EMBnet.news

Volume 15 Nr. 2
June 2009

- European Strategic Forum on Research Infrastructures (ESFRI) projects
- Course reports
- Constructing and working with protein interaction graphs

and more ...

Editorial

This issue carries the standard sections, with plenty to read as usual. We have asked for contributions on the specific area of ESFRI projects. Research infrastructures are important to each and every member of our community. At this time ESFRI is presenting final proposals to the European Commission and we decided that it would be the right time to ask the coordinators of six of the ongoing projects to let us know some details on what is being prepared. The six letters to the editor are the responses that we obtained. In one way or the other, Bioinformatics is in all of them.

The Editorial Board of EMBnet.news wishes to thank all the authors but a special reference must be made to the authors of these letters to the editor, that accepted to write them in such a short period of time and yet have brought us a considerable wealth of information on those projects.

EMBnet.news is here again making available reports and reviews, announcing events, etc. Please do not forget that, as a member of this community, regardless of the nature of your connection to EMBnet and even if you do not have one, you are more than welcome to contribute to this publication.

The editorial board: Erik Bongcam-Rudloff, Domenica D'Elia, Pedro Fernandes, Andreas Gisel and Lubos Klucar.

Protein Spotlight (ISSN 1424-4721) is a periodical electronic review from the SWISS-PROT group of the Swiss Institute of Bioinformatics (SIB). It is published on a monthly basis and consists of articles focused on particular proteins of interest. Each issue is available, free of charge, in HTML or PDF format at
http://www.expasy.org/spotlight

We provide the EMBnet community with a printed version of issue 105. Please let us know if you like this inclusion.

Cover picture: *Taraxacum vulgare* (Dandelion), 2009 [© Erik Bongcam-Rudloff]

Contents

Editorial Board:

BBMRI

The Pan-European research infrastructure for Biobanking and Biomolecular Resources: managing resources for the future of biomedical research





Heli Salminen-Mankonen[1], Jan-Eric Litton[2], Erik Bongcam-Rudloff[3], Kurt Zatloukal[4] and Eero Vuorio[1]

[1] University of Turku, Centre for Biotechnology, Turku, Finland
[2] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
[3] The Linnaeus Centre for Bioinformatics, Swedish University of Agricultural Sciences and Uppsala University, Sweden
[4] Institute of Pathology Medical University of Graz, Graz, Austria

Public website: www.bbmri.eu
BBMRI Coordinator: kurt.zatloukal@meduni-graz.at
BBMRI Executive Manager: eero.vuorio@utu.fi

Summary

Biobanks are a key resource for unravelling the molecular basis of disease subtypes, identification of new targets for therapy and reduction of attrition in drug discovery and development. The broad spectrum of existing biobanks is considered as a specific strength of European research. Unfortunately the diversity - lack of standardisation - of these biobanks and the differential ethical and legal landscape across Europe have prevented their effective use. Development of common IT infrastructure and sustainable funding schemes are key features for large transnational projects interlinking different national and regional biobanks. Agreement on common standards is equally important for all *de novo* biobanks. In 2008, a pan-European infrastructure BBMRI (Biobanking and Biomolecular Resources Research Infrastructure) was established to bring cohesion to the European biobanking community and to make the existing and new high quality biological resources available for health research in Europe (Fig. 1).

Introduction

The sequencing of the human genome, completed in the 21st century, allows researchers to integrate new data on genetic risk factors with demographic and lifestyle data collected via traditional cohort studies or via modern communication technologies. The technical prerequisites now exist in Europe for merging large volumes of molecular genetic data obtained by using new high throughput DNA analysis platforms with clinical, epidemiological and national health registry data. One of the widely recognised strengths of European research is our long-term tradition to conduct large-scale clinical and epidemiological studies and to collect biological samples from the study subjects for analyses and storage. Combined with data on nutrition, life style and environmental exposure, such longitudinal cohorts have already demonstrated their strengths [1-2]. Therefore it is not surprising that human-derived biological sample collections and related biomolecular resources were identified as one of the most promising research infrastructures in the first ESFRI (European Strategy Forum on Research Infrastructures) roadmap published in 2006 [http://cordis.europa.eu/esfri/roadmap.htm] [3]. In addition to BBMRI (Biobanking and Biomolecular Resources Research Infrastructure, http://www.bbmri.eu) five other biological and medical sciences (BMS) infrastructures were selected for the roadmap of 34 mature proposals, representing bioinformatics, structural biology, mouse biology, translational research and clinical research. Biobanked samples and information about the samples obviously provide important material also for the other BMS infrastructures.

Figure 1. BBMRI logo.

Although the existing biological sample collections, biobanks and biomolecular resources are a recognised European strength, biobank-based research suffers from the characteristic European weakness – fragmentation. The samples available have been collected and stored nationally using different techniques and IT-solutions, and under different ethical and legal frameworks, which has made pan-European collaboration in large-scale projects quite challenging. Such a situation has resulted in duplication of effort in different EU Member States, although at the same time large research projects funded by the EC or other sources have demonstrated the usefulness and feasibility of collaboration between large European population-based biobanks [4-5]. Without doubt, efficient use of existing European biobanks has been hampered by their heterogeneity and lack of standardisation and has overall resulted in their underutilisation. For the study of common multifactorial diseases, a major health problem of the ageing European population, collaboration between existing national resources should create the synergism, the gain of statistical power and the economy of scale needed. The existence of population isolates is another European strength by facilitating analysis of predisposing factors in more homogeneous genetic background.

This is the landscape where BBMRI is expected and prepared to integrate the existing quality controlled biobanks, biomolecular resources and enabling technologies into a novel pan-European biomedical research infrastructure, and to guide the way towards establishment of high quality *de novo* European biobanks adhering to the guidelines drafted by BBMRI [6]. The European Commission has granted 5 Mio. € funding for 27 months (2008-2010) to the preparatory phase of BBMRI to conceptualise and secure funding for the construction of the European research infrastructure for biobanking and biomolecular resources.

The aims of BBMRI

The objectives to be addressed by the BBMRI consortium during the preparatory phase are to develop a plan to integrate existing quality controlled biobanks, biomolecular resources and enabling technologies into a novel pan-European biomedical research infrastructure. BBMRI will not only provide a comprehensive source of information about existing biological sample collections and biomolecular resources, but will also provide an operational concept for a sustainable infrastructure, deliver standard operational procedures for future biobanking and codes of conduct for European biobanks. A particularly challenge is the generation of an IT infrastructure capable of linking the existing biobank-derived genetic and molecular phenotyping data with data from clinical phenotyping and health-related registries. Furthermore, BBMRI will evaluate the heterogeneous European ethical and legal frameworks to find solutions how to implement a pan-European infrastructure, as well as to elaborate sustained funding solutions for European biobanking.

