Inflammation is an essential biological process that encompasses the response of pluricellular organisms to environmental stresses, such as physical damage to tissues, infection and other immune challenges. The inflammatory response seeks to avoid, contain, reverse and heal tissue damage provoked by such stresses. Inflammation is central to the pathology of major complex chronic conditions such as autoimmune diseases, arthritis, chronic lung disease, inflammatory bowel disease and psoriasis. More recently, inflammation was also linked as a major component of the pathophysiology of atherosclerosis, coronary heart disease, Alzheimer’s disease (Wyss-Coray 2006), several types of cancer (Clevers 2004) and metabolic syndrome (Maiti & Agrawal 2007). Therapeutic approaches based on thwarting the inflammatory response have proven to be of great clinical benefit for many of these chronic ailments. On a different time scale, acute severe inflammatory illnesses such as sepsis, trauma and acute pancreatitis also have broad societal and economic impacts (Angus et al. 2001). Unfortunately, and somewhat surprisingly, modulation of inflammation has met with disappointing results for illnesses severe enough to justify admission to the modern intensive care unit (ICU) (Remick 2003).

Sepsis is a clinical syndrome representative of acute, complex inflammatory diseases. Sepsis is defined as the systemic host response to infection with clinical manifestations that span a broad set of inflammation-related signs and symptoms, such as fever, tachycardia, tachypnoea and decreased arterial blood pressure leading to clinical shock, a state of insufficient oxygen delivery to tissues characterized by hypotension and acidosis (Levy et al. 2003). Although the host’s response to sepsis strives to contain infection and promote repair, the intensity of the inflammatory response often leads to compromised tissue function, organ failure and death. Severe sepsis accounts for 2–11% of all admissions to hospitals, approximately 750 000 cases a year, with an associated case-fatality mortality of 35 per cent (Angus et al. 2001). Mortality rates from sepsis and septic shock have not changed significantly over recent decades (Martin et al. 2003), and it affects the very young (Watson et al. 2003) and the very old disproportionately (Sands et al. 1997). The Centers for Disease Control and Prevention (CDC) reports that...
the incidence of sepsis increased from 7.4 per million patients in 1979 to 17.6 per million patients in 1987. Multiple organ dysfunction syndrome, associated with 80 per cent of deaths in modern ICUs (Angus et al. 2001), is causally related to the inflammatory response (Bone 1996), and it is also a common complication of other causes of acute severe inflammation, such as multiple trauma. The process of death typically includes progressive organ system shutdown requiring support of the circulation, ventilation and renal function.

In sepsis, acute inflammation is initiated by the recognition of danger signals using pathogen-associated molecular pattern (PAMP) receptors on the surface of dedicated surveillance cells, typically dendritic cells and tissue macrophages. Such PAMPs, associated directly to pathogens or their products (e.g. lipopolysaccharide, lipoteichoic acid, flagellin or bacterial RNA) activate dendritic cell and tissue macrophage Toll-like receptors (TLRs) (Abreu & Arditi 2004). Activation of TLRs initiates intracellular signalling of specific cascades leading to enhanced expression of early pro-inflammatory proteins, such as tumour necrosis factor (TNF) and interleukin-1 (IL-1), active in an autocrine, paracrine and endocrine fashion, with the purpose of mass mobilization of innate immunity. Regulatory anti-inflammatory proteins, such as IL-1 receptor antagonist (IL-1RA) and IL-10, almost synchronously follow. TLRs also have the ability to recognize molecular patterns, that, although not foreign, should not be accessible to those receptors in health. These damage-associated molecular patterns (DAMPs) are often the signature of cellular disruption resulting in the failure of intracellular containment of these DAMPs (Matzinger 2002). As disease progresses, it appears probable that there is perpetuation of injury, as spreading tissue damage promotes further inflammation, and inflammation contributes to tissue damage, although competing theories exist as to how this exactly happens (Prince et al. 2006). A pictorial representation of this dual contribution to the initiation and perpetuation of inflammation is shown in figure 1. Other important aspects of the acute inflammatory response have been described. Acute inflammation is associated with cell-type-dependent modifications of programmed cell death, appropriately retarding cell death in pathogen-fighting neutrophils, and hastening cell death in most other cell types. Inflammation promotes the production of inducible nitric oxide synthase (iNOS) leading to local vasodilatation promoting metabolite delivery and export, and key components of the complement cascade important for antimicrobial activity. Furthermore, trafficking of dendritic cells to local lymph nodes initiates specific target recognition by T-cells and clonal expansion of these cells. These steps are necessary for an appropriately controlled host response. However, the types of health challenges faced by patients in ICUs trigger inflammation at a scale unlikely to be compatible with survival, and, therefore, there is little guarantee that the evolutionary adaptation to severe infections is indeed appropriate. In other words, what is a ‘well-oiled machine’ in response to a typical challenge might not be well suited for maximal challenges. How this initial response leads to undesired effects of cellular dysfunction and organ failure remains a challenging problem from a reductionist perspective: in fact, a multi-scale systems approach may be the only fruitful approach because of the several mechanisms at different scales (figure 2) that appear to contribute to the clinical disease:

— First, an overly exuberant pro-inflammatory response can injure tissues directly. For example, the alarm-phase cytokine, TNF, is capable of inducing programmed death of functional cells in several tissues (Meldrum et al. 2006). Similarly, high-mobility group B1 (HMGB1), a cytokine-like protein that is released later in response to sepsis, has been shown to cause gut epithelial and hepatocellular injury (Sappington et al. 2002).

— Second, injury can be due to cellular hypoxia secondary to impaired tissue perfusion. Systemic inflammation causes vasodilation, increased microvascular permeability, and impaired cardiac contractility. Since flow to many vascular beds is pressure-driven below a tissue-specific threshold (Coleman et al. 1971), vasodilation will create inequalities in blood flow distribution with an overall decrease in blood perfusion of some tissues (through hypotension) and wasteful perfusion (through decreased resistance of the vascular bed) in other tissues. Vasodilation is thought to be caused primarily by increased production of nitric oxide (NO), which in turn is caused primarily by upregulated iNOS expression. Microvascular hyper-permeability also results from an excess of NO and various eicosanoids, as well as direct effects of TNF and IL-1. Increased permeability leads to extravasation of cell-free fluid from the circulation resulting in insufficient circulating blood volume and impaired cardiac output. Intravascular fluid loss is typically important, and preventing or
correcting this loss might be a very important therapeutic action. Cardiac pump failure in sepsis is also thought, from animal experiments, to be compounded by the combined effects of TNF, IL-1 and reactive nitrogen and oxygen species (free radicals, RNS and ROS, respectively) directly altering cardiac muscular cell function (Flierl et al. 2008).

Collectively, the effects of maldistribution of blood flow, hypovolaemia and impaired cardiac function lead to cellular hypoxia in several tissues. Since most cells depend on mitochondrial oxidative phosphorylation to generate energy, hypoxia inevitably leads to cell dysfunction or death.

— Third, activation of intravascular coagulation and thrombosis pathways may lead to organ dysfunction through thrombosis in the microcirculation. Inflammatory mediators, such as TNF, initiate the coagulation cascade (Esmon 2004), as demonstrated experimentally in large human cohorts of patients with sepsis (Dhainaut et al. 2003), irrespective of the causative microorganism (Kinasewitz et al. 2004). Inflammation also stimulates fibrinogen synthesis, further promoting coagulation. Contributing to the problem is potential occlusion of vascular beds by platelets and inflammatory cells themselves, either physically or by the formation of extracellular nets (Ma & Kubes 2008), the evolutionary goal of which may be to trap circulating microbial entities (Hickey & Kubes 2009).

— Fourth, processes other than hypoxia may result in cellular energetic failure. With increased NOS activity during inflammation, energy production might be impaired despite adequate intracellular PO₂. NO directly competes with O₂ as a substrate for respiratory chain complex III. Large quantities of intracellular NO, resulting from iNOS activity, may lead to significant competitive inhibition of electron transport, and thus inadequate ATP production. Inappropriate decreases in oxygen utilization have been described consistently under...
inflammatory stress. It may also be that learned mechanisms of reallocation of cellular resources under stress might be prioritizing cell survival over function in the face of limited energy supply. Oxidative stress injury could also be a major contributing mechanism to cell and organ dysfunction (Granger et al. 1981). According to the ‘classical model’, ischaemia leads to conversion of xanthine dehydrogenase into the enzyme xanthine oxidase (XO). In addition, during ischaemia, ATP may ultimately be degraded to xanthine. During reperfusion, O2 becomes available to support XO-dependent oxidation reactions that generate ROS. These free radicals, by reacting with a myriad of substrates and enzymes, interfere with normal cellular metabolism. Although ROS formation is directly related to the degree of ischaemia and the duration of hypoperfusion, its role in causing ischaemia/reperfusion (I/R) injury has been questioned (Brealey et al. 2004).

In summary, hypothesized and documented inflammatory response pathways interact in a complex manner, with numerous feedback loops, thereby motivating a systems-level approach to inflammation modelling. It does appear, however, that energy failure is a likely common pathway for cell dysfunction, organ failure and ultimately death.

2. TREATING SEPSIS

2.1. Clinical successes

Modulation of immunity has led to remarkable progress in transplant medicine, where the chief objective of immune suppression of the host response to grafted immunogenic solid organs is balanced against toxic side-effects of immunosuppressive regimens and the associated propensity to opportunistic infections. Years of research and observations have broadened the standard concept of immune suppression to the more subtle notion of immunotolerance. In other words, how can the immune system be fooled in accepting what is clearly non-self (Weissman & Shizuru 2008)? Several successful clinical trials have propelled the use of anti-TNF antibodies and IL-1 receptor antagonists as classes of agents for rheumatoid (Waugh & Perry 2005) and psoriatic arthritis (Punzi et al. 2007), as well as Crohn’s disease for the former agent (Carter et al. 2003; Camilleri 2007), where the clinical rationale of interrupting an inflammatory loop perpetuated by an overly sensitive immune system appears sound. In addition to the well-documented side-effect of increased risk of infections, a worsening of psoriatic-like conditions (Fiorino et al. 2009) as well as an increased incidence of malignancies (Bongartz et al. 2006) have been reported, plausibly related to impaired immune surveillance (Beyer et al. 2009). Nevertheless, it does appear remarkable that administration of a single agent, following a pre-prescribed dosage regimen, is effective in alleviating symptoms of complex chronic inflammatory diseases where hundreds of molecular species and cell types participate in phenotypic expression, and where the body has ample time to compensate in a number of ways to both excessive TNF expression and its treatment.

