SECTION II

2. Molecular Building Blocks in Context

Background

Over the past half century, molecular biology has enabled a deep understanding of the mechanisms that underpin the propagation and evolution of living matter. EMBL has played a major part in many of the advances that have unravelled how the molecular building blocks of life (DNA, RNA and proteins) are used. For example, EMBL research has enabled the scientific community to make significant progress in sequencing and analysing genomes, solving protein structures, understanding processes such as gene regulation and body pattern formation, and deciphering the molecular machinery that viruses use to propagate themselves. These remarkable discoveries have revolutionised our understanding of life. Even so, they constitute only a small fraction of the molecular principles that underlie the diversity of the living world.

A central challenge in molecular biology today is the comprehension of genotype–phenotype relationships. This means understanding not just the contribution of genetic variation, which has been the main focus for molecular biologists to date, but also the environmental factors influencing phenotypes and how life responds and adapts to its environment. To enable life, it is now evident that genetically encoded information must on the one hand be integrated with environmental signals to maintain cellular integrity and activity, and on the other hand to permit adaptation. From unicellular to complex multicellular organisms, living systems must adapt to constantly changing contexts and environments. Most cellular and organismal functions exhibit an inherent regulatory flexibility, which enables them to adapt to external change. This capacity to sense and respond ensures homeostatic function, which is favoured by evolution. How the environment drives the adaptation of organisms and ultimately their evolution remains one of the biggest questions in biology, be this at the level of a microbial community evolving antibiotic resistance, of cells in a tumour responding to their microenvironment, or of changes in an organism's morphology or behaviour due to the availability of food. Numerous examples of such phenotypic plasticity exist in nature, yet our molecular understanding of how they work and how they evolved remains rudimentary or non-existent.

While the impact of the environment on phenotypic outcome is often well described at the organism and population levels, the underlying molecular mechanisms remain relatively uncharacterised. In particular, the extent to which phenotypic variation is due to genetic variation and/or environmental influences has been difficult to assess. The major challenges today are to **determine systematically which environmental signals influence phenotypic variation, how different responses to a changing environment are mediated at the molecular level, and to understand mechanistically how these responses translate into conditionally adaptive or deleterious phenotypes.**

In this new Programme, EMBL researchers will leverage their strengths in molecular biology to explore the **molecular basis of life in context**. This means investigating how an organism integrates its genetically encoded information and intrinsic and extrinsic environmental signals to produce various molecules, macromolecular complexes, subcellular compartments, organelles, cell states, or cell types, which give rise to different phenotypes. Inferring changes in different contexts requires an understanding of how cellular components and processes change over time, how they are interconnected, and how they feed back on one another. This extends from understanding how the genome is differentially expressed in various environmental contexts, right through to how metabolites and gene networks are integrated during metabolic reprogramming of cells (Figure MO1). This knowledge is fundamental to our understanding of life in its natural environment.

In addition to studying how molecular processes are affected by acute changes, whether physiological (e.g. developmental, nutritional, hormonal), toxic (e.g. exposure to chemical pollutants, unusual temperatures), or pathological (e.g. infection), it will be necessary to interrogate how adaptive molecular changes propagate throughout an organism's lifetime, and the longer-term implications of this across generations. In EMBL's new programme, EMBL researchers will take longitudinal approaches to trace adaptive cellular functions in natural settings, in the lab, and upon perturbation. EMBL's current research applies state-of-the-art technical approaches to visualise and measure molecular and cellular heterogeneity, coupled with multiscale technologies to explore molecular processes. This will enable EMBL to tackle the challenge of dissecting how cellular functions adapt to changes in environmental context.

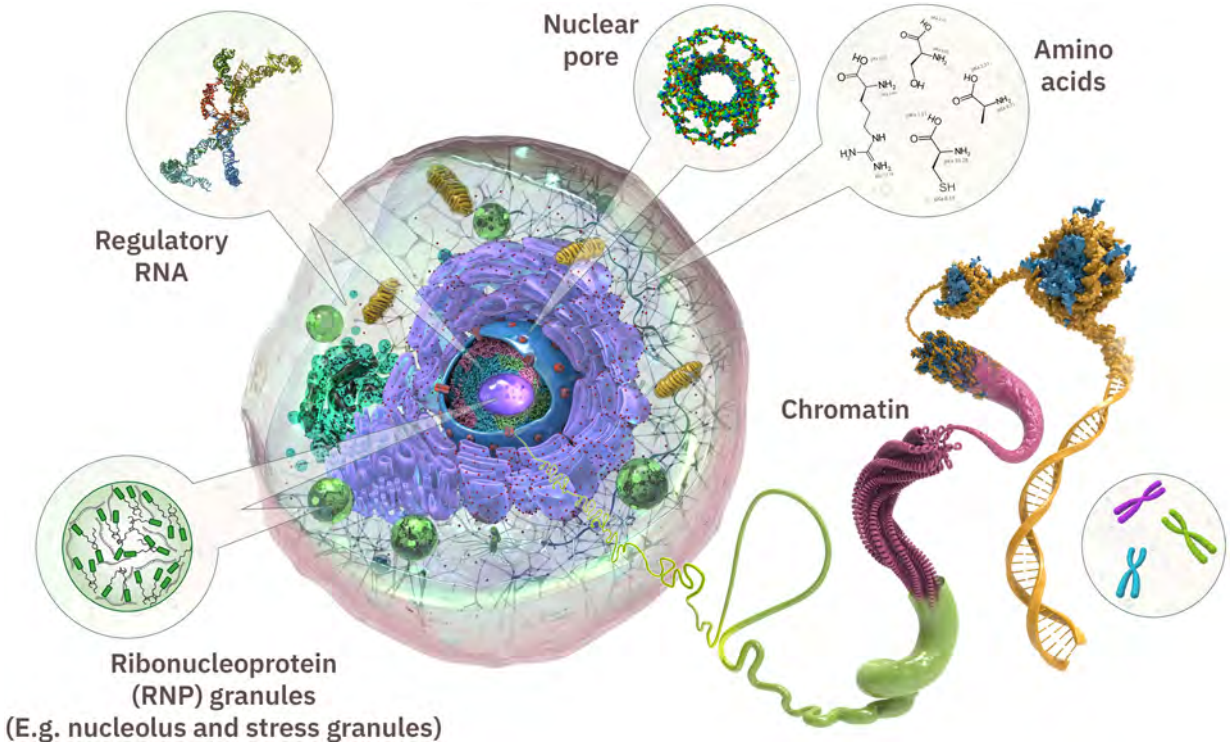


Figure MO1 | A cell contains a diverse range of molecular building blocks.
Illustration of a prototypic eukaryotic cell containing a selection of primary organelles, internal structures, compartments and molecules, which illustrate areas of ongoing research at EMBL that are relevant to the study of responses to biotic and abiotic environmental inputs.

