Disease modelling, biomarkers and virtual physiology - the role of jcpex! and U-CEP

Extended abstract

Gabor Szekely (ETH Zurich) / Andreas Schuppert (RWTH) / Christian Stary (JKU) / Stefan Oppl (JKU) / Nils Meyer (Metasonic) / Rainer von Ammon (CITT)

This paper will be based on Chapter 5 of
VPH-FET: Future and Emerging Technologies for the Virtual Physiological Human
Support Action in FET Proactive - Reference: 258087
VPH-FET Research Roadmap
Advanced Technologies for the Future of Virtual Physiological Human
Editors: Marco Viceconti, Gordon Clapworthy

Editorial Team of chapter 5

Name	*Affiliation*
Pietro Lio	Computer Laboratory, University of Cambridge, UK
Carel Meskers	Leiden University Medical Center, The Netherlands
Andrea Remuzzi	Instituto di Ricerche Farmacologiche Mario Negri, Bergamo, Italy
Andreas Schuppert	Aachen Institute for Advanced Study in Computational Engineering Science, RWTH Aachen University, Germany
Gabor Szekely	Computer Vision Laboratory, ETH Zurich, Switzerland
Erwin de Vlugt	Biomechanical Engineering, Delft University of Technology, The Netherlands

The vision of this paper is to bring together a team of biologists and from medicine regarding the definition of biomarkers with specialists of process or event modelling methodology to implement such biomarkers and with IT specialists of Ubiquitous Complex Event Processing of accordant real-time processing platforms for a tremendous vast amount of events or signals per second. In the final paper we will combine and transfer the process modelling approach based on the example of the THESEUS project jcpex! [??, http://theseus-programm.de/en/936.php] with the U-CEP idea as discussed in this workshop series. A human body is actually a collaboration of around 50 or more trillion cells which do event processing and collaborating via the receptors of the cell membrane and the related effectors which "manage" or control the processes of a protein machinery. To model and control the event processing of specific biomarkers can be the basis to avoid and heal diseases. Especially interesting are the dynamic aspects of a not strongly fixed and in advance defined collaboration of processes. The processes can be combined dynamically as it is perhaps also typical in the fields of biology and medicine. This also includes the aspects of Uncertainty modeling as also mentioned in Ch. 5 of the VPH-FET roadmap. We discuss modeling approaches from UAI [??, http://www.auai.org/, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.24.2746&rep=rep1&type=pdf, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.113.5144&rep=rep1&type=pdf) as well as the suggestion of a reference model for non-deterministic U-CEP applications[??, see

workshop paper Danilov/Tyler/Ammon/Etzion] and special aspects of uncertainty in the disciplin of CEP [??, Etzion/Niblett 2010].

VPH-background according to Ch. 5 of the VPH-FET roadmap:

The Vision

During the past decade we witnessed a tremendous progress in creating powerful tools for the simulation and modelling of living systems. Systems biology offers today numerous approaches capable to quantitatively characterize cellular and molecular regulatory processes in unprecedented detail. Systematic efforts on the field of VPH research resulted in complex, often quite comprehensive mechanistic models of human physiology on the tissue and organ level.

Unfortunately the transition between a physiological and a pathological state is still poorly modelled in general terms, in spite of its vital importance for all clinical applications as well as drug research. This is mostly caused by the inherent complexity of disease development, which is a multi-factorial process which can only be understood by accounting for the interplay of all related aspects, like genetic and epigenetic factors, the physiological state of the effected organism and the influence of numerous environmental parameters.

The fundamental target of the envisioned research is the development of essential components of a generic framework enabling the comprehensive characterization of the set of diseases, the *diseaseome*, in its complexity and heterogeneity by allowing to integrate modelling and simulation approaches available today in all aspects mentioned above. We will restrict our vision for the VPH-FET roadmap on the impact of diseases on the human physiology and its consequences for clinical applications. This will allow to fully utilize the current state and future progress resulting from the ongoing investigations on molecular, genetic or pathogen-driven causes of the diseases emerging from numerous public and privately funded research programs, both on EU and national level.

