"4D Biology for health and disease" workshop report

Abrahams, J.P.; Apweiler, R.; Balling, R.; Bertero, M.G.; Bujnicki, J.M.; Chayen, N.E.; ... ; Taussig, M.J.

ELSEVIER

Meeting Report

"4D Biology for health and disease" workshop report☆

Jan-Pieter Abrahams[1], Rolf Apweiler[2], Rudi Balling[3], Michela G. Bertero[4], Janusz M. Bujnicki[5], Naomi E. Chayen[6], Patrick Chène[7], Gary L. Corthals[8], Tomasz Dyląg[9,*], Friedrich Förster[10], Albert J.R. Heck[11,12], Peter J.F. Henderson[13], Ralf Herwig[14], Philippe Jehenson[9], Sasa Jenko Kokalj[9], Ernest Laue[15], Pierre Legrain[16], Lennart Martens[17,18], Cristiano Migliorini[19], Andrea Musacchio[20], Marjetka Podobnik[21], Gebhard F.X. Schertler[22], Gideon Schreiber[23], Titia K. Sixma[24], August B. Smit[25], David Stuart[26], Dmitri I. Svergun[27] and Michael J. Taussig[28]

The ''4D Biology Workshop for Health and Disease'', held on 16–17th of March 2010 in Brussels, aimed at finding the best organising principles for large-scale proteomics, interactomics and structural genomics/biology initiatives, and setting the vision for future high-throughput research and large-scale data gathering in biological and medical science. Major conclusions of the workshop include the following. (i) Development of new technologies and approaches to data analysis is crucial. Biophysical methods should be developed that span a broad range of time/spatial resolution and characterise structures and kinetics of interactions. Mathematics, physics, computational and engineering tools need to be used more in biology and new tools need to be developed. (ii) Database efforts need to focus on improved definitions of ontologies and standards so that system-scale data and associated metadata can be understood and shared efficiently. (iii) Research infrastructures should play a key role in fostering multidisciplinary research, maximising knowledge exchange between disciplines and facilitating access to diverse technologies. (iv) Understanding disease on a molecular level is crucial. System approaches may represent a new paradigm in the search for biomarkers and new targets in human disease. (v) Appropriate education and training should be provided to help efficient exchange of knowledge between theoreticians, experimental biologists and clinicians. These conclusions provide a strong basis for creating major possibilities in advancing research and clinical applications towards personalised medicine.

Europe is facing multiple challenges in the health sector, ranging from problems associated with an ageing population to the shortage of new, more efficacious and cost-effective treatments. Healthcare today is expensive and largely based on the treatment rather than the prevention of disease. Drug development is focused largely on a "one size fits all" approach, where sometimes only a minority of the patients treated benefit, from for example certain cancer treatments. Health research, therefore, has the ambitious task of improving our understanding of disease mechanisms and of translating them efficiently into clinical prognosis/diagnosis and prevention/treatment.

To approach these problems, advanced -omics technologies (e.g. genomics, proteomics, structural biology, interactomics and metabolomics) are increasingly being utilised with the aim of studying properties on a suitable scale to obtain a global, integrated view of cellular and organismal processes. The dramatic reduction in cost of these high-throughput studies has the potential to provide a basis for much more directed, personalised and predictive approaches to medicine than is currently possible, for example using a combination of high-throughput sequencing and more traditional diagnostic procedures.

To apply these developing technologies successfully, we need to be able to turn the exponentially growing amount of generated biological data more effectively into new, readily usable knowledge. Therefore, the European Commission Directorate General for Research organised the "4D Biology workshop for Health and Disease" that was held in Brussels, 16–17th March, 2010. The workshop aimed at finding the

☆ The views expressed are purely those of the authors and may not in any circumstances be regarded as stating an official position of the European Commission.

*Corresponding author: Dyląg. T (Tomasz.DYLAG@ec.europa.eu)

best organising principles for large-scale proteomics, interactomics and structural genomics/biology initiatives and setting the vision up to 2020 for high-throughput research and large-scale data gathering in the biological and medical sciences.

Major conclusions of the workshop include the following points.

• Development of new technologies and approaches to the analysis of the data is crucial, being the origin of, and the driving force for, the majority of scientific advances in biomedical research and medical practice.

(i) There is a clear need to use modern mathematics, physics, computational and engineering tools more effectively in biology, for example statistics, non-linear dynamics, information theory, systems theory and systems control, and new tools should be developed.

(ii) Biophysical methods should be developed that span a broad range of time/spatial resolution and characterise structures and kinetics of interactions. Many examples were discussed during the workshop involving the use of X-ray free electron lasers, nano-crystallography, NMR, (molecular) imaging (X-ray, electron microscopy/tomography, fluorescence microscopy, optical projection tomography, nanoscopy, etc.), small angle X-ray scattering, binder-based assays and microfluidics. Methods for studying single molecules *in vivo* are particularly needed.

