5. Infection Biology

Background

Infectious diseases are among the most prevalent causes of human illness and death in the world. Global warming, a decrease in biodiversity, antimicrobial resistance (AMR), modern lifestyles (travelling habits, high urban density, mass food production, and globalisation), and socio-economic factors, such as disparities in public health, make combating infectious diseases a more urgent challenge than ever. Besides humans, infection impacts all life forms on Earth. Pathogens can cross species boundaries, creating alarming possibilities for future epidemics and adding to rising concerns about changes in biodiversity and planetary health. Further increasing the challenge is the diversity of pathogens, which include viruses, bacteria, fungi, protozoa and other parasites. Within each of these groups, there is staggering diversity, both in their fundamental biology and ecology.

Disease mechanisms and transmission routes are also varied. For instance, both bacteria and viruses use diverse mechanisms to evade their host's immune response and to hijack the host's cellular processes to use them for their own benefit, while residing outside or inside different cell types. Many prominent infectious diseases in humans represent zoonoses showing successive propagation in diverse susceptible new hosts. These include HIV/AIDS, influenza, Ebola virus disease, plague, salmonellosis, and most recently SARS-CoV-2. Other pathogens (e.g. Epstein-Barr virus or *Mycobacterium tuberculosis*) persist in the host and can establish a dormant infection. In all infectious diseases, the host response is key to disease development and manifestation. Pathogens have developed a wide range of mechanisms to modulate both the innate and adaptive parts of host immunity to facilitate infection and transmission. In many infectious diseases, it is the failure of the immune system to respond appropriately that is a substantial cause of mortality.

Human pathogens now pose a greater threat due to their increasing resistance to available therapies. Multidrug-resistant (MDR) pathogens are found in rapidly growing numbers worldwide, increasing mortality caused by once treatable diseases and reducing our quality of life. Bacterial pathogens account for more than 3.5 million deaths per year globally, with an increasing fraction coming from MDR infections. Similarly, in more than 10% of adults living with HIV, the virus has developed resistance to the main first-line antiretroviral drugs (nevirapine and efavirenz), according to the World Health Organization. Recently, both WHO and the UN have declared antimicrobial resistance (AMR) a major threat to global public health. Awareness has finally permeated all layers of society, but appropriate action on the development of effective anti-infective strategies is still lacking. For instance, only one novel broad-spectrum antibiotic class has entered the market since the 1990s. The repeated failures of target-driven drug discovery have led most big pharmaceutical companies to close their antibiotics R&D departments. This, together with the rapid development of resistance to existing compounds, creates a highly challenging situation.

The increased risk of transmission of disease from wildlife to humans and between humans is likely a hidden cost of human economic development and disruption of ecosystems, which has catalysed the (re-)emergence of viral pathogens such as Zika virus and SARS-CoV-2. With our ability to treat these serious global public health threats being challenged, infectious diseases can re-emerge as a major factor of human morbidity and mortality, even in countries with developed public health systems. Insights into pathogen spread, evolution, and transmission, as well as characterising and understanding the host response, are critical to tackling infectious diseases (Figure IB0).

Figure IB0 | Mapping infection biology.

The study of infection biology needs to take into account multiple factors such as the unique mechanisms of various pathogens (e.g. viruses, bacteria, fungi, and protozoa), the host organism (e.g. bats, pigs, and chickens), and situational contexts (e.g. farmed land, food chain, and urban areas). EMBL aims to combat this at several biological scales (e.g. from molecules to communities) with a view to impacting a number of applications (e.g. antimicrobial resistance, diagnostics, drug discovery, and outbreak surveillance).

The Opportunity

With infectious diseases having immediate and devastating consequences globally with long-term repercussions, research on the biology and mechanisms of infection, as well as on diagnostics and treatment, is vital and urgent. Pathogen biology and disease are fundamentally linked to the interaction with the host, which can take on many perspectives, from the conflict with the immune system to the parallel genomic evolution of pathogens on different timescales and under different selective pressures (e.g. drugs, immune system, and competing microbes).

Technology developments over the past decades provide an increasing array of atomic, molecular, cellular, and physiological assays that can be performed on pathogens, on cellular systems modelling humans, or on humans directly. This produces remarkably rich datasets, which can be used to probe the human (or other host) response to infection. The opportunity awaits to understand the molecular machineries of pathogens and the host response in more detail than ever before, to discover new interactions at the host–pathogen interface, and thereby to provide new means for treating infection.

To tackle the AMR crisis, an improved molecular understanding of host–pathogen biology is not enough. This needs to be coupled with more specific diagnostics methods to optimise the use of existing antimicrobials, a better understanding of the mechanisms of AMR selection and spread to devise innovative resistance prevention strategies, and new antimicrobial therapies which are less prone to resistance development. Modern technologies can enable effective tracking of pathogens between hosts, across national boundaries and ecological habitats (Chapter 7: Planetary Biology).

The SARS-CoV-2 pandemic has revealed the urgent need for transparent, secure, and scalable sharing of pathogen data for the global management of infectious disease. Pathogens do not respect national, ecological, or species boundaries. Thus, infectious diseases of humans and animals are studied across a wide range of scales, ranging from longitudinal studies of single infections or individual outbreaks to indefinite global surveillance, host context, and historical and evolutionary perspectives. Currently, scientists have limited tools to access structured infectious biology data. Unified and systematic views into pathogen biology, allowing the user to navigate the many different dimensions of datasets, curated metadata, and literature will be essential to allow scientists, public health agencies, and clinicians to connect, access, and use this information in a meaningful way, so as to fundamentally change the way infectious diseases are understood and treated.

Research Aims

For several decades, EMBL has undertaken significant research activity on selected pathogens and their interaction with their host. Much of this work has been driven by cellular and structural biology, which has unravelled detailed mechanisms of fundamental interest, as well as potential drug targets. More recently, high-throughput genetics- and proteomics-based approaches, as well as cutting-edge imaging technologies, have enabled more systematic views of the host–pathogen interface. In parallel, a concerted effort to tackle the AMR crisis has been forming. In the new Programme, EMBL will integrate multidisciplinary experimental and computational approaches to dissect the complex molecular mechanisms behind the interplay of pathogens and their host ecosystems:

I. *In situ* information about the cellular organisation of pathogens and the structural basis of their essential biological processes can give insight into new drug targets and inspire therapeutic strategies. With a **focus on pathogen-specific protein machineries**, EMBL aims to extend beyond *in vitro* structure determination into visualising these machineries in their cellular environment, using cryo-electron tomography (cryo-ET) and other techniques that build on EMBL's strength in integrating molecular information across resolution scales.

II. The mechanisms of pathogen adhesion, invasion, and effector translocation in the host are still poorly understood. Equally elusive remains the understanding of the targets and mechanisms of effector proteins which hijack diverse host cellular processes to manifest infection. EMBL aims to use its diverse cutting-edge technologies to **systematically map, view, and model host–pathogen interfaces** at the atomic, molecular, and tissue level. This will enrich the understanding that scientists have of diverse infection processes, opening avenues for new anti-infectives and immunomodulatory strategies.

III. Capitalising on the growing volume of population-scale human biological datasets, EMBL aims to develop sophisticated statistical models to **map the human genetics of infection susceptibility**. These models can be deployed to combat endemic and epidemic outbreaks, for example in the context of national data within EMBL member states. In collaboration with

clinicians, EMBL will also employ multimodal approaches for cellular phenotyping of patient samples and machine learning methods to characterise the human response to infection.

