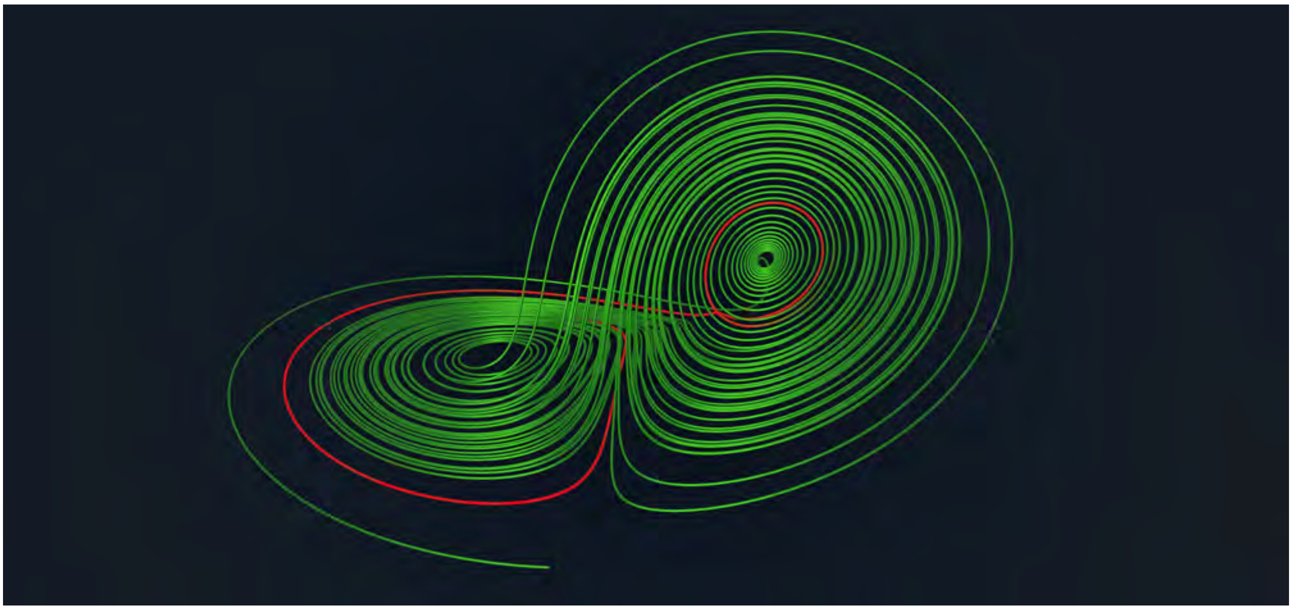
9. Theory at EMBL

Background

As EMBL moves into exploring life in context, there is now a great opportunity and need for a revitalised inclusion of theoretical approaches in the life sciences. Interest in the use of theory in biology has been repeatedly expressed in the past. However, limitations of the experimental possibilities to query biological systems have restricted the use - and impact - of theory to selected, simplified systems. This situation has fundamentally changed in recent years. Researchers now have unprecedented capabilities to study living systems in their holistic context.

It has become possible to perturb living systems in more dynamic and versatile ways than ever before and to extract precise quantitative data using a host of different assays from omics to imaging. Life scientists are now able to experimentally address the dynamics and complexity of living matter across many scales, at high-resolution, and at systems-wide coverage. The multiple scales of living systems (e.g. molecules, organelles, cells, tissues, organs, organisms, ecosystems) are a fundamental aspect of their nature. They are full of emergent properties: properties that only manifest themselves at the integrated systems-level view and that are not directly evident from the properties of the system’s components. Because of that, the working principles of living systems can often not be understood through purely intuitive approaches, but deeper understanding can be gained through formal mathematical reasoning and modelling. A theory- and modelling-guided approach is thus needed and timely, and can now be integrated with improved mathematical techniques, better computational power, and new experimental possibilities in order to **reveal the essential features and general principles of living systems in their natural environment**.



The Opportunity

EMBL aims to build a new conceptual theory programme which will complement the data-driven computational approaches that are already, and will remain, one of the strengths of EMBL (Chapter 8: Data Sciences). Modern biology, with its recently acquired capabilities to produce quantitative data across all scales and to generate unprecedented, dynamic perturbations, is now well-positioned to fully benefit from and partner with theory-guided approaches. Conversely, these developments enable for the first time concrete theoretical advances towards treating the unique complex properties of living systems, with major implications across scientific disciplines. The interplay between theory and experiments aims to uncover general principles in living systems on Earth. An integrated approach builds on EMBL's previous successes of identifying molecular players and interactions and will be essential to **turn data into understanding**.