2.2. Clinical failures

In contrast, with the exception of the proper use of antibiotics (Kumar et al. 2006) and human recombinant-activated protein C (Bernard et al. 2001), an anticoagulant with anti-inflammatory properties (Joyce & Grimm 2002), trials at immunomodulation have largely failed for acute severe inflammatory illness, and of sepsis in particular. This failure has triggered much consternation and soul-searching in the critical care community (Marshall 2000; Sweeney et al. 2008). Presumably, major reasons for this dismal record relate to expectations as to how one should measure success and failure of an intervention in sepsis, and a process of insufficient reasoning in translating abundant pathophysiological knowledge acquired from extensive in vitro exploration and successful preclinical profiles of many lead compounds into clinically and genetically diverse human beings (Eichacker et al. 2002). Clinical trialists have also been obfuscated by the current burden of proof of efficacy, which still remains an improvement of all-cause mortality at a fixed time horizon, typically 28 days, rather than evidence of biological efficacy. Favourable biological activity is suggested in several trials by improvement in physiological markers (Annae et al. 2002; Bakker et al. 2004; Watson et al. 2004; Sharshar et al. 2006), typically most pronounced in the sickest patient (Bernard et al. 2001; Marshall et al. 2003; Lipiner-Friedman et al. 2007). As an example, a review of the first 7500 patients receiving TNF-targeting interventions suggests a significant overall benefit of 4 per cent in reducing 28-day mortality. However, no single phase III randomized clinical trial quite reached significance (Panacek et al. 2004). An in silico exploration of anti-TNF intervention in sepsis suggests that the overall mortality difference is a balance between patients truly helped (Bernard et al. 2001; Marshall et al. 2003; Lipiner-Friedman et al. 2007). As an example, a review of the first 7500 patients receiving TNF-targeting interventions suggests a significant overall benefit of 4 per cent in reducing 28-day mortality. However, no single phase III randomized clinical trial quite reached significance (Panacek et al. 2004). An in silico exploration of anti-TNF intervention in sepsis suggests that the overall mortality difference is a balance between patients truly helped (Bernard et al. 2001; Marshall et al. 2003; Lipiner-Friedman et al. 2007).

Experimental data of the necessity of TNF for a successful response to treatment (Pfeffer et al. 1993) support the concept that anti-TNF treatment could prevent a subgroup of patients from mounting an appropriate challenge to infection. Because of the lack of availability of reliable biomarkers for infection and the septic process, no amount of clinical data could prove harm, but this is nevertheless suggested by the severity—response relationship and in silico explorations. This conundrum is compounded by the nonlinear and time- and context-dependent interplay among multiple molecules, cell types, tissues and organs in vivo. The processes of cellular recruitment, proliferation, mode of death (necrosis versus apoptosis) and energetics manifest clinically as organ dysfunction; the complexity of this web of interactions has rendered sepsis a fertile field of often promising preclinical studies based on a reductionist approach to treatment, which failed to translate into successful therapies in actual patients. An engineer or mathematician would say that sepsis is a problem of
dealing with transients: the added difficulty in the context of acute illness is that the clinician is attempting to interfere with a highly dynamic process, rather than one that has reached a ‘biological steady state’, loosely defined as one where the time scale of the intervention is much faster than that associated with disease evolution. (It should be noted that the term ‘biological steady state’ is an equilibrium state—strictly different from steady or asymptotic in the mathematical sense.) Furthermore, there is considerable variation between the onset of disease and encounter with the healthcare system; the relative contributions of variation in stage of disease at the time of encounter and individual host–pathogen idiosyncracies to clinical observations are difficult to separate. These considerations suggest that sepsis treatment, conceptualized as interfering with a dynamic process, may be more successful if titrated, and also based on a comprehensive disease model, which is currently lacking.

2.3. The future

The National Institutes of Health (NIH) in its Roadmap Initiative—implemented to foster transformative and cross-cutting research beyond the scope of a single institute—promotes computational modelling as a tool of knowledge discovery and possibly the preferred approach to integrate such knowledge across biological scales (http://nihroadmap.nih.gov/). Similarly, the Food and Drug Administration (FDA) in its ‘critical path’ document (Food and Drug Administration 2004) has called for the use of in silico models to augment preclinical studies in animals in order to develop novel therapeutic agents and devices. At present, there are no clear guidelines as to how such initiatives should be benchmarked and no external incentives to develop such initiatives within the pharmaceutical industry. Furthermore, there has been little progress in redefining or modifying the criteria of efficacy for new biologicals. The experience of the last three decades of human clinical trials, combined with a high bar for success, has resulted in fewer registrations of new randomized trials for sepsis. Therefore, the critical care community has promoted the Surviving Sepsis Campaign, a long collection of recommendations to practising clinicians as to the state-of-the-art of sepsis support and treatment. These recommendations are the result of an extensive review of existing evidence regarding diagnosis, fluid management, infection source control, antibiotic treatment and organ support for victims of severe sepsis. Remarkably absent are agents that are truly modifiers of the host response (Dellinger et al. 2008).

It would appear that model-based combination approaches constitute the best hope for leapfrog progress in sepsis therapy. Such approaches would clearly integrate a disease model, in the mathematical sense, existing preclinical data and base recommendations on biological readouts, such as biomarkers, which are of pathophysiiological importance. Proof of concept that such an approach would shift our approach to intervention strategies and clinical trial design remains elusive for a number of legitimate reasons (Marshall 2004), including: (i) lack of a universally accepted disease model, (ii) lack of relevant data to calibrate and validate such models against, (iii) lack of biomarkers of disease activity, (iv) lack of point-of-care methods to measure such biomarkers, and (v) lack of understanding of the concept of model-based interventions by end-users. Enormous progress has been made at the basic science level, and sepsis, as a rapidly evolving disease with high case fatality, offers a unique opportunity of using systems medicine and transdisciplinary collaboration to achieve translational success.

3. MODELS FOR DECISION SUPPORT

As healthcare professionals, clinicians in practice make informed diagnostic and therapeutic decisions for patient care; this is implicitly based on a ‘model’ composed by the clinician’s understanding of the relationships between interventions and expected outcomes of such interventions, in the context of the patient disease state. Similarly, engineers working to solve a process problem in an industrial setting use their knowledge to diagnose the problem and make appropriate changes. When feasible, this decision-making is automated via a control system or algorithm; in regulatory (setpoint tracking) mode, this system compares measurements with desired values (setpoints) and calculates the manipulated variable changes that need to be made to bring the measured quantities to their respective setpoints (figure 3). The current ‘state-of-the-art’ in the industrial setting is (non)linear model-based control, where a linear (Muske & Rawlings 1993) or nonlinear (Morari 1994; Qin & Badgwell 1999) mathematical model of the process is used explicitly in the calculation of the input changes needed to return the process to its desired operating mode. A tutorial introduction to nonlinear model predictive control can be found in Rawlings et al. (1994). When translating these tools to medical practice, a formal loop-closure in the engineering sense is unlikely, given that clinicians are responsible for patient safety and are justifiably hesitant to allow a computer algorithm to make a critical decision that has traditionally been made by a human. However, using a model-based algorithm to recommend a treatment intervention that is either accepted or over-ridden by a clinician may provide improvements in patient treatment by systematically exploring the potential treatment space and identifying the intervention best suited to return the individual patient to a healthy state.

An open-loop approach to treatment design and hypothesis evaluation is the focus of the PhysioLab platform of Entelos, Inc. (www.entelos.com), which uses a top-down approach to construct disease-specific in silico models that capture both the biology and the dynamics of the biological response. Simulation of the PhysioLabs in the areas of asthma, obesity, diabetes (types 1 and 2), rheumatoid arthritis, cholesterol metabolism, cardiovascular/atherosclerosis and skin sensitization (allergy) can be used to: (i) test a proposed intervention quickly on a simulated individual or population, (ii) hypothesize and evaluate interventions that
target specific receptors or pathways, and (iii) identify potential targets for a drug or antibody intervention that may modulate the disease state. Since these studies are in silico, it is possible to perform a much broader and deeper analysis, for far less money and at far greater speed, than an equivalent series of studies in vitro or in animals.

A critical element in this algorithmic framework is the quality of the patient model; in fact, the quality of the underlying model correlates directly with theoretically achievable performance from a model-based control system (Morari & Zafiriou 1989). The model, and its corresponding control system, take on the characteristics of the problem to some degree. In the context of a drug administration decision-support system, models of pharmacokinetics (PK; drug concentration versus time) and pharmacodynamics (PD; disease and toxicity responses to drug administration) need to be constructed. The building of such models is certainly not a new concept; PK models of drug distribution were discussed conceptually as early as 1937 (Teorell 1937a, b). An entire section below is devoted to the classes of models often employed and their respective merits/detractors. The recent advance, driven by the decrease in computer size and simultaneous increase in calculation speed, is the ability to use such models in real-time decision support, as depicted in figure 4. In the forward path lies the patient or animal undergoing treatment; this is the component for which clinicians have outstanding linear intuition. As this is often statistically motivated, the ability to alter one or two manipulated inputs, having linear effects on the outputs, would not require the level of systems understanding or computational complexity we advocate here. In contrast, consider the following: there are many \( n > 3 \) measurable variables, with particular variables changing importance based on the operating (or disease) state of the patient; there are multiple potential interventions, and the effects of these inputs on clinical observables may not be a linear relationship; finally, the possible interventions may interact either positively or negatively in the determination of patient outcome. When a model is coupled with a suitable systems engineering optimization or control algorithm, a formal decision-support system may be used in a variety of modes, including simulation of candidate hypotheses and interventions as well as making treatment intervention recommendations.