IV. To improve treatment efficacy, reduce side-effects, and prevent the development of drug resistance, EMBL proposes to develop individualised, microbiome-based diagnostic and surveillance tools. EMBL will combine predictive models, automated screening, medicinal chemistry, and structural biology to identify new anti-infective targets and compounds. Systematic genetics and proteomics approaches will expedite drug mode-of-action identification, which is a current bottleneck in antimicrobial drug discovery. EMBL further aims to **create a better understanding of antimicrobial resistance mechanisms and transmission paths** in patients and in the environment, so as to devise **new ways to delay, prevent, or revert this resistance.**

V. EMBL has a long history in collating and openly sharing genomic and proteomic data of pathogens to promote fundamental science. In recent years, EMBL has also started working closer to the frontline of public health, collaborating with public health agencies on a platform for secure and controlled data-sharing for pathogen genomics and AMR surveillance. This includes the provision of **international data hubs** for the SARS-CoV-2 pandemic. EMBL will develop further **data platforms, computational frameworks, and infrastructure** to allow scientists across multiple domains to access and process multi-dimensional pathogen data. These resources will support rapid response to future outbreaks of novel pathogens and facilitate target identification for anti-infective research.

Given EMBL's unparalleled resources, technologies, and multinational research networks as well as its broad relevant expertise from informatics to molecular and systems-level sciences, EMBL is in a unique position to provide new knowledge and innovative solutions to combat microbial pathogens.

EMBL's Approaches

Pathogen Specific Molecular and Cellular Machineries

Pathogens employ a range of specific protein machineries in their replication and invasion of host cells. Several examples are described below to illustrate how EMBL aims to move beyond *in vitro* structure determination to **visualise these dynamic machineries** in their cellular environments.

Replication Machinery of RNA Viruses

Many important human viral pathogens are **RNA viruses**, including Ebola, measles, rabies, SARS-Cov-1 and 2, Dengue, Zika, hepatitis C, polio, rhinovirus, rotavirus, HIV, influenza, and Lassa. All RNA viruses, except retroviruses, encode an evolutionarily related RNA-dependent RNA polymerase (RNAP) that is able to transcribe and replicate the RNA genome. As the key viral transcription/replication machine, the viral RNAP is a primary target for antiviral drug development.

EMBL researchers aim to obtain a detailed structure-based understanding of how the **influenza virus RNAP** functions, and specifically the distinct processes of transcription and replication of the segmented, negative-sense RNA genome. In the past, crystal structures of the 270 kDa heterotrimeric RNAP have made it possible to pioneer a new generation of anti-influenza drugs that target unique features of the influenza RNAP. Most

recently, multiple high-resolution cryo-EM snapshots of all stages of transcription by the influenza RNAP – initiation, elongation, termination/polyadenylation, and recycling – have been determined, leading to a model of the complete transcription cycle with numerous novel features (Figure IB1 A–C). Similar methods can be used to investigate the functional dynamics of RNAPs from other emerging viruses, including arenaviruses, bunyaviruses, and coronaviruses.

In the future, the main aim is to go beyond *in vitro* mechanisms. In an influenza-infected cell, for example, transcription and replication of the viral genome occur in the context of viral ribonucleoprotein particles (RNPs), in which most of the RNA is coated by protective viral nucleoprotein in the nucleus, aided by specific host factors. The goal will be to understand the mechanisms operating at these two extra levels of complexity: (i) how are nucleoproteins dynamically remodelled around the RNAP during transcription? and (ii) how is the influenza RNAP able to robustly pirate nascent host transcripts to prime its own transcription, a process known as cap snatching? With the rapid development of cryo-ET workflows, visualising viral RNPs in the chromatin context will likely become feasible. Additional lines of investigation include applying techniques developed to probe the structure of long non-coding RNAs to study the structure of viral RNA such as the 5' untranslated region of SARS-CoV-1, which possibly functions as an internal ribosome entry site to initiate protein translation on the human ribosome.