(iii) Correlative approaches that allow the study of a specimen in different ways, such as correlative fluorescence and electron microscopies, or mass spectrometry and cryo-electron microscopy, are needed to "reconcile" different technologies and permit simultaneous, multiparametric analyses. Correlative fluorescence and electron microscopy could, for example, fill the existing resolution gap between light-microscopy-based imaging and electron microscopy. It is desirable that analogous correlative approaches be developed, for example between structural and cellular biology, proteomics and cellular biology, or proteomics and structural biology.

(iv) To understand complex systems, quantitative and dynamic *in vivo* data are required that could answer numerous questions: what are the number and location of specific molecules per cell at given time points? How fast does a particular reaction occur? How strong is an interaction? *In situ* and *in vitro* methods need to be further improved and used together to investigate macromolecules and their interactions.

• New technologies and high-throughput studies have produced an enormous amount of system-scale datasets for many diverse types of cellular components in various organisms, defining their identity, interactions and functional states. The immediate challenges towards making the data usable and on the way to full understanding include data validation, integration, storage and sharing in a biologically meaningful way.

Although there is an obvious strength in data diversity, database efforts need to focus on improved definitions of ontologies and standards so that data, and most importantly their associated metadata, can be understood and shared. Priority should be given to the establishment of data quality control to (i) share and exchange data, (ii) compare different types of data, (iii) facilitate non-expert users to make sense of the data, (iv) avoid ambiguities and (v) facilitate the access to validated reagents (e.g. antibodies).

To address the current challenges in data gathering and handling, the terms 'hard' and 'soft' data were used. Nucleotide sequences and protein structures can be defined as 'hard data', relatively easy to validate, store and represent in a unique manner. The gathering, storage and sharing of 'soft data', such as protein–protein interactions, spatial and temporal changes in protein concentrations or the kinetics of interaction, are particularly challenging because of the variation in the information content of data and metadata and because of their potentially variable quality or reproducibility.

The challenge ahead is validated (soft and hard) data gathering, integration, storage and sharing in a biological meaningful way. The worldwide Protein Data Bank (wwPDB) can be seen as an exemplary resource for central hard data storage, where all experimentally determined structures of proteins, nucleic acids, and complex assemblies are deposited and can be accessed.

The abundance of information presents many hurdles to the investigators who need to interpret the data and derive new biological insights. Therefore, efforts need to focus on data integration to obtain comprehensive models of biological systems, to understand the overall behaviour of systems in response to various environmental stimuli and to gain evolutionary insights and translate new applications into clinical practice. Because biologists need to benefit from these datasets without being overwhelmed, data should be accessible for interpretation in a user-friendly manner, for example through differently targeted portals integrating knowledge. One relevant field would be the integration of all the available information to understand human variability, for example the functional and structural relevance of single nucleotide polymorphisms (SNPs). Further examples are self-assembling systems (e.g. kinetochores and centrioles) or self-organising systems (e.g. mitotic spindle). The creation of 4D atlases or a 'Google cell' portal may be viewed as ways of achieving such goals.

• Research infrastructures should play a key role in fostering and strengthening multidisciplinary research, maximising knowledge exchange between different disciplines, facilitating access to diverse technologies, and being the centres for forging new alliances and collaborations. Researchers could therefore address scientific questions to such multidisciplinary technology centres, receive the right expertise and training, and access expensive, state-of-the-art technologies. As these cutting-edge infrastructures will become reference technological platforms for biological data production, they should take responsibility for defining the required standards and curate, disseminate and preserve the data through long-term repositories.

• Understanding disease on a molecular level is crucial in the search for biomarkers and new druggable targets in human cellular pathways, the areas in which industry is currently facing a shortage of inspiration and resources. Nowadays, high-throughput gene, protein and metabolite measurements (-omics) can dramatically accelerate the hypothesis generation and testing in disease models, and the emerging field of chemical genomics has the potential to bridge the gap between genomics/proteomics and therapeutics. Computer simulations integrating knowledge of molecules, organs and tissue responses will help to prioritise targets, predict the effects of combinatorial therapy and facilitate the design of clinical trials. Therefore, systems approaches promise to improve decision making in pharmaceutical development and may represent a new paradigm in the search for biomarkers and new targets in human disease. Because of the high costs and pressure to deliver new products, pharmaceutical companies are frequently not venturing into innovative and risky drug discovery efforts. For this reason the early stages of drug development will increasingly be a close collaboration between academia, industry and hospitals. To bring the latest advances in cellular and molecular biology to the clinic, translational research needs to be promoted and high-throughput research needs to be conducted on patient samples. Animal models will still be useful for several specific applications, but more emphasis should be given to human derived model systems.