In one previous example of the successful use of theory at EMBL, researchers in cell biology used a theoretical biophysical model to study the properties determining self-organisation of microtubule and motor proteins. The researchers were able to experimentally test and verify model predictions because of the development of a very simplified and controllable *in vitro* system. Today, the limitations of using theory in only a small number of experimental settings have fundamentally changed because of breakthroughs in experimental and data-generating capacities. Theoretical approaches can now be applied to more problems and in models of higher complexity. For instance, following the successful path of using theory in conjunction with experiments in neuroscience research, researchers at EMBL are building theoretical models of the retina, not only to recapitulate its visual response patterns but also to derive predictions on the neuronal circuit structure and function. These predictions are then experimentally tested to better understand the neuronal principles and mechanisms of visual information processing (Figure TH1).

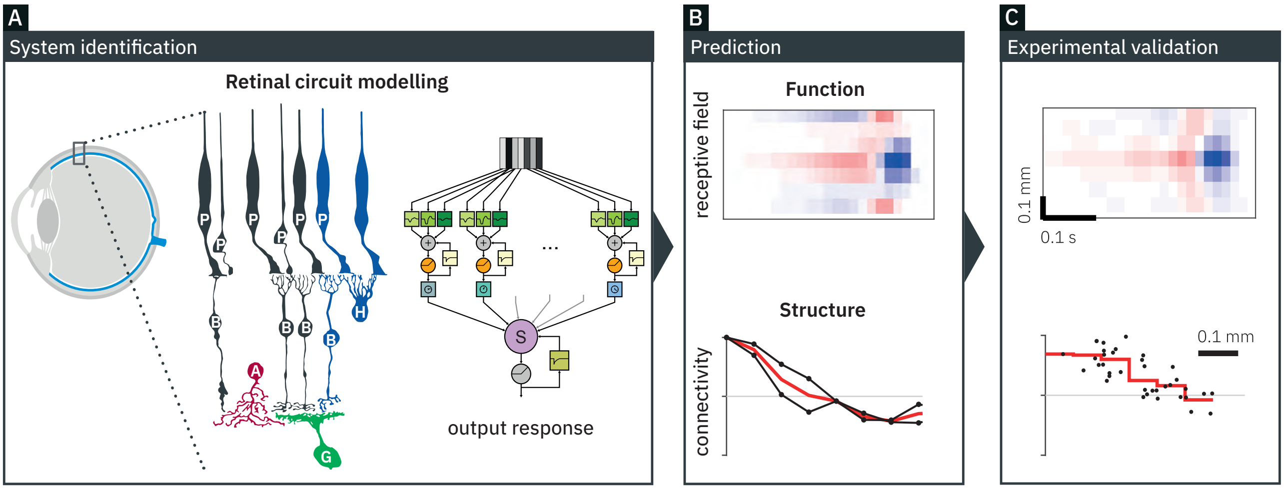


Figure TH1 | Theory-driven approach for neural circuit characterisation, followed by experimental validation.
Neuroscience research faces a need to link big data on brain anatomy and physiology, as high-throughput measurements of these become increasingly feasible. Neural circuit models **(A)** (e.g. of the retina) can provide predictions of such links **(B)**, which can be subsequently tested by experiments for validation **(C)**.

EMBL researchers also use a theory-driven approach to study complex systems; a theoretical framework is used to investigate the origin and function of oscillatory gene expression dynamics in embryonic cell ensembles. Building on concepts from **synchronisation theory**, an entrainment strategy was established and has successfully provided, for the first time, predictive and precise control over gene expression rhythms and oscillation in mouse embryos. This theory-guided approach provides the researchers with an abstracted phase-oscillator model representation of a network that is in reality much more complicated. It offers entirely new insights and reveals a functional role of oscillation rhythm in pattern formation (Figure TH2).