While the earlier reference to model quality versus performance skewed heavily in favour of developing detailed models (Morari & Zafiriou 1989), a qualification is that the ability of the model-based decision-support system to return useful (or even meaningful) recommendations is highly dependent on the appropriate complexity and structure of the underlying model, in addition to its accuracy (Parker & Doyle 2001). In other words, disease models and decision-support systems should be designed with the end-user in mind; clinical observables should be included as relevant readouts in an easy-to-use interface to the underlying computational system. Understanding some of the potential model scales, benefits and potential pitfalls is important prior to selecting a model for use in a decision-support system.

### 4. MODELLING TOOLS

A commonly employed tool, as alluded to above, is the statistically derived model based on observed associations between available measurements and an outcome, perhaps at a fixed time point (e.g. 28 day mortality). A key feature of statistical models is that predictions are based on prior observations. Consequently, such models lack sufficient predictive accuracy to provide good predictions in settings broader than those used to generate the data used in model development. This is a concept different from that of external validity, which refers to the preservation of predictive accuracy under similar experimental conditions: for example, in cohorts of patients with the same disease process, but originating from a different study. A different experimental setting would be, for example, to predict the effect of treatment from data.
that do not include treated patients. It is therefore outside the scope of this class of models to predict effect based solely on known mechanism of action in the absence of prior clinical observations, a deal-breaking limitation in their helpfulness in assisting the development of new therapeutic strategies. Statistical models, by construction, are designed to describe cohort behaviour and not individual behaviour. Efforts at constructing data-driven individual models have met with mitigated success and are not used in clinical practice (Chang & Bihari 1994). An additional shortcoming of most statistical models is their static nature. The dynamics of pathogenic infection and the inflammatory response cascade play a vital role in disease response, and efforts to apply more sophisticated statistical models, such as microsimulation methods, in the prediction of time-dependent outcomes, are complex and restricted to cohort behaviour (Clermont et al. 2004a).

4.1. Phenomenological models

Perhaps the most commonly encountered dynamic models, phenomenological models are designed to capture the observed biological response using a small number of equations and parameters. This modelling approach is ubiquitous in biomedicine: cancer chemotherapeutics are commonly modelled using linear compartmental models, using popular software packages such as ADAPT II (D’Argenio & Schumitzky 1997) and NONMEM (Sheiner & Beal 1980; Beal & Sheiner 1982; Beal 1984); in the diabetes field, the most commonly encountered model is the Bergman ‘Minimal’ Model (Bergman et al. 1989). The quality of this model has been debated (Quon et al. 1994; Finegood & Tzur 1996), but as long as its limitations are acknowledged (Weber et al. 1989; Steil et al. 1994) the original model and its extensions, including contributions from fatty acids (Roy & Parker 2006) and exercise (Roy & Parker 2007) among others, continue to be used successfully. Low-order model representations have also been employed in inflammation (see §7 for model details). The underlying structure of phenomenological models is the compartment, with each compartment represented mathematically by an ordinary differential equation (ODE), and the rates into and out of a compartment being either linear or nonlinear depending on the phenomenon requiring capture. For example, the following phenomenological model of sepsis has been previously proposed (Day et al. 2006a):

\[
\begin{align*}
\frac{dP}{dt} &= k_{pg}P \left(1 - \frac{P}{p_m}\right) - \frac{k_{pm} s_m P}{\mu_m + k_{mp} D}, \\
\frac{dN^*}{dt} &= -s_m R - \mu_n N^*, \\
\frac{dD}{dt} &= k_{dp}(f(N^*)) - \mu_d D, \\
\frac{dC}{dt} &= s_c + \frac{c_{uf}(n^* + k_{ud} D)}{1 + f(n^* + k_{ud} D)} - \mu_d C,
\end{align*}
\]

(4.1)

Here \( P \) is pathogen, \( N^* \) is activated phagocytes, \( D \) is a marker of cellular damage and \( C \) is a canonical circulating anti-inflammatory mediator providing overall stability to the inflammation system. Though highly abstracted, as stated by the original authors (Day et al. 2006a), it is representative of the dynamics necessary to capture the systemic inflammatory response. Nonlinearity is a common characteristic encountered in biological systems, with saturation phenomena perhaps the most common of the nonlinear terms. Michaelis–Menten relationships are used to capture saturating effects of \( P \) on its own removal (second term on right-hand side of equation (4.1)) and the effect of \( R \) on \( N^* \). Higher order effects are captured using Hill-type nonlinearities, as in equation (4.6), which can represent two-ended saturation at both low and high \( V \), as well as being used as a delay approximation when employed in dynamic equations as a function of state variables. Other common nonlinearities in biological systems include bilinearity (as in the last right-hand-side term of equation (4.1)) and inhibition, a second-order variant of which is shown in equation (4.5) where \( C \) downregulates the function \( f(V) \).

The shortcoming inherent in phenomenological models is captured by their name—they model the observed response or behaviour at the scale of interest, but lack mechanistic understanding and description. And while the phenomenological model structure is often chosen to simplify model parameter estimation from the available experimental data, the nonlinearities can interfere with identification as well. The calibration of parameters to experimental data is markedly easier if the dynamic model is linear in structure, although there are minimum data requirements (see Ljung 1999). Tools based on computer algebra (Audoly et al. 1998) have been developed to establish the property of \textit{a priori} identifiability—the ability to uniquely identify (and quantify) all model parameters from the available experimental data. While this is a theoretical property, in that measurement uncertainty and biological variability may negatively affect quantitative accuracy of the parameter estimates, it is a valuable test that can establish the need for more measurements or a redesign of the experimental protocol prior to executing the experiment. Extending this analysis to general nonlinear systems remains an open problem, but some tools for specific nonlinearities (e.g. polynomial structures) have been developed (Audoly et al. 2001; Polisetty et al. 2006; Belin et al. 2007). The challenge in both parameter calibration and \textit{a priori} identifiability provided by ratios, common in Michaelis–Menten and Hill equations, motivates the exploration of more biologically motivated model structures. While these mechanistic descriptions are typically not
linear, and of potentially higher dynamic (state) order, the ability to incorporate mechanistic information through detailed biological studies and measurements may provide an improved mapping between measurements and the model, while simultaneously assisting in the model parameter identification process. The eventual use of a model is further discussed in §5, below, but a brief foreshadow about the relationships between model structure, accuracy and utility is warranted here. When using a model to make a treatment decision, the concept of control-relevance—a model property defined as the ability to provide accurate predictions while being mathematically useful (e.g. invertible, quickly simulated) in a control or optimization algorithm—will contribute to the choice of model structure.

4.2. Stochastic modelling

The number of cells or molecules participating in a process that models wish to simulate may be insufficient to allow cyclical resurgence of bacterial cells (Kumar et al. 2004), a non-physiological physiotype permitted by the persistence of infinitesimal (<1) quantities of microbes. In such cases, noise and stochastic effects play important roles in realistic descriptions of system behaviour. There has been a great deal of interest in applying stochastic simulation algorithms (SSAs) to cellular processes (Arkin et al. 1998); most of these methods rely on the Gillespie algorithm (Gillespie 1977), which is an exact simulation of a stochastic reaction process. Extensions of the Gillespie method provide the core for a number of recent simulation tools (Adalsteinsson et al. 2004) and methods to improve the performance of these computationally intensive methods (Rathinam et al. 2004) review the advantages and disadvantages of such methods relative to the standard ODE approach outlined above. Algorithms aimed at improving the Gillespie model may still be computationally prohibitive when the number of reactions or molecules are large. In this case the net reaction rate becomes very big so that the time between events is so small that millions of steps must be taken to advance in time, with no clear prioritization of a subset of reactions. A now classic approach to this difficulty is to make an approximation in the form of the chemical Langevin equation (Gillespie 2000) (in the mathematical literature, a stochastic differential equation for Brownian motion of a particle).

Stochastic models are also important because their solution may yield qualitatively different behaviours from deterministic ODEs. For example, the ODE solution of a bistable system will evolve to either solution depending on initial conditions, while the richer stochastic solution will typically flip back and forth between the two steady states. The mean time spent in each steady state can be calculated using probability theory, and is involved in solving certain partial differential equations. Stochasticity probably plays a major role in determining which programmes will be activated in a given inflammatory cell under immunological challenge and therefore its future phenotype. Lipniacki et al. explored stochastic formulations of inflammation-relevant intracellular signalling (Lipniacki & Kimmel 2007) and of stochastically governed cell differentiation of T-lymphocytes (Lipniacki et al. 2008) in a bistable model.

4.3. Populations, individuals and cells, and back again

Population models are commonplace in the field of pharmacokinetics (Beal & Sheiner 1992; Sheiner & Ludden 1992; Carson & Cobelli 2001). A large number of patients, each having a small number of data points collected, are used to characterize the population average behaviour as well as the key covariates (e.g. age, race, body weight) that contribute to interpatient variability. Generally, this approach modifies a phenomenological model, like those of equations (4.1)–(4.6), by using model parameters with multiple dependencies, as follows (Sheiner & Ludden 1992):

\[ p_i = \theta_{pi} + \eta_i + \theta_i C_i. \]  

(4.7)

Parameters \( p_i \) for patient \( i \) are a function of the (sub)population mean value of the parameter \( \theta_{pi} \), interindividual variability in the parameter \( \eta_i \), and any known correlative effects \( C_i \) scaled by their (sub)population mean correlation \( \theta_i \). With sufficient data, both the underlying model structure (through the observed dynamics) and the population variability (through the need for \( \eta \) and \( \theta \), or \( C \), to describe individual responses) can be characterized simultaneously. Software tools are available for constructing these models (including NONMEM (Beal & Sheiner 1992) and SPK (Resource Facility for Population Kinetics 2008)), but these tools often provide dramatically better performance with linear model structures. Given the nonlinearity often present in mechanistic representations of biological systems (especially at the cellular level for enzyme kinetics and saturations), another method may provide a more accurate model at the cellular (population) scale.