Strategy

BBMRI will improve the accessibility and interoperability of the existing comprehensive collections of population based and disease orientated biological samples from different (sub)populations of Europe, including the attached data on health status, nutrition, lifestyle and environmental exposure of the study subjects. As the existing biobanks have a strong national character and background, a distributed hub-and-spoke structure has been suggested for BBMRI. This structure should provide great flexibility so that new members and partners can be connected at any time and the structure can be adapted to the emerging needs of biomedical research. Combined with the expertise of the clinicians, pathologists, bioinformaticians and molecular biologists involved, a globally unmatched, Europe-wide platform for translational medical research is envisaged to develop personalised medicine and disease prevention to the benefit of European citizens. To reach this goal, also

biotech and pharmaceutical industry must have a possibility to collaborate with academic researchers in order to fully realise the enormous potential of European biobanking. An important strategic goal is to create guidelines for better interoperability of *de novo* biobanks. Such guidelines should also help to overcome the current obstacle created by the heterogeneous ethical and legal landscape in Europe. In addition to clinical, ethical and legal experts, patient communities will be involved to achieve standards and guidelines which properly balance individual values, such as protection of privacy and informed consent, with shared values of facilitated access and progress in health care development and prevention.

Action plan

The action plan of the preparatory phase BBMRI is defined in the grant agreement with the EC. The seven Work Packages (WP) of BBMRI (Table 1) are responsible for the specific deliverables aimed at integrating the existing quality controlled biobanks, biomolecular resources and enabling technologies into a novel pan-European biomedical research infrastructure. The operational concept of BBMRI for the next stage will be developed based on the experience gained during the preparatory phase. The distributed hub-and-spoke structure will facilitate generation of technological platforms in areas such as biological resources, high-throughput techniques, bioinformatics and other advanced analytical tools for data analysis. Such platforms will also foster collaboration between academia and industry.

Hubs are coordinated and directed by an executive management, which is supported by a governance council as well as by a scientific and ethical advisory board and receives input from the stakeholder forum (Fig. 2). The IT infrastructure which employs a federated database architecture will integrate the complex network of hubs, members and associated partners.

One of the challenges of setting up large scale studies is the heterogenous legal and ethical frameworks within the EU. The new European legal entity (ERIC) that is currently being developed by the European Commission particularly to support the needs and operation of research infrastructures, foresees the establishment of operational sites in different Member States under one legislation. Such an entity would provide an outstanding opportunity to generate an integrated and harmonized biomedical research area and joint policies for legal and ethical frameworks within the EU. If ERIC is delayed, an interim solution for a legal entity (an international society or an association) will be considered.

Successes

All work packages started their work in February 2008 although funding from the EC only became available in June 2008. Due to the large size of the project the management and organizational structure may appear complex, but has shown its effectiveness in practice. The three-level management structure provides an open and transparent decision-making process which can also be applied for the implementation stage of BBMRI.

Table 1. BBMRI Work Packages (WP's) and governance structure and leaders/chairs.

Work Packages (WP)	**Leader(s)**
WP1: Management and Coordination	K. Zatloukal (AT), E. Vuorio (FI)
WP2: Population-based Biobanks	L. Peltonen (FI/UK), A. Metspalu (EE)
WP3: Disease-orientated Biobanks	E. Wichmann, (DE), T. Meitinger (DE)
WP4: Biomolecular Resources and Molecular Tools	U. Landegren (SE), M. Taussig (UK)
WP5: Database harmonisation and IT-infrastructure	J-E. Litton (SE), M. Fransson (SE)
WP6: Ethical, Legal and Societal Issues	A. Chambon-Thomsen (FR)
WP7: Funding and Financing	G. Dagher (FR), J. Ridder (NL) C. Brechot (FR)
Governance Council Chair	L. Peltonen (FI/UK)
Advisory Board Chair	G-J. van Ommen (NL)
Coordination Board Chair	K. Zatloukal (AT)
Stakeholder Forum Chair	M. Griffith (IR)

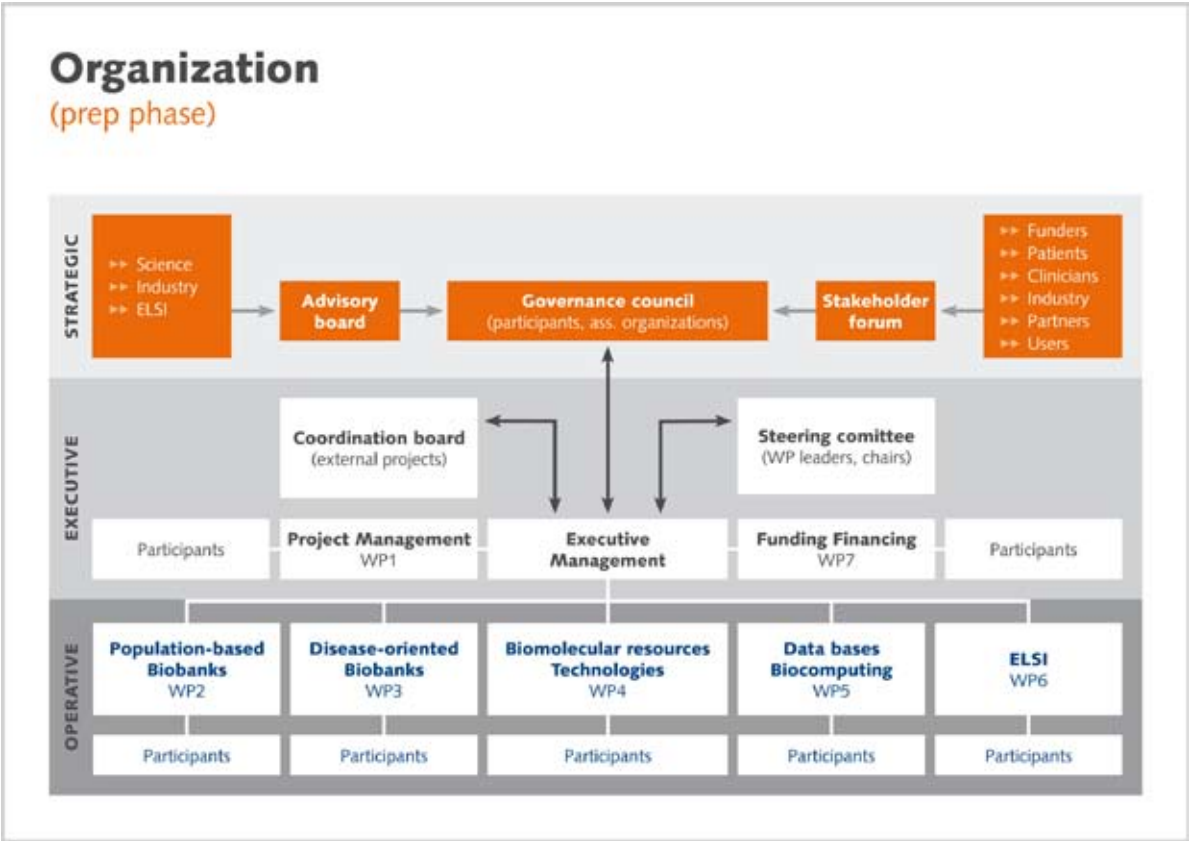


Figure 2. BBMRI organisation and responsibilities.

Initially the EU-funded BBMRI preparatory phase project comprised 50 participants and about 150 associated organisations from 24 countries. One indication of the increased visibility and acceptance of BBMRI is the fact that within the first year the number of associated organisations has increased to 190 representing 29 European countries (Fig. 3). Partners of BBMRI have placed major emphasis on the proper embedding of BBMRI in the global biobanking community. Political and public recognition of BBMRI has increased through presentations in the European Parliament and in a large number of journals and public media. Work on biobank harmonization is coordinated with the P^{3}G consortium (The Public Population Project in Genomics, www.p3gconsortium.org), the strategic research agenda of the Innovative Medicines Initiatives (IMI), the WHO, and the OECD initiative on a global network of Biological Resource Centres. Key representatives of these projects and initiatives have contributed to the goals of the preparatory phase as participants in work packages, members in various boards or as external experts.