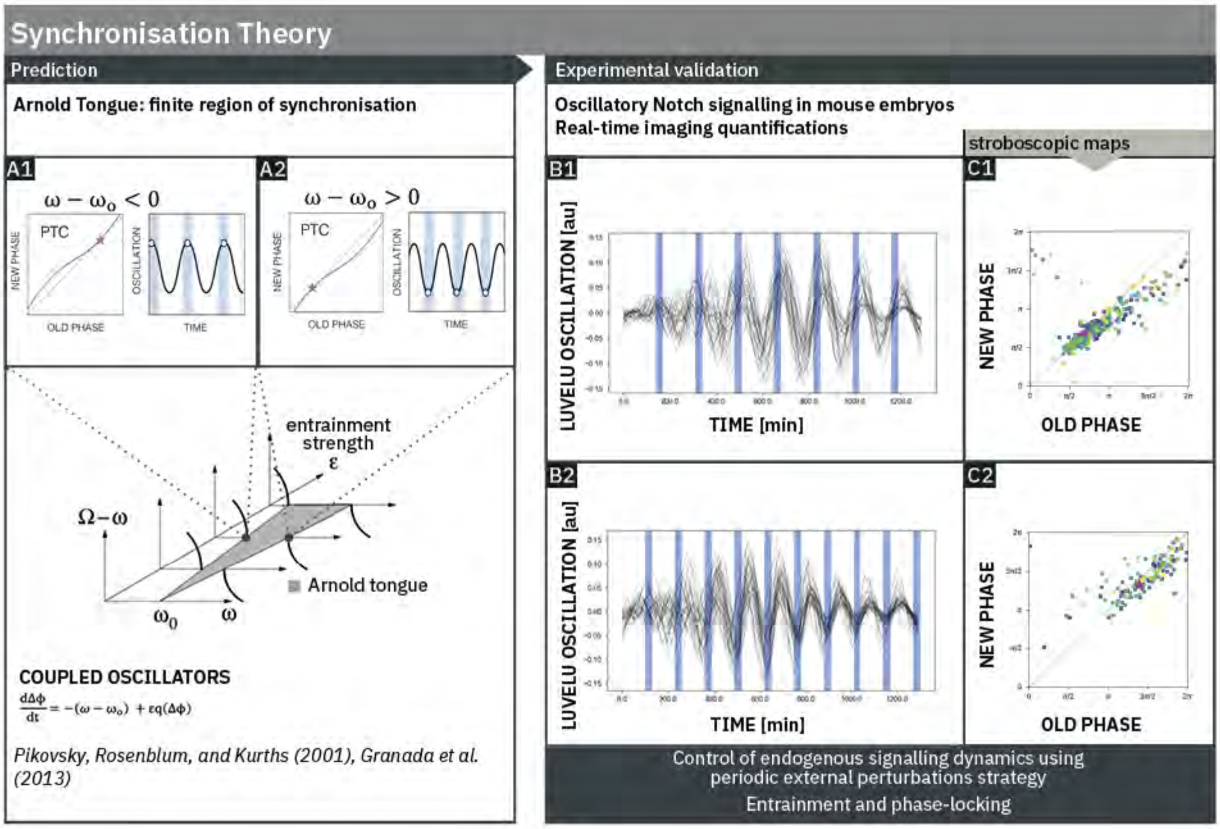


Figure TH2 | Synchronisation theory guides experimental strategy to predict and control segmentation clock oscillations in mouse embryos.

Theory (Arnold tongues) makes universal predictions about the dynamic adjustment of coupled oscillators during entrainment **(A)**. In particular, the mismatch in frequencies $(\omega\text{-}\omega_o)$ directly impacts on the phase-locking after synchronisation i.e. compare **A1** with **A2**. Experimental validation reveals that indeed mouse embryonic oscillators can be controlled (entrained) to different frequencies/periods (**B**, **C**: **B2** = 170 min, **B2** = 130 min). In **B**, individual time series correspond to real-time quantifications of Notch-signalling oscillations in mouse embryonic cells. Importantly, as predicted by theory, phase-locking changes as a function of frequency mismatch (compare stroboscopic map in **C1** vs. **C2**).

Research Aims

The goal of EMBL's new conceptual theory theme is to build up first-principle approaches to gain a novel understanding into the complexity of living systems. It will build on EMBL's strong culture of interdisciplinarity and collaboration, and aims to bring together **the entire community of experimentalists and theoreticians**. This linking aspect is key to reach the goals and to assemble a research programme that integrates theoretical approaches established in other fields, such as physics, mathematics, and information theory, to solve challenging biological problems which, in turn, are expected to inspire new theory.