An alternative method for representing population behaviour, now in the context of cells rather than patients, is to capture the population response as a distribution, i.e. to use a population balance model (Ramkrishna 2000). The heterogeneous responses of the cells to systemic perturbations are captured through internal cellular properties (e.g. cell sensitivity, cell-cycle stage, inflammation response level, oxygen, substrate or cytokine concentration) and the definition of kernel functions. Drawbacks of this model structure are the development of the kernel functions (which are often statistical distribution driven rather than derived from biological mechanism) and its computational complexity; the common form of a population balance model is a set of partial integro-differential equations. Depending on the number of intracellular parameters and the boundary conditions, numeric solution varies from complicated to intractable (Mantzaris et al. 2001a–c).
Recognizing that system-level response to cellular events is composed of the collective response of individual cells, a model structure explicitly recognizing both the cellular model complexity and the existence of population phenomena is the cell ensemble (Domach & Shuler 1984; Shuler 1999; Henson 2003). Here, a single (intra)cellular model is specified, but the parameters of the model are recognized as coming from distributions (which are specified by the user). A population of \( N_c \) cells is generated via Monte Carlo sampling from the parameter distributions (Ogunnaike 2006), but with the intracellular mechanistic equation structure held constant for each cell. Overall system response is then generated by simultaneously simulating \( N_c \) cells. To capture intercellular dynamics and interactions with the physiological system, the cellular models are coupled to extracellular equations representing key nutrients or cytokines, similar to extracellular dynamics in population balance models (Daoutidis & Henson 2002; Zamamiri et al. 2002). By simulating the large number of individual cells (recognizing that the large number of equations will have an impact on model simulation speed and the ability to perform analytical analysis), the interplay between challenges (e.g. pathogenic infection) and system response at the whole-organism and cellular level, through the potentially different activation responses and cytokine production of individual cells to a stimulus, can be observed. A potential drawback of this model class lies in the difficulty in handling cell division, which may or may not be critical in models of the inflammatory response: neutrophils and macrophages, key effectors of the inflammatory response, are terminally differentiated cells (cells that do not divide); on the other hand, clonal expansion is a critical aspect of cell-mediated and humoral immunity, thought to play an important role in chronic inflammatory processes and the later phases of acute inflammation.

### 4.4. Rules-based models

There are situations where continuous quantitative models (e.g. ODEs) are less relevant because of the inability to precisely quantify state or variable values. For example, the exact peak value of circulating IL-6 may not correlate with outcome, but prolonged ‘high’ levels of IL-6 in the bloodstream may signal uncontrolled inflammation. Likewise, the ability to exactly quantify the intracellular reaction kinetics is not currently possible given the inability to measure all the necessary reactive intermediates. The reactants and products are well characterized, however. As a result of these structural uncertainties, the exact nature of quantitative models can be replaced with a more ‘fuzzy’ or rules-based approach.

Fuzzy (or fuzzy logic) modelling is used to capture imprecise or uncertain events for systems, where a precise model may be too difficult to construct or may not exist, while fuzzy control addresses the fact that human decision-making operates with approximate data and implicit objective functions (Mahfouf et al. 2001). In this approach, a set of tabulated rules would be constructed from the underlying physics, biology and medical understanding for the process in question. Quantitatively, variables can be classified as ‘low’, ‘medium’ or ‘high’, through the use of activation functions (conceptually similar to the formulation of neural network models, the use of mathematical formulations such as the Hill relationship or hyperbolic tangent is common). As a result, decision trees and fuzzy logic control can be used to make clinical decisions. A key advantage of this framework is the ability to specify both modelling and control (treatment) rules in clinical terms. Drawbacks include the training of models when fitting to data is required, as well as the structural selection of the underlying mathematics and solution (e.g. what is the shape of the activation function, how many levels to use for variables and interactions). It can be seen from Mahfouf and co-workers (2001) that there appears to be significant medical adoption of fuzzy modelling tools, although it can be argued whether this is a result of their ease of adoption and comprehension for clinicians or because they are the best tool for the job.

An alternative rule-based framework is the BioNetGen language (BNGL; http://bionetgen.org), a model-building paradigm that has been developed for biological systems (Blinov et al. 2004; Faeder et al. 2009). Rule specification is accomplished by defining the reactants and products with the corresponding reaction rates. The existence of specific intermediates—such as the phosphoforms along a multi-step kinase pathway or the activation of a molecule requiring multiple individual bindings to activate—need not be specified in the overall reaction, as they are handled automatically by the software in the construction of the simulation equations. While the underlying model remains of ODE or SSA types for BNGL models, the specification of such models is simplified dramatically through the rules-based engine and the simulation can be done in either deterministic or stochastic mode. A further advantage is the automated code generation by BNGL compilers, thereby reducing the typographical errors common in specifying what could be hundreds of equations for a cellular pathway or cell–cell interaction dynamics.

### 4.5. Physiological modelling

Anatomy and physiology describe, respectively, the connectivity and the physical and chemical interactions of the organs and tissues in the body. While modelling in this manner is not a new idea, in that the concept was recognized by Teorell as early as 1937 (Teorell 1937a,b), advances in computational power have made these models attractive in that they provide a greater degree of biological insight than what is available from phenomenological models. Furthermore, tissue-specific information and mechanism, when available, can be included in an intuitive and mechanistically accurate manner. Physiological models of drug pharmacokinetics are commonly employed in the cancer field (e.g. Chen & Gross 1979; Doyle et al. 2007; Florian et al. 2007), and physiological PK/PD models have been used in a variety of fields including diabetes (Sorensen 1985; Parker et al. 2000; Roy 2008), anaesthesiology (Wada & Ward 1994), and cancer (Jusko & Ko 1994; Florian et al. 2007).
Mass balances around tissues of interest are used to construct a physiologically based model of cytokine or intervention dynamics; each tissue is therefore represented mathematically as a set of (ordinary differential) equations describing the rate of change of substances of interest within the tissue space, which can be represented as total tissue space or subdivided into vascular/extravascular (and further extravascular subdivision to interstitial/intracellular is possible). A candidate model with tissue compartments is shown in figure 5. Average values for many of the flow and volume parameters in PBPK models across multiple species are available in the literature (US EPA 1984; International Life Sciences Institute 1994). A comforting fact in using such literature parameters is the fact that circulating concentrations and measurable levels are often more sensitive to the parameters in metabolic terms than tissue flows and volumes (discussed for a diabetes example in Parker et al. (2000)). The associated drawback is that the remaining metabolic or reaction parameters need to be estimated from experimental data. The most commonly available data, systems-level measurements of circulating cytokines for example, can provide guidance in model specification, but these data are often insufficient to fully parameterize a physiologically based model because of failures in a priori identifiability—some parameters are simply not uniquely identifiable from circulating measurements. Tissue- or cellular-scale measures are often required to fully characterize this high-order model, which may require the scaling of in vitro experimental results to the in vivo scenario or the interspecies scaling of preclinical animal-derived parameters to the human patient case. Although mechanism, when identified from cellular experiments, is generally conserved, the interactions of cells with the in vivo environment often leads to changes in parameter values when scaling across these scenarios.

Hence, the most significant uncertainties that manifest in a physiological model involve the characterization of the intra-tissue dynamics, which are closely related to the availability and confidence in experimental measurements. As a result, the absence of tissue- or cellular-scale detail forces the modeller back to a phenomenological description of some aspect of the model. Despite this potential shortcoming, the ability to represent the body using physiological and anatomical accuracy and connectivity may still provide further benefit in understanding the system dynamics in response to disruptions. For example, a large recruitment signal for neutrophils at an injury site would not necessarily induce systemic recruitment of neutrophils.

Figure 5. Compartment-based physiological model schematic (cancer case study—note the tumour compartment). Arrows between compartments are flows (typically blood, carrying drugs or cytokines of interest quantified as concentrations). Tissues are single compartments (well perfused) or subdivided into vascular/extravascular spaces (if significant transport resistance exists). Metabolic effects, such as clearance, DNA incorporation or irreversible intracellular binding are represented by diagonal arrows exiting particular (sub)compartments. RBC, red blood cell.
Throughout the body. In contrast, high circulating cytokine levels resulting in maladaptive recruitment of neutrophils or macrophages to remote uninjured sites could negatively affect the ability of the host to fight off pathogen invasion and cause undesired damage in otherwise healthy tissue. These effects that could not result from a strictly phenomenological treatment, but can be observed in physiologically based simulations, even with incomplete mechanistic descriptions of cellular events.

5. Decision Support Tools

While mathematical models can be used to characterize and understand the dynamics of inflammation, a key reason to construct such a model is to use it explicitly in the construction of a decision-support system for real-time use in a clinical setting. Hence, an accurate model is desired (recall that model quality affects theoretically achievable performance (Morari & Zafiriou 1989)), but a model that can be efficiently and effectively employed in a systems engineering framework, known as a control-relevant model (Parker & Doyle 2001), is perhaps more useful. As a result, an imperfect model that captures the key observable behaviour may be the best model. Once such a model, in one of the frameworks discussed in the previous section, is constructed, the following tools can be used to design treatment interventions.