One of the early tasks of BBMRI was to prepare an inventory of existing population-based (WP2) and clinical (or disease-orientated, WP3) biobanks in Europe using a survey questionnaire based on an existing P^{3}G questionnaire. This core questionnaire was supplemented with a total of ten supplementary questionnaires dealing with issues such as standardization of procedures, data collection and handling, IT solutions as well as legal and ethical issues and funding. Data obtained from the survey has been added to the BBMRI web site, but additional information is expected to be collected by the WPs throughout the preparatory phase.

WP4 has reviewed existing resources for affinity reagents and other biomolecular resources as analytical tools applicable to biobanking. This has led to a new community standard of affinity reagents (MIAPAR) (submitted for publication), designed to tackle the problems of scattered information and imprecise descriptions and facilitate database implementation. In addition, a new database for molecular methods (MolMeth, www.molmeth.org) [7-8] has been established,

Figure 3. BBMRI Participants and associated organisations. Participants are co-applicants of the project and full members. They have an official vote on formal issues in the Governance Council that is responsible for the definition of the appropriate strategy and processes, and is required for the approval of reports and any changes of the work plan. Associated Members do not have an official vote, but they receiving all relevant information on BBMRI (e.g. forthcoming events, achievements, media reports) and they have the right to participate in several activities, such as Work Package meetings and the BBMRI Governance Council meeting.

providing best practice-based protocols for molecular analyses of different types of samples. Key to any large assembly of data, be it biological, clinical, epidemiological, or behavioral, is the system for information management. WP5 coordinates and supervises all processes of the IT, informatics and infrastructure in the project. Success has been made in finding consensus on a general information management system for maintaining unique and secure coding systems for specimens, subjects and biobanks. WP5 has initiated a searchable BBMRI catalogue of disease-oriented biobanks together with WP3. The catalogue provides a high-level description of Europe's biobanks, including contact details, background and objectives, descriptions of available samples and data and the possibilities for access according to informed consent, links to the biobanks websites etc.

Analyses on the ethical, social and legal issues of the infrastructure have resulted in a conceptual paper on ethics related policies for biobanks and biomolecular resources. Furthermore, a WIKI+ platform for legal aspects of biobanks has been designed and presented as planned. The next step will be its launch on the project website and the involvement of experts.

WP7 has progressed with the challenging tasks of surveying the current and prospective financial needs of biobanks in order to develop a sustainable funding concept for BBMRI. In addition to support from the EC, the biobanks participating in the BBMRI preparatory phase have received commitment from 23 research and health ministries and funding agencies in 13 different European countries. Successful negotiations for sustainable funding for joint BBMRI activities form a critical and vitally important effort as the future of BBMRI is dependent on an efficient central unit coordinating all key activities of the pan-European biobank and biomolecular resource.

Acknowledgements

The preparatory phase of BBMRI has been financially supported by the European Commission (grant agreement 212111).

References

1. Muilu J, Peltonen L, Litton JE (2007) The federated database--a basis for biobank-based postgenome studies, integrating phenome and genome

data from 600,000 twin pairs in Europe. Eur J Hum Genet. 15(7):718-23.

2. Benyamin B, Perola M, Cornes BK, Madden PA, Palotie A, Nyholt DR, Montgomery GW, Peltonen L, Martin NG, Visscher PM (2008) Within-family outliers: segregating alleles or environmental effects? A linkage analysis of height from 5815 sibling pairs. Eur J Hum Genet. 16(4):516-24.
3. ESF SCIENCE POLICY BRIEFING - 32 - May 2008. 5. Key Challenges of European Biobanking. http://www.esf.org/fileadmin/links/EMRC/SPB32Biobanking%5B1%5D.pdf
4. Aulchenko YS, Ripatti S, Lindqvist I, Boomsma D, Heid IM, Pramstaller PP, Penninx BW, Janssens AC, Wilson JF, Spector T, Martin NG, Pedersen NL, Kyvik KO, Kaprio J, Hofman A, Freimer NB, Jarvelin MR, Gyllensten U, Campbell H, Rudan I, Johansson A, Marroni F, Hayward C, Vitart V, Jonasson I, Pattaro C, Wright A, Hastie N, Pichler I, Hicks AA, Falchi M, Willemsen G, Hottenga JJ, de Geus EJ, Montgomery GW, Whitfield J, Magnusson P, Saharinen J, Perola M, Silander K, Isaacs A, Sijbrands EJ, Uitterlinden AG, Witteman JC, Oostra BA, Elliott P, Ruokonen A, Sabatti C, Gieger C, Meitinger T, Kronenberg F, Döring A, Wichmann HE, Smit JH, McCarthy MI, van Duijn CM, Peltonen L; ENGAGE Consortium (2009) Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat Genet. 41(1):47-55.
5. Sabatti C, Service SK, Hartikainen AL, Pouta A, Ripatti S, Brodsky J, Jones CG, Zaitlen NA, Varilo T, Kaakinen M, Sovio U, Ruokonen A, Laitinen J, Jakkula E, Coin L, Hoggart C, Collins A, Turunen H, Gabriel S, Elliot P, McCarthy MI, Daly MJ, Järvelin MR, Freimer NB, Peltonen L (2009) Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nat Genet. 41(1):35-46.
6. Yuille M, van Ommen GJ, Bréchot C, Cambon-Thomsen A, Dagher G, Landegren U, Litton JE, Pasterk M, Peltonen L, Taussig M, Wichmann HE, Zatloukal K (2008) Biobanking for Europe. Briefings in Bioinformatics. 9(1):14-24.
7. 't Hoen PA, Ariyurek Y, Thygesen HH, Vreugdenhil E, Vossen RH, de Menezes RX, Boer JM, van Ommen GJ, den Dunnen JT(2008) Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. Nucleic Acids Res. 36(21):e141.
8. Altman RB, Bergman CM, Blake J, Blaschke C, Cohen A, Gannon F, Grivell L, Hahn U, Hersh W, Hirschman L, Jensen LJ, Krallinger M, Mons B, O'Donoghue SI, Peitsch MC, Rebholz-Schuhmann D, Shatkay H, Valencia A (2008) Text mining for biology--the way forward: opinions from leading scientists. Genome Biol. 9 Suppl 2:S7.

EATRIS

Infrastructure bridges Basic Research and Medical Innovation

Regina Becker and Rudi Balling

Scientific directorate, Helmholtz Centre for Infection Research, Braunschweig, Germany

www.eatris.eu

The "**E**uropean **A**dvanced **T**ranslational **R**esearch **I**nfra**S**tructure in Medicine", EATRIS, is a new research infrastructure initiative which will help Europe to fulfil its potential in the strategically critical area of translational medical research. EATRIS is a unique framework linking European countries to accelerate the development of new medicinal products by facilitating access to a new pan-European infrastructure. EATRIS is one of the biomedical infrastructure projects initiated by the European Strategy Forum on Research Infrastructure (ESFRI) and is currently funded by the 7th Framework Programme of the European Commission.