Research in the theory theme aims to explain biological phenomena using mathematical formalism and models and will cover the whole spectrum of EMBL's research. Theory is **more than collections of models.** It aims to uncover the **essential and general principles that transcend details of a particular system**. Theory aims to **make experimentally testable predictions** and can hence guide, and not just follow, experimentation. Through iterative cycles of theory and experimentation, the goal is to establish a theory-guided pattern to discovery that lifts the quality and efficiency of the scientific process in life sciences at EMBL and beyond.

EMBL aims to promote theoretical research in the life sciences, and as a pan-European organisation it is ideally placed to lead this effort. The goal is to ensure theory is embedded in active, vibrant experimental research and that both approaches are complementary and relevant to each other. One mechanism to achieve this is via the EMBL Interdisciplinary Postdocs (EIPOD) programme which fosters collaborations between interdisciplinary groups at EMBL (Figure TR2 in Chapter 11: Training). For theoreticians from outside the life sciences community, EMBL is an attractive place to immerse themselves and benefit from its wealth of both experimental and computational biology. As an intergovernmental organisation, EMBL is well-positioned to serve as a European exchange node, helping to promote the landscape of ongoing theoretical research in biology across member state institutions, as well as strengthening and growing the field as a whole. Finally, EMBL's wide-ranging and ambitious future research vision outlined in the Molecules to Ecosystems Programme is suitably connected and interlinked by theory and modelling. A theoretical research programme is of strong strategic importance to all aspects of EMBL research, including planetary biology, human ecosystems, infection biology, microbial ecosystems, cellular and multicellular dynamics of life, and molecular building blocks in context, with several examples highlighted below.

EMBL's Approach

EMBL aims to establish a new, dedicated theory theme that will be tightly embedded within the Molecules to Ecosystems Programme. Below are a number of examples which highlight how theoretical and experimental approaches will be integrated to tackle the fundamental questions outlined in this EMBL Programme.

Planetary Biology: The aim of Planetary Biology research is to study how organisms - and populations - respond to and integrate environmental cues and function in ecosystems. Addressing these complex questions requires an integrated theoretical and experimental approach, as highlighted also by the examples described above. For instance, the development of a new theoretical, conceptual framework was instrumental to move forward the biodiversity in ecosystem functioning (BEF) field and allowed it to extract mechanisms from experimental data, revealing a key role for species-complementarity in ecosystem function. With the advancement in the acquisition - and complexity - of data across spatial and temporal scales, as laid out by the Planetary Biology theme, there is an even greater need to complement experimental approaches with advanced **data integration and the development of new theoretical frameworks**, to gain an understanding of underlying mechanisms. **Multiscale modeling** will also be essential to guide **ecological engineering** efforts to develop ecosystems in controlled laboratory conditions, such as mesocosms, and to reveal principles and critical motifs underlying terraformation and, more generally, ecosystem function (Chapter 7: Planetary Biology).

Human Ecosystems: The central question of this new research area at EMBL is how environmental factors can precipitate disease and, more generally, how genotype and environment influence human phenotypes, and how living systems maintain, or fail to maintain, homeostasis. Such questions are being addressed by, for example, **control theory and systems engineering**, which have been developed for human-made devices to make sure systems work robustly at desired operating points. Despite the far-reaching implications that control theory could have for understanding human ecosystems, work in this area is still scarce. Moreover, the study of living systems holds fundamental advances for the field of control theory itself. Theoretical advances in control theory open novel experimental and data-analytical scenarios that would provide deep insights into the universal principles through which homeostasis is achieved, or lost. Importantly, a better understanding of these principles has the potential to open up entirely new strategies, guided by control theory, to steer living systems away from a disease state back to a stable, healthy equilibrium. The theory of **critical fluctuations** and phase transitions might also provide insight into underlying principles for observations in different diseases where large fluctuations precede the onset of the disease, with possible uses for early detection and prevention.