5.1. Optimization and optimal control

Optimal control is a tool commonly described in the literature in the solution of biomedical systems problems. Here an objective is coupled with a set of constraints, and the entire problem is solved either analytically or via programming techniques. Optimal control problems are mathematically posed as follows (Bertsekas 1995):

$$\min_{u(t)} J(y(t), u(t), r(t)),$$

subject to

$$\dot{x} = f(x(t), u(t), p),$$

$$y(t) = h(x(t), u(t), p),$$

$$u_{\text{min}} \leq u(t) \leq u_{\text{max}},$$

$$y(t_f) = y_{\text{fin}}.$$  \hfill (5.1)

(5.2)

(5.3)

(5.4)

(5.5)

While the objective, $J$, is general, the most commonly employed engineering objective function (equation (5.1)) is the least-squares deviation of a variable of interest, $y(t)$, from the user-supplied target trajectory, $r(t)$. This minimization (or equivalently maximization if the objective is multiplied by $-1$) is performed over a time horizon from 0 (taken as the present time, by convention) to $t_f$, a user-specified final time. The optimization variable is $u(t)$, which can be independently specified over the entire horizon. The constraints are typically model descriptions, incorporated as ODEs (equations (5.2) and (5.3)), and the acceptable bounds on the optimization variable (equation (5.4)) are either explicitly determined from physical constraints (e.g. pump rates, non-negativity) or implicit limits resulting from physical phenomena (e.g. flow rate limits imposed by haemolysis). Finally, an endpoint for the variable(s) of interest is specified, as in equation (5.5).

As an example, consider a haemoadsorption device that would be used to remove cytokines from the blood. An optimal control problem might have the following specification. The target values ($r(t)$) could be the basal cytokine levels, and the model-predicted values would be cytokine levels from a mathematical model, which are compared with actual patient-specific values using point-of-care measurements (about an hour delayed from real-time for measurement processing). Manipulating blood flow through the haemoadsorption device would alter the rate of cytokine removal by the device, and, as such, the concentration of cytokines at the device exit (where blood is returned to the body). The lower bound on flowrate, the optimization variable, would be just greater than the flow at which coagulation occurs in the device; the upper bound is the point just before blood haemolyses in the device. Finally, a desired end-of-haemoadsorption-treatment cytokine level would be specified (again, perhaps the basal levels). The result is an optimal control problem for cytokine regulation via haemoadsorption in response to sepsis.

Solving this problem is equivalent to solving a two-point boundary value problem. Under a set of assumptions, analytical solutions are available. However, a programming solution is more common, and a commonly employed tool is control vector parameterization (Martin & Teo 1994). This involves a discretization of the time axis into $k$ equal-length segments, over which the optimization variable(s) (e.g. blood flow rate through the device) is held constant, and the levels of $u(k)$ are the decision variables in the resulting optimization problem. The exact method for solving this optimization is up to the user; gradient search is perhaps the most common. The input profile generated by the optimization often has a characteristic ‘bang-bang’ shape, where the input profile over the horizon of interest switches back and forth between the applied constraints, $u_{\text{min}}$ and $u_{\text{max}}$. This technique is good for solving nonlinear control problems with constraints, but there are inconsistencies with medical practice—primarily, there is no accommodation made for real-time feedback adjustment of the treatment profile in the presence of new measurement data (that are collected while the optimal control-generated therapy is being administered).

5.2. Receding horizon control

An alternative approach to model-based treatment intervention is to use receding horizon, or model predictive, control (Muske & Rawlings 1993; Allgöwer et al. 1999). This formulation poses and solves an open-loop optimization each time a measurement is collected, the result of which is a sequence of input values or changes to be implemented over a user-specified time horizon. Structurally, the problem is similar to the optimal control formulation above; a cytokine or
damage-associated objective is minimized (again, typically a sum-of-squared-error goal (5.6)) over a horizon, haemoadsorption flowrate (or other intervention) is held constant between measurements, models are used explicitly in the algorithm solution (5.7 and 5.8), and constraints on the inputs can be rigorously enforced (5.9 and 5.10). The departure from an optimal control formulation comes in two places: (i) the endpoint constraint is not included and (ii) the problem is reposed and the solution recalculated each time a measurement is collected. A typical model predictive control (MPC) problem may be formulated as follows (Muske & Rawlings 1993; Morari & Ricker 1994):

$$\min_{\Delta u(k|k)} \| \Gamma_y R(k + 1|k) - Y(k + 1|k) \|^2_2 + \| \Gamma_r \Delta u(k|k) \|^2_2,$$

subject to

$$x(k + 1|k) = f(x(k|k), u(k|k), p),$$
$$y(k|k) = h(x(k|k), u(k|k), p),$$
$$u_{\min} \leq u(k|k) \leq u_{\max}$$

and

$$| \Delta u(k|k) | \leq u_{\text{rate}}.$$

The mathematical formalism employs statistical notation: $Y(k + 1|k)$ is the vector of predicted measurable outputs (length $N_y$) at time $k + 1$ given information up to time $k$. The desired trajectories or values for the cytokines or damage surrogate(s) is $R(k + 1|k)$, and the optimization variables (the degrees of freedom for the optimization) are the input changes $\Delta u(k|k)$ (length $N_u$). The change in the input, rather than the input value, is commonly employed in the MPC formalism to help eliminate steady-state offset, a condition where $Y \neq R$ at stable operation. Trade-offs between the need to keep measured cytokines or other outputs near their desired values, versus minimizing the undesired effect of small variations in the measurement (due to measurement variability inherent to the measurement device or method) altering the intervention, are accomplished by altering the weighting matrices $\Gamma_y$ and $\Gamma_r$, respectively.

As in optimal control, an optimization routine is called in the solution of the MPC problem, and, as a result, input constraint incorporation is straightforward in both magnitude-constrained (equation (5.9)) and rate-constrained (equation (5.10)) forms. Output and state constraints may lead to infeasibilities in the optimization problem, and as a result these are often omitted or incorporated in a ‘soft’ form by replacing the hard constraint with corresponding penalty terms in the objective function (Zafiriou & Chiou 1993). A schematic of MPC implementation is shown in figure 6.

Revisiting the example from §5.1, at the present time $k$, the current cytokine levels are measured and compared with the model-predicted cytokine values. To address (expected) differences between the patient and the model, an additive correction ($d(k)$) is made to the model predictions over the $N_p$-length prediction horizon. The difference between the desired cytokine values ($R(k + 1|k)$) and the model-predicted ones ($Y(k + 1|k)$) over the $N_p$-length prediction horizon is used to calculate the $N_p$-length series of haemoadsorption flow rate changes that minimize the objective (typically driving $Y$ as close as possible to $R$). The first flow rate change is implemented, possibly requiring confirmation from a clinician, and the process repeats when the next measurement is collected.

In the case of time-variable, or temporally mismatched, measurements of different entities (e.g. cytokines and damage surrogates), a multi-rate MPC formulation is available (Lee et al. 1992; Gopinath et al. 1995). Here, the algorithm makes decisions using model-predicted values until new measurement information is available, with each measured quantity updated independently. While the MPC problem formally returns a solution that is suboptimal, in that it is a locally optimal solution resolved at each measurement time point, the algorithm has shown excellent performance and a respectable degree of robustness in the industrial application setting (Qin & Badgwell 1999). The structural advantages of MPC, including the on-line solution of an optimization problem that incorporates available patient information, are significant, but a key drawback of this control structure is the challenging analysis of algorithm performance. Algorithm stability guarantees, which may be required by the FDA before deployment in a clinical setting, often require the use of endpoint constraints or long prediction horizons (i.e. large $N_p$ or ‘infinite-horizon’ formulations (Rawlings et al. 1994)), which may limit the achievable performance.

5.3. Mixed-integer programming

To this point, the modelling and control theoretic tools have focused on ODEs and continuous-valued variables. However, some measurements (e.g. DNA arrays and other colorimetric methods) provide a less numerically precise estimate. Alternatively, a modeller may choose not to use real-valued variables (because of experimental uncertainty or atypical natural distributions), but rather to discretize the variables as ‘high’ or ‘low’. ODE-based models are poorly posed to handle this qualitative information because these lumped models are founded on the continuum assumption. This is the same reason that ODE models do not work well for small (i.e. countable) numbers of molecules—the continuum assumption loses validity in this regime.

In a similar manner, treatment or outcome variables are not necessarily continuous. Most common are the ‘yes/no’—the decision to treat or not to treat, response or no response, survival or death. These are examples of binary [0, 1] variables, yet highly relevant in the clinical setting. Oral drug administration is often explicitly or implicitly quantified, in the manner of fixed-dose pills (e.g. one aspirin or two?) or infused drugs mixed at specific concentrations or dose levels. Treatment decisions are also made on quantified observations, such as ‘high’, ‘moderate’ or ‘low’ levels of TNF. Rather than resorting to fuzzy logic, mixed-integer programming (MIP) is a candidate formal
mathematical structure for handling these discrete values. Here discrete quantities are assigned corresponding integer values, so increasing doses might be represented as levels 1, 2 and 3, for example, with no treatment assigned a value of zero. If all variables were discrete, an integer programming problem would result; this further reduces to a binary program if all variables—not just the decision variables—are integer-valued in the range [0, 1]. However, drug dose levels or other quantified interventions interact with continuous biosystems models, in general; an administered drug dose results in a PK profile that is continuous. This continuously valued information is the ‘mixed’ part of the MIP problem. Based on the user-selected objective, and the structure of the underlying system model, the nature of the MIP problem is specified: a linear objective and model yields a mixed-integer linear programming (MILP) problem; replacing the objective with a quadratic form, as in the MPC problem above (equation (5.6)), yields an MIQP (Q for quadratic); and the explicit inclusion of other nonlinear terms (e.g. Michaelis–Menten or Hill saturations, product inhibition) requires the solution of an MINLP (NL for nonlinear).