Translational Medicine

"Translational medical research" or "translational medicine" is the crucial step between basic laboratory research and practical, clinical applications. At its best, translational medicine is a two-way street: the discoveries of basic research are developed into new tools for clinical care, and observations made in clinical care can inspire new approaches in basic research (Figure 1).

Everywhere in Europe basic researchers are hitting the same obstacles: they explore disease mechanisms, they find ways to influence these mechanisms and find new targets. They find ways to influence the target. All this contributes to the understanding of the biology of human disease. But the next step, the practical application of their basic work and the advancement into preclinical and clinical trials, is not available

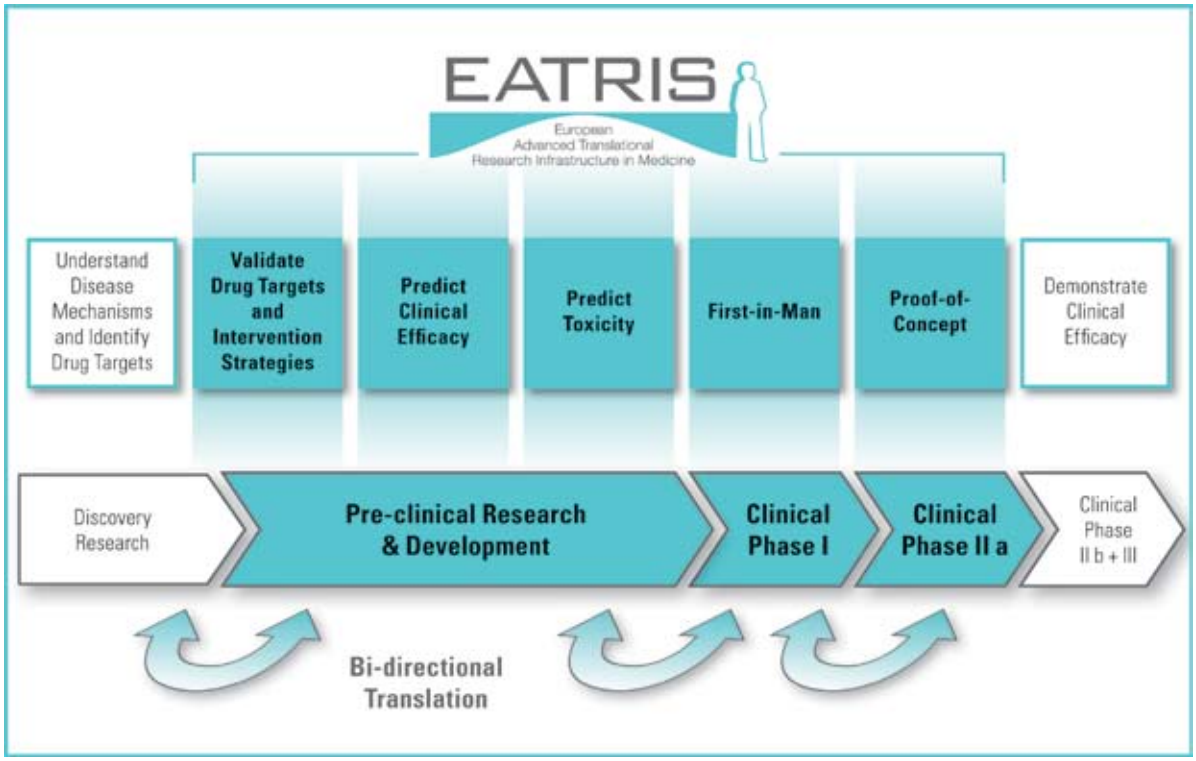


Figure 1. The infrastructure provided by EATRIS will cover the complete chain necessary to bring translational projects for which a first proof of principle has been established to the first proof of concept in human (clinical phase IIa).

to them. Despite this productive and strong biomedical research in Europe, the research stays in laboratories and journals and does not result in the groundbreaking new therapies and diagnostics that should follow from this work.

EATRIS – the Mission

EATRIS serves the translation of both diagnostics and therapeutics with a three-way approach, focusing on: product, people and population. EATRIS provides the necessary means to the research community to further develop their research results into products, EATRIS trains scientists to think beyond their discipline and provides researchers with knowledge about clinical needs and regulatory requirements, and finally EATRIS contributes to public health by improving diagnostics and treatment for the population as a whole.

The EATRIS infrastructure will provide the following types of infrastructure: physical facilities, expert knowledge and training programmes. It uses the unique approach of opening the doors to the best comprehensive translation centres to provide access for external users with promising discoveries.

EATRIS will enable academic institutions to translate their research and gain greater control and greater returns on early research investments. EATRIS translational research centres are the instruments to fulfil this mission but EATRIS will be more than a network of centres. The supply of research support is harmonised, facilities for research built up in EATRIS are largely complementary and a central governance structure coordinates the internal management and serves at the same time as an entrance portal for external scientists.

The EATRIS Centres

Building on excellence

Each EATRIS centre will consist of one or more institutions which are excellent in translational research and will be capable of handling the entire development chain for one or more medicinal products. The EATRIS centres will specialise according to their core expertise in products such as diagnostics, small molecule drugs, biologics, vaccines or novel therapeutics. The initial disease fields envisaged are the most pressing

ones: cancer, infection, cardiovascular, metabolic and neurological diseases.

The strategic goal is to have all necessary disciplines, such as those needed in pre-clinical development labs and study centres, as a strong innovation core close together, and complementing them with large-scale technology facilities like screening platforms. Building on existing excellence and knowledge will avoid unnecessary duplication of research environments and limit the need for the construction of new research facilities.

Comprehensiveness

EATRIS centres comprise the technological facilities, expertise and the clinical environment needed for all aspects of translational research. In an integrated environment involving different disciplines, the following will all be part of EATRIS centres: GLP animal facilities, labs for medicinal chemistry, libraries and the associated screening facilities, 'omics'-screening platforms, imaging facilities, both pre-clinical and clinical, cyclotrons to produce tracers and GMP facilities, as well as hospitals and early clinical trial units. The EATRIS goal is to take up discoveries and develop them to a stage where a proof of concept in human (clinical phase IIa) can be demonstrated.

Professional Management

In each EATRIS centre, project managers experienced in medicinal product development will oversee and steer the translational projects. Their experience in product development ensures an optimal progress on the development chain and compliance with regulatory requirements. Another key success factor is the multidisciplinary team which will be assembled for each project in order to provide input for each stage of the development chain from the outset.

Capacity building

Education is another essential element of EATRIS to improve translational medicine. Training programmes such as PhD programmes and trainings for technicians in order to bring valuable knowledge about translation both back from the clinic into the basic research laboratory and in the other direction to train Europe's next generation of clinical scientists and basic translational researchers. This will lead to more flexible and open career paths and a better exchange between the still too separate worlds of the clinic and research labs.

Figure 2. EATRIS is a new research infrastructure initiative consisting of a network of well-renowned biomedical translation research centres across Europe. Currently ten European countries are partner countries in the EATRIS consortium (turquoise).

Uniqueness of EATRIS

- The variety of platform technologies which EATRIS engages in will be unrivalled in any other framework. EATRIS closely intertwines development and basic research to provide a deeper understanding of the underlying biology and thus facilitates innovative solutions.
- The operational connection of hospitals and technology centres will encourage the development of a more personalised medicine. Carried out alongside therapeutics development, the development of novel and robust biomarkers will make therapies both more specific and less toxic.
- Combining technology and therapy development will lead to a faster drug development: Labelling therapeutic agents so that they can be traced *in vivo* with imaging methods can

speed up the assessment of the success of the therapeutic approach.