The challenge

In general, diseases can be represented by variations of physiological parameters leading to a reduction of the functionality of organs or their mutual control, both accompanied with a significant negative impact on the functionality of the overall system, in extreme cases with fatal consequences. The reduction can have direct or indirect consequences:

- a decreased performance of an organ in normal life, such as chronic obstructive pulmonary disease (COPD) leading to reduced capacity of the lung for oxygen uptake resulting in significant reduction of quality of life.
- a significant reduction of the range of homeostasis, leading to a reduced stress tolerance of the system, such as allergies, but no impact on stress-free life. In the extreme case, some diseases based on genetic disorders, such as Gilberts disease, have no clinically relevant impact. However, they reduce physiologically relevant parameters, such as excretion or metabolization of drugs. Therefore they lead to toxic adverse drug effects for otherwise uncritical drugs or administration regimes. As these disorders do not reduce quality of life in the unstressed state, the classification as disease may be unclear, although neglect of the disorders may lead to unforeseen life threatening effects in stressed situations.
- increase of risk for fatal system failure without visible impact on systems performance (such as hypertension increases the risk of stroke, but does not directly affect the performance of the system)

Many well-known diseases, such as cancer, infections or neurodegenerative disorders like Alzheimer's disease lead to severe consequences for life, such that there is common sense to characterize these disorders as disease. Other diseases (like ADHS), however, lead to impacts on life depending on cultural conditions, others (like increased cholesterol levels) have unclear consequences on overall lifespan or quality of life.

Because of the tremendous complexity of the diseasome and its interaction with physiology, modelling offers a multitude of challenges.

a) Modelling the interface between –omics based disease models and mechanistic models of human physiology

- Genetic / proteomic / metabolic aberrations from "healthy" status affect physiological parameters resulting in clinically abnormal disease state. Disease modelling has to provide the quantitative interface between the physiological parameter set $\underline{p}$ describing the virtual physiology on a macroscopic level and the set of abnormal –omics parameters. E.g., a genetic mutation can result in an overexpression / suppression of protein levels or in reduced enzymatic activity of proteins which may be compensated by changed expressions of other proteins or result in a shift of a physiological parameter. If the parameter exceeds a threshold value, then a disease may develop on a clinical level. Hence the sensitivities of the physiological model parameters $\underline{p}$ on –omics parameters have to be quantified with good accuracy in the range of the thresholds.
- Models quantifying the impact of disease-related shift on the –omics level to the human physiology on the macroscopic level across cell – tissue – organ levels are intrinsically multi-scale models across a multitude of scales. Hence efficient multi-scale modelling approaches which are adapted to the special structures of scale interactions in biological systems are required.
- Special emphasis must be laid on the heterogeneous time scales, which are relevant in the dynamics of disease development and response to therapies. Processes across a multitude of time scales may be relevant for disease development and therapeutic effects, ranging from seconds and minutes on phosphorylation level to months and years for shifts of clinically relevant parameters in chronic diseases. Moreover, non-linear dynamics plays an eminent role resulting in dynamic features like phase locking, time lags, or hysteresis. Reliable disease models have to reflect these dynamic phenomena in an efficient mathematical formulation.
- An appropriate definition of homeostasis in terms of physiological model parameters will be a crucial challenge for disease modelling. As mechanistic models based on quantitative biomedical understanding of the underlying mechanisms will rarely be available, data analysis and parameter identification tools are required which allow an effective characterization of the homeostatic range for a physiological model allowing to separate sub-clinical from clinically relevant physiological states.
- Special attention should be lead on the impact of genetic variation of populations on disease development. Genetic variation can change the sensitivity of physiological parameters with respect to pathogens, toxins or drugs. They have significant impact on health care workflows. Hence a modelling framework is required which allows to simulate and analyse the effects of population heterogeneity on diseases using the established disease models. Most straightforward might be Monte Carlo sampling of –omics parameter combinations resulting in a distribution of disease models instead of a single model. The distribution of disease models could be used as an underlying prior for Bayesian modelling approaches.

