The collection and storage of biomolecular and clinical information is more accessible and cheaper than ever. A major challenge now is to make sense of the vast volumes of data being produced, which is where complex computational models can play a vital role.

Advances in technology and the development of new experimental methods have had a significant impact on the study of disease. This has led to new research directions, including: the acquisition of detailed ‘molecular fingerprints’ from patients containing information, for instance, on genotype, gene or protein expression, or metabolite levels; the study of intracellular processes in healthy and diseased tissue via the manipulation of gene activity within cells; and the construction of comprehensive disease-specific databases that combine patients’ medical history with laboratory and clinical data as well as saving relevant tissue samples.

Computational analysis and modelling can help make sense of the data being collected. Initially, molecular fingerprints can be used to identify biomarkers that signal an elevated risk of acquiring a disease or to confirm diagnosis. Information about intracellular processes can be used to construct artificial networks of the molecular interactions involved and evaluate their role in the disease, while more complex quantitative dynamic models can track the underlying molecular processes over time. Such networks might also help to predict both the likely course of a disease and its response to treatment.

Maintaining and analysing databases of disease-specific information will require significant computational resources. Although this will be a gargantuan data collection and analysis effort, the results will help guide clinical decision-making and tailor therapeutic approaches for the individual rather than the ‘average’ patient. For instance, doctors can use measurements of a patient’s biomarkers in addition to their medical history and clinical when choosing which treatment regimen to prescribe.

DATA DIFFICULTIES

Although these strategies are promising, research is at a relatively early stage and there are many issues to be resolved. The range of disease-specific data being collected on a systematic basis needs to be increased. More information should come from longitudinal studies, in which samples or patients are followed over time. These data should be of high quality and suitable for comparison; many newer technologies are prone to experimental variation, limiting their utility in the clinical setting. Accordingly, it is important to have computational models that can distinguish technological variation from true biological difference.

Another major challenge relates to the difficulties researchers can encounter when seeking access to clinical data and relevant biological samples. Numerous legal and ethical issues related to privacy and personal data protection must be addressed — along with many organizational issues — if this situation is to improve.

Researchers need reliable, accurate and trustworthy statistical tools to deal with the information being generated. It is too easy for false-positive results to be produced from small samples of complex data. Similarly, without sophisticated statistical techniques, it is impossible to incorporate elements of chance into dynamic processes. The technologies used to visualize the results of modelling activities are equally important; they must be user-friendly by displaying results in a way that can be readily understood by fellow scientists and medical colleagues. The intuitive nature of these visual presentations is critical when studying complex biological systems.

MODELLING OPPORTUNITIES

Diseases such as HIV/AIDS, tuberculosis, hepatitis C, influenza and malaria affect large patient populations, making this a particularly important use of computer modelling. The models must be updated regularly to keep up with the rapid evolution of infectious agents.

Interactions between a pathogen and its host. A modelling approach based on molecular networks can reveal information about the relationship between a pathogen and its host. The development of dynamic models that show how infectious agents replicate within cells will be an important step forward, as will quantitative descriptions of pathogenic spread throughout tissue and organs.

Vaccine design. Computational models can predict the likely outcome of improved strategies for vaccine design, such as the best combination of an antigen with an agent to boost the immune response. Data analysis

The Max Planck Institute for Informatics has developed bioinformatic models of HIV infection that support clinical treatment of AIDS patients. In collaboration with virologists and physicians throughout Germany, the models have been derived with statistical learning techniques, based on clinical data collected over two decades. They can be queried freely over the Internet (www.geno2pheno.org) and are helpful in the latter stages of infection when therapy options are limited (Lengauer, T. et al. J. Nat. Rev. Microbiol. 4, 790-797 2006).
Computational models of disease are invaluable for interpreting large swathes of patient-specific data.

More sophisticated statistical, algorithmic and numerical methods are needed to analyse datasets with complex biological and clinical information.

Significant progress will be made if separate models of the disease process — operating at different levels and spanning a range of timescales — can be combined.

**Pathogen evolution.** Infectious diseases evolve quickly to evade the immune response. This process can be modelled in order to select the best viral strain to use in batches of a vaccine. These analyses of viral ‘evolutionary drift’ are already influencing the design of vaccines protecting against seasonal influenza (Fig. 1).

Similarly, rapid evolution of pathogens can lead to the development of resistance to the drug therapy. By modelling this process using a statistical program, it is possible to predict the relative success of various treatment strategies (see Highlights box, below left). Further development of computational tools will incorporate patient details and therapeutic history into the analysis.

**Epidemiology.** Models that extend beyond biology to include demographic data and the social aspects of human behaviour can predict the risk of infection, and how the disease will spread, both locally and globally.

**PULLING TOGETHER**

Most recent disease-modelling efforts focus on only a limited aspect of the disease process. In order to create more realistic and useful tools, it is important to integrate these approaches and develop more dynamic, large-scale models. This will be a challenge, not least because of the different timescales involved: models of disease progression and healing span days or even years, whereas chemical reactions can be completed in a matter of microseconds.

Many of the applications outlined here are at an early stage; however, there is scope for a great deal of progress over the next 5 years. Ultimately, information based on computational disease models will benefit patients by improving diagnosis and prognosis, helping to develop new treatments and substantially reducing the risk of inadequate or even damaging therapy.

---

**For references see pages 70 and 71**

---

**Fig. 1** Phyllogeny of influenza viruses and annual variation.

Most recent disease-modelling efforts focus on only a limited aspect of the disease process.

---

above | Computational analysis and modelling can help make sense of the vast amounts of clinical data collected.

left | Gathering clinical data can lead to the identification of ‘molecular fingerprints’ for patients, concerning genotype, gene or protein expression, or metabolite levels.