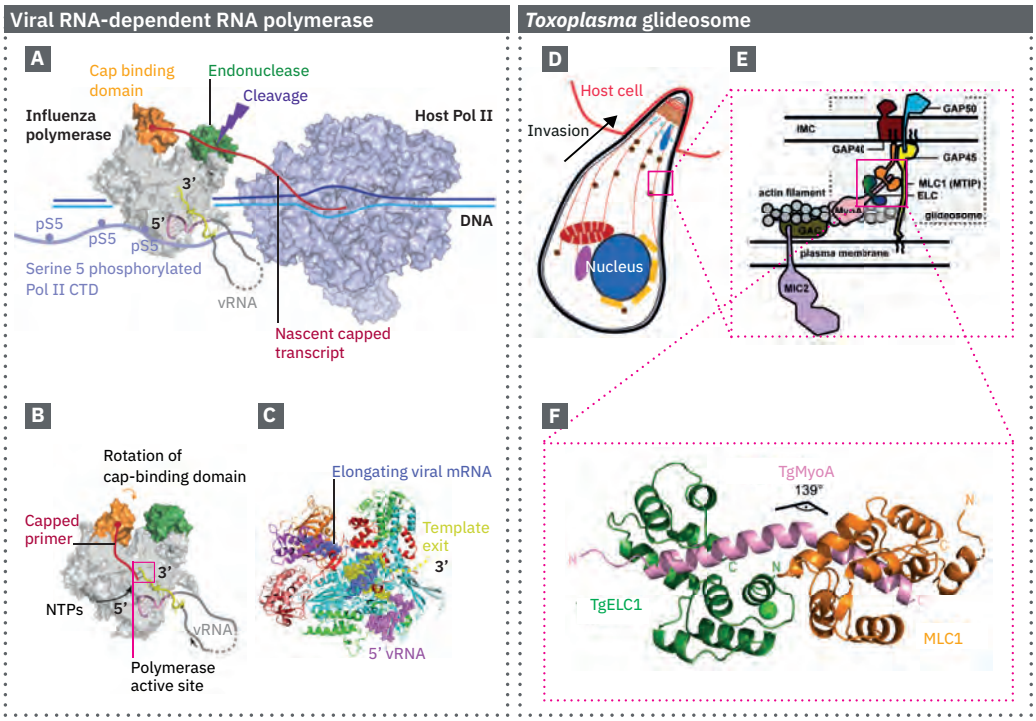


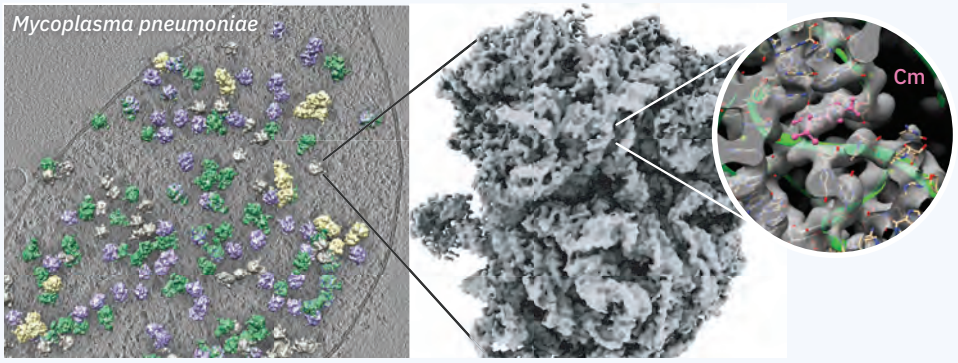
Figure IB1 | Molecular machines: the influenza polymerase and the parasitic glideosome.
(A) Schematic of 'cap-snatching' by influenza RNA-dependent RNAP (Flupol). In the infected cell nucleus, FluPol (grey) associates with host RNAP II (Pol II – light blue) by binding to its phosphorylated CTD. This gives Flupol access to nascent capped transcripts (red) that it cleaves with its endonuclease activity (green domain) and uses to prime transcription of viral mRNAs. The chromatin environment in which FluPol robustly manages to compete at the right moment for capped RNAs can be visualised by cryo-ET in infected cells. **(B)** Schematic of cap-dependent transcription initiation. **(C)** *In vitro* cryo-EM structure of FluPol actively elongating viral mRNA. **(D)** Schematic depiction of the invasion of an apicomplexan parasite into a red blood cell. **(E)** Schematic representation of the current model of the glideosome. Actin is immobilised to the plasma membrane whereas myosin A is part of the glideosome, which binds the essential light chains ELC and myosin light chain MLC1. Myosin A and its light chains further interact with glideosome associated proteins GAP40, GAP45 and GAP50, which anchor the glideosome in the outer membrane of the inner membrane complex. **(F)** High-resolution X-ray structure of a trimeric sub complex (MLC1, ELC1, fragment of MyoA) of the glideosome.

Molecular Machines *In Situ* Within *Mycoplasma pneumoniae*

Pathogenic species of *Mycoplasma* are involved in a number of diseases targeting the human respiratory system. One of these, *M. pneumoniae*, has become a model for systems biology and multi-omics approaches, due to its very small genome size. EMBL researchers have used *M. pneumoniae* to establish a workflow that allows key molecular machineries to be visualised in action at near-atomic resolution inside the cell by cryo-electron tomography (Tech Dev Box TD1_IB). This has yielded the first *in vivo* structural view of a central supramolecular complex that directly couples transcription and translation, a mechanism unique to bacteria. Building on these technical advances, the next tier is to investigate membrane protein regulation and organisation. Although many of the pathogenic activities of *M. pneumoniae* are localised to the cell membrane, the organisation of the membrane proteome is largely unknown. EMBL will aim to structurally elucidate the organisation of the *M. pneumoniae* membrane proteome using integrated in-cell cryo-ET and proteomics approaches. As a similar level of resolution has not been achieved for *in situ* imaging of any organism to date, EMBL's approach will likely reveal general principles of membrane proteome organisation.

Technology Development Box TD1_IB | Towards creating whole-cell near-atomic resolution structural models.

Imaging is a powerful tool for both fundamental research and biomedical diagnostics. The Mahamid Group has been pioneering cryo-electron tomography (cryo-ET) pipelines that enable the visualisation of an entire cell of a bacterial pathogen, *Mycoplasma pneumoniae* (left, background). Focusing first on the ribosome (left, foreground; middle), the protein production machinery of all living cells and a major antimicrobial target, has already yielded fundamental mechanistic insights into its coupled function with RNA polymerase in bacteria (O'Reilly *et al. bioRxiv* 2020). Albeit a model system with a reduced genome, *M. pneumoniae* still encodes 688 proteins. Assigning each protein and DNA and assembling protein complexes like in a puzzle in these electron microscopic images will allow composition of the first full 3D molecular picture of a cell. This is a computationally intensive process but has immense implications both for basic biology and for novel screening concepts. For example, treating this human pathogen with chloramphenicol, a well-known protein synthesis inhibitor, reveals the precise binding site of the drug on the ribosome (right) *in situ*. As more protein complexes are mapped inside the cell, this setup can be used as a screening platform to identify drug targets and their off-targets in the cell, as well as to view the downstream cellular responses they elicit at a nearly atomic resolution.



Cryo-ET slice of a *M. pneumoniae* cell (left), overlaid with annotated ribosomes: grey, single ribosomes; green, ribosomes interacting with putative RNA polymerase; yellow, ribosome assemblies engaged in protein production from a single mRNA; purple, a super complex of ribosomes directly engaged with RNA polymerase, unambiguously demonstrating for the first time transcription-translation coupling *in vivo*. A high-resolution reconstruction of the *M. pneumoniae* ribosome (middle) at near-atomic resolution (3.5 Ångström) with a zoom into the ribosome active site (right) where chloramphenicol binding (pink) stalls protein production.

Unique Gene-expression Systems in Trypanosome Parasites

The majority of human African trypanosomiasis (sleeping sickness) is caused by the *Trypanosoma brucei gambiense* and is transmitted by the tsetse fly. Trypanosomes possess specialised RNA processing pathways, which are distinct from those of their mammalian hosts and hence provide ideal drug targets. EMBL researchers will study the molecular basis and RNA interaction dynamics of two key trypanosomal RNA processing pathways. First, the mitochondrial RNA (mtRNA) editing pathway will be targeted with structural biology approaches and parasite cell biology. mtRNA editing in trypanosomes is conducted by the enzymatic RNA editing core complex together with a second protein complex, the RNA editing substrate binding complex (RESC). No structural information on RESC or any of its subunits is available. The aim will be to use structure-function approaches to understand the assembly mechanism of RESC, its composition, its specificity for mRNA and gRNAs, and its role in mtRNA editing. A second unique process in trypanosomes is trans-splicing, which separates the functionally unrelated coding clusters with several RNA–protein complexes that are unique to trypanosomes. EMBL groups will characterise these dynamic complexes in atomic detail and will explore their function in trans-splicing. This objective is highly relevant with respect to future drug targeting, as trans-splicing and the related complexes are essential for the parasite.

The Apicomplexa Invasion Machine (Glideosome) of *Plasmodium*

Plasmodium species are intracellular, parasitic single cell eukaryotes of the family Apicomplexa and are the causative agents of malaria, which causes ~400,000 deaths per year. Another apicomplexan parasite, *Toxoplasma gondii*, is responsible for toxoplasmosis in humans. Proliferation and transmission of these obligate endoparasites in their host organisms rely on efficient cell invasion. This active process is based on parasite motility that is referred to as gliding and is empowered by an actin/myosin motor. This motor is localised within the intermembrane space between the parasite's plasma membrane and inner membrane complex (IMC), an additional double layer of membranes that is unique for these single-cell organisms. While motility is achieved by the interaction of the myosin with actin filaments, the myosin is linked to the IMC by a membrane-embedded multi-protein complex referred to as the glideosome (Figure IB1D-F). The structures of a few individual glideosome components are known; however, how these proteins assemble into an active complex and perform their function remains elusive. To obtain structural and functional understanding of the glideosome, two approaches will be followed: i) individual glideosome components, subcomplexes of known intermediate assemblies, and the entire complex of all known glideosome members will be structurally and functionally characterised using biochemistry, X-ray crystallography, small-angle X-ray scattering (SAXS), and cryo-EM; and ii) analysis of the endogenous membrane complex will be analysed by mass spectrometry to identify so far unknown glideosome components followed by single-particle cryo-EM. This work will help to understand the underlying mechanisms of motility in apicomplexan, which is crucial for any invasion process of the parasite.

Systematically Mapping and Modelling Host–Pathogen Interfaces

Mapping Host–Pathogen Interfaces

EMBL will combine systems-based approaches, computational biology, and structure–function analyses to map the machines and the effector arsenals of pathogens, their host targets, and the mechanisms of action that pathogens exploit to survive and proliferate in the intracellular context.