b) Modelling the effects of co-morbidities resulting from joint genetic background of diseases

Human diseases result from abnormalities in an extremely complex system of molecular processes. In these processes, virtually no molecular entity acts in isolation and complexity is caused by the vast amount of dependencies between molecular and phenotypological features. In large scale meta-analysis (e.g. Schadt et al., Nature Genetics, doi:10.1038/ng1589) the mutual involvement of genes and diseases have been analysed on a genome-wide level. Apparently one-to one relations between genes and diseases showed to be exceptions, restricted to the Mendelian diseases. Complex diseases, however, like cancers, metabolic disorders, autoimmune diseases or psychotic disorders, cannot be associated to single genes only. Although associations between the genotype and physiology may neglect relevant biological mechanisms such as epigenetic control, transcriptomics, protein phosphorylation etc., the studies showed that complex diseases are in any case associated with large sets of genes. Moreover, the relations between diseases or groups of diseases and the respective gene sets are not one-to one. Most of the disease-related gene sets showed significant overlaps indicating relationships of complex diseases on the genome level. Using these overlaps, both on the level of the diseases and on the genome level, a relationship network can be established, where either two diseases are connected if the respective gene sets show a significant overlap, or two genes are connected if they are related to the same disease. However, the resulting networks showed a very high degree of connectivity such that it proved to be hard to extract clinically relevant conclusions from networks connecting only genotype and clinical phenotype. Moreover recent results show that GWAS studies linking complex diseases and genotype could not lead to results of therapeutic relevance [J. Couzin-Frankel, Science, DOI: 10.1126/science.328.5983.1220]. Hence, more advanced multi-level network approaches linking genomics, proteomics metabolomics with clinical phenotypes may be required.

In particular, it has been stressed that the combination of genomic, proteomic, metabolomic and environmental factors may provide insights into pathogenomic mechanisms and lead to novel therapeutic targets. For example, psychiatric disorders seem to lend themselves in particular for a systems based analysis approach. It is well known that schizophrenia has a strong genetic compound with concordance rates in monozygotic twins reaching approximately 50 per cent. This increased risk is conferred by a multitude of different genes with the most important genetic polymorphisms accounting for only one per cent of increased risk. It seems likely that the disease is ultimately precipitated by a complex interplay of genetic predisposition and a broad spectrum of environmental as well as nutritional factors. In this context, epidemiological factors such as urbanicity, geographical distribution and migration behaviour but also maternal risk factors such as infections, malnutrition and adverse life events during pregnancy have been suggested to be associated with the risk of schizophrenia onset. The relationship between these factors and the interplay with genetic determinants remains unknown and integrated, system based investigations seem to be a promising approach to obtain deeper insights into the disease aetiology. The network-based description of complex diseases can be structured in the form of network layers. The top layer is formed by human beings and the connections between them, such as family relationships. The second layer is generated by human diseases. Noteworthy, the diseases are linked to each other as many illnesses have related pathologies or even causal relationships. An intricate example is the increased prevalence for diabetes in schizophrenia patients (approximately 15 per cent), which

seems to be directed as diabetes patients have not been reported to have an increased prevalence of schizophrenia. Known links between diabetes, obesity and asthma have been brought into context with the representation of disease commonalities in networks. Interestingly, an increased risk of obesity can also be found in schizophrenia patients and the resulting cardiovascular diseases mainly account for the higher mortality rate. The third and bottom layer is formed by molecular systems and the network represents interactions between these molecules such as involvement in common pathways, protein-protein interaction or co-expression. The different layers of networks are highly dependent on each other. The network of patients and the network of molecular interactions are closely linked depending on which molecular abnormality is present in which patient; this information can be represented in a directed graph. As the structure of the graph depends on the disease under investigation, the properties of the graph may give important clues about the aetiology of the disease. This network concept has been applied in the context of human cancers and showed that networks can be used to infer biologically relevant dependencies between risk conferring genes.