The Opportunity

Building on EMBL's past and current research and strengths in technology development, and taking advantage of exciting new advances in structural biology, multi-omics, genetic engineering, imaging, and computational biology, EMBL can now investigate biological complexity in a way that takes account of environmental variation at the subcellular and molecular levels. EMBL researchers now have the tools to measure inherent variability within and between individuals (e.g. in gene expression), together with the impact of intrinsic signals (e.g. growth factors, hormones, metabolites) and various extrinsic environmental inputs or stresses (e.g. toxicity, nutrition).

At the level of DNA, dynamic protein binding, epigenetic modifications and mutations can be measured at the single cell level; at the level of RNAs, their dynamic processing, modifications, folding, and protein associations can now be captured and linked to physiological and pathological contexts; and, at the level of proteins, their production, folding and allosteric interactions, within macromolecular complexes or subcellular

compartments, such as stress granules, can be followed with a precision that was not even imaginable a few years ago. Furthermore, metabolic and signalling pathways can now be measured dynamically in living systems. These approaches will provide profound insights into the interplay between genetic variation and environmental variables at the molecular level, and will reveal how adaptation to rapidly changing environments is also achieved at the molecular level (e.g. genetic and epigenetic changes, RNA processing, translation alterations, and stoichiometric or allosteric changes in proteins).

Research Aims

In the new Programme, EMBL will strive to provide a deeper understanding of molecular and subcellular processes in the context of the cell, during development and disease, and ultimately in ecosystems. Countless research questions can be asked to investigate the impact of the environment on biological processes. For example, how is the same genome exploited to generate the diverse cell types that make up an organism? How does a single genotype give rise to multiple phenotypes after exposure to environmental signals in the context of phenotypic plasticity? How can exposure to the same environmental factors result in different responses from individuals with different genotypes? How do complex molecular machines orchestrate processes over the lifespan of a cell or an organism, or even across generations? How does metabolism change in response to changes in cellular context, and what effect does this have on subcellular functions? To explore these questions, EMBL will continue to investigate the molecular building blocks of life in different contexts, and will apply systematic and controlled alterations of the environment to study the impact on molecules, their modifications, and the macromolecular complexes they form, as well as on subcellular structures. EMBL aims to:

I. Understand and Predict Function from DNA sequence

The genome is the blueprint of life, but the environment (intrinsic or extrinsic) is key to shaping phenotype. Variations in genomic sequence, epigenetic states, 3D architecture, as well as protein and RNA functions, will be investigated with a view to understanding cellular heterogeneity and phenotypic variation. Specific aims will be:

- To define the physical and dynamical properties of the genome and the factors that influence its functions, particularly gene expression.
- To unravel the structure and function of transcription complexes and other nuclear machineries.
- To explore the many roles of chromatin and the degree to which epigenetic modifications can influence gene expression states.
- To study chromosome folding, its role in genome function, and the mechanisms underlying the folding process.
- To investigate regulatory RNAs at different levels.

II. Develop New Transformative Technologies and Methods

Crucial for the dissection of molecular mechanisms are the technological developments and conceptual models that will lead to a truly molecular understanding of the complex interactions between the environment and biological functions across scales. High-throughput molecular data from (meta)genomics, transcriptomics, and metabolomics, including at the single-cell level, are becoming routine. Following the revolution in imaging technologies and structural biology, molecules can now be visualised with unprecedented spatial and temporal resolution. These technologies must be combined with perturbation strategies (whether genetic, chemical,

or physical) to interfere with the spatio-temporal precision of biological processes. EMBL aims to **develop novel approaches to manipulate molecules and macromolecules**, particularly to enable combinations of approaches (e.g. optogenetics, dCas9, and single-cell omics) to perturb and measure molecules.

For a more mechanistic understanding of the complex interactions between environmental factors and subcellular machineries, molecular structures need to be studied *in situ* to understand how they perform their functions in distinct contexts. The challenge is to capture as much molecular information as possible *in situ* and across timescales while cellular and organismal processes (e.g. cell division, cell differentiation, or reproduction) are actually occurring. In this regard, EMBL aims to develop **new tools for dynamic *in situ* structural biology** to follow genomic information as well as molecular structures and functions over time and in response to intrinsic and extrinsic environmental cues. Machine learning and other AI approaches will be applied to make predictive models of the complex molecular processes that underlie changes in molecular and subcellular structures and functions in different *in vivo* contexts.

III. Understand Subcellular Function in Context

Subcellular systems have been extensively investigated *in vitro*. Many of the specific molecules and complexes responsible for cellular processes have been identified from biochemical purification and *in vitro* manipulations. EMBL scientists will now try to understand the molecules and the subcellular components that they are part of while they are carrying out their functions **in a cellular or *in vivo* context**. EMBL aims to explore several key components and processes to understand their roles in responding to environmental variation:

- **Nuclear organisation**, including nuclear transport, trafficking, and the structure and function of nuclear compartments.
- The **nucleolus** as a key player in sensing and responding to cellular stresses, and **Cajal bodies**, which appear to play a role in key RNA-related metabolic processes.
- Membrane-bound **organelles** such as the endoplasmic reticulum, the Golgi apparatus, vacuoles, lysosomes, and mitochondria play key roles in metabolic pathways and environmental responses.
- **Metabolites** as the functional read-out of cell physiological states and metabolism.
- **Liquid–liquid phase-separated compartments**, such as stress granules, nucleoli, and **intrinsically disordered proteins** (IDPs), which may mediate responses to different types of stress and extrinsic environmental factors.
- The molecular details of **symbiosis**, which may be key to many aspects of cellular function and responsiveness.

Ultimately, EMBL researchers need to understand how organisms evolve and adapt at the molecular level, in the face of environmental variability and in the context of ecosystems. This can only come from a deep understanding of the molecular processes involved. The combination of mechanistic studies, discovery research, and technology, underpinned by a strong interdisciplinary culture, are particular strengths of EMBL that will be leveraged to provide spatial and temporal insight into biological processes in individual cells, organisms, communities, and ecosystems.

EMBL's Approach

Understand and Predict Function From DNA

Even in organisms as diverse as flies, worms, plants, mice, and humans, the number of genes is very similar, ranging from around 15,000 to 30,000 in most cases. What is the basis of biological complexity, if it is not due to the number of genes? One component is the number of different ways the coding genome is utilised within an organism, for example to produce different protein variants through alternative splicing, and different activity states through post-translational modifications. There is also huge variation in the non-coding genome, with humans having 3 billion base pairs in their genome, compared to only 180 million in flies. This vast sea of non-coding DNA is made up in large part of the remnants of transposable elements (mobile genes), which over evolutionary time have provided many of the regulatory elements (e.g. enhancers) that instruct when and where genes should be expressed, as well as structural elements that help to organise the genome in three dimensions within the nucleus. It is becoming increasingly clear from human genome-wide association studies (GWAS) and quantitative trait locus (QTL) studies that most disease-associated variants are in the non-coding portion of the genome, impacting the function of regulatory elements, such as the binding of transcription factors to their target sites.