The protein machines that bacterial pathogens use to adhere to cells and translocate effector proteins into the host cell have been a focal point of infection biology for decades. Effector proteins are used to usurp host defenses and to hijack various host cellular processes so that the pathogen can survive and proliferate. Past EMBL research has focused on structural analysis of a specialised secretion system of *Mycobacterium tuberculosis*, which secretes a number of key virulence proteins required for infection (Chapter 4: Microbial Ecosystems, Figure ME3B). In the next EMBL Programme, EMBL will continue to contribute cutting-edge structural insights into the **function of diverse secretion systems**, including type III systems, which cross even the host membrane to translocate effectors directly inside the host. These efforts will be complemented by computational analyses (e.g. Hidden Markov Model-based sequence analysis and machine learning approaches) and structure-function follow-ups to identify, classify, and understand the diverse protein domains that pathogens use to adhere to host surface-exposed sugars, lipids, and proteins, the so-called adhesins. The identification of conserved species-specific adhesins can be used as likely vaccine targets. Together with collaborators, EMBL's next goal is to provide a better overview of the structural and functional diversity of secretion apparatuses and adhesion protein families across different pathogens. Going beyond classical structural biology, the aim will be to visualise entire systems *in situ* using cryo-ET and correlative light and electron microscopy (CLEM) approaches.

Furthermore, machine learning and bioinformatics will be combined with proteomics-based labelling approaches to systematically identify the **secretion arsenals** of intracellular pathogens, which can reach more than 300 proteins for some species. EMBL will also employ a suite of computational and experimental approaches to **map the host targets of viral, bacterial, and eukaryotic effectors** (Figure IB2A; Tech Dev Box TD2_IB). Proteomics-based approaches (e.g. affinity purification coupled with quantitative mass spectrometry, thermal proteome profiling (TPP), proteomic phage display, and phosphoproteomics and ubiquitin proteomics) in *in vitro* infection contexts, and integrative computational approaches (homology models, structural information on interfaces, and human protein–protein interactions) will be used to identify host targets of effector proteins. Altogether, these efforts will provide a comprehensive understanding of how diverse pathogens interact with their host, revealing points of convergence and divergence in pathogen biology.

A common strategy for all intracellular pathogens is to **intercept host signalling pathways** and use them for their own benefit, for example to reorganise lipid transport and exploit cytoskeletal elements to build protective intracellular compartments; to avoid lysosomal killing and to move within the cell; to blunt immune responses; or to deregulate cell death pathways. To do this, pathogens utilise a myriad of approaches, from effector proteins with intrinsically disordered regions that mimic binding motifs of signalling proteins to effector proteins that carry their own signalling enzymatic activities (e.g. kinases, phosphatases, E3 ligases, DUBs, Rho GTases, and GEFs). Interestingly, within this vast repertoire of activities, pathogens have come up with novel enzymatic activities for post-translational modifications (PTMs), such as serine ubiquitination and glutamylation, to control the activity of host signalling cascades (Figure IB2B). EMBL will utilise its unique computational and experimental expertise on signalling networks and PTMs (Tech Dev Box TD2_IB), such as phosphorylation, ubiquitination, and Rho signalling, to dissect the mechanisms that pathogens use to intercept host signalling networks and the structural and functional ramifications those PTMs have on the human proteins. Thereby, EMBL researchers will not only gain a better understanding of both host and infection biology, but will also identify new intervention points for drugs and tools for bioengineering.

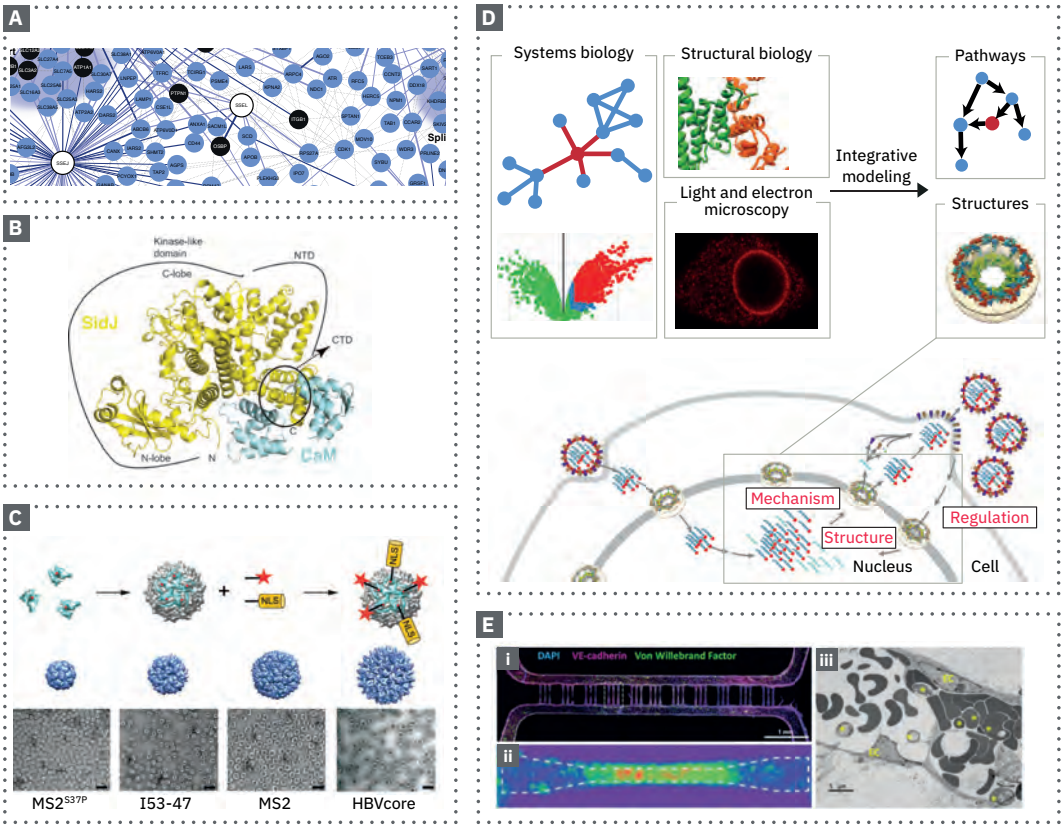


Figure IB2 | Probing host–pathogen interfaces.
(A) Systematically mapping physical interactions of secreted *Salmonella* effector proteins with their host targets during infection; effectors cooperate to hijack host processes (Walch *et al. bioRxiv* 2020). **(B)** Cryo-EM structure of SidJ- Calmodulin (CaM) glutamylation complex (Bhogaraju *et al. Nature* 2019). Legionella effector SidJ counteracts the action of SidE, another bacterial effector that catalyses serine ubiquitination of host proteins to promote Legionnaires' disease (Bhogaraju *et al. Cell* 2016; Kalayil *et al. Nature* 2018). Bacterial effector repertoire also includes DUB proteins that add another layer of control for SidE activity (Shin *et al. Molecular Cell* 2020). **(C)** Labelling of different viral capsid proteins with maleimide reactive NLS peptide and reactive fluorescent dye allows for building capsids with different labelling ratio. Capsid structures (top) and their EM images (bottom, scale 50 nm) are then used to define the biophysical parameters for their nuclear entry though the Nuclear Pore Complex (Paci and Lemke, *bioRxiv* 2019). **(D)** Integrative pathway and structure modelling of the nuclear export of viral genomes during influenza A virus infection cycle. **(E)** *In vitro* vascular systems, **(i)** engineered capillary-size vessels can be used to spatiotemporally track *P. falciparum*-infected red blood cell sequestration in the microvasculature. Sequestration in the capillaries can be quantified with **(ii)** light or **(iii)** electron microscopy. Sequestration of infected red blood cells is shown in **(ii)** with a heat map (blue, no cytoadhesion; red, high cytoadhesion) and in **(iii)** with an asterisk. EC is endothelial cells.