Consistent with clinical practice, and using the MPC problem in equations (5.6)–(5.10) as an example, it is possible to repose this problem to explicitly address quantified information using a MIQP or MINLP problem (Harrold & Parker in press), depending on the system model used, as follows:

$$\min_{u(k), T(k)} \sum_{w=1}^{N_p} (R(w + 1|w) - Y(w + 1|w))^2$$

$$+ \Gamma_m \sum_{q=1}^{N_q} u(q|q),$$

subject to

$$x(k + 1) = f(x(k), u(k), T(k)),$$

$$y(k) = h(x(k), u(k), T(k)),$$

$$T(k) \in \{ T \},$$

$$u_{\text{min}} \leq u(k|k) \leq u_{\text{max}},$$

$$|\Delta u(k|k)| \leq u_\Delta.$$

The receding-horizon control problem independent variables, $u(k|k)$, and the integer-valued decision variables, $T(k)$, are used to minimize the objective function (5.11). The additional equations represent discrete-time model states (5.12) and outputs (5.13), the quantized variable treatment definition (5.14), continuous input bounds (5.15) and rate of change constraints (5.16), as well as integer variable Big-M constraints (not explicitly shown). An advantage of this formulation is that new patient information can be incorporated as a measurement update and an improved solution computed. While such an approach...
appears daunting, experience with simulated cancer treatments indicates that clinically relevant problems (1894 equations having 865 continuous and 168 discrete variables) can be solved within a second on a desktop computer (Harrold & Parker 2009).

6. ISSUES IN MODELLING BIOLOGICAL PROCESSES

Constructing plausible descriptive or predictive models of biological processes faces issues that are commonly not encountered in physics/engineering models, or are generally less severe. Sources of uncertainty and variability are routinely encountered in biological data, the notion of the existence of biological laws is on much softer grounds, and the very concept of the existence of clear causal chains loses clarity in complex systems with numerous redundancies, feedback mechanisms and levels of regulatory control. Impressive advances in data acquisition and processing techniques have and will continue to help alleviate some of these obstacles, while others await further advances in information, mathematical and systems theory.

6.1. Variability, uncertainty, similarities and differences

Whether a biological process exists or how it actually happens are often significant sources of uncertainty. The concept is quite intuitive when applied to the existence of the exact wiring of specific signalling pathways, where elegant specific solutions typically requiring extensive experimental— theoretical collaborations have helped (Kuepfer et al. 2007). Some of the core processes involved in the pathophysiology of sepsis are clearly multi-scalar, however—hypotension causes organ dysfunction in a number of ways. Cellular energy failure is undoubtedly a major driver of cellular dysfunction, but inability to clear toxic metabolites or direct organ—organ crosstalk, in the form of mechanical or neural interactions between organs themselves, may also play major roles. The modelling implication of uncertain mechanisms and their relative importance, as broadly described above in the pathophysiology of acute inflammatory diseases and sepsis, is that the process of designing disease models should not be limited to the construction of a single model. Rather, the focus should be the inclusion and evaluation of competing hypotheses, as often done in weather forecasting (Gneiting & Raftery 2005), or some balanced representation of such hypotheses according to some prior likelihood (obtained from the existing empirical evidence or expert opinion) embodied in a consensus model. At the present time, the examination of competitive models has been mostly restricted to machine-learning approaches to competitive wiring of signalling pathways. Couched in dynamical systems language, these factors translate to uncertainty in model structure or rules, and, within a given structure, to uncertainty in parametrization. Compounding the problem for processes with unclear starting points, such as sepsis, uncertainty in the initial conditions and variation in experimental conditions must also be considered. This is particularly disconcerting when controlled experimental conditions result in wide inter-individual variability of measurements across cohorts of genetically identical mice (a common observation in preclinical cancer chemotherapy studies). These two factors may in fact be crucial to our lack of success in translating promising preclinical information into actionable therapies in humans (Salluh & Bozza 2008; Dyson & Singer 2009).

Measurement uncertainty, both in accuracy (reproducibility of data given identical circumstances) and bias (how close is the measurement to the true value), is typically of greatest severity in biological data. For example, several fluorescent antibody-based assays, such as enzyme immunosorbent-linked assays (ELISA) or multiplex bead-based capture techniques (e.g. LumineX) measure fluorescence units and assume either that the binding reaction between the fluorescent antibody and molecule of interest has reached equilibrium or that the degree of completion of the binding reaction is exactly known at the time of the readout. Transposition of the fluorescence value to a standard curve yields the desired measurement, which is typically an average between (often very different readouts of) duplicates or triplicates. In our experience, we often had to reconstruct such standard curves and adapt them to our specific experimental conditions. Reproducibility and standardizations of experimental protocol and measurements remain significant sources of variability that modellers must take into account in their model calibration efforts. Thus, interventions can range from reappraising data, as described above, to the construction of models that take into account such variability in the process of structure and parameter estimation in a statistical or probabilistic fashion (Bertrand et al. 2001; Zenker et al. 2007). In spite of such caveats, there is also remarkable regularity in the inflammatory response across individuals and across species. The responses to acute endotoxin administration in rodents and humans are quite similar in their timing and resolves conflicting reports inhibiting progress in translation. The variance observed in data collected from animal studies is, to some degree, a function of the measurement device and method. However, this does not capture the entirety of the spread observed. While this is one of

J. R. Soc. Interface (2010)
the driving factors behind the use of statistical methods in the design and analysis of animal and human trials, it also calls into question the use of ODEs in the modelling of population average data—or more accurately, the benefits and drawbacks of constructing such models. An advantage of this simplified model structure at the whole-organism level is that uncertainties need not be captured mechanistically at every level of detail (e.g. organ, cell, nucleus); instead, these can be lumped into parameters characterizing the observed data at the organism scale. The corresponding drawback is that the loss of mechanistic accuracy may inherently limit the utility of the average ODE model in the treatment design for an individual (Morari & Zafiriou 1989).

As discussed above, one method for handling this variability is in the context of population models (Beal & Sheiner 1992; Sheiner & Ludden 1992; Resource Facility for Population Kinetics 2008). Here, the variability is incorporated into an ODE model as parameter variability, with the model parameters belonging to distributions informed from (pre)clinical data. As the resolution is increased to organs, cells and intracellular dynamics, the ability to inform population-style models from data decreases because of the general lack of reliable in vivo measurements at the smallest scales. Furthermore, the continuum assumption may begin to break down as properties such as heterogeneity, well-mixedness and large numbers of molecules may be violated. At highly resolved scales, interactions are not deterministic—the probability of a meaningful interaction is dependent, using molecular reactions as example, on the reactant molecules, energy and perhaps enzyme availability. Although the mechanistic structure of the (intra)cellular process may still be unknown, this variability can be incorporated at the correct scale through the use of stochastic ODEs or other appropriately abstracted stochastic simulators (see §4.2, above, for more detail on the modelling methods).

6.3. Collaboration between experimentalists and theoreticians

It is interesting to note from discussions with academics and industrial PhDs the degree to which academic control theory research has not made the translation to the industrial sector. Rigorous tools that provide performance guarantees, such as robust control and control systems designed using Lyapunov stability analysis, are not widely employed in the chemical process industry. Likewise, nonlinear control is a powerful tool, but its use has also been limited to those applications clearly requiring its capabilities (i.e. where linear control tools are insufficient on performance or stability grounds) and also of a state dimension to admit solution. In contrast, the model predictive control paradigm has been highly successful in industry and enjoys wide acceptance. However, the studies of robust performance guarantees and the necessary mathematical tools again do not often translate beyond the academic realm. The reason for this translation failure is at least twofold: (i) a control system that requires exclusive expertise to maintain after deployment is judged ‘too expensive’ unless the industry has a very high margin (e.g. pharmaceuticals) and (ii) the solution to the control problem is both theoretically complex, making it hard for the non-expert to understand the need for such a system, and computationally complex, such that the completion of the calculations cannot be guaranteed in a real-time industrial setting. These obstacles extend beyond the chemical industry.

Engineers and mathematicians have been addressing biomedical problems for decades. In diabetes, the first models were published in the early 1960s (Bolie 1961), with the first truly successful translation of model to clinic being the ‘Minimal Model’ by Bergman and co-authors in 1981 (Bergman et al. 1981). Likewise, cancer models date back at least to the 1970s, and only in the 2000s have model-derived treatment decisions started to manifest (Fornier & Norton 2005). Interesting to note in both of the clinical successes cited here is that the primary author is a medical doctor. Both Dr Bergman and Dr Norton have worked over the years with mathematicians and engineers (Prof. C. Cobelli being the engineer and driving force in the diabetes modelling area; he has continued to work on this problem (Nucci & Cobelli 2000; Bellazzi et al. 2001; Cobelli et al. 2007; Pedersen et al. 2008)). In contrast, the plethora of papers published by engineers and mathematicians claiming to have ‘solved’ such problems as glucose control, HIV treatment, cancer therapy and anaesthesia delivery—among other examples—address only the ‘necessary’ part of the problem. In other words, a technological solution has been presented that, if provided suitable information, can successfully achieve a stated objective. The missing piece is the clinical and translational aspect, the ‘sufficient’ component of a necessary and sufficient condition, where a clinician would participate in posing the objective and constraints of the problem to make sure that a proposed solution is sufficient for deployment in a clinical setting. As the case study below will demonstrate, the collaboration of clinicians with engineers and mathematicians leads to mathematical models that are both efficient (in their use of states, parameters, etc.) and allow the evaluation of clinically relevant questions. Furthermore, the ability to use these mathematical models in the context of treatment design, or their control relevance (Parker & Doyle 2001), is a design criterion as well, as discussed above.