Infection Biology: The potential impact and importance of theoretical models became apparent during the SARS-CoV-2 pandemic of 2020, when decision makers around the world, who had to make some of the most far-reaching political and economic decisions for generations, looked to mathematical models for guidance. These models are highly complex, depend on large numbers of parameters, and integrate many diverse fields of expertise. In this area, theory researchers will be able to contribute components and building blocks for various models e.g. on pathogen diversity and evolution, transmission mechanisms, and the spread of antimicrobial resistance. There is also a great need for better quantitative and predictive understanding of the immune system, including at the individual level, using data from modern single-cell omics assays. Such models could, for instance, help explain, and even predict, the high variance of disease severity after a virus infection.

Microbial Ecosystems: Theory and modelling build our understanding of microbial ecosystems at multiple levels. First, these approaches can help tackle the vast genetic variation of individual microbial species by combining current knowledge on gene function and effects of sequence variation to predict strain traits, such as antibiotic resistance and virulence, based on genome sequence information. Second, coarse-grained molecular or ecological models and/or organism-scale metabolic models can be used to understand species interactions: chemical warfare, co-dependencies, and metabolic cross-feeding. This will open the path for the design of artificial communities, the understanding of emerging behaviours, and the rational modulation of complex real communities, like the gut microbiota of an animal, to acquire specific traits or members.

Cellular and Multicellular Dynamics of Life: Life is multi-scale and dynamic in nature; it emerges from the interplay of countless molecular building blocks and their dynamic interactions, integrating biochemical, physical, metabolic, and environmental cues, and bridging many temporal and spatial scales. As such, living systems differ fundamentally from the problems that classical theoretical fields have focused on previously. The treatment of out-of-equilibrium systems - a defining property of biological systems - is a central challenge of modern theoretical physics. Researchers at EMBL and beyond now have unprecedented experimental possibilities to obtain quantitative, dynamic, large-scale datasets across scales, for instance in developing embryos. Theory is central to extract generalisable meaning from such data (Figure MD4 in Chapter 3: Cellular and Multicellular Dynamics). For instance, **model reduction methods** need to be further developed to identify effective degrees of freedom, state variables, and the important parameters. The goal here is to extract the essence of complex systems in simplified, yet meaningful and predictive models.

Morphogenesis of multicellular systems poses challenges due to pattern and organ formation in space and time. The revolution in microscopy imaging technologies provides an unprecedented wealth of information in four dimensions, posing new challenges to carry out meaningful comparisons and quantify spatially- and temporally-structured information. **Geometrical modelling** studies the formal description of shapes and can provide novel strategies to describe morphological changes across scales and experimental microscopy modalities (Figure TH3). Thanks to theoretical abstraction, the exploration of morphological differences and similarities is reformulated as characterising the geometry of the manifold of shape. It becomes possible to model and quantify morphogenetic processes from different sources of image data and gain insight into mechanisms that allow living systems to develop and adjust their shape over time and under changing conditions.