• Modern biology is interdisciplinary. Bridging of previously distant fields such as physics and

cellular biology has led to unprecedented insights into cellular functions. However, theoreticians, experimental biologists and clinicians are often speaking different languages and the flow of information is hampered by different types of background and expertise. Appropriate education and training should be provided to bridge the gap in mutual understanding and help efficient knowledge exchange. Biomedical researchers should receive integrated training in several disciplines, including mathematics, physics, engineering and computer science. For this, training opportunities should be created and interdisciplinary programmes should be promoted at every level of education and career stage.

• The most efficient way to support translational and multidisciplinary research in Europe is by an appropriate mix of large and medium size projects which are focused on specific biological and/or methodological questions. Scientific creativity should be fostered and should be paramount in smaller high risk and exploratory projects. An important conclusion is that one should aim at understanding the complexity and dynamics of cellular processes at the molecular level, as this promotes the discovery and clinical testing of new diagnostic and therapeutic procedures and helps determining individuals' predisposition to particular diseases or conditions. The -omics technologies are producing invaluable amounts of biological data that need to be validated, stored and organised in a comprehensive and user-friendly manner. Integration of information at different scales is crucial. Through a systems approach, datasets become directly relevant for further biological and medical investigations, can be checked for self-consistency, and are constrained to have uniform and relevant data gathering protocols. Therefore, collecting data in ways that are consistent with further system analysis provides a strong organising principle for choosing what data to collect and how to organise and store it for analysis. High quality data that are well curated will be important at many different levels of biological and medical research. The ability to integrate and explore from the atomic to systems level is the challenge of the future.

Taken together, these conclusions provide a strong basis for creating major possibilities in advancing research and clinical applications towards personalised medicine, by establishing new technologies for pipelining, gathering, storage and analysis of vast quantities of experimental data, into the development of new applications for the clinic.

***Jan-Pieter Abrahams*[1]**
***Rolf Apweiler*[2]**
***Rudi Balling*[3]**
***Michela G. Bertero*[4]**
***Janusz M. Bujnicki*[5]**
***Naomi E. Chayen*[6]**
***Patrick Chène*[7]**
***Gary L. Corthals*[8]**
***Tomasz Dyląg*[9]**
***Friedrich Förster*[10]**
***Albert J.R. Heck*[11,12]**
***Peter J.F. Henderson*[13]**
***Ralf Herwig*[14]**
***Philippe Jehenson*[9]**
***Sasa Jenko Kokalj*[9]**
***Ernest Laue*[15]**
***Pierre Legrain*[16]**
***Lennart Martens*[17,18]**
***Cristiano Migliorini*[19]**
***Andrea Musacchio*[20]**
***Marjetka Podobnik*[21]**
***Gebhard F.X. Schertler*[22]**
***Gideon Schreiber*[23]**
***Titia K. Sixma*[24]**
***August B. Smit*[25]**
***David Stuart*[26]**
***Dmitri I. Svergun*[27]**
***Michael J. Taussig*[28]**

[1]Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands
[2]European Bioinformatics Institute, Hinxton, Cambridge, UK
[3]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Luxembourg
[4]Centre for Genomic Regulation, Universitat Pompeu Fabra, Barcelona, Spain
[5]International Institute of Molecular and Cell Biology, Warsaw, Adam Mickiewicz University, Poznan, Poland
[6]Imperial College, London, UK
[7]Novartis Oncology Research, Basel, Switzerland
[8]Centre for Biotechnology, University of Turku, Åbo Akademi University, Turku, Finland
[9]Research Directorate-General, European Commission, Brussels, Belgium
[10]Max-Planck Institute of Biochemistry, Martinsried, Germany
[11]Bijvoet Center and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, The Netherlands
[12]Netherlands Proteomics Centre, The Netherlands
[13]Institute for Membrane and Systems Biology, University of Leeds, Leeds, UK
[14]Max Planck Institute for Molecular Genetics, Berlin, Germany
[15]Department of Biochemistry, University of Cambridge, UK
[16]Commissariat à l'Énergie Atomique et aux Energies Alternatives, Paris, France
[17]Department of Medical Protein Research, VIB, Ghent University, B-9000 Ghent, Belgium
[18]Department of Biochemistry, Ghent University, B-9000 Ghent, Belgium
[19]F Hoffmann-La Roche, Basel, Switzerland
[20]European Institute of Oncology, Milan, Italy
[21]National Institute of Chemistry Slovenia, Ljubljana, Slovenia
[22]Paul Scherrer Institut, Villigen, Switzerland
[23]Weizmann Institute of Science, Rehovot, Israel
[24]Netherlands Cancer Institute, Amsterdam, The Netherlands
[25]VU (Vrije Universiteit) University, Amsterdam, The Netherlands
[26]University of Oxford, UK
[27]European Molecular Biology Laboratory, Hamburg, Germany
[28]Babraham Bioscience Technologies, Cambridge, UK