However, predicting the functional impact of a genetic variant (e.g. a single-nucleotide polymorphism (SNP) or indel) is still incredibly challenging and depends on multiple considerations, including the protein machineries that accompany transcription factors, as well as the chromatin state and 3D organisation of regulatory elements relative to the gene in question, all of which can change in different cell types, tissue types, and environmental contexts. Regulatory elements function in the context of chromatin and the three-dimensional organisation of the genome, folded at various scales, ranging from chromosomal territories to gene loops and nucleosomal arrays. Although each scale is being actively studied, including seminal contributions from EMBL researchers, **scales are typically studied in isolation with little integration across them**. To understand how the physical and molecular properties of the genome are utilised during processes such as transcription, DNA replication, and nuclear transport, a more holistic approach is required, integrating *in vitro* and *in situ* techniques with different resolutions. With cutting-edge expertise and new methods in genomics, cellular imaging, and structural biology, EMBL is particularly well suited to apply such an approach in the context of the next Programme.

The Physical and Dynamical Properties of the Genome

The dynamic, transient, and apparently stochastic way in which transcription factors interact with their cognate DNA binding sites makes **establishing the mechanistic links between transcription factor binding and transcriptional outputs** extremely challenging. This is confounded by the influence of protein–protein interactions, chromatin avidity and function, 3D nuclear 'hubs' of activity, and the emerging concept that nuclear factors can change their physical state and undergo phase separations to function. Transcriptional networks, including redundancy and feedback regulation, have clearly evolved to buffer the deleterious impact of genetic variation or environmental changes. Predicting function from sequence is therefore a huge challenge, and it is a pressing, open question as to how much can be predicted from sequence data alone.

EMBL researchers are tackling this challenge from multiple directions. Approaches from structural biology are being used to understand how transcription factors recognise DNA. This is complemented by high-resolution genomics approaches pioneered at EMBL, such as single-molecule footprinting. This method reveals which base pairs are occupied by transcription factors, and – importantly – which combinations of transcription

factors are bound to the same molecule of DNA at the same moment in time. The integration of these two approaches will facilitate more accurate modelling of DNA–protein interactions and how they are modulated in different contexts, and will form a basis for new machine learning approaches.

Transcription factors also bind to each other, to co-factors, and to multi-protein transcriptional complexes, such as the Mediator and SAGA complexes, forming very large protein–nucleic acid assemblages that are estimated to be in the megadalton range in some cases. Importantly, the makeup and interactions of such complexes are thought to vary dramatically at different stages of development or in the context of environmental fluctuations such as hormonal responses or nutritional conditions. Such complexes are now visible due to recent advances in single-particle cryo-EM and cryo-ET, which EMBL researchers are optimising to **visualise transcriptional complexes in the context of the nucleus**. Some of these large assemblages can form biomolecular condensates, and have the physical property of being able to phase separate *in vitro* and to form punctate areas of increased concentration *in vivo*. EMBL researchers are dissecting the molecular properties of phase separation using biophysical techniques (e.g. optical tweezers, atomic force microscopy), biochemical approaches (e.g. solubility mass spectrometry) and high-resolution fluorescent imaging approaches (e.g. fluorescence correlation spectroscopy, fluorescence recovery after photobleaching, or lattice light-sheet microscopy). The dynamics of these large condensates, or transcription factor 'hubs', is being quantified using live-imaging approaches, for example in the context of embryonic development. Such microenvironments contain localised concentrations of transcription factors and cofactors, and are thought to support robust transcription at enhancers that contain low-affinity sites. In the next EMBL Programme, researchers will test this hypothesis by exploring the function of microdomains and phase separation in gene expression or other genome functions (e.g. DNA repair), by integrating genetic engineering and optogenetics with cutting-edge lattice light-sheet microscopy and mass spectrometry.

Transcription Complexes

Eukaryotic transcription initiation requires the recruitment of DNA-dependent RNA polymerases to the transcription start site to form very large, multi-subunit transcription initiation complexes. EMBL researchers have contributed to unravelling the molecular mechanisms of eukaryotic transcription initiation by determining structures of RNA polymerase I and III, and providing molecular insights into how viral RNA polymerases – in particular influenza RNA polymerase – interact with and hijack the eukaryotic transcription machinery to transcribe their own genes (Chapter 5: Infection Biology). The next challenge is to **locate eukaryotic and viral RNA polymerases in their cellular context and study their function** using a combination of high-resolution imaging approaches (e.g. *in situ* cryo-ET, correlative light and electron microscopy (CLEM), and super-resolution microscopy) and complementary genomics approaches. Using such an integrated approach, EMBL researchers will also study how chromatin modifiers and remodellers affect chromatin architecture and contribute to gene activation or repression. Using structures of nucleosome remodellers bound to single nucleosomes (such as the INO80–nucleosome complex) as starting points, the interaction of remodellers with the chromosomal landscape, for example around gene promoter regions, will be explored first in reconstituted systems, and subsequently *in vivo*, in the nucleus.

Chromatin States and Functions

Alongside transcription factors, chromatin plays a key role in regulating gene activity and helps to maintain expression states during development, conferring an epigenetic cellular memory. Epigenetic mechanisms can also respond dynamically to external cues to guide genome outputs. Chromatin-based systems can thus represent molecular mediators of cellular responses to intrinsic and extrinsic signals. The future challenge

is **to determine whether epigenetic information can be instructive in creating specific gene expression states, and to understand how epigenetic information interacts with genetic information, particularly in the face of environmental change**. EMBL scientists are utilising multimodal synthetic approaches to model the functional consequence of precise chromatin perturbations. For example, by exploiting epigenome editing and chemical approaches as perturbation tools, the hierarchy and causal function of chromatin changes will be investigated and used to understand the extent to which chromatin forms an allosteric network of interactions that encode regulatory information. This has direct implications for understanding cellular plasticity and information processing within biological systems, and will make it possible to test hypotheses about the direct functionality of chromatin and the extent to which it acts as an effector of past and present cellular context. These strategies are complemented by approaches to understand allosteric networks using whole-genome reconstitutions in conjunction with cryo-EM, integrating the wealth of sequencing-based data with atomic resolution information from structural studies. EMBL scientists will also develop tools to probe next-generation molecular dynamics, allowing allosteric communication of factors embedded in native chromatin environments to be studied.

One pioneering project aims to systematically determine the functional role of chromatin-based information *in vivo* by generating an extensive resource of precision-edited mice programmed with specific functional mutations in a large cohort of genes involved in epigenetic processes, which are of particular interest in the context of responsiveness to environmental change (Chapter 10: Scientific Services). This exploits EMBL's optimised pipeline for generating mammalian CRISPR–Cas9-edited genomes, which is available to scientists within all EMBL member states. This pan-EMBL project enables researchers to move beyond conventional loss-of-function studies, which typically obscure whether molecular consequences are due to the absence of a protein or complex or to its enzymatic activity per se, to provide a deeper mechanistic understanding of the contextual function of chromatin states *in situ*. EMBL has a longstanding track record of pioneering studies in multi-omics platforms, and will integrate these platforms with emerging areas of mathematical image processing. The engineered mice and reagents will form a framework platform to further elucidate the interactions between epigenetic and genetic function in specific cell types and cell contexts, and will be freely accessible to the scientific community as an extensive resource.