7. A CASE STUDY IN INFLAMMATION

7.1. Low-order approximants

Our initial attempt to model the acute inflammatory response comprised a very high level, three-variable model: a growing pathogen ($P$); a pathogen predator ($M$) that multiplies in the presence of pathogen and destroys pathogen; and a late inflammatory mediator ($L$) triggered by $M$ and which can itself grow ($Kumar$ et al. 2004). The system of ODEs below represents this process where the $k$'s are rate constants
and $\theta$ and $W$ are form factors for the hyperbolic tangent function

$$\frac{dP}{dt} = k_P P(1 - P) - k_{PM} MP,$$

(7.1)

$$\frac{dM}{dt} = (k_{MP} P + L)M(1 - M) - k_{PM} MP - M$$

(7.2)

and

$$\frac{dL}{dt} = k_{LM} \left(1 + \tanh \left(\frac{M - \theta}{W}\right) \right) - k_{LM} MP.$$

(7.3)

Analysis of this system of equations reveals the possibility of a number of intricate behaviours, ranging from elimination of $P$ and resolution of both $M$ and $L$ (health), to elimination of $P$, but persistence of $M$ and $L$ (aseptic sepsis), to persistence of all three variables at high levels (septic sepsis), to persistence of $P$, but low persistent levels of $M$ and $L$ (immunosupression), to cyclic behaviour. All behaviours result over different ranges of initial conditions (dose of $P$) and system parameters. Lessons from this low-order model include the possibility of identifying conditions for the existence of self-sustaining inflammation and the possibility of parameter and initial condition-dependent existence of tristability (Kumar et al. 2004). As a first extension of this model, the inflammatory response is in fact the result of a balance between effectors that amplify the inflammatory cascade and a parallel anti-inflammatory response effect by counter-regulatory mediators that prevents the pro-inflammatory cascade from getting out of hand, so that a minor local infection does not result in a massive mobilization of resources. Active counter-regulation is a widespread feature of acute biological response mechanisms, blood coagulation being another prominent example (Adams et al. 2007). In abstract form, the three-state model of acute inflammation is extended by the introduction of an anti-inflammatory mediator ($C$) and by modifying the meaning of $L$ to represent tissue damage ($D$) (and thus a marker of health/death)—these changes characterize a variety of four-variable models (Day et al. 2006a; Reynolds et al. 2006) (see also figure 1). These were also formally analysed mathematically and can give rise to a wide range of behaviours. Simulations explored the impact of modulating exogenous administration of $M$ or $C$ on the asymptotic level of $D$. In other words, manipulations of the immune response will change the trajectory and final state of the system (Day et al. 2006b; Reynolds et al. 2006).

Our first attempt at calibrating a model of acute inflammation was performed with a rather large (22-equation) model (Chow et al. 2005). This work introduced the hypothesis that acute inflammation was using the same mechanisms irrespective of the specific stressor causing inflammation. Because a single model with a unique set of parameters was used to describe the inflammatory response to endotoxin, trauma caused by a laparotomy and haemorrhagic shock, we believe that this model provides strong support for a universal wiring of the acute phase of the inflammatory response. All experiments in support of the model were conducted in mice. Serum TNF, IL-6, IL-10 and nitrates/nitrates were measured at four time points within a 24 h period following the insult. Because the experiments were conducted in mice, animals had to be sacrificed at each time point, and therefore no longitudinal time series were obtained from individual animals. Furthermore, only four of the 22 state variables were measured, resulting in a highly under-determined model. Contrary to a widespread misconception, even severely under-determined models such as this one cannot be forced to fit the data if the mechanisms embodied in the equations are biologically unsound (mathematically this would be interpreted as a dynamic structural mismatch between the data and the model formulation). Finally, we obtained a calibration of the model by convex optimization. Using a standard local/convex optimization routine such as Newton–Raphson or Levenberg–Marquardt will identify, from a starting guess of parameter values, a set of parameters that minimizes the error between the simulated model and observed data. Such algorithms will typically converge to a local minimum, which will yield different model parameterizations if different initial guesses are used. Indeed, as a result of the underspecified and underconstrained local optimization, it is typically possible to produce many candidate models that fit the empirical data acceptably well, while (possibly) possessing very different behaviours in unobserved states, or simply relying on different mechanistic drivers to produce the same results.

In the case presented, there is undoubtedly a large collection of parameter sets that also could have fitted the dataset (Chow et al. 2005). This limitation does not invalidate the conclusion of universality of the acute inflammatory response, but affects the potential application of this particular parameter set to predictions of subsequent experiments if those were to be substantially different. In other words, this naïve approach lacks robustness and carries a high risk of the model not translating to subsequent experiments.

We were not satisfied with the ratio of observed variables to actual model variables, and with the relative complexity of the models. Therefore, we scaled back to the minimum meaningful number of variables that could still describe the data; conceptually akin to the successful ‘Minimal’ modelling approach used by Bergman and Cobelli for diabetes (Bergman et al. 1981). The result was a generation of models that have between eight and 10 variables, including pro-inflammatory mediators TNF and IL-6, anti-inflammatory mediator IL-10, serum neutrophil counts and serum creatinine, a marker of organ (renal) dysfunction (Daun et al. 2008). Calibration was performed on doses of endotoxin of 3 mg kg$^{-1}$ and 12 mg kg$^{-1}$ administered to 250 g Sprague–Dawley rats, four per dose, with seven time points per animal over 24 h. A prediction was provided for an intermediate dose of 6 mg kg$^{-1}$, as shown in figure 7 (Daun et al. 2008).

One challenge in model calibration is the solution of the least-squares problem for parameter estimation. We have adopted a systematic approach to solving this inverse problem.

— First, constrain parameter space by constructing lucid bounds on parameter values. This is often possible from literature and (pre)clinical data/knowledge.
— Second, construct a set of heuristics that penalize parameter sets for producing model behaviours that are deemed unrealistic. For example, models that fail to ‘heal’ in finite time with negligible stimulation or display peaks in unobserved variables of unacceptable/unbelievable duration or magnitude lead to further parameter space constraints.

— Third, we proceeded to explore formal model parameter reduction techniques (Daun et al. 2008) (not to be confused with model reduction techniques, which generally focus on reduction of the state—not parameter—dimension (Moore 1981; Glover 1984; Chiang & Safonov 1992)). This process involves calculating the sensitivity of the simulation results to small changes in parameter values (Yue et al. 2006). The analysis is typically performed for each variable in the model at each time point where data are available, although we are also applying methods that perform global sensitivity (Carter et al. 2003; Bongartz et al. 2006). We then proceed to an iterative process of fixing parameters to which the model is insensitive and recalculating sensitivities. This process of parameter reduction is stopped when there are no more insensitive parameters. The ‘price to pay’ is then evaluated by recalibrating the model only using sensitive parameters. This analysis also yields another extremely important piece of information: the process of identifying sensitive parameters, on the one hand, and the difficulty of constraining them through literature searches, on the other, identify where critical knowledge gaps exist. It was clear from our modelling efforts that simulations were extremely sensitive to pathogen burden and rate of growth, and also to the specific distribution of immune cells. Experimentally, we had virtually no information on either of those domains of data. Consequently, the outlook for building truly predictive models of sepsis to explore pathophysiological hypotheses remains grim until longitudinal quantification of key drivers are better described experimentally or clinically.

— Finally, given animal variability and data uncertainty, there are legitimate reasons to carry a large number of parameter sets as representative for our animal population. Accordingly, our consensus model is actually an ensemble of models, obtained in one of the two ways: (i) multi-start convex optimization of the reduced model, where the ensemble may include several hundred model parameterizations (Waugh & Perry 2005), or (ii) a fully probabilistic approach where parameter space is explored using Markov chain Monte Carlo algorithms (e.g. Metropolis-Hastings (Bertrand et al. 2001; Zenker et al. 2007)) and parameter sets are given a probability density based on their ability to fit experimental data.

This approach, in our estimate, represents the most satisfactory solution to the inverse problem, and may also
be the most practical in terms of providing probabilistic predictions (Hancioglu 2007).

The observation that acute inflammation is compartmentalized is of high clinical significance (Bonjoukos et al. 1993). Compartmental ODEs are quite useful when modelling processes involving tissue localization or transport across membranes and when the number of elements belonging to each species (e.g. molecules of a specific protein) is large, but where spatial processes such as diffusion are not critical. We are currently working on a systematic process for developing such models and provide an example here. Using a reduced model that we developed for sepsis (Reynolds 2008) as the starting point, we created models with varying numbers of compartments, typically two or more tissue compartments communicating via a blood compartment. These models included pathogen, macrophages and neutrophils as cell types, a pro-inflammatory effector, an anti-inflammatory effector and tissue damage. Macrophages are confined to the tissue, while the rest of the components are diffusible. In particular, neutrophils diffuse into compartments following pro-inflammatory gradients (see fig. 20 and subsequent in Reynolds (2008)). Simulations are initiated by seeding one or more organs with pathogen and admit a minimum of three fixed points, as do simple models: restoration of health; high pathogen, high damage death; and no pathogen, high damage death. Typically, linked organs either heal together or die together. Simulations suggest that relative tissue volumes, relative diffusivity and resident macrophage populations are strong determinants of outcome. Compartmentalization of the inflammatory response may also solve apparent contradictions observed in the literature, including situations where animals dying of a severe infection mount an appropriate response to infection at the primary site, while inflammatory cells are diverted to peripheral tissues, and why interventions not primarily directed at the source may help improve outcomes (Alves-Filho et al. 2005).

7.2. A case study in therapeutics: blood purification for sepsis

The Cytosorb blood purification device consists of polystyrene di-vinyl benzene copolymer beads of 300–800 μm diameter in a biocompatible coating packed in a 10 g cartridge with a total surface area of 850 m² g⁻¹. When inserted in parallel with the animal’s circulation starting 20 h after cecal ligation and puncture (an inflammatory insult resulting in native bacteria invading the peritoneal cavity) for a duration of 4 h, we observed a significant improvement in short-term mortality (figure 8), mortality at one week (data not shown) and a remarkable clearance of TNF, IL-10 and IL-6, arguably key cytokines in the early pathophysiology of sepsis. Blood purification is a feasible intervention in humans, and optimizing the conditions under which a blood purification device should be used seems a desirable goal. One can also assume that those conditions, like the titrated care offered by an ICU physician, would vary dynamically and thus could be represented in the form of a sophisticated systems engineering problem, if a suitable model of sepsis, and of the control intervention, were to exist.

To characterize cytokine-specific ability of the bead material to adsorb a molecule, cytokine-rich serum was filtered through the column ex vivo (Dileo et al. 2009). Other parameters characterizing the intervention are either known (e.g. bead mass and density) or are intervention dependent (e.g. timing and duration of intervention, blood flow). The blood purification intervention model shown in figure 9 can be coupled to a compartmental model ensemble, for evaluating individual response to the treatment intervention. This last step is the current work in progress.