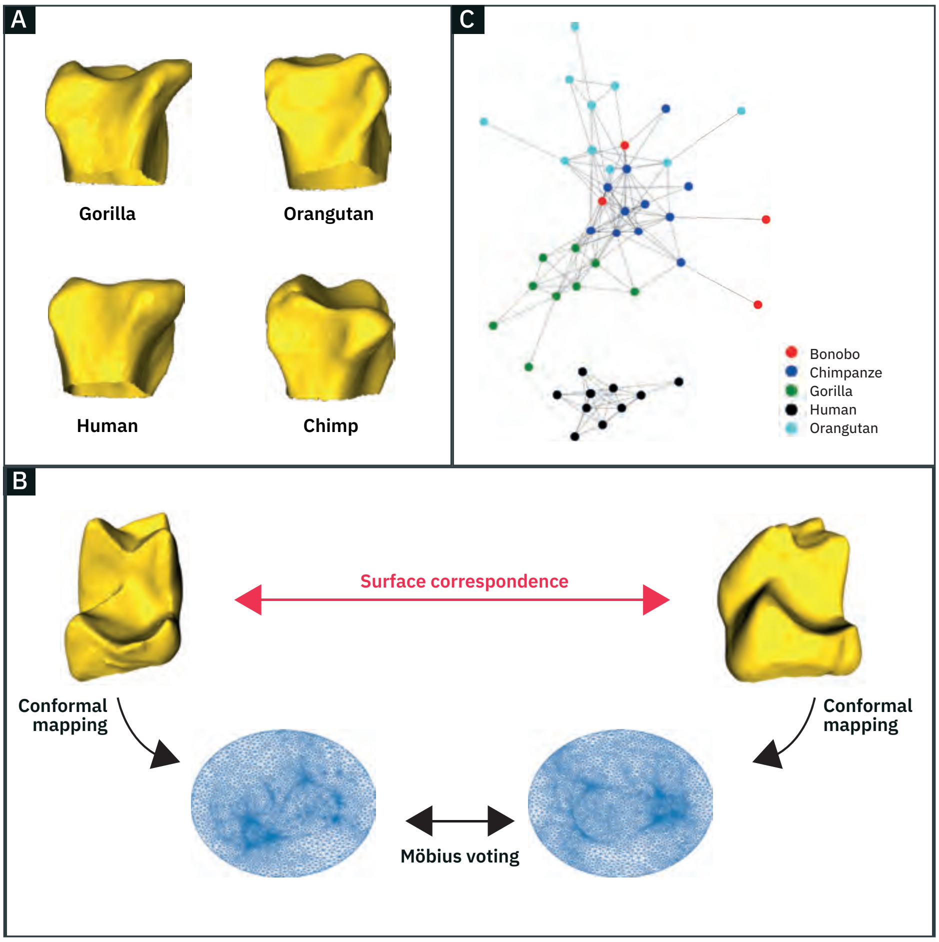


Figure TH3 | Geometrical modelling.
This type of theory is used to quantitatively explore morphological differences, here in tooth development across different ape species and humans. Meshes are extracted from 3D images of samples from each considered species **(A)**. Correspondence is established between meshes with diffeomorphic mapping **(B)**. Inter- and intra-species variations in tooth morphology can be investigated and modelled from these maps **(C)**.

Another important characteristic of all biological systems is time. The rich variety of frameworks for dynamical systems in physics and applied mathematics enables researchers to conceptualise biological phenomena such as cellular decision-making and development in terms of symmetry and symmetry-breaking, bifurcations, effective dimensions, manifolds, attractors, limit cycles, stability, determinism versus stochastics, perturbations, or self-organisation. Excitingly, experimental approaches are now available to link to **dynamical systems theory**. These theories can be instrumental, when combined with experimentation, in revealing the fundamental and universal principles that guide dynamic cellular, multicellular and tissue function (Figure TH2).

Molecular Building Blocks in Context: Molecular building blocks self-organise and exhibit collective properties that can serve to counter the variability seen in living systems. Examples of such collective processes are the functioning of the transcriptional machinery of eukaryotic cells involving tens or hundreds of protein subunits, or the build up and maintenance of cytoskeleton components out of large numbers of monomers. **Statistical mechanics** shows that the joint operation of such interacting elements leads to collective effects including phase separation and phase transitions. **Phase separation** has recently been identified as a fundamental organisation process in cell biology. This provides scientists with an opportunity to use theoretical concepts to address questions such as scaling (e.g. how does the size of membraneless organelles depend on the global physiological state of cells) and the derivation of phase diagrams. In a different context, **phase transitions** can be related to cellular differentiation events. Again, tools of statistical physics are very important to study those processes. In particular, continuous phase-transition points are characterised by critical states in which fluctuations of all scales coexist. Critical behaviour has been associated with optimal information transfer in different physical and biological systems, and it appears to be a fundamental organising principle in biology, although this is still an open question. Additionally, critical phenomena are known to be a signature of universal properties and provide a basis for the **coarse-graining** of collective phenomena. Therefore, identifying critical states within cells and tissues in a univocal manner will allow researchers to establish universal unifying principles of living processes and to determine their correct scale of description.

Spatial modelling and computational simulations are already used to determine structures of biological macromolecules and analyse their dynamics. Better **theoretical understanding of the experimental processes and associated noise models** that generate the data used for structure determination, for example in cryo-electron microscopy, would facilitate building more accurate structures and assessing their uncertainty. Establishing expertise in **molecular dynamics** and other **simulation approaches** would enable understanding the dynamics of molecular systems at multiple scales, from conformational changes in macromolecules to the spatial dynamical behaviour of subcellular systems.