Viewing the Host–Pathogen Interface

EMBL will employ its unique imaging and structural capacities to view intracellular pathogens at different stages of infection and characterise their interactions with the host at an atomic-resolution level. FIB-SEM, cryo-EM, and cryo-ET will be used to capture subcellular changes during the course of infection and to view compartments that intracellular pathogens create or reside in. Pilot studies are underway, with an initial focus on different viruses, including SARS-CoV-2, and will be performed in collaboration with the University of Heidelberg. EMBL will also exploit its expertise in nuclear pore complex (NPC) biology and couple biophysical measurements with imaging and electron microscopy structural information to gain a better mechanistic insight into how large viral capsids enter the nucleus (Figure IB2C), how viral pre-integration complexes facilitate nuclear import, and how viral ribonucleoproteins (RNPs) exit the nucleus through the NPC at later stages of infection (Figure IB2D). EMBL will also capitalise on its wide interests in nuclear biology and shed more light into how viruses take over nuclear processes. For example, cryo-ET studies of chromatin,

chromatin remodellers, and viral pioneering transcription factors will facilitate an understanding of how host cell chromatin is hijacked by viruses to directly integrate and/or modulate gene expression.

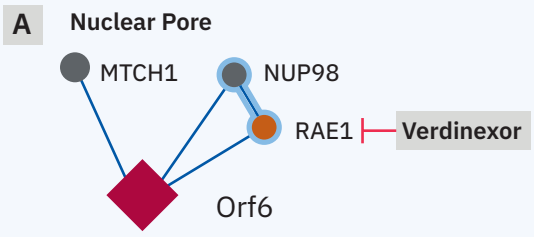
Modelling Host–Pathogen Interfaces

Beyond mapping and viewing interfaces, EMBL will go a step further to develop computational models of these interfaces, integrating diverse data types and engineering the next generation of sophisticated experimental models to study host–pathogen interactions. To create computational models of different granularity, from signalling network models to whole-cell models, EMBL will integrate experimental information on host and host–pathogen protein–protein interactions, genome-wide data on protein expression, localisation, activity, and PTMs during infection, and structural data on interface points. For example, EMBL researchers will integrate structural and systems-based information to model the way the influenza RNPs exit the nucleus via the NPC and assemble viral particles when leaving the cell (Figure IB2D). The models will point to new intervention points for viral and bacterial infections and will guide future experimentation at the host–pathogen interface.

Technology Development Box TD2_IB | Integrated systems-based approaches to tackle emerging pathogens.

Pathogens use a myriad of mechanisms to trick host defences and to hijack host processes for their own benefit. These mechanisms range from intercepting host signalling and epigenetics to directly recruiting necessary protein machines and enzymes to build compartments, protect from host defences, scavenge nutrients, and move within the cell. EMBL is home to a suite of unique systematic technologies that can provide paradigm-shifting insights into host-pathogen interfaces.

Quantitative proteomics-based approaches (see Chapter 5: Microbial Ecosystems, Tech Dev Box TD3_ME), such as thermal proteome profiling (TPP), are unique in providing the state of host and bacterial proteins during infection, illuminating the active interfaces and pathways at different stages of infection, and pinpointing possible drug targets. Systematic protein-protein interaction (PPI) profiling can unravel the direct targets of effector proteins and identify drugs that disrupt these interactions. Recently, EMBL has pioneered methods to map PPIs in the context of infection (Figure IB2A). Profiling proteome turnover (Savitski *et al. Cell* 2018), phosphorylation (Potel *et al. Nature Methods* 2018), and localisation (Selkrig *et al. Nature Microbiology* 2020) can provide further insights into mechanisms pathogens use to intercept host responses. This versatile and systematic profiling of proteome states can be combined with computational approaches to assess the functional consequences of these events at different levels, from an atomic view of a PPI interface (Bradley & Beltrao *Plos Biology* 2019) to dissecting signalling networks (Müller *et al. Nature Cell Biology* 2020; Ochoa *et al. Nature Biotechnology* 2020), to identify downstream druggable host proteins/pathways, and to predict pathogenic traits, such as AMR (Galardini *et al. eLife* 2017; Bradley *et al. Nature Biotech* 2019). In collaboration with others, EMBL has piloted research using these integrative approaches to identify how repurposed drugs may block the infection cycle of SARS-CoV-2 (Gordon *et al. Nature* 2020; Bouhaddou *et al. Cell* 2020).



B

Protein		Interaction motif	
SARS-CoV2 Orf6	56	QPMEID	61
SARS-CoV Orf6	57	EPMELD	62
VSV M protein	49	DEMDTHD	55
KSHV Orf10	414	EPMQS	418

SARS-CoV-2 Ofr6 protein interacts with an interferon-inducible mRNA nuclear export complex. **(A)** Small molecule inhibitors shown for RAE and **(B)** the NUP98-RAE1 interaction motifs from several viral species (Gordon *et al. Nature* 2020).

EMBL will also engineer novel 3D vascularised *in vitro* tissues and establish relevant organoid systems to simulate better infection contexts and study host–pathogen tissue interactions. These will include models to recreate human cerebral malaria pathology with cutting-edge *in vitro* bioengineering approaches. EMBL researchers will develop 3D blood-brain-barrier (BBB) models (Figure IB2E) with tubular geometry that incorporate multiple cell types (brain microvascular endothelial cells, astrocytes, and pericytes) to study how parasites, platelets, neutrophils, T cells, and cytokines potentially contribute to BBB dysfunction. Future applications of these models could extend to other parasites, viruses, or bacteria that cross the BBB and cause meningitis. EMBL will also develop vascularised cardiac models capable of mimicking endothelial barrier function seen *in vivo* to study viral or bacterial infections linked to cardiomyopathy. Altogether, these models will facilitate understanding of critical, hard-to-simulate host–pathogen interfaces and provide a unique platform for probing relevant host-targeted therapies.

Molecular Biology of the Host Response

Human Genetics of Susceptibility to Infection

EMBL is in an excellent position to use human genetics, coupled with molecular and cellular phenotyping of patient samples collected by collaborators, to characterise the human response to infection. The availability of more than six million genotyped individuals across Europe, many of them also with exome or genome sequences (Chapter 6: Human Ecosystems), enables comprehensive analyses to find host susceptibility loci that are involved in differential response to infection. EMBL scientists are also developing more sophisticated joint geographic and statistical models, which can more accurately integrate over the heterogeneous nature of infection exposure (e.g. by age, social contacts) and both traditional epidemiological and genetic risk factors. These models can potentially leverage encrypted anonymised proximity tracking, either as a way to estimate population contact maps or – more ambitiously – to fully model infectious agent exposure. These models can then be deployed to understand both endemic and epidemic disease outbreaks. The outputs of these models will provide more finely-grained risk population models, and will result in a better understanding of the pathways of infectious disease and potentially new drug targets or drug repurposing opportunities. These research efforts from EMBL complement worldwide efforts to study infectious disease genetics, which are expected to be brought together into the Genome-Wide Association Studies (GWAS) Catalog, a data resource aggregating the results of all published GWAS, which is run jointly by EMBL and the National Human Genome Research Institute (NHGRI), part of the US National Institutes of Health. By integrating these data with an extensive molecular understanding of host–pathogen interfaces, more complex models for genetics are also possible, for example involving epistatic interactions between genes in the human genome. Looking ahead, with a substantial proportion of European populations having genotype or genomic information available, such models will be important in optimising future healthcare delivery in the context of infectious diseases. With each national endeavour in EMBL member states, EMBL scientists will work in partnership with the genomic and epidemiology groups of the member state to best exploit EMBL's scientific expertise within the context of national datasets and healthcare delivery structures.