Chromosome Folding

The genome is non-randomly organised within the nucleus. Genome folding can create proximity between genes and regulatory elements, leading to the accumulation of regulatory factors and the formation of local microenvironments. Short-range chromatin looping occurs in the context of larger chromatin domains, which operate in chromosome territories. Understanding the roles of the complex and dynamic organisation of the genome for its various functions, and the **interplay between DNA sequence variation, genome folding, and gene expression**, will be key to understanding phenotypic variation and DNA damage responses in the context of environmental change.

Recent years have seen a true revolution in the understanding of higher-order genomic structure, due to the development of new molecular approaches such as chromosome conformation capture (3C and genome-wide Hi-C) and advances in super-resolution microscopy. Recent structural, perturbation, and imaging approaches have provided profound insights into the macromolecular structures involved in genome folding in well-studied model organisms such as humans, mice, flies, yeast, and bacteria. Examples of such pioneering research from EMBL include solving the structure of chromosomal replication domains *in vivo*; elucidating the dynamics of mitotic and meiotic chromosome formation; understanding the structure and timing of chromatin domains during early mouse and *Drosophila* embryogenesis; dissecting the functional roles of chromatin topology during X-inactivation, and of chromatin looping during developmental enhancer–promoter communication; and the first demonstration of DNA loop extrusion by condensin. In the next

Programme, EMBL researchers from diverse disciplines will collaborate to address longstanding fundamental questions related to chromosome organisation and function. High-resolution imaging (e.g. cryo-ET, CLEM, super-resolution light microscopy) and complementary single-locus localisation techniques (e.g. OligoDNA-PAINT, DNA FISH) will be combined with genomics approaches to gain structural insights into specific chromosomal loci that are well characterised at the genomic level.

Such an interdisciplinary approach will be used to elucidate the native structure of activated and repressed genes, the molecular mechanisms controlling 3D organisation and gene expression changes, and their interplay during epigenetic processes such as X-inactivation, as well as the nature of interactions between enhancers and promoters. The interactions of promoters with their cognate enhancers and, more generally, between remote chromatin sites will be studied in the context of embryonic development using model systems. Building on seminal contributions from EMBL researchers in elucidating when, where, and how these contacts are first made, future directions will focus on their real-time kinetics and physical properties, and their regulation and functional role in transcriptional initiation. Whole-genome reconstitutions in conjunction with cryo-EM will enable EMBL researchers to link the wealth of sequencing-based chromatin folding data with atomic-resolution information. This integrated approach will be combined with perturbations to the system, either mutants in *cis* or depletion of proteins in *trans*, to determine the contribution of specific chromatin modifiers to changes at the structural and transcriptional level.

Regulatory RNAs

The functional consequence of transcription is to produce RNA, but there is accumulating evidence that RNA plays diverse roles in the regulation of transcription. Not all RNAs are translated – non-coding RNAs (ncRNAs) play fundamental roles in the regulation of chromatin structure, epigenetic processes, and translation. EMBL scientists recently discovered that a huge repertoire of proteins bind to RNA, and that this RNA binding may have a regulatory role in altering the functional state of the protein in many cases. Many of these interactions are through non-conventional RNA binding domains, and may involve intrinsically disordered regions. Future research at EMBL will include the study of the **functional role of these new RNA–protein interactions in the context of transcriptional regulation in the nucleus and regulatory processes occurring in the cytoplasm**, including autophagy and cellular metabolism. Emerging evidence suggests that ncRNAs are also involved in mediating a cell's ability to respond to environmental perturbations. This may be mechanistically linked to their role in epigenetic processes such as gene silencing, by the regulation of chromatin structure and DNA methylation. EMBL researchers study ncRNAs in the context of Polycomb-dependent gene silencing, X-inactivation, embryonic development, p53-dependent stress response, and response to viral infections. Solving the 3D structure of long non-coding RNAs will provide key information for understanding the functional mechanisms of this unexplored group of non-coding transcripts (Chapter 5: Infection Biology).

In addition to DNA and chromatin epigenetic modifications, RNA is also post-transcriptionally modified, and recent studies indicate that there are hundreds of different chemical modifications in various RNA species throughout all domains of life. RNA modifications thereby add an additional, potentially reversible, and dynamic layer of information that can fine-tune the functional properties of RNA in response to environmental cues. To understand the principles of **RNA editing and its functional significance**, researchers at EMBL are using crystallography and single-particle cryo-EM to study the structure–function relationships of complexes involved in RNA editing and processing. They are also using genomics-based approaches to uncover all genome-wide RNA modifications and perturb them through genome engineering. RNA editing complexes such as APOBEC1 will also be exploited using synthetic biology approaches for ‘transcriptome editing’, akin to CRISPR–Cas systems, which is one exciting avenue EMBL researchers are exploring.

Develop New Transformative Technologies and Methods

Novel Approaches to Manipulate Molecules and Macromolecules

Many tools are available to manipulate genomes, RNA, or proteins under different conditions and to visualise and measure the impact of these changes both *in vitro* and *in vivo*. Manipulation of the genome and other molecules can be achieved with unprecedented precision and ease, thanks to CRISPR–Cas9, transcription activator-like effector nucleases (TALEN), and other technologies. EMBL researchers can explore higher-order genomic structures thanks to super-resolution microscopy and molecular approaches such as DNA FISH and chromosome conformation capture (3C and genome-wide Hi-C). These technologies can be applied at the single-cell level and provide spatio-temporal and allele-specific information on genome dynamics. Researchers can genetically engineer genomic tags for imaging or barcoding DNA, RNA, or proteins in cells or whole organisms. Introduction of precise genetic variants, and of DNA sequences that enable the inducible expression (drug-induced) or deletion (LOX-, viral-, or CRISPR-mediated) of a gene, or the degradation of a protein (via a degron) are possible. Optogenetic approaches allow light-mediated manipulation of proteins *in vivo*, often translocating a protein from one subcellular compartment to another (Tech Dev Box TD1_MD). Furthermore, the integration of high-throughput data in (meta)genomics, transcriptomics, and metabolomics facilitates the move from structure to function, and enables biological discoveries and the modelling of processes in various areas of biology. EMBL's unique strength in **technology development** will allow the generation of the tools needed to drive discovery.

New Tools for Dynamic *In Situ* Structural Biology

EMBL will continue to develop the tools that enable dynamic structural cell biology *in situ*, providing scientists with an understanding of how molecular structures perform their functions in distinct contexts. This will include new correlative methods and approaches for dynamic super-resolution microscopy that can be applied to larger, more physiological specimens. New technologies are also needed to image the physical and chemical properties of biological matter and its environment, including forces, tension, density, and ionic strength, and EMBL needs to combine measurements with quantitative sampling of dynamic 3D cellular morphology. For many biological models, penetrating deep into the tissue is key to studying cells in their native environment, so EMBL researchers will continue to advance deep tissue imaging with subcellular resolution and molecular and physical readouts (Tech Dev Box TD2_MD). Many molecular processes happen on very short timescales that cannot currently be sampled. High-speed live imaging with a time resolution on the order of nanoseconds will open a completely new world, allowing scientists to watch molecular mechanisms unfold in live cells. The next generation of detectors, required to make this a reality, is now starting to emerge. Underpinning many of these imaging technologies are new chemical biology approaches to label and highlight molecules, their modifications, and activities. EMBL will also continue to build tools to bridge spatial scales. EMBL has world-leading instrumentation and tools for crystal harvesting, along with advanced synchrotron X-ray technologies. These also include integration of imaging approaches from *in vitro* molecular imaging by nuclear magnetic resonance (NMR), X-ray crystallography, and single-particle cryo-electron microscopy (cryo-EM); and *in situ* imaging using cryo-electron tomography (cryo-ET) of macromolecular complexes, subcellular structures, cells, and organisms.