The structure of the links in the network layers can give insights into disease complexity. On the molecular level, disease related abnormalities could either arise from single biochemical systems as reflected in alterations of only few molecules. In multi-genic disorders, this is normally not the case and the high complexity is reflected in abnormalities in a large number of genes, molecules and the accompanying clinical symptoms. In psychiatric diseases, the clinical features of different patients are rarely very similar; this lack of homogeneity is directly reflected in the patchiness of the graph linking patients and disease related abnormalities. In schizophrenia, for example, there is a broad spectrum of clinical manifestations, which resulted in the hypothesis of the existence of diverse underlying etiologies. It is essential to be aware of this degree of patient heterogeneity to be able to determine disease intrinsic molecular abnormalities. The known relationships between different diseases are ultimately reflected in similarities of the underlying molecular pathologies. Therefore, the network of disease relationships is intrinsically linked with the network of molecular functions. The disease network can then be seen as a collection of disease specific clusters that interact with each other depending on how related the biochemical underpinnings of the different diseases are. Information from different diseases and commonalities as well as differences of the molecular information may thus give important leads about pathological mechanisms, diagnostic applications and novel drug targets.

c) Modelling of disease development on physiological level

Modelling of disease development requires to establish quantitative dynamic models for the shift of the physiological parameter set $\underline{p}$ as a function of time: $\underline{p} = \underline{p}(t)$. This function can either be modelled explicitly or as a function of the root cause of the disease. E.g., the root cause of the disease may be a tumour, such that the growth of the tumour size s affects a physiological parameter p of an organ. Then the time course of p can either be identified as a function $p=p(t)$ or as a function $p=p^*(s(t))$. In the latter case the time course of the physiological parameter is modelled via nesting of the impact of tumour size s on the physiologically relevant parameter p and the time course of tumour size s. This might be beneficial if both function can be established independently, e.g. if reliable growth models of the tumour exist and if the quantitative effect of tumour sizes on the physiology parameters p is established independently by the function $p^*(s)$. In practice, even single parameters p may depend on multiple tumour parameters such that the identification of p^* may require estimators for

multivariate nonlinear functions. In case of high complexity of p* it might be beneficial to represent p directly in terms of an empirical function p(t) which has to be learned from data.

A special challenge for modelling of disease propagation will be the optional discontinuity of the function p(t). Even if the root cause of the disease on the physiological level, e.g. the tumour size or the viral load in a viral infection, are smooth and monotonic functions in time, the effect on the physiology, expressed by the parameter p, may become discontinuous if the function p*(s) shows discontinuities e.g. representing organ failure due to tumours exceeding a critical size. Because of the high redundancy of biological functionality and its intrinsic robustness against failure, a reliable characterization of these critical parameters may require very exact models on a detailed level. Detailed studies on networks have shown that random deletion of up to 70% of the nodes in a network does not necessarily affect the overall functionality of the network [e.g. Stelling et al., *Nature* **420**, 190-193 (14 November 2002) | doi:10.1038/nature01166]. However, targeted deletion of a very few "hub" nodes can result in fatal network failure. However, as detailed modelling can result in very high costs for data acquisition and validation of the models, efficient shortcuts using reduced models have to be developed which allow reliable estimates for critical parameter bounds which are not too far from the true vales.

d) Model based identification of Biomarkers for early identification of disease

A great challenge for modelling of disease propagation is the identification of biomarkers for detection of early phases of disease development. Identification of early disease states is crucial for any preventive approach in health care and could have a tremendous impact on health care systems. However, whereas identification of fully developed diseases can be performed by classification of well separated physiological or –omics parameter sets, a clear separation of the parameters between healthy and early disease states cannot be expected. Hence early identification requires a reliable characterization of the bounds of the physiological parameter range, which is associated with the "healthy" state. However, because of the complexity of the "critical surface" in the multidimensional physiological parameter space, a reliable characterization may not be feasible for complex diseases. Alternatively the identification of specific signals, which indicate the crossing of critical parameters, may be sufficient as well. As the identification of signals for criticality does not require the a priori characterization of a critical surface in the physiological parameter space, the identification of generic signals for criticality appears to be a probably feasible workaround. This approach may be based on the established models for phase transitions in statistical thermodynamics, where generic signals for criticality, based on specific features of the systems noise, can be identified.