Cellular Phenotyping of Host Immune Response

Host responses can also be studied by molecular phenotyping of patient samples, in particular from accessible tissues or samples such as blood, urine, or sputum, usually acquired at the point of hospitalisation. In collaboration with clinical partners, EMBL can provide the technology and analytical components needed for clinical studies. A particular strength of EMBL is multimodal phenotyping integration, such as integrating measurements of single-cell RNA expression, DNA methylation, and cellular microscopy with computational analyses using innovative machine learning techniques.

In patient-derived samples and in model cellular systems, EMBL will extensively study key immunological cell types using various technologies. An intriguing opportunity would be finding evidence of O-GlcNAcylation of local chromatin involved in retroviral and retrotransposon suppression in the host genome. This is one example of how fundamental mechanistic biology can provide new insights into infection biology and, conversely, how the study of infectious agents provides insight into fundamental biology. EMBL will also study the fundamental properties of immune cells, using various genomics tools such as single-cell expression and chromatin accessibility, as well as novel microscopy techniques, such as Brillouin microscopy, to measure surface stiffness in migrating immune cells (Tech Dev Box TD2_MD).

By working with collaborators in the developing world, in particular via the Human Heredity and Health in Africa (H3Africa) Initiative and via global collaborations such as the CRyPTIC consortium for *Mycobacterium tuberculosis* treatments, EMBL scientists will apply experimental and computational techniques in areas of the world with a wider variety of endemic and epidemic diseases. This has many ramifications. Firstly, higher levels of many endemic diseases and greater variation in human genetics, in particular in Sub-Saharan Africa, mean statistical studies can be powered sufficiently to find biological effects and generate associations with specific genetic variants. Secondly, EMBL can work in partnership with scientists in the developing world to deliver innovative field site pathogen genomic testing schemes using, for example, nanopore sequencing.

New Anti-infective Strategies

EMBL aims to generate new knowledge, resolve current bottlenecks, and provide innovative solutions to the rapid increase of AMR, one of the biggest public health challenges of the 21st century. In the next EMBL Programme, EMBL will build on its automated screening platforms, data analytics, and extensive molecular expertise to develop diagnostics and anti-infective strategies, to map AMR reservoirs and transmission routes, and to devise new ways to delay, prevent, or revert AMR.

Developing Diagnostic and Surveillance Tools

To foster antimicrobial stewardship and reduce the risks for development of resistance, EMBL proposes to **develop microbiome-based diagnostic and surveillance tools** for identifying pathogens or pathobionts and their resistance capabilities. The current first-line treatment for infection-associated symptoms is broad-spectrum antibiotics. Only if there is no improvement are more precise but time-consuming microbiological tests performed to identify the pathogen and its drug susceptibilities. To avoid the misuse of antibiotics, which promotes AMR, and to mitigate the collateral damage of antibiotics on commensal microbial species, rapid diagnosis of specific pathogens and their AMR potential is urgently needed. This knowledge can be used to guide tailored treatment options.

Gut microbiome research at EMBL (Chapter 4: Microbial Ecosystems), enables the identification of bacterial pathogens and viruses based on metagenomic or metatranscriptomic fingerprints of known infectants. This includes profiling of non-invasive samples such as stool or saliva, down to the resolution of individual genes and even residues (e.g. those that are known to confer AMR through modification of target). Leveraging EMBL's collections of public data and emerging platforms on individual pathogens, as well as the increasing knowledge of antimicrobial resistance and mobile genetic elements (Chapter 4: Microbial Ecosystems, Figure ME4B), a supervised list of hundreds of infectious agents together with their resistance risk will be developed. This list will be amended with rules derived from machine learning to summarise the worldwide distribution and abundance of pathogens and their associated resistance. This approach will not only provide strain-specific, biogeographic knowledge on infectious agents for more precise diagnosis, but will also reveal the functional repertoire of pathogens, including potential AMR genes and mechanisms.

Devising New Targets, Molecules, and Strategies to Fight Pathogens

To prevail in the fight against microbial infections, new effective targets, molecules, and strategies are urgently needed. EMBL's vision is to build on its multidisciplinary expertise, platforms, and capabilities in informatics, pathogenesis, structural biology, high-throughput screening, and drug discovery to reduce current bottlenecks in mode-of-action (MoA) understanding of anti-infectives and to discover alternative strategies and new targets to combat bacterial and viral pathogens.

A major bottleneck in anti-infective discovery is the identification of the MoA of new compounds. This is vital to improve drug efficacy, reduce the potential for resistance development, guide combinatorial drug use, and mitigate adverse effects of drugs on other microbes or on the host. EMBL has pioneered systematic genetic (chemical genetics) and biochemical approaches such as TPP (Figure IB3B) to **identify the MoA of novel anti-infective compounds**. In the new EMBL Programme, these will be complemented with machine learning approaches applied to experimental data and public chemoinformatics and drug databases (ChEMBL, Open Targets, STICH, DrugBank; many developed or maintained by EMBL), computational docking approaches taking into account natural sequence variation, and *in situ* imaging of drug action (Tech Dev Box TD1_IB).

The second way in which EMBL can contribute is by providing novel cutting-edge approaches in drug discovery. This work will leverage EMBL's comprehensive crystallographic fragment screening and chemical biology platforms (Chapter 10: Scientific Services) to **identify new candidate molecules for rapid optimisation**. These targeted approaches will be geared towards antiviral and antibacterial drugs that specifically target facets of the host–pathogen molecular interface, such as candidate secretion systems, adhesins, and secreted proteins.

Moreover, EMBL aims to **develop new strategies that break with the current trends of antimicrobial drug discovery**. Previous efforts have concentrated on target-driven discovery of broad-spectrum monotherapies and often failed, as identified molecules were prone to rapid resistance development, lacked *in vivo* efficacy, or had adverse effects. EMBL will seek strategies that follow a new path. The first goal will be to identify molecules that target specific pathogenic traits (e.g. biofilm formation and intracellular growth) rather than conserved processes that are essential for the growth of all bacteria. This will yield more specific therapies, less prone to resistance development and less detrimental to the resident microbiota. One such direction will be to identify molecules that ectopically trigger suicidal systems of bacterial pathogens, such as the toxin-antitoxin pairs prevalent in bacteria (e.g. *M. tuberculosis* carries over 100 pairs). These consist of a 'toxin' protein that inhibits bacterial growth and an interacting 'antitoxin' (RNA or protein) that neutralises the toxin. EMBL researchers will strive to uncover the mechanisms of these pairs and the molecules that they sense. The ultimate goal will be to exploit these mechanisms to turn these systems against the pathogens that carry them.