The technologies applied will generate vast amounts of various data types, which will need to be integrated into predictive models that can be tested and refined (Chapter 3: Cellular and Multicellular Dynamics). To process multiple biological data types on a large scale, a variety of computational approaches will be applied, including molecular dynamics, dynamical simulations, advanced data science, and machine learning. These models will be further refined and validated through an iterative cycle of experimental perturbations and

modelling to reach accurate, predictive models at different scales. By integrating models, a comprehensive atlas of cellular processes can be established with the goal of understanding the cellular basis of life. The Theory at EMBL theme will be particularly relevant for developing the conceptual frameworks that will underpin these models (Chapter 9: Theory at EMBL).

Understand Subcellular Function in Context

Eukaryotic cells contain subcompartments and organelles that carry out distinct functions. Organelles are generally surrounded by membranes and contain specialised transport systems for the exchange of macromolecules and metabolites between the organelle and the cytosol. The nucleus is the largest organelle in eukaryotic cells and, despite intensive research, is in many ways still poorly understood. Multiple cellular compartments also exist throughout the cell, which facilitate spatio-temporal regulation of biological reactions but do not possess membranes. These compartments include stress granules in the cytoplasm and nucleoli, and Cajal bodies in the nucleus. Understanding the functional role of nuclear architecture and the dynamic structure of nuclear membraneless organelles will be major topics in EMBL's next Programme.

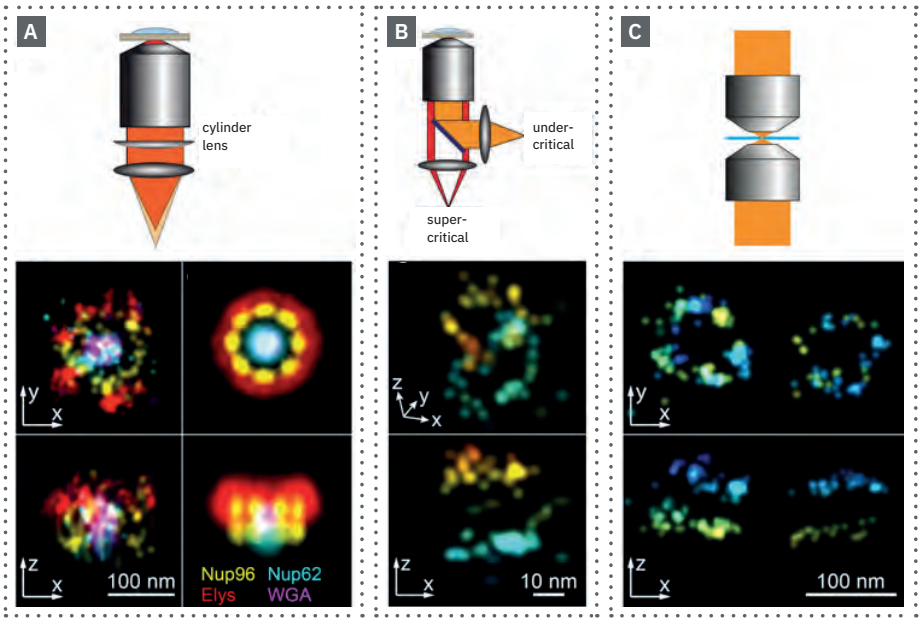
Nuclear Organisation

The nucleus not only contains the genome of the cell, but is also the environment within which several major cellular processes must be orchestrated. These include transcription, DNA replication, chromosome assembly in preparation for cell division, and DNA repair in response to damage. The dynamic organisation of the nucleus is currently a major research focus at EMBL, and several insights into the architecture and dynamics of very large molecular assemblies in their nuclear context have been made. A recent example has been the collaborative work of EMBL researchers to analyse the architecture and dynamics of the nuclear pore complex (NPC), which spans the nuclear membrane and coordinates the exchange of macromolecules between the nucleus and cytoplasm. The elucidation at EMBL of the NPC structure in humans and in an alga, using structural biology techniques, super-resolution microscopy (Tech Dev Box TD1_MO), and modelling, has revealed its intricate molecular structure, with more than 1,000 proteins in the human NPC. EMBL researchers are now exploiting *in vivo* systems to investigate the dynamic assembly and turnover of the NPC in context. Using yeast genetics and *in situ* structural biology, an essential role for a subset of nucleoporins (the protein building blocks of the NPC) in both mRNA export and NPC turnover has recently been defined.

While detailed insights into the structure, dynamics, and function of the nuclear pore at the periphery of the nucleus are emerging, a full understanding of **how multimolecular complexes contribute to shaping the functional architecture of the nucleus** has not yet been achieved. From genomics studies, insights into the molecular profiles of specific chromosomal loci and their changes between active and inactive states are available. In addition, key molecular players in the nucleus and their activities have been defined, and nuclear architecture has been studied extensively at various scales. However, very little is understood about how the **different levels of nuclear architecture affect various functional processes, how functions are controlled across space and time, and how the environment influences function**, including DNA replication, transcription, RNA processing, and nuclear transport. By investigating the interfaces between these processes using a combination of *in vitro* and *in situ* techniques across scales, major breakthroughs in understanding structure–function–environment relationships can be made.

Technology Development Box TD1_MO | 3D Super-resolution microscopy.

Single-molecule localisation-based super-resolution microscopy (SMLM) reaches nanometre resolution and can provide structural insights into cell biological questions. However, such high resolution has been limited to 2D measurements. To enable 3D super-resolution imaging with structural resolution, the Ries Group is developing three complementary approaches: **(A)** A new data analysis workflow that takes into account specific optical properties of the microscope allows the extraction of precise 3D positions from the shape of the single-molecule images. This approach works in up to four colours simultaneously, can capture entire cells, and is compatible with automated high-throughput SMLM. **(B)** The group is developing 'supercritical-angle localisation microscopy' in which the near-field emission of fluorophores is evaluated to obtain a *z*-resolution of a few nanometres in the vicinity of a glass cover slip. **(C)** 4Pi-SMLM is being implemented collaboratively and interferometry is being used to obtain nanometre isotropic resolution, including in thick samples. These technologies allow the Ries group and collaborators to extract precise positions of proteins in complex molecular machines in all three dimensions, which cannot be achieved with other imaging methods. This will help to unravel the structural arrangement of the proteins driving clathrin-mediated endocytosis and, in collaboration with the Ellenberg group, the structure of the nuclear pore complex.