8. DISCUSSION

8.1. Key clinical needs and barriers to adoption of methodology

There has been considerable focus on methods to improve the difficult and costly process of validating the positive impact of new interventions on outcomes in phase III randomized clinical trials (Ammane 2009). The need for such an approach is particularly urgent in the field of acute inflammation secondary to trauma and sepsis (Yaffe & Fink 2000; Cobb et al. 2005). As systems scientists attempt to bridge the gaps between massive data streams offered by high-throughput methods and the existing analytical methods, and to develop a coherent multi-scale framework of interpretation, the need to prioritize data and computational methods that will help accomplish this goal is essential (Ye et al. 2005). A significant effort is therefore present to articulate concepts underpinning the discipline of systems medicine, as a translationally relevant extension of systems biology (Auffray et al. 2009).

A number of obstacles along this path remain. First, there is a perceived dichotomy between the traditionally inductive approach of the scientific method and the deductive approach of model-based hypothesis generation (Kitano 2002). In fact, there is no such dichotomy—they are complementary, in that model predictions, obtained from simulations, must still be subjected to experimental verification. Still, there is no consensus as to what constitutes a reasonable burden of proof that a prediction is actually verified, while there is the intuition that it may not always be
possible to reduce the problem of prediction validity, and thus model validity, to the $p$-value of a statistical test. Second, clinicians and scientists not trained in quantitative methods, which are ultimately responsible for completing the bench to bedside realization, are typically agnostic of many of the quantitative methods discussed above. This is unlikely to change, and thus efforts at promoting interdisciplinary exchange and collaborations should be encouraged. Furthermore, the tools developed, potentially through such collaborations, should be packaged to facilitate the practice of medicine and treatment—potentially a tall order for what is otherwise ‘research-grade’ code that requires an expert to execute. Third, there are no sufficiently detailed in silico models of human sepsis because of: (i) existing residual uncertainty in the causal chain of outcomes, thus yielding uncertainty in model structure and measurement accuracy and bias, (ii) large gaps in data at relevant modelling scales to calibrate such models, (iii) lack of availability of point-of-care testing of relevant analytes, and (iv) theoretical and algorithmic deficiencies in the existing methods of maximizing information derived from incomplete and disparate data. The support of organizations such as the International Society for Systems Biology (http://www.issb.org), the Society for Complexity in Acute Illness (http://www.scai-med.org), several national funding agencies and private foundations to articulate these difficulties, to promote communication across disciplines and importantly to maintain a focus on the translational relevance of this effort are commendable.

8.2. On translating animal results, which are not human, to humans

The study of specific biological mechanisms of sepsis, especially if cell type specific, is best done in vivo under controlled conditions. When the objective is to vary entire host conditions (e.g. knock-out genes) or interfere with a disease’s natural course (e.g. therapeutic attempts), then one has to resort to in vivo disease models. Animal models of sepsis are still too heterogeneous with regard to type of insult, duration and supportive therapy to be regarded as representative of the human condition. Using standardized animal models may eliminate some of the differences between animal and human studies, allowing a greater degree of translation (Dyson & Singer 2009). There have been attempts at standardizing animal models of sepsis and septic shock over the last two decades (Fink 1990). Many will argue that caecal ligation and puncture, where the colon is ligated at the hepatic flexure and the caecum perforated with a needle allowing bacterial efflux into the abdomen, represents a reasonable experimental construct of sepsis (Buras et al. 2005; Fink 2008; Nemzek et al. 2008). However, the advantages of such an animal disease model are balanced by the fact that human sepsis is obviously more heterogeneous, with several potential local sources of infection (pneumonia, renal system, abdominal, etc.). Although the basic wiring of the acute inflammatory response appears very similar across species, the relative contribution of the myriad processes may be very different. For example, the lung has a high resident macrophage population to aid in infection clearance and it is also a highly perfused organ. Therefore, the lung is easily inflamed when compared with other organs, whether it is the primary site of infection or not. One would then expect sepsis therapies to have differential success depending on the primary site of infection, which has been confirmed clinically (Bernard et al. 2001). Adding to the difficulty is that timing of therapy is critical in a rapidly evolving disease, also eloquently demonstrated in sepsis (Rivers et al. 2001; Bernard et al. 2004). Unfortunately, time zero of an infection is truly not known in a typical patient, in contrast to an animal experiment. Some effort has to be put into biomarker-based outcome stratification algorithms.

Figure 9. Clearance of (a) TNF, (b) IL-6, (c) IL-1B and (d) IL-10 by the Cytosorb blood purification device. Asterisk indicates $p < 0.05$ between control and treated groups.
or estimation routines that provide calculated projections of time and magnitude of initial infection. Such strategies are promising in animal series (Osuchowski et al. 2009), and also in human cohorts (Kellum et al. 2007; Rivers et al. 2007), in support of previous in silico work (Clment et al. 2004b).

Animal models that have demonstrated efficacy have typically been used to demonstrate a reduction of mortality in otherwise healthy animals, where the cause of death was obviously attributable to sepsis. In humans, however, the baseline mortality is lower and, given the demographics of the target population, sepsis is a contributing factor to mortality rather than the unique cause. Therefore, the extrapolation of animal studies to humans could well be improved if less lethal animal models were used, preferably in older animals. Genetic factors also play an important role in both the risk of sepsis and its outcome, and several polymorphisms have been associated with either susceptibility or outcome (Kellum 2003). Factors such as sex and race also influence susceptibility and severity of sepsis, plausibly through complex genetic factors (Dombrovskiy et al. 2007; Dyson & Singer 2009). Only recently have animal experiments focused on sex-specific markers and outcomes. While there is progress in understanding the underlying physiological correlates of age and sex, there is no clear parallel to the notion of race in animals. Within species, strains is a vastly more radical concept than that of race. Therefore, the translation from basic science in animals to clinical outcomes in humans at present remains a difficult proposition. Careful planning of future animal experiments, in concert with intensive efforts at developing mechanistic computational disease models that include known physiological correlates such as age and sex, may offer the best chance at correctly predicting outcomes and the impact of specific interventions based on biomarker profiles in humans.

8.3. Systems medicine: how systems approaches drive advances in clinical practice

The systems biology boom has resulted in the continuing generation of tremendous amounts of published data (a 10/2009 search for ‘systems biology’ returned more than 10 000 hits on www.pubmed.org), as well as systems-derived analysis tools for addressing this complex dataset (representative reviews include Doyle & Stelling (2006); Polpitiya et al. (2009)). At the heart of this field is the collaboration between basic science, in this case biology, and systems science, including engineering, mathematics and computer science. But what about translational science? The application of biological knowledge to improve the human condition is the practice of medicine; hence, systems medicine is the coupling of systems science with medical treatment decision-making (Auffray et al. 2009). This field is borne of the recognition that disease diagnosis and treatment is a systems response problem, incorporating multiple inputs, measurements and unobservable effects, and that simple one-pathway or one-effect methods are unlikely to succeed on a broad scale. The key advance observed over the past 10–15 years that opened the possibility of translating systems medicine to the bedside is the transition of mathematical and engineering approaches to clinical problems from ‘theoretically interesting’ (and generally devoid of concrete clinical practice knowledge) (Martin 1992; Parker et al. 1999; Zurakowski & Teel 2006) to practicable methods that explicitly incorporate clinical knowledge and constraints in diseases such as cancer (Lee & Zaider 2008; Harrold & Parker 2009), diabetes (Bondia et al. 2009; Du et al. 2009; Kumareswaran et al. 2009) and glucose control in critical care (Plank et al. 2006; Hovorka et al. 2008; Blaha et al. 2009; Cordingley et al. 2009).

From a systems engineering perspective, the field of systems medicine is rife with challenges. Model structures are neither fixed nor mechanistically characterized in many cases, so model selection and reconciliation with data from multiple studies (and collection methods) is needed. Measurements and analytical methods have inherent uncertainty in the reported values, meaning that parameter calibration and model structure selection have conflicting objectives of accuracy and robustness to small perturbations. An advantage to having a high-fidelity mathematical model, however, is the ability to use simulation as an analysis tool—examples include: parametric and model structure sensitivity analysis (Huang et al. 2001; Yue et al. 2006); evaluation of current ‘standard of clinical practice’ to the base model and any parametric or structural perturbations; a priori identifiability analysis and experimental design (Audoly et al. 1998, 2001; Bellu et al. 2007); and the development of ‘what if’ scenarios via Monte Carlo methods (for perturbations) or for the evaluation of novel intervention strategies. The development of model-based treatment decision-support systems relies heavily on the underlying model (Morari & Zalirioiu 1989; Parker & Doyle 2001), thereby asking for both predictive accuracy and robustness when the model is used to make decisions as part of a formal algorithm such as model predictive control. Unfortunately, performance and robustness are properties that must often be balanced (Skogestad & Postlethwaite 1996), so it is important to involve clinicians when tuning algorithm response to achieve clinically desired performance and robustness levels. To address the theoretical challenges, in the context of the variety of possible clinical case studies, is a challenge to the theoretical and clinical participants of the transdisciplinary collaborations advancing systems medicine.

Funding for this work was provided by the National Institutes of Health (USA)/NHLBI (grant no. R01-HL080926, J. Kellum, PI).

REFERENCES


Pfeiffer, K. et al. 1993 Mice deficient for the 55 kd tumor necrosis factor receptor are resistant to endotoxic shock, yet succumb to L. monocytogenes infection. Cell 73, 457–467.

Plank, J. et al. 2006 Multicentric, randomized, controlled trial to evaluate blood glucose control by the model predictive control algorithm versus routine glucose management protocols in intensive care unit patients. Diabetes Care 29, 271–276. (doi:10.2337/diacare.29.02.06.dc05-1689)