EMBL will also continue exploring **drug combinations**, which remain largely unexplored for antibacterial treatments. Combinations can offer new solutions, allowing for reuse of neglected antibiotics or for repurposing of other drugs and even food additives, to inhibit AMR pathogens (Figure IB3A). More importantly, combinations offer ways to prevent, reduce, and revert antibiotic resistance. Recently, EMBL researchers discovered that combinations have species-specific activity and can thus be used to decrease the collateral damage of antibiotics to commensals (Chapter 4: Microbial Ecosystems, Figure ME4A). This reduces pressure on the entire community to develop resistance. To provide more natural treatment options, the combinatorial effects of bioactive molecules present in food will be explored. Efforts to use phages or probiotics to selectively remove looming pathogens from human microbiome will provide a further innovative strategy (Chapter 4: Microbial Ecosystems).

Preventing the Development and Spread of Antimicrobial Resistance

Technological advances in DNA sequencing, telecommunications, and artificial intelligence provide unrivalled opportunities for monitoring AMR in health systems, industry, agriculture, sanitation, and transportation. EMBL can advance this revolution by developing tools and knowledge for **AMR detection and discovery**, based solely on genomic information. These tools will promote precise diagnostics, guide medical decisions, and support epidemiology and environmental surveillance. EMBL's previous contributions have opened new paths for bacterial GWAS and predictive models. Ensuing goals are to: (i) improve the cataloguing of resistance gene families in databases; (ii) improve genotype-phenotype associations in diverse strains of pathogens and in microbiome communities for confident discovery of new AMR genes; (iii) track the evolution of resistance in infected patients and the environment; and (iv) build robust predictive models for AMR development and evolution. Extending on genomics-based efforts, EMBL aims to integrate omics, imaging, and phenotypic data and to pursue molecular mechanistic and structural insights (Figure IB3C) into the functioning of **new AMR elements** (e.g. transporters and ribosome modulation).

Particular attention will be devoted to **understanding and preventing the development and spread of AMR**. In bacteria, resistance and virulence genes can move within or between genomes carried on mobile genetic elements (Figure IB3D; Chapter 4: Microbial Ecosystems, Figure ME4B). This is a critical issue for human health, as it leads to rapid development of difficult-to-treat MDR superbugs. By combining expertise in bioinformatics, metagenomics, microbiology, genetics, chemical biology, and structural biology, EMBL aims to begin filling this knowledge gap, mapping the transmission routes, the underlying mechanisms, and its Achilles heel (e.g. environments where AMR confers a fitness cost). This understanding is essential to identify strategies to limit or even prevent the development and spreading of AMR, and to revert it in cases where it is already advanced. Here, EMBL aims to systematically explore paths that exploit the use of a second drug or chemical (inducing collateral sensitivity), microbial species interactions, and host responses.

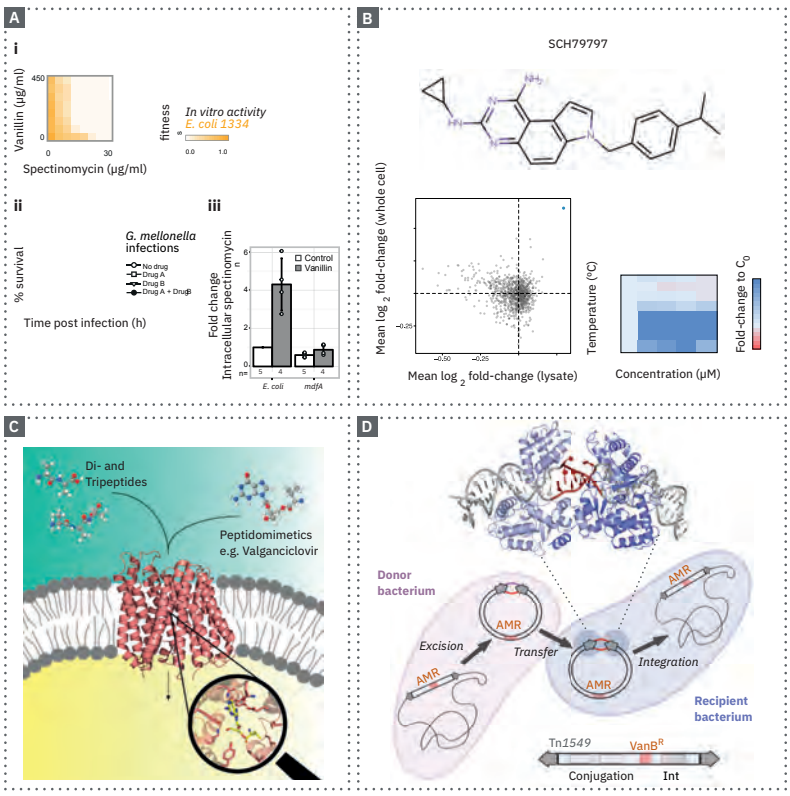


Figure IB3 | Mechanisms for new anti-infectives.

(A) Drug combinations provide new solutions – a neglected antibiotic, spectinomycin and a food additive, vanillin act synergistically against MDR *E. coli* **(i)** *in vitro* and **(ii)** *in vivo*. **(iii)** Vanillin promotes spectinomycin uptake by *E. coli* via the MdfA transporter (Brochado *et al. Nature* 2018). **(B)** Thermal proteome profiling (TPP) can be used to identify the MoA of antibiotics (Mateus *et al. Molecular Systems Biology* 2018; Imay *et al. Nature* 2019). Here, an example of a new molecule with dual activity is shown. One of its moieties targets the essential dihydrofolate reductase (FolA) in *E. coli*, causing its strong thermal stabilisation (Martin *et al. Cell* 2020). **(C)** A structure of the peptide transporter DptA with the antiviral drug valganciclovir provides insights into how the drug enters human cells and what sequence variants may lead to resistance (Ural-Blimke *et al. JACS* 2019). **(D)** Transposase-DNA complex structure reveals mechanism for mobile genetic element-mediated AMR transfer (Rubio-Cosials *et al. Cell* 2018). The Tn1549 transposon transfers vancomycin resistance from a donor (pink) to a recipient bacterium (blue), using self-encoded transposition (Int) and conjugation proteins.

Data Platforms

Platforms for Outbreaks and Monitoring of Global Threats

EMBL has a track record of curating, annotating, and sharing multi-omics data from pathogens, along with contextual metadata, primarily from scientific experiments. In recent years, collaborative projects with European public health agencies have aimed to develop platforms for data-sharing and analysis of outbreaks. The experience in the provision of data platforms and portals (e.g. the Pathogen Portal and the European SARS-CoV-2 Data Platform for the coordination of data from the current pandemic) has emphasised the need to **consider public health as a further arm of the scientific community that provides and uses molecular data**. Extending EMBL's operations to support public health brings important data for EMBL's traditional scientific services and research, but also allows for EMBL to engage in new science to predict, understand, and monitor outbreaks and global threats.

EMBL researchers have also developed novel genome workflows and DNA search tools which enable real-time global surveillance based on clinical sequence-based diagnostics for tuberculosis (TB; Tech Dev Box TD3_IB). Current collaborations with the WHO supranational TB laboratory in Argentina will pilot this platform as an early-warning system for national public health organisations (e.g. automated outbreak detection or cross-border transmission alerts) and as a means for measuring prevalence of AMR to current and repurposed drugs. Furthermore, this collaboration is intended to facilitate the adoption of sequencing by low- and middle-income countries.