- Biomedical researchers and clinical scientists located at universities, research institutions or SMEs will be attracted to EATRIS for the support it offers to their research projects. EATRIS will act like a funnel drawing in a multitude of excellent discoveries, select the top ones and translate them for the benefit of the patient.

Let the EATRIS vision come true

The Preparatory Phase (2008-2010)

A strong consortium of excellent research centres in the area of translational biomedical research and relevant national and regional research policy makers has gathered to work on this vision of flourishing translational biomedical research in Europe (Figure 2).

The current Preparatory Phase of EATRIS aims at defining the benchmarks to establish a cutting-edge infrastructure for translational medical research, comprehensive training programmes and the requirements for implementation. A business plan will present the financing strategy and the governance scheme of EATRIS. Access conditions, a legal framework and a competitive IP management regime are being drawn up during this phase. The results of the Preparatory Phase will form the basis for the next implementation stage.

Construction Phase (2011-2015)

During a Construction Phase the different EATRIS sites will see their capacities expanded and a full coverage of the necessary technological facilities established. During this phase will already support a limited number of user projects within the framework of the existing infrastructure in the centres. By the end the EATRIS will be fully operational and offer support on a regular basis.

ECRIN

Integrating clinical research in Europe: the European Clinical Research Infrastructures Network

Jacques Demotes-Mainard, Roxane Brachet, Christine Kubiak
INSERM, Institut de Santé Publique, PARIS

www.ecrin.org

Summary

ECRIN consists of integrating national clinical research facilities into a EU-wide network, able to provide support to clinical research in any medical field, and for any type of clinical research through information and consulting, and through a set of flexible services for the conduct of multinational clinical studies. This distributed infrastructure, based on the integration of competence centres, provides access to clinical research projects after assessment by its scientific board. A team of European correspondents working in the coordinating centre of each national network is the key actor in the provision of consulting and decentralised services. These services are particularly relevant for academic clinical research, especially under circumstances where international cooperation is required (ie. in rare diseases), or for clinical trials sponsored by biotechnology SMEs who often lack the capacity to act as a sponsor in EU-wide studies.

Introduction

Development of diagnostic and therapeutic innovation, and delivery of improved health care to EU citizens requires clinical research during the whole process extending from understanding the mechanism of disease, genetic studies or identification of biomarkers, clinical development and evaluation, and to post-marketing strategy trials.

Figure 1: Map of countries participating in ECRIN.

The recent development of therapeutic innovation is mainly based on biopharmaceuticals and on personalized treatments, on pharmacogenetics and toxicogenetics, on the use of biomarkers, and requires access to large populations of patients, enabling clinical trials adapted to these new therapeutic strategies with a need to focus on specific patient subpopulations (1-5). Further, a very large number of rare diseases are without effective interventions. In addition, the quality of clinical trials and other clinical investigations, the quality of clinical and biological data, and the rate of enrolment of patients into clinical trials are all requiring urgent improvements. Hence, the quality of the clinical research infrastructure is one of the main factors determining the competitiveness of European clinical research. European academic research (6), as well as the pharmaceutical and biotechnology research and development need an efficient, integrated, and professionalized organization of clinical research, based on competence centres able to provide efficient support through a consistent set of services for clinical trials. Infrastructures supporting clinical trials, data management, quality assurance, monitoring, ethics, and regulatory submissions are required for improving the quality and raising the credibility of data. An integrated, EU-wide infrastructure allows the conduct of multinational studies in Europe, taking advantage of the EU population and competencies, unlocking latent expertise and patients currently scattered across the EU member states.

ECRIN: a three-step project

The European Clinical Research Infrastructures Network (ECRIN) is designed to improve the capacity of the European Union to perform high-quality clinical research, and to promote innovative pharmaceutical and biotechnology development as well as development of other interventions (7). This integrated clinical research infrastructure bridges the fragmentation of clinical research in Europe through the interconnection of national networks of clinical research centers and clinical trial units. ECRIN participants are currently Austria, Belgium, Denmark, Finland, France, Germany, Hungary, Ireland, Italy, Spain, Sweden, Switzerland, and the United Kingdom (Figure 1). ECRIN plans extension to other existing national networks in other member states, and stimulates the set-up of new national networks for further connection through its capacity building programme.

Coordinated by the National Institute of Health and Medical Research in France (INSERM), ECRIN

started with a first project (2004-2005), funded by the European Commission under the FP6 Health programme, and helped identify bottlenecks to multinational collaboration (mostly the poor capacity of academic institutions to act as sponsors in multinational studies) and define a strategy for provision of services (mostly support to sponsors across the borders) (8 -10)

This conclusion served as a basis for the second phase of the ECRIN project (2006-2008, FP6 Health Programme) in which transnational working groups prepared guidelines and procedures to support clinical studies in any medical field, in any patient population, and for any type of study.

These working groups covered: interaction with ethics committees, interaction with competent authorities and regulatory requirements, adverse event reporting, data management, study monitoring, and quality assurance.

In its third step (2008-2011, FP7 Health priority - Infrastructures programme), ECRIN enters into the preparatory phase of the European strategy forum on research infrastructures (ESFRI) roadmap infrastructures. During this current third phase, the team of European Correspondents, a specialised staff located at the coordinating centre of each national network (Figure 2), provides decentralised support to multinational studies through a set of flexible services to investigators and sponsors (interaction with ethics committees, with competent authorities and support on regulatory submissions, in adverse event reporting, in drug dispensing, in the circulation of blood and tissue samples, in study monitoring). ECRIN also provides centralized services, including data management through accredited data centres.

ECRIN may also help with consulting and practical information on ethical and regulatory requirements, insurance, centre selection, cost evaluation, and funding opportunities during the preparation of the clinical research project.

An independent scientific board is in charge of providing access to ECRIN, based on a set of eligibility and acceptance criteria (see www.ecrin.org).

Impact and users

ECRIN has a substantial impact on the structuring of clinical research in the European Union, through the debate on the *legislative framework* for clinical research in the EU by contributing to the discussion on the 2001/20/EC Directive (12, 13) and to the FP7 ICREL (Impact on Clinical Research of European Legislation) project (14). ECRIN also promotes the active participation of patients and citizens, and transparency in clinical research, and has launched the *International Clinical Trials Day* (each 20th of May, see www.jameslindlibrary.org) as a yearly communication event on the challenges raised by clinical research. In addition, ECRIN initiated, with the other ESFRI-biomedical research infrastructures and the pharmaceutical companies as participants, the FP7 Innovative Medicines Initiative (IMI) (15, 16) EMTrain project to develop a pan-European *education* platform (www.emtrain.eu).

Such an integrated infrastructure will benefit mainly the academic scientific community but also SMEs or pharmaceutical companies, and public-private partnership programs, as well as series of projects developed by disease-oriented scientific networks. As access to patients is a limiting factor, ECRIN will promote research on rare diseases, improving diagnostics and treatment strategies. This will enable translation of medical innovation into healthcare, hence to the benefit of patients and citizens.