e) Quantitative modelling of the dynamics of disease development and its interplay with physiological response and therapy

Many diseases, like infections, are characterized by external pathogens which are the root cause of the disease. The pathogen will become clinically relevant if the intrinsic physiological control systems, e.g. the immune system, cannot compensate the effects caused by the pathogen. Hence a special challenge for modelling this type of diseases will be a quantitative, reliable modelling of the dynamics of pathogen growth, optional change of pathogen population under immune / drug stress and the resulting dynamics of the immune response. Paying attention to autoimmune diseases could offer an additional pathway towards better understanding of such transition from a disorder to

a disease and identifying potential failure mechanisms in the underlying regulatory dynamics.

f) Modelling of diseases based on loss of control

These diseases are characterized by either genetic or epigenetic changes on the –omics level which result in a critical shift of physiological parameters. Due to positive feedback loops, which aim to rapid compensation of unexpected shifts on physiological level, the disease-induced shifts of the parameters on the physiological level may result in a continuous, accelerated increase of the underlying changes on the –omics level. Then, normally beneficial positive feedback loops lead to lack of control if specific disorders on the –omics level occur. Examples for relevant disease types of this class are:

- Inflammations, characterized by overshooting of immune response on pathogens
- Allergies, characterized by response of the immune system on non-pathogens
- Autoimmune diseases, characterized by a loss recognition of own cells by the immune system
- Malignancies, characterized by loss of proliferation control of cells, accompanied by efficient mimicry of the malignant cells to hide from immune attack.

Challenges for modelling of these diseases will be a reliable quantification of the respective feedback loops, which are involved in the regulatory system which is out of control.

The VPH motivation

The major motivation for this type of research from a VPH perspective is to avoid the treatment of physiology and pathology models as two separate domains. While this might be acceptable in the classic reductionist approach, the VPH integrative approach imposes to overcome this partitioning. At the same time we expect major improvements in understanding the dynamic interplay between healthy physiology, pathogens and therapeutic actions. Developing tools capable to account for both the discrete and continuous nature of the transition between the healthy and diseased state will open fundamentally new understanding of disease formation and, ultimately, will allow to revise the current, rather phenomenological ontology of the diseasome and derive some general principles from this more systematic and principled view.

Impact on Biomedicine

The fundamental motivation for the envisioned framework is biomedical: defining the transition to the diseased state means *diagnosis*; the dynamics of this transition and the phenotype transformations associated to it means *prognosis*; the external interaction with the system in the attempt to bring it back to the un-diseased state, or to a diseased state whose associated phenotype is less disabling for the patient, is *treatment*. Diagnosis, prognosis, treatment: by modelling the transition to the diseased state we explore the foundational aspects of the biomedical reasoning, and we open up to an infinite number of technological applications in virtually all branches of biomedicine.

General Impact

The human body is the most complex autonomous system ever created, making it the ultimate challenge for every modelling and simulation effort. Its functioning principles provide adaptability to properly react to external threats, like pathogens, which has always been an inexhaustible source for inspiration in science and engineering. A lot of what will be learned during the envisioned research can be used to recognise the insurgence, predict the effects, and plan the recovery of dysfunctional conditions in complex systems. Such knowledge can be applied for example in developing realistic models for macro-economies, to ecosystems, to transportation, to information management, just to name a few. If we can develop modelling strategies that can represent the dynamics of dysfunctional conditions in the human body, there are very good chances that these methods can be successfully applied also to less complex dynamic systems.

References:

…