Public health microbiology is now a major global force in terms of generating sequences and associated metadata, and in terms of its importance for human health. Currently, the uses to which public health agencies put genomic data are relatively limited, and not all data are deposited openly. EMBL will deliver platforms and analytics that make pathogen genomics data and microbiome metagenomics data more accessible, facilitating global collaboration. These data will be integrated with existing EMBL open data resources and will incorporate new types of data (e.g. socio-economic, travel, meteorological, climate change, food chain, social media, healthcare, and biodiversity).

EMBL will leverage existing pathogen data platforms to deliver key global infrastructure to address future rapid outbreaks and surveillance of pathogens via the **4D pathogen genome, variation, and phylogenetic map**. This will be directly accessible by users via a navigable map, underpinned by structured omics datasets, tools to manage and support data sharing, and an analytical machine to render raw sequences into genomes, variations, and phylogenies. The system will connect pathogen data to host data, such as vector distributions; host transcriptome, immunome, and microbiome; and clinical or epidemiological data. The system will drive open data sharing in infection biology and will enable early controlled access, allowing public health agencies to collaboratively pre-analyse data before making it public, and offering dedicated cloud computing.

Technology Development Box TD3_IB | Real-time genome-based global epidemiology service for TB.

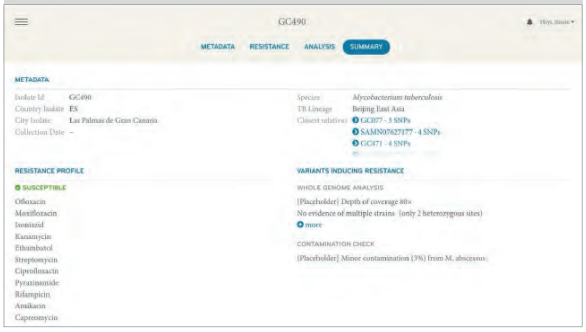
There are more than 10 million new cases of TB and 1.5 million deaths every year. One in five deaths is due to MDR *Mycobacterium tuberculosis*, the causative agent of TB. Solving two key problems would greatly aid global management of TB. First, comparing a bacterial genome from a patient with a live database of global infections is currently not possible, partially due to the sheer challenge of building such a system in a scalable manner. The second challenge is that the majority of the TB burden is borne by low- and middle-income countries, who have many challenges to the adoption of routine sequencing of TB and whose frontline staff have limited incentive to share data. Towards these ends, the Iqbal group is developing a platform, Mykrobe Atlas TB. The first problem is solved by developing a DNA search index capable of storing millions of genomes. The second is solved by providing state-of-the-art TB diagnostics connected to upload of data (if consented, and with embargo). Upload is rapidly followed by information on outbreaks, drug resistance, and monitoring, providing an incentive. An early version is being trialled in Argentina at the Malbran Institute in 2020, with potential for follow-up in Madagascar, India, and South Africa.

MYKROBE ATLAS TB

Data upload	Intended users	Features	Benefits
M. tuberculosis raw sequence data	National TB reference labs	Secure access	Outbreak detection
Geography (fuzziness supported)	TB control officers	Sharing within/ between organisations	Cross-border movement detection
No patient identifiable data	W.H.O	Customisable dashboards	Fast analysis: incentive for data submitter
		Filter by time/space/ genetics/resistance	

Real-time surveillance

Drug resistance profiles



Open Targets for Microbes

The Open Targets Platform is an integrated EMBL data resource that allows users to identify and prioritise targets for the development of safe and effective medicines for human disease (Figure IT4 in Chapter 12: Innovation and Translation). The platform integrates evidence from various databases at EMBL's European Bioinformatics Institute (EMBL-EBI) and elsewhere to identify and score likely targets, and provides additional data to allow prioritisation based on target tractability, known safety liabilities, and tissue selectivity. Currently the platform focuses on chronic human disease with a genetic contribution. Building

on its experience of developing the Open Targets Platform, EMBL will develop a sister platform for microbes, serving both academic and pharmaceutical communities, to tackle the challenges of infection biology. EMBL plans to liaise with relevant European institutes and industrial partners to develop an informatics platform that can be used to identify protein targets for anti-infectives, and to quantify disease relevance, tractability, and resistance evolution potential. The resource will promote the prediction and modelling of small-molecule target engagements (on pathogen or host), foster the prediction of resistance profiles, and help scientists to understand the MoA of novel compounds.

The infrastructure already available through the Open Targets Platform could be restructured to include pathogen proteins as targets, alongside the host proteins with which they physically interact. This information would be extracted from published experimental research, or would be predicted based on data from biological pathways and known interactions. Additional descriptions of disease ontologies and of AMR pathways would also be necessary. Such a database could be provided from the cloud as an open access resource to allow the following questions to be answered:

- What virulence proteins do pathogens encode? What information or data are available on these? What are their close orthologues from related pathogens or human genes?
- Which human proteins or receptors interact with the pathogen or its proteins? What networks are they part of? What drugs or compounds are available? How do these active compounds work? What bioactivity and AMR data are there?
- What clinical trials exist for compounds binding these targets? Are there drug repurposing opportunities? If drugs are being used in clinical trials, what are the known targets of these drugs? Are there known safety risks? What classes of mechanisms do these drugs fall into?

Impact

The recent SARS-CoV-2 pandemic has exposed the gaps that exist not just in our knowledge of pathogen biology and human susceptibility to infection, but also in our abilities to rapidly detect infections, monitor transmission routes, and develop therapeutics. EMBL in close collaboration with its member states can help address many of these limitations.

Firstly, there is a large societal impact that stems from the lack of broad and integrated approaches for mechanistic analyses of the biology of pathogens and their interactions with humans and other hosts. The use of cutting-edge technologies such as proteomics, structural biology techniques, high-resolution imaging, and computational methods to map host–pathogen interfaces across different levels and scales stands to provide a deeper understanding of pathogen biology, new therapeutic intervention points, and new means and mechanisms to modulate host cell biology. Within these approaches lies the opportunity to gain new knowledge of the principles by which pathogens invade cells and evade the immune system. From these principles, the global community will have a far better knowledge to react to new emerging pathogens. Cross-disciplinary collaborations such as the Centre for Structural Systems Biology (CSSB) at EMBL Hamburg, a joint initiative comprising three universities and six research institutes, will be important to focus on mechanistic questions in infection research, including concrete applications for drug discovery and new therapies.

Currently, AMR is shaking modern medicine to its foundations: invasive surgery, transplantation, chemotherapy, premature infant care, and care of the critically and chronically ill are some of the areas that depend on the ability to control infections. MDR pathogens are a major threat to global public health and a priority to be urgently addressed, both according to the WHO and the UN. The current SARS-CoV-2 pandemic might further expose the negative consequences of our shortage of working antibiotics as hospitalised

patients are regularly treated with multiple antibiotics to prevent pneumonias, which will in turn promote further AMR development. Building on the largest genomic databases, unique tools, multidisciplinary skills, and international research networks, EMBL will have unmatched possibilities to make a difference towards combatting resistance dissemination and the threat of MDR infections. The knowledge and intervention approaches devised at EMBL will benefit the entire European research landscape and society with a long-term medical and fiscal benefit.

The new infectious disease-related computational tools will improve pathogen and AMR detection. Dedicated data platforms are intended to deliver immediate utility to the EMBL member states, focusing on the needs of public health users. By gaining broad engagement, these will amass hugely valuable open datasets for science. For fundamental scientists in the member states, the 4D pathogen genome, variation, and phylogenetic map will provide a uniquely rich view into public data, integrating pathogen, host, geography, and time. Finally, Open Targets for Microbes will be a foundational resource that can underpin the discovery of future anti-infectives.