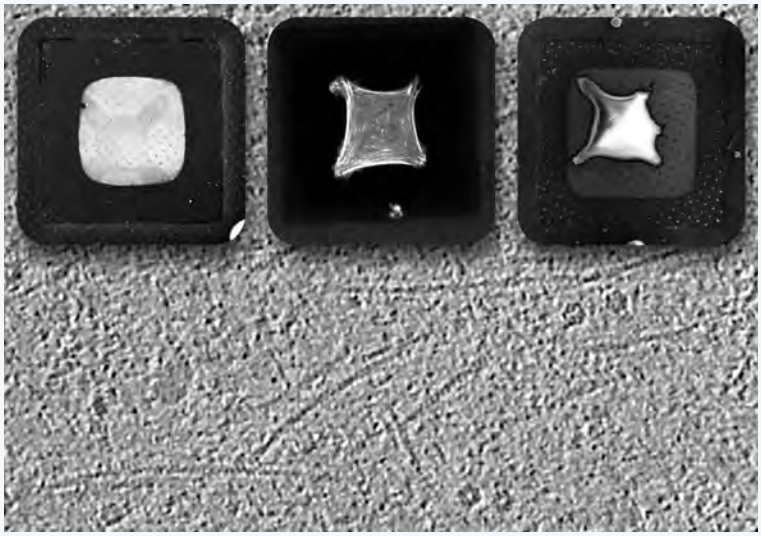
Nucleolus and Cajal Bodies

The nucleolus plays a key role in sensing and responding to cellular stresses such as hypoxia, pH fluctuations, and DNA damage, as it pursues the energy-intensive process of ribosome biogenesis. Inhibition of RNA polymerase I transcription is the first step of ribosome assembly, and leads to changes in nucleolus morphology. **What regulates the shape and size of nucleoli and their massive reorganisation during environmental perturbations like stress? What are the functional consequences of these morphological changes?** To address these questions, it is imperative to connect live imaging data to structural data. To do this, EMBL researchers are developing a framework to locate the three nucleolar subcompartments that specialise in the various steps of ribosome assembly, so they can track proteins by live imaging. Subsequently, macromolecular complexes (e.g. pre-ribosomes or transcription complexes) will be visualised *in situ* by cryo-EM and CLEM (Tech Dev Box TD2_MO; Tech Dev Box TD_SS2), and their molecular contexts will be probed by proximity labelling.

Cajal bodies have been implicated in RNA-related metabolic processes such as biogenesis, histone mRNA processing, and recycling of splicing small nuclear ribonucleoproteins (snRNPs). Cajal bodies also contain large protein complexes involved in the expression of genetic information, such as the Integrator complex. Paralleling work on the nucleolus, EMBL researchers aim to understand the *in situ* structure of macromolecular complexes such as Integrator, and how the processes leading to snRNP assembly are linked to the subcellular architecture of Cajal bodies. Ultimately, this approach aims to provide a **comprehensive and dynamic picture of pre-mRNA processing machineries at the multiscale level** as they carry out their functions in their native cellular environment. Finally, as both nucleolus and Cajal bodies are membraneless, phase-separated organelles, the extent to which phase separation contributes to and influences the molecular functions of each compartment will be explored.

Technology Development Box TD2_MO | Photo-micropatterning for cryo-EM.

The Mahamid Group, together with collaborators in France, have developed a new technique called photo-micropatterning for applications in molecular-resolution cryo-EM of intact cells. Photo-micropatterning allows spatially controlled cell adhesion and the manipulation of cell shapes on cryo-EM grids. This technology overcomes one of the technical challenges presented by cryo-ET, whereby only cells that are positioned in the centre of a grid square are available for processing and therefore imaging. Micropatterning enables the position of cells on the grid to be controlled with a high degree of spatial accuracy, increasing the number of cells available for imaging from only a few to about 30. Micropatterning can also be used to manipulate the shape of cells and study their mechanical behaviour. What's more, numerous different patterns can be generated on the same grid, allowing direct comparison of different intracellular architectures, including the cytoskeleton, the nucleus, or the Golgi apparatus. By understanding the three-dimensional architecture of macromolecules, the collective behaviours that give rise to new mechanical properties can be explained. For example, in the image below, cryo-ET of an RPE1 cell (background image) reveals branching actin filaments and hexameric densities related to cell adhesion. The insets show a cell grown on a cross-shaped micropattern. This technical advance will help to streamline the cryo-EM pipeline and facilitate automation of the process. It's a significant bridge between structural and cell biology that will enable this technology to become a routine method in the future.



Membrane Trafficking and Organelle Biosynthesis

Beyond the nucleus, the eukaryotic cell contains many membrane-bound organelles, including the endoplasmic reticulum, the Golgi apparatus, vacuoles, lysosomes, and mitochondria (Figure MO1). The basic processes involved in membrane trafficking and organelle biogenesis in the secretory pathway, such as vesicle budding, fusion, or transport, are well defined. **How these processes are integrated and give rise to the size and shape of specific organelles and how this relates to diseases such as cystic fibrosis, lung fibrosis, or cardiovascular disease-related cholesterol homeostasis** remain major questions in the field. EMBL researchers use and continue to develop live-cell imaging and systematic genetics on a systemic scale to achieve an understanding of these processes. Quantitative data can then be used to test mechanistic models. This has already revealed a number of interesting factors that respond to stimuli such as cellular cholesterol depletion, growth factor-induced growth control, DNA damage, or cell differentiation, by changing the localisation of membrane proteins or membrane traffic-associated proteins from the cytoplasm to the nucleus or vice versa.