References

1 Innovation or Stagnation? Challenge and Opportunity on the Critical Path to New Medical Products. http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html

2 Zerhouni E. The NIH Roadmap. Science 302 :63-72 (2003).

3 Zerhouni E. Translational and clinical science – time for a new vision. NEJM, 353 :16211623 (2005)

4 The Royal Society. Personalised medicines : hopes and reality. September 2005, pp 1-52. Access through http://www.royalsoc.ac.uk/document.asp?id=3780

5 Rawlins MD. Cutting the cost of drug development ? Nature Rev. Drug Discov. 3 :360-364 (2004)

6 Remuzzi G, Schieppati A, Boissel JP, Garattini S, Horton R: Independent clinical research in Europe. Lancet 364: 17232-6 (2004).

7 Demotes-Mainard J, Ohmann C : European Clinical Research Infrastructures Network : promoting har-

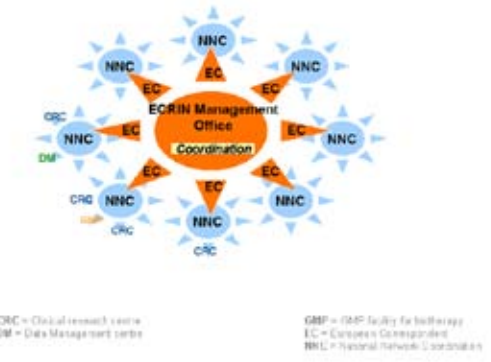


Figure 2: Organisation of the network

monisation and quality in European clinical research. Lancet 2005; 365, 107-108.

8 Demotes-Mainard J, Chêne G, Libersa C, Pignon JP : Clinical research infrastructures and networks in France : report on the french ECRIN workshop. Thérapie, 60:183-199, 2005.

9 See reports and comparative analyses on www.ecrin.org

10 Demotes-Mainard J, Ohmann C, Gluud C, Chene G, Fabris N, Garattini S, Carné X, Lafolie P, Collet JP, Crawley F. European Clinical Research Infrastructures Network Meeting report : 'Towards an integration of clinical research infrastructures in Europe', Brussels, Feb 14-15th, 2005. Int J Pharm Med, 19:43-45, 2005.

11 Krleža-Jeric K, Chan A-W, Dickersin K, Sim I, Grimshaw J, Gluud C for the Ottawa Group. Principles for international registration of protocol information and results from human trials of health related interventions: Ottawa Statement (part 1)[1]. BMJ 2005;330:956-8.

12 EC-EMEA conference on the Operation of the Clinical Trials Directive, Oct.2007, London.

13 ESF-EMRC Consensus Conference Investigator-Driven Clinical Trials, Sep. 2008, Strasbourg.

14 ICREL conference on Impact on Clinical Research of European Legislation, Dec. 2008, Brussels http://www.efgcp.be/Conference.

15 Donnelly F, Jehenson P. European Technology Platform on Innovative Medicines. Int J Pharm Med, 19 :153-161 (2005).

16 The Innovative Medicines Initiative (IMI) Strategic Research Agenda: Creating Biomedical R&D Leadership for Europe to Benefit Patients and

1 http://bmj.bmjjournals.com/cgi/content/extract/330/7497/956

ELIXIR

Data for Life: from information to the Medicines and Bio-industries of the Future

Andrew Lyall
ELIXIR Project Manager,
European Bioinformatics Institute (EBI), Hinxton, UK

www.elixir-europe.org

In May 2007 EMBL was awarded a Framework 7 grant of 4.5 Million Euro to run the Preparatory Phase of a project called ELIXIR of which Professor Janet Thornton, Director of EMBL-EBI is the coordinator. The purpose of ELIXIR is to construct a sustainable infrastructure for biological information in Europe. The purpose of the Preparatory Phase is to make the plan for the construction phase which will follow.

ELIXIR is one of six bio-medical projects that are part of the European Strategic Forum on Research Infrastructures (ESFRI) Roadmap. It is very significant that bio-medical projects are part of the ESFRI Roadmap as this is the first time that it has been recognized at this level that biology needs infrastructures in the same way that the physical sciences do. This is necessary because the nature of biological research is changing due to the availability of new high-throughput technologies such as next-generation sequencing.

Biology is changing from an activity engaged in by individuals and small groups to one in which large coordinated projects will make a much larger contribution. The intensive nature of the new technologies means that teams of peoples are required to generate the data and then other teams to process and understand it and to translate that understanding into improvements in healthcare and so on. This in turn is going to need to be a change in the way in which the infrastructure for biology research is funded. This is because the new technologies, as well as being extremely powerful, require much more substantial capital investment than did previous technologies. In particular, they produce vast

amounts of data and there is an urgent need for substantial investment in biological computing infrastructure if Europe is gain maximum benefit from these advances. It is not going to be enough to reorganize the way in which research funding is allocated; there is not enough money there for that to work. Biology has to coordinate its activities and access the sources of funding in a structured way that has, until now, only been employed in the physical sciences. This is the reason for ELIXIR and for the inclusion of biology projects in the ESFRI process.

The Human Genome Project is hardly completed and we are embarking upon the Thousand Genome Project. Further down the line, the Cancer Genome Project is proposing to sequence the genomes and tumor genomes of 25,000 cancer patients. How long will it be before we see the announcement of the Million Genome Project; and then what? Further more, it is not just sequencing technologies that are set to produce prodigious quantities of data; advances in structural biology, proteomics and biological imaging are also poised to add to the deluge. All these data need to be curated, archived and made available to the biological and medical community in useful and relevant ways; hence the need for ELIXIR. All of this activity will also need to be coordinated to ensure that everything is covered in the most appropriate way and that there is no duplication that wastes resources; hence again, the need for ELIXIR.

Although the amount of money that is going to be needed to achieve ELIXIR goals may seem very large, it is in fact quite small compared with the amount of money that will be spent generating the data in the first place. Further more, it is also relatively small in comparison with the amount of money that Europe has spent in the past on infrastructure for the physical sciences. When one thinks that the Grand Challenges facing Europe are biological ones, namely, healthcare for an aging population, the need to protect the environment, the need for increased food production, the need to sustain a viable European Biotechnology and Pharmaceutical Industry and so on, the question we should actually be asking is "Can Europe afford NOT to make this investment?".

ELIXIR consists of 32 organizations and institutes from across all of Europe. The Preparatory Phase started in November 2007 and will run for three years. Its purpose is to produce a plan for building and funding the construction phase which will follow. The plan is intended to cover a ten year period, something unprecedented in the usual grant funding processes used in biology, requiring commitment from the politicians and funding agencies alike. Physicists have been walking the corridors of power for several generations and know how to coordinate their actions to lobby for very large resources; the time has come for us to do the same, but we cannot take another generation to do it, for it we do then Europe will be left behind. Europe already invests less in centrally-organized biological-data management than Japan and the USA and we must not fall behind the emerging economies of India and China.

ELIXIR needs the support of every biologist in Europe, particularly in convincing politicians and funding agencies of its importance - find out more from the ELIXIR web site (www.elixir-europe.org), attend the stakeholders meetings, tell your friends about us and do what you can to make sure that the funding agencies in your country understand the need to contribute to ELIXIR.

The ease with which you will be able to access the fundamental data of biology in the future depends on the success of ELIXIR now!

