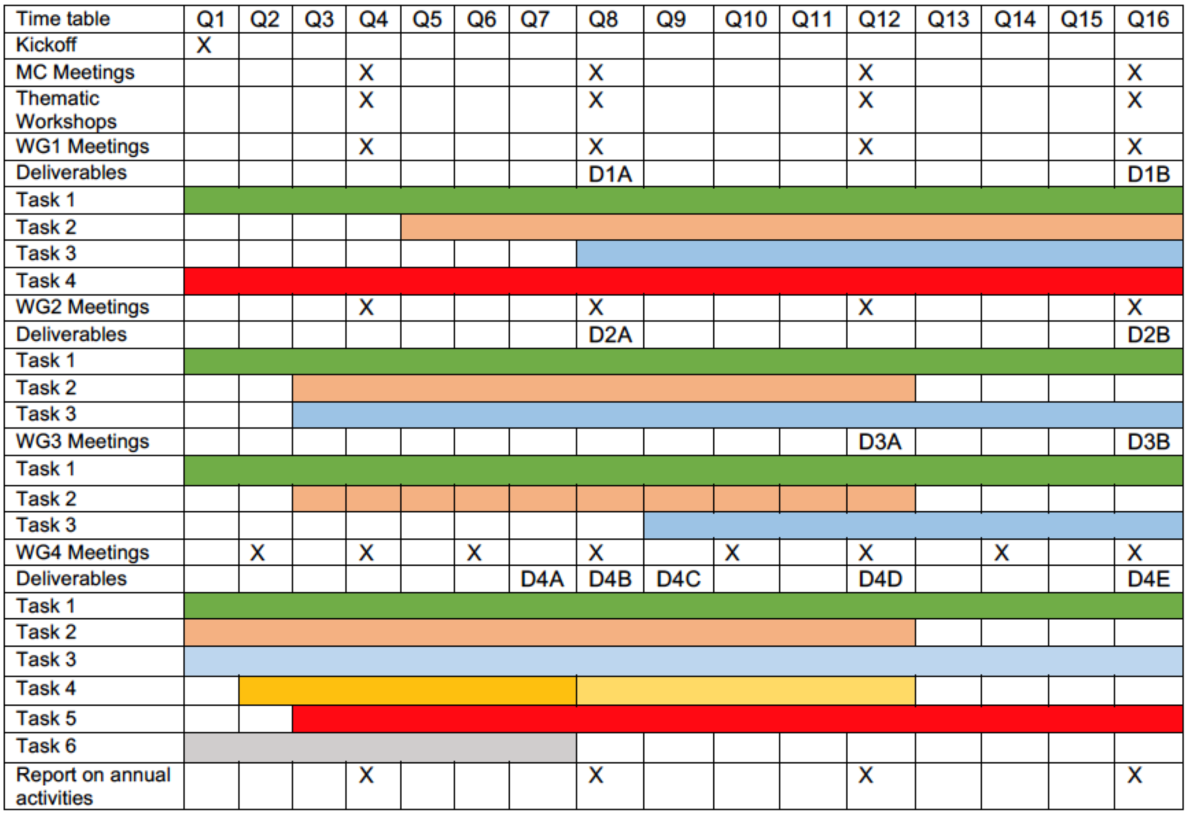
WG4, Milestones: M4.1: Implementation of a prototype of National Platform for Dissemination in the field of Informational Biology. **Deliverables**: D4A: Video Documentary 1. D4B: Report on the prototype of the National Platform for Dissemination. D4C: Report on activities of the 2 first years. D4D: Video Documentary 2. D5E: Report on activities of the 2 final years.

4.1.3. RISK ANALYSIS AND CONTINGENCY PLANS

Risk description	**A:** Prob.	**B:** Impact	Contingency plans
Uncertainty regarding the current corona pandemic situation	3	2	Deliver scientific meetings and workshops in hybrid mode. Promote feasible STSMs and networking activities on a smaller scale and build a platform for dissemination of the results between Working Groups.
Lack of interest from Young researchers and innovators and difficulties to keep the gender balance and the involvement of COST Inclusiveness target countries.	2	3	The Action has almost 50% of Inclusiveness Target Countries and this will ensure the bi-directional knowledge exchange and capacity building in ITC. Gender and age representation will be subject to constant monitoring in order to ensure inclusiveness and geographic balance. The Action will be promoted among participating universities and synergy with other events to raise awareness will be deployed by also using social media. Access to workshops, Short-Term Scientific Missions (STSMs) and conferences will be promoted, especially for Young researchers and innovators, under-represented gender, and participants from IT countries with less developed structures.
Lack of multidisciplinary interaction.	3	3	The Network of Proposers of DYNALIFE includes members with a wide ranging expertise covering all the relevant disciplines for the Action. The multidisciplinary character of the Action will be promoted and monitored through periodic meetings and dedicated activities such as i) the organisation of multi-disciplinary workshops to enlarge and strengthen the research network; ii) the promotion of cross-disciplinary STSMs; iii) the organization of sessions dedicated to the Action in third-party conferences; iv) dedicated dissemination activities.
Limited interest from industries. Difficulties in finding financial resources to carry out computational tasks and biological experiments.	3	3	Upon joining the Action, every member will be invited to get in touch with relevant industrial partners and discuss their involvement within WGs. The collaboration with industries will be subject to constant dedicated monitoring actions. Special invitations to companies to participate to workshops and meetings will be issued. Dedicated dissemination activities aimed at stakeholders will be deployed. A database of potentially interested stakeholders will be created and maintained. At the same time, all the partner universities and research centres both from participating and countries and cooperating IPC countries, such as USA and Canada, will be involved. The synergy between public and private sectors will allow to overcome the difficulties to ensure a proper funding for experiments.