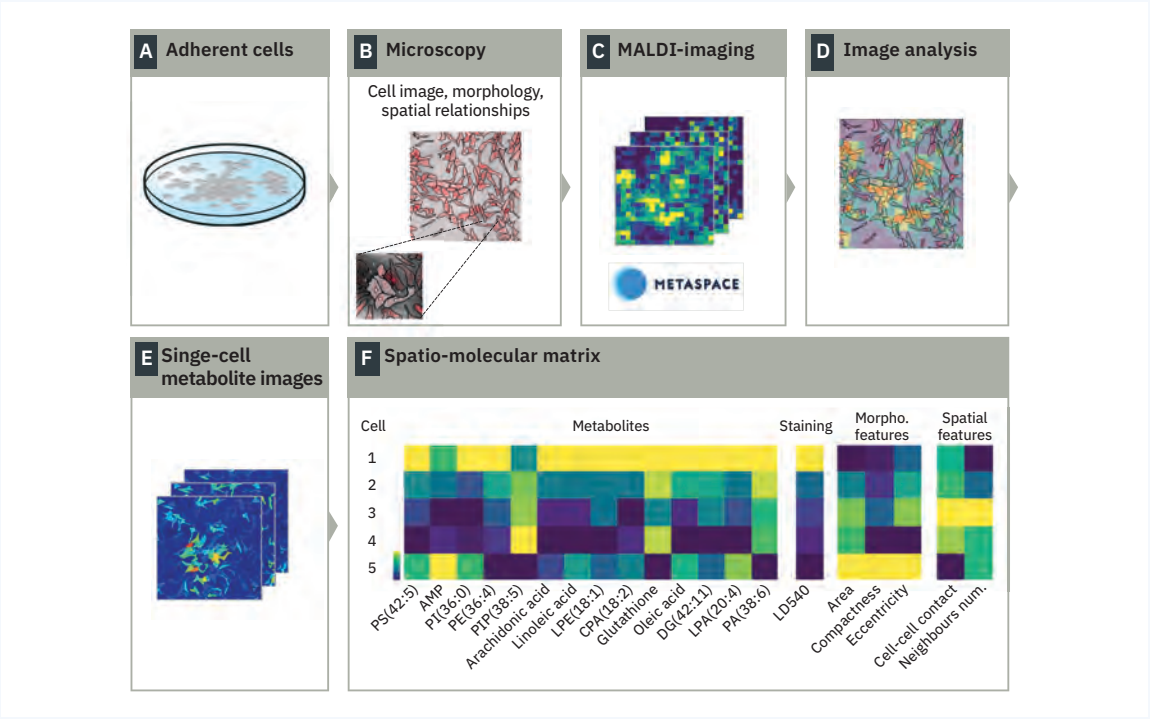
Metabolites

The metabolic profile of a cell provides a functional read-out of its physiological state. Cells and tissues have unique metabolic fingerprints that provide organ or tissue-specific information. This also holds true for many disease states. The subcellular distribution and function of metabolism and metabolites are still not completely known, and many mechanistic links await discovery. Metabolism also lies at the core of organismal survival in the environment. In the next Programme, EMBL will study the **links between intracellular metabolomes, cellular phenotypes, and spatial organisation of cells**, and the influence of the environment on metabolic profiles at the subcellular, cellular, tissue, and organ levels will be explored.

To map metabolic pathways, EMBL researchers will exploit thermal proteome profiling (TPP). The power of this technology is that it makes it possible to detect the effects of metabolites on proteins that manifest in changes in protein properties, and hence opens a new window on allostery. EMBL will also use its in-house expertise in spatial metabolomics, based on MALDI imaging mass spectrometry, to investigate the metabolome across scales, from single cells and tissues to whole organisms. This technology enables the detection of metabolites including amino acids, lipids, fatty acids, and the products of glutaminolysis and glycolysis. Many recent studies have explored the adaptability of metabolic pathways in animals based on comparative analyses of networks aided by theoretical and computational predictions. It should now be possible to integrate this comprehensive background knowledge with direct experimental observations. At EMBL, MALDI imaging mass spectrometry (Tech Dev Box TD3_MO) will be used for the first time for high-throughput metabolic screening of thousands of *Drosophila* embryos to characterise how their metabolism is reprogrammed upon genetic and environmental perturbations. This approach will allow the network structure at genomic and population-wide scales to be studied. Metabolic profiles and changes will be detected and linked directly to the uptake of exogenous, environmentally critical molecules such as pollutants, herbicides, and unwanted drugs. This will serve as a model to evaluate how organisms, including humans, are affected by the molecules in their environment. The high-throughput nature of this technology will allow EMBL scientists to set up screens for thousands of exogenous molecules that pose risks to humans.

Technology Development Box TD3_MO | Spatial single-cell metabolomics (SpaceM).

Measuring levels of metabolites in single cells is crucial for understanding metabolism, its heterogeneity, and its links to cellular phenomena and to cellular and transcriptional programs. However, conventional metabolomics is only feasible for bulk analysis where tissues need to be homogenised. Recently, imaging mass spectrometry (MS) was shown to be a successful method for acquiring spatially resolved metabolic fingerprints from tissues and cell cultures. These fingerprints can be measured directly from each probed location with high molecular specificity and sensitivity, and with a spatial resolution as small as 10 µm. The Alexandrov Team has a particular focus on furthering the development of novel imaging MS methods by using **matrix-assisted laser desorption/ionisation (MALDI) MS.** The team has recently developed a workflow for spatial single-cell metabolomics and lipidomics by co-registering microscopy and MALDI imaging MS data. **(A)** Cultured cells are **(B)** first imaged using bright-field and fluorescence microscopy. **(C)** MALDI-MS is performed in a raster pattern, and **(D)** the resulting laser ablation marks are imaged and analysed separately. **(E)** The microscope image is integrated with the MS data, and metabolic profiles are normalised based on the overlap of laser ablation marks with cells. **(F)** The resulting spatio-molecular matrix integrates information about metabolism and phenotype of the individual cells and enables single-cell metabolic investigations of cell types, cell states, and cell–cell interactions**.**



Signalling from the Environment to Cellular Compartments

Liquid–liquid phase-separated compartments, such as stress granules, nucleoli, and Cajal bodies, are increasingly being recognised as dynamic entities that may mediate responses to external factors. Their highly dynamic nature and the fact that the components within them are in constant exchange with the surrounding cytoplasm or nucleoplasm may allow a cell to rapidly respond to its environment. Consistent with this view, the organisation of the cytoplasm has been shown to change considerably in response to environmental stimuli such as stress, with many proteins and RNAs being sequestered to phase-separated structures. EMBL researchers have recently shown that, upon food deprivation, P-bodies containing certain

RNAs and proteins form in the ovarian germline cells of *Drosophila* during oogenesis. When this happens, protein synthesis and egg production halt, indicating environmental sensitivity (Figure MO2). Future studies will investigate how the properties of liquid–liquid phase-separated entities relate to their dynamic biological functions in response to their environmental context and the underlying mechanisms involved.

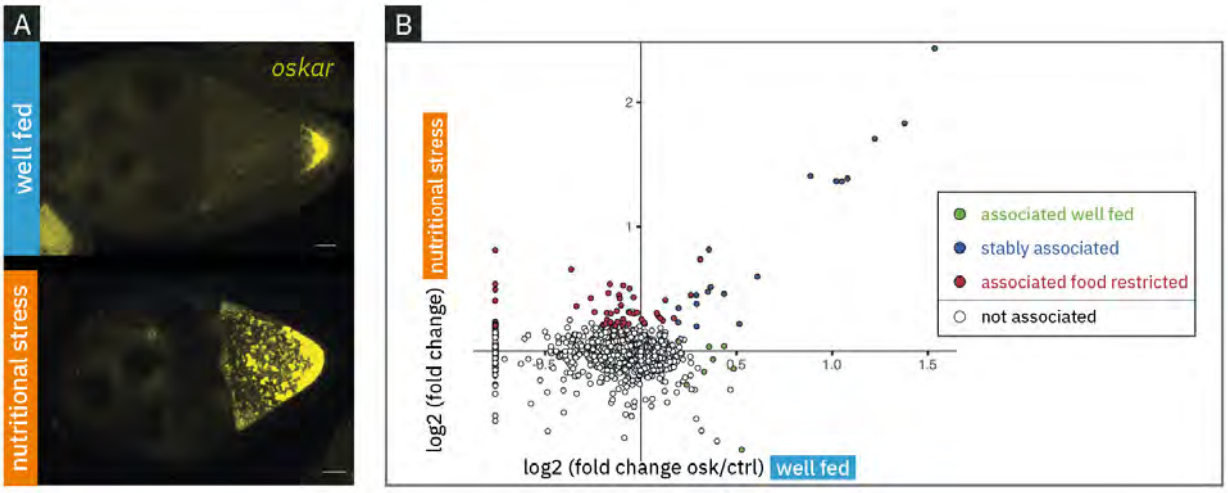


Figure MO2 | Changes in RNA-associated proteome upon nutritional deprivation.
(A) Fluorescent *in situ* hybridisation showing the distribution of oskar mRNA in an egg chamber of a fruit fly, either well fed or subjected to nutritional stress. Scale bar 10 µm. **(B)** Analysis of the changes in mRNP composition upon nutritional stress. oskar mRNPs from well-fed flies and flies that were deprived of protein-rich food for 4.5 hours were analysed by quantitative mass spectrometry.