Infrafrontier

Mouse models and phenotyping data for the European biomedical research community

Michael Raess and Martin Hrabé de Angelis

Institute of Experimental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health (GmbH), Germany

www.infrafrontier.eu

infrafrontier@helmholtz-muenchen.de

Mice as models for human diseases

Mouse and man share 95% of their genetic make-up. Mice are easy to keep and breed in the laboratory and researchers have developed a comprehensive toolbox for altering the mouse genome. This is why mice are ideal models for human diseases such as Diabetes, Osteoporosis, Asthma, Alzheimer's disease or Depression, as documented by an exponentially increasing number of scientific publications on mouse models and the Nobel Prize in Medicine in 2007. To provide a source for new mouse models and support the research activities in the field of functional genomics, the systematic mutagenesis of the approximately 25.000 genes in the mouse genome is currently underway, coordinated by the International Mouse Knockout Consortium (IKMC). These developments create an enormous demand for access to a systematic functional and molecular characterisation of the mouse mutants. Furthermore, new mouse models for human diseases must be made available to entire European mouse genetics, biomedical and translational research community [1].

Infrafrontier

It is clear that this tremendous task cannot be fulfilled by individual research facilities or on the national level alone. This is the rationale for the European project *Infrafrontier* (The European infrastructure for phenotyping and archiving of model mammalian genomes, www.infrafrontier.eu), which is coordinated at the Helmholtz Zentrum München by Prof. Hrabé de Angelis. *Infrafrontier* is on the European roadmap for research infrastructures of ESFRI (European Strategy Forum for Research Infrastructures, http://www.cordis.europa.eu/esfri) and receives funding from the EC's Seventh Framework Program. It will organise a pan-European research infrastructure to increase the capacities for systemic phenotyping and archiving of mouse models. The *Infrafrontier* consortium currently contains 22 partners (representing 14 phenotyping and archiving centres, 1 bioinformatics institute and 12 European ministries and funding agencies) from 10 different European countries. Six new partners will join *Infrafrontier* in the near future, extending the project to Austria, Czech Republic and Canada (Fig.1).

Phenomefrontier and Archivefrontier

In *Infrafrontier*, the four existing European primary phenotyping centres (or mouse clinics, see below) in Germany, France and the U.K. will associate with three emerging phenotyping centres in Spain, Italy and the Czech Republic, and with the Toronto-based Centre for Modeling Human Disease (CMHD) to form *Phenomefrontier*, a sustainable pan-European research infrastructure, providing capacities and access to mouse model phenotyping in Europe and around the globe.

The second major part of *Infrafrontier*, called *Archivefrontier*, aims to increase the capacities for the archiving and distribution of mouse models on a sustainable basis. To achieve this, EMMA, the European Mouse Mutant Archive (www.emmanet.org, [2]) will be extended and upgraded. Several new archiving nodes in Europe and one in Canada will be added to the EMMA network (Fig. 2).

Figure 1. The members of the *Infrafrontier* consortium are phenotyping and archiving facilities as well as ministries, research councils and funding agencies in 12 European countries and Canada.

Systemic phenotyping and archiving / distribution of mouse models

Mouse Clinics use a whole-system approach to obtain a comprehensive picture of the systemic effects of mutations in the mouse genome: mutant mice are screened for alterations in bone and cartilage development, neurology and behaviour, clinical chemistry, immunology, energy metabolism and many more, comprising all essential organ systems [3]. At least 320 parameters are measured per mutant mouse line. This *systemic phenotyping* approach broadens our view of the functions of individual genes. Indeed, as recently reported by the German Mouse Clinic [1, 4], in the mouse lines characterised so far most mutations affected more than one organ system. Novel phenotypes were discovered in more than 95% of the cases, even in mouse lines that had been used in research for many years. This clearly shows the value of the approach. The existing European mouse clinics collaborate in the European pilot project EUMODIC (The European Mouse Disease Clinic, www.eumodic.org) to systemically phenotype 650 mutant mouse lines, following standardised phenotyping protocols that are detailed in the EMPReSS database (European Mouse Phenotyping Resource of Standardised Screens, www.empress.har.mrc.ac.uk, [5]).

Mouse models are archived and distributed to the biomedical research community by EMMA. The currently ten European partners of EMMA run an archive of together more than 1600 cryo-preserved mouse lines. Frozen sperm or embryos of about 300 mouse lines are added each year. In 2008 EMMA distributed upon request more than 360 mouse lines (as frozen material or life mice) to research institutions around the world.

Managing phenotyping and archiving data acquisition

High-throughput phenotyping and archiving would not be possible without the development of sophisticated computational tools for the management of complex logistics and the handling of large amounts of data. The phenotyping and archiving centres organised in *Infrafrontier* use several different software solutions, most of them developed in-house by dedicated bioinformatics groups.

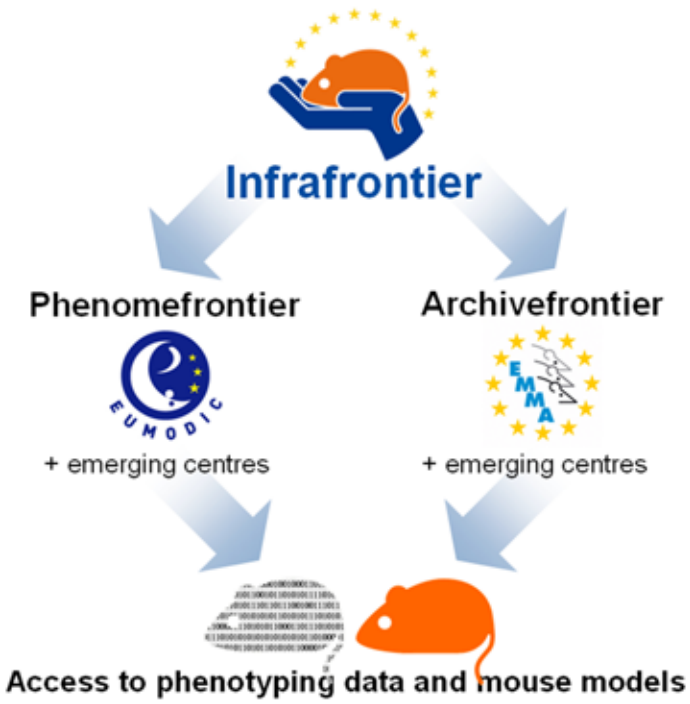


Figure 2. *Infrafrontier* organises two complementary pan-European research infrastructure networks: Phenomefrontier for large-scale systemic phenotyping, Archivefrontier for archiving and dissemination of mouse models. Both networks build on existing European initiatives and will include new facilities.

- **Mouse and facility management systems:** Management of mouse keeping, mouse breeding programs including pedigree information, cage capacities, specific housing or dietary requirements, as well as the documentation of all individual mouse fates have to be accomplished.
- **Workflow management systems:** High-throughput systemic mouse phenotyping requires the logistic management of multi-parallel test pipelines following a standardised workflow. Similarly, cryo-preservation and re-derivation of frozen material by the mouse archives involves complex parallel workflows. An additional task is the management of intellectual property or material transfer agreements associated with the different mouse lines.
- **Laboratory information management systems, storage device and sample management:** In the systemic phenotyping pipelines at least 320 different parameters are measured for each mouse line and their results recorded in a relational database. These results have to be matched with SOPs and metadata describing e.g. experimental and environmental conditions. In the archiving centres the storage capacities of the cryo-tanks or freezers, as well as the storage locations of each individual cryo-sample and its aliquots (including the back-up systems) have to managed.