A: Probability of occurrence, 1–5, B: Impact, 1–5. 1: very low; 2: low; 3: moderate; 4: high; 5: very high

4.1.4. GANTT DIAGRAM



Time table	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16
Kickoff	X															
MC Meetings				X				X				X				X
Thematic Workshops				X				X				X				X
WG1 Meetings				X				X				X				X
Deliverables								D1A								D1B
Task 1	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Task 2					🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧
Task 3								🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦
Task 4	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
WG2 Meetings				X				X				X				X
Deliverables								D2A								D2B
Task 1	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Task 2			🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧				
Task 3			🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦
WG3 Meetings												D3A				D3B
Task 1	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Task 2			🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧				
Task 3									🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦
WG4 Meetings		X		X		X		X		X		X		X		X
Deliverables							D4A	D4B	D4C			D4D				D4E
Task 1	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Task 2	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧				
Task 3	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦
Task 4		🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨				
Task 5			🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
Task 6	▨	▨	▨	▨	▨	▨	▨									
Report on annual activities				X				X				X				X

References

- E. Schrodinger, What is Life?, Cambridge University Press, Cambridge, U.K., 1944.
- C.E. Shannon, A Mathematical Theory of Communication, Bell System Technical Journal, Vol 27, 379-424, 623-656, 1948.
- R.W. Hamming, Error detecting and error correcting codes, Bell System Technical Journal, Vol. 29, 147-160, 1950.
- J. D. Watson and F.H.C. Crick, A structure for deoxyribose nucleic acid. *Nature* Vol.171, 737–738,1953.
- H.P. Yockey, R.P. Platzman, and H. Quastler, Symposium on Information Theory in Biology, New York, Pergamon, 1958.
- S.W. Golomb, B. Gordon, and L.R. Welch, Comma Free Codes, Canadian Journal of Mathematics, vol. 10, 202-209, 1958.
- G. Gamow, What is lIfe?, Trans. Bose Rea Inst., Vol. 24, 185-192, 1961.
- Von Neumann, Theory of Self Reproducing Automata, University of Illinois Press,1966.
- C. H. Waddington, Towards a Theoretical Biology, Nature, Vol. 218, 525, 1968.
- H.A. Johnson, Information theory in biology after 18 years, Science Vol. 168, 1545–1550,1970.
- H.P. Yockey, Information Theory and Molecular Biology, Cambridge University Press, UK, Cambridge, 1992.
- L.E. Kay, Who Wrote the Book of Life?, A history of the genetic code, Stanford University Press, California, 2000.
- H.P. Yockey, Information theory, evolution, and the origin of life, Cambridge University Press, 2002.
- E. May, Special Issue on Communication Theory, Coding Theory and Molecular Biology, IEEE-EMB Magazine, 69-81, 2006.
- National Research Council, Globalization, Biosecurity, and the Future of the Life Sciences, Advances in Technologies with Relevance to Biology: The Future Landscape, Washington, DC, The National Academies Press, 2006.
- J. Bryant, Design and information in biological systems, WIT Transactions on State of the Art in Science and Engineering, WIT Press, Vol 27, 2006.
- R.A. Gatenby and B.R. Frieden, Information Theory in Living Systems, Methods, Applications, and Challenges, Bulletin of Mathematical Biology Vol. 69, 635–657, 2007.
- S.I. Walker and P.C.W. Davies, The algorithmic origins of life, J. R. Soc. Interface, 10, 2013.
- S. Chen, A. Ansari, W. Sterrett, K. Hurley, J. Kemball, J. Weddell, and P.I. Imoukhuede, Current State-of-the-Art and Future Directions in Systems Biology, Progress and Communication in Sciences, UNSYS digital, 2014.
- M. Barbieri, Code Biology, A New Science of Life, Springer, 2015.
- A.G. Dimitrov, F. Fekri, A.A. Lazar, S.M. Moser, and P.J. Thomas, Special Issue on Biological Applications of Information Theory in Honor of Claude Shannon’s Centennial—Part 1, IEEE Transactions on Molecular, Biological, and Multi-scale Communications, vol. 2, no. 1, 2016.
- J. H .E. Cartwright et al., DNA as Information, Theme Issue of Phil Trans Royal Soc A, Vol. 374, Issue 2063, 2016.
- V.I.Jurtz, et al., Sequence analysis: An introduction to deep learning on biological sequence data: examples and solutions, Chapter 10, Bioinformatics Vol. 33 (22), 3685-3690, 2017.
- S.I. Walker and C. Mathis, Network Theory in Prebiotic Evolution, Publishing AG, C. Menor-Salván (ed.), Prebiotic Chemistry and Chemical Evolution of Nucleic Acids, Nucleic Acids and Molecular Biology Vol. 35, 2018.
- E. Papakonstantinou, K. Pierouli, E. Eliopoulos and D. Vlachakis, Systems Biology Consolidating State of the Art Genetics and Bioinformatics, Intech Open, 2019.
- S. Brenner, Life’s code script, Nature, Vol. 482, 461, 2012.
- P. Nurse, Biology must generate ideas as well as data, Nature, Vol. 597, 305, 2021.