EMBL will address one aspect of this by interrogating **intrinsically disordered proteins** (IDPs), which are implicated in a wide range of cellular functions, including the generation of membraneless organelles formed by liquid–liquid phase separation. This phenomenon is directly related to the intrinsic properties of IDPs, and has far-reaching consequences for cellular regulation that are as yet barely explored. IDPs are able to gain functional advantages by remaining natively unstructured, either completely or partially. EMBL is well positioned to contribute significantly to this area with its existing research and technologies, including investigating the relationship of IDPs to one another, their biological significance, and their response to environmental cues, especially in organisms in different environments as part of projects under the Planetary Biology theme (Chapter 7: Planetary Biology). EMBL aims to use a variety of approaches to characterise IDPs, such as small-angle X-ray scattering (SAXS), fluorescence resonance energy transfer (FRET), various NMR-related techniques (paramagnetic relaxation enhancements, residual dipolar couplings, chemical shift perturbations, and relaxation analysis). Cellular cryo-electron tomography yields *in situ* information across various scales on both assemblies and individual macromolecules. In-house expertise in TPP approaches will also form an integral part of EMBL's strategic approach. EMBL researchers will use TPP to explore protein–protein solubility and phase separation properties on a genome-wide scale, providing insights into the roles of enzymes, metabolites, post-translational modifications, and phase transitions in the cell under changing environmental conditions. Bioinformatics and sequence analyses will allow the prediction of disorder and also generate ensembles of conformers, which can be assessed in light of the experimental evidence. Finally, as the (fully or partially) disordered models must be validated, annotated, stored, and disseminated, proper curation and presentation of these ensembles via EMBL's data resources is important. This will ensure that the results are discoverable and interpretable by a broad user community.

Looking Ahead: Delving into the Molecular Details of Symbiosis

Beyond intra-organism and biophysical context, inter-organism symbiosis is central to many aspects of cellular function, and indeed is at the root of eukaryotic evolution via mitochondrial co-option. EMBL researchers are engaged in large-scale projects to study the molecular underpinnings of symbiotic relationships, for example by perturbing individual species in gut microbial communities and engineering synthetic and *ex vivo* communities to dissect functional interactions (Chapter 4: Microbial Ecosystems). These approaches will guide and empower further studies by EMBL to investigate the complex molecular interface between (gut) microbial communities and host physiology over an organism's lifetime, and the intergenerational or evolutionary consequences. The recent construction of two complementary EMBL gnotobiotic facilities will support this. Similarly, mapping and observing host–pathogen interfaces to engineer the next generation of sophisticated experimental models to study host–commensal–pathogen interactions coupled with host genetics will be pivotal in understanding susceptibility to infections (Chapter 6: Human Ecosystems and Chapter 5: Infection Biology). The technology to integrate imaging and molecular profiling data will also play an important role in the identification and molecular understanding of symbioses as they occur in marine and coastal samples to be collected by the TREC project (Chapter 7: Planetary Biology). EMBL already collaborates to understand photosymbiosis between single-celled hosts and microalgae in oceanic plankton, applying a combination of quantitative single-cell structural, chemical, imaging, proteomics, and metabolomics techniques to study cell–cell interactions and subcellular mechanisms. The genetic and epigenetic basis of such interactions can be tested using model systems that extend to animal communities. For example, the genomics, transcriptomics, and epigenomics of *Platynereis* and *Drosophila* and their symbionts will be coupled to environmental context to characterise cellular responses. Working across models and scales, EMBL will focus on symbiotic connections between organisms and on the common molecular underpinnings that explain them.

Impact

EMBL has a strong track record in the discovery and mechanistic dissection of the molecular foundations of biological complexity. EMBL researchers have made seminal contributions to understanding the differential usage of DNA, the dynamic 3D organisation of the genome, and the roles of RNA processing and localisation, post-translational modifications, and protein transport and structure. The new Programme will leverage and significantly extend this expertise, opening up new areas of discovery by gaining a holistic view of the dynamics of large macromolecular assemblies in their *in vivo* context, while integrating information across scales to understand the complex interactions between all of these processes. A universal approach, integrating multiscale *in vitro* and *in situ* information with real-time kinetics is imperative to understand how our genome is used and how genes, proteins, and metabolites interact. But genetics is only one part of the equation. An understanding of the relationship between genes and their environment is fundamental to dissecting and modelling biological complexity.

The need to bridge the anatomical, molecular, cellular, and environmental scales requires not only data integration but a true merging of disciplines to form a new era in the molecular life sciences. EMBL has the expertise and tools to make these fundamental links and to understand the molecular basis of biological complexity and how it is influenced by context. For example, major EMBL projects will generate freely available resources, such as cohorts of mouse models created with precision genome engineering, which can serve as valuable resources to member state scientists, enabling them to tackle central questions in genomics and epigenomics that link development and disease. Using such resources and tools will enable scientists to monitor the nature of responses to environmental fluctuations (e.g. human exposomes in human cohort studies, or soil microbial communities) and will allow the formulation of hypotheses that can be tested in the laboratory context. A deeper molecular understanding will also pave the way for precision

interventions to mitigate the harmful effects of our ever-changing world on living systems, including humans, and will help inform policymakers on issues relating to planetary and human health. With its well-established expertise in molecular biology and technology development, combined with truly interdisciplinary and innovative thinking, EMBL is exceptionally well placed to pioneer technology development programmes, both in modelling the molecular impact of environmental change and in discerning its phenotypic consequences. These developments can also be leveraged by member states and the wider scientific community by sharing knowledge and technologies.

The influence of environmental context on cellular and organismal biology, from molecular processes occurring on timescales of nanoseconds to the lifetime of whole organisms or even multigenerational timescales, is only just beginning to be understood, yet has enormous repercussions. Environmental influences permeate all aspects of life, including both our susceptibility to disease and that of our offspring. EMBL scientists are in an excellent position to discern the underlying mechanisms of **contextual responses** at the molecular level, connecting responses across subcellular scales and compartments. This will provide a basis for decoding the complex interactions across molecular layers (for example, genetic and epigenetic layers) that manifest as specific outcomes in natural or artificially perturbed environments, and which are crucial for almost all aspects of life – from bacterial ecosystems, to organism development, to human health. This research will create a roadmap for understanding how molecular architecture is responsive to context.