The **core elements** of EMBL's new theory theme:

I. A new **Theory Transversal Theme** (Chapter 14: People, Processes, and Places) will be established at EMBL. This pan-EMBL structure will allow the recruitment of group leaders in the theory area and be an organising structure for visiting scientists (see below). Establishing theory as a research theme at EMBL is essential to make further efforts viable to recruit theoretically-inclined staff across the organisation, from PhD students to group leaders. An environment composed of a critical mass of people following a coherent paradigm is crucial for such research to flourish and acquire visibility. Within this new Programme, EMBL aims to build local alliances between EMBL and European institutes.

A pilot project is being established at EMBL Barcelona: the EMBL-CRG Collaborative Environment for Data-Driven Predictive Modelling. This is an exciting new initiative where visiting and local theoreticians will work together to develop computational modelling from the molecular to the tissue scale, initially focusing on the multicellular dynamics of developing organs, tumour growth, and synthetic tissues.

II. A new **Theory Visitor Programme** will be implemented at EMBL to help bring seemingly disparate scientific disciplines together. Visiting theory researchers will be able to apply for financial support to cover visits at any EMBL site for a period of weeks to months, in order to initiate or deepen collaborations with EMBL groups. It will complement the newly established Theory research theme at its beginnings and it will help strengthen it through time.

A pilot project was established in February 2020 with the aim that three-week to six-month fellowships will help promote theory-based approaches across EMBL through the formation of new contacts and collaborations.
(https://www.embl.org/about/info/scientific-visitor-programme/theoryembl/)

III. EMBL will offer new **training opportunities** in theory-driven approaches for the next generation of life scientists both at EMBL and in the member states. A common language and reciprocal understanding (and appreciation) are crucial to enable the new generation of life scientists to lead a constructive dialogue with theoreticians. Interdisciplinary training will also make it easier for theoreticians to enter into biology and vice versa.

IV. EMBL will hold **conferences and workshops**, including smaller, highly intensive scientific meetings as well as broader European strategy meetings. Some meetings have already been planned, such as the ‘Perspectives of Theory in Biology’, and the EMBO | EMBL Symposium ‘Theory and Concepts in Biology 2022’.

Theory at EMBL will be treated as a fundamental addition not present in the current configuration of EMBL science. It raises the need for, and at the same time will catalyse, even deeper interdisciplinarity. It will be firmly linked to EMBL’s research and training missions, as theory will be supported by these cornerstone missions and will drive them further.

Impact

The proposed new emphasis on a firm interplay between theory and experimentation will be an integral requirement for achieving EMBL’s overall goals outlined in this EMBL Programme. EMBL’s new research directions, from molecules to ecosystems, will only be possible by establishing a firm link between theory and experimentation to uncover general principles and gain predictive understanding. Theory is needed to achieve a sound understanding of the multi-scale hierarchical organisation of biological systems and to design experiments and data acquisition at each relevant scale and across scales. These efforts conversely hold the potential to unlock substantial research advances in theoretical fields ranging from applied mathematics to statistical physics.

Formulating general principles enables researchers to make testable and quantitative predictions, which in turn guide future biological experiments. Predictive models enable systems engineering, including biotechnologies (e.g. synthetic biology for energy, food, or chemicals production), medicine (targeted, curative therapies), geoengineering (understanding and mitigating greenhouse gases emissions and the effects of climate change), and planetary biology (modelling biodiversity loss and resistance spread).

Theory helps to create interdisciplinary contacts and can serve as an accelerator for scientific collaborations, granting theoreticians greater access to data and experiments. A new theoretical research track will also strengthen and complement existing, data-driven, computational, and service activities across EMBL. Embedding a dedicated theory research programme within EMBL will generate outstanding opportunities for fundamental research as well as for the life sciences community. To collect and incorporate input from the community, EMBL researchers invited 16 international theory researchers with a strong link to life sciences to participate in a strategic discussion in which the participants will outline the benefits of bringing more theory to the life sciences and map out the best ways to do so. The participants will share their views on the future potential of theory in biology and will discuss the implementation of theory in the context of EMBL. The outcomes and recommendations that result from this workshop will constitute additional input that will be integrated at a later stage.

With the explosion in data-generating and measurement technologies over the past three decades, biology risks becoming mostly descriptive, with vast amounts of data that are hard to navigate, further increasing the risk of subdividing into evermore specialised subfields that are siloed. Such directions will be contrary to the primary purpose of scientific research. Strengthening theory, coupled with experimentation, will provide a feasible mechanism to support high-quality, resource-efficient, and conceptually advanced science.