An example of such a software solution is MausDB. It is used in the German Mouse Clinic and the German EMMA node. MausDB is realised as a LAMP system (**L**inux as operating system, **A**pache web server, **M**ySQL database, **P**erl as programming language) and accomplishes most of the tasks listed above [6]. MausDB is an open-source software that is freely available from the download section of the Institute of Experimental Genetics of the Helmholtz Zentrum München (http://www.helmholtz-muenchen.de/en/ieg/downloads/index.html).

Accessing phenotyping and archiving data

Phenotyping data from mutant mouse lines is available at EuroPhenome (www.europhenome.org, [7]), an open-access platform which contains the data collected by the EUMODIC project. This MySQL relational database offers a data browser

based on PHP, JSP and AJAX and several tools for data visualisation and data mining. An important future step will be the mapping of the phenotyping parameters onto phenotype ontologies that can be included into genome databases like Ensembl, Gene Ontology (GO) and Mouse Genome Informatics (MGI) [8, 9]. Information on mouse lines available from EMMA can be found on the archive's website (www.emmanet.org). The mouse line descriptions contain information on the genetic background, the phenotype and genotype descriptions that are cross-linked to the MGI database.

Bioinformatics activities in Infrafrontier

Infrafrontier's bioinformatics activities have, besides the maintenance of the project's web resources, two major objectives. The first is a comprehensive survey on IT systems for managing facilities, mice and data that are used by phenotyping and archiving facilities around the world. This will lead to a report with recommended IT solutions for new facilities. Moreover, suggestions for a common nomenclature will be made that will facilitate exchange of mouse line data between the facilities. The second objective is the definition of minimum information standards required for an integration of the existing databases for phenotyping, archiving and mouse production data (EuroPhenome, EMMA, EUCOMM [the European part of the IKMC]).

Since data is a major asset of *Infrafrontier*, it is important that the project keeps strong links with other information-oriented European initiatives, such as the coordination action CASIMIR (Coordination and sustainability of International Mouse Informatics Resources, www.casimir.org.uk), the EMBRACE Network of Excellence (A European Model for Bioinformatics Research and Community Education, www.embracegrid.info), and the ESFRI project ELIXIR (European Life Sciences Infrastructure for Biological Information, www.elixir-europe.org), as well as the other ESFRI research infrastructure projects in the biological and medical sciences.

Conclusions

In order to understand the fundamental processes governing living organisms in health and disease, modern life sciences employ increasingly sophisticated technologies and produce increasingly complex and large datasets. High-throughput phenotyping and archiving of mouse models of human diseases are good examples for this development. Infrafrontier will provide access to a sustainably funded research infrastructure to meet the growing demand for these services in the biomedical research community. Infrafrontier in concert with the other ESFRI biological and medical research infrastructure initiatives will promote cutting-edge research ranging from basic science to the translation of results into novel drugs and treatments. Their success is likely to change the face of biomedical research in the European Research Area.

References

1. Beckers J, Wurst W, Hrabé de Angelis M (2009) Towards better mouse models: enhanced genotypes, systemic phenotyping and envirotype modelling. Nat Rev Genet 10: 371-380.
2. Hagn M, Marschall S, Hrabé de Angelis M (2007) EMMA - the European mouse mutant archive. Brief Funct Genomic Proteomic 6: 186-92.
3. Brown SDM, Hancock JM, Gates H (2006) Understanding mammalian genetic systems: The challenge of phenotyping in the mouse. Plos Genetics 2: 1131-1137.
4. Fuchs H, Gailus-Durner V, Adler T, et al. (2009) The German Mouse Clinic: A Platform for Systemic Phenotype Analysis of Mouse Models. Current Pharmaceutical Biotechnology 10: 236-243.
5. Brown SDM, Chambon P, Hrabé de Angelis M, et al. (2005) EMPReSS: standardized phenotype screens for functional annotation of the mouse genome. Nature Genetics 37: 1155-1155.
6. Maier H, Lengger C, Simic B, et al. (2008) MausDB: An open source application for phenotype data and mouse colony management in large-scale mouse phenotyping projects. BMC Bioinformatics 9.
7. Mallon AM, Blake A, Hancock JM (2008) EuroPhenome and EMPReSS: online mouse phenotyping resource. Nucleic Acids Research 36: D715-D718.
8. Hancock JM, Adams NC, Aidinis V, et al. (2007) Integration of mouse phenome data resources. Mammalian Genome 18: 157-163.
9. Beck T, Morgan H, Blake A, et al. (2009) Practical application of ontologies to annotate and analyse large scale raw mouse phenotype data. BMC Bioinformatics 10: S2.

INSTRUCT

an Integrated Structural Biology Infrastructure for Europe




Rosemary Wilson[1], Claudia Alen Amaro[2], Susan Daenke[2], David Stuart[2],

[1] Structural Biology Unit, EMBL Hamburg, Germany
[2] Wellcome Trust Centre for Human Genetics, University of Oxford, UK

http://www.instruct-fp7.eu/

Introduction

Structural biology is a branch of molecular biology concerned with the molecular structure of biological macromolecules such as proteins and viruses. Resolving a molecule's structure can give us insights into the function of a molecule and eventually lead to potential drug targets. Structural biologists use many different techniques (Figure 1) including X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, electron microscopy (EM), light microscopy and a range of other imaging techniques. Whereas these disciplines have traditionally been tackled as single entities, this is no longer sufficient to gain a detailed and quantitative understanding of the dynamic structure and biological context of the cell. Each technique resolves structures at different resolutions, and over the past few years, scientists have recognized the need and advantage of collaborating on projects, integrating several different techniques to resolve a single biological problem in all its complexities.

The future

Over the next few decades, structural biology will face major scientific challenges, and the need to integrate existing techniques and develop emerging technologies, has never been more important. Advances in structural biology will produce other problems and scientists also recognize the need for developments in supporting technologies such as the automation of data collection and structural determination and data management. These challenges are all more than any one lab can manage individually, and coordinating efforts and equipment will be paramount for achieving these goals and pursuing cutting-edge science in Europe. The recent instruction from President Obama to significantly increase the ambition and scope of scientific research in the USA highlights the importance of an equal ambition for INSTRUCT in coordinating a truly integrated link between structural and functional research in Europe.

The project

Coordinated by Prof. David Stuart from the University of Oxford, INSTRUCT is one of the biomedical projects in the European Strategy Forum on Research Infrastructures within the Framework 7 of the EU, and aims to establish an integrated structural biology infrastructure in Europe to support the development of cellular structural biology research. A number of centres across Europe will allow access for scientists to state-of-the-art equipment in core and complementary techniques, thereby opening new scientific horizons in biomedical research. Alongside these scientific challenges, INSTRUCT also plans to advance and disseminate technologies and methodologies, and to train scientists across Europe in using state-of-the-art and newly developed infrastructure. This vision requires a major, long term and focused investment and will revolutionize the way biomedical research is done in Europe. A biological infrastructure on this scale is unprecedented in Europe, but is crucial in order for European science to maintain its competitive edge and play a leading role in such a vital research area. To date, European research has often been hindered by the short-term status of European funding. Securing a long-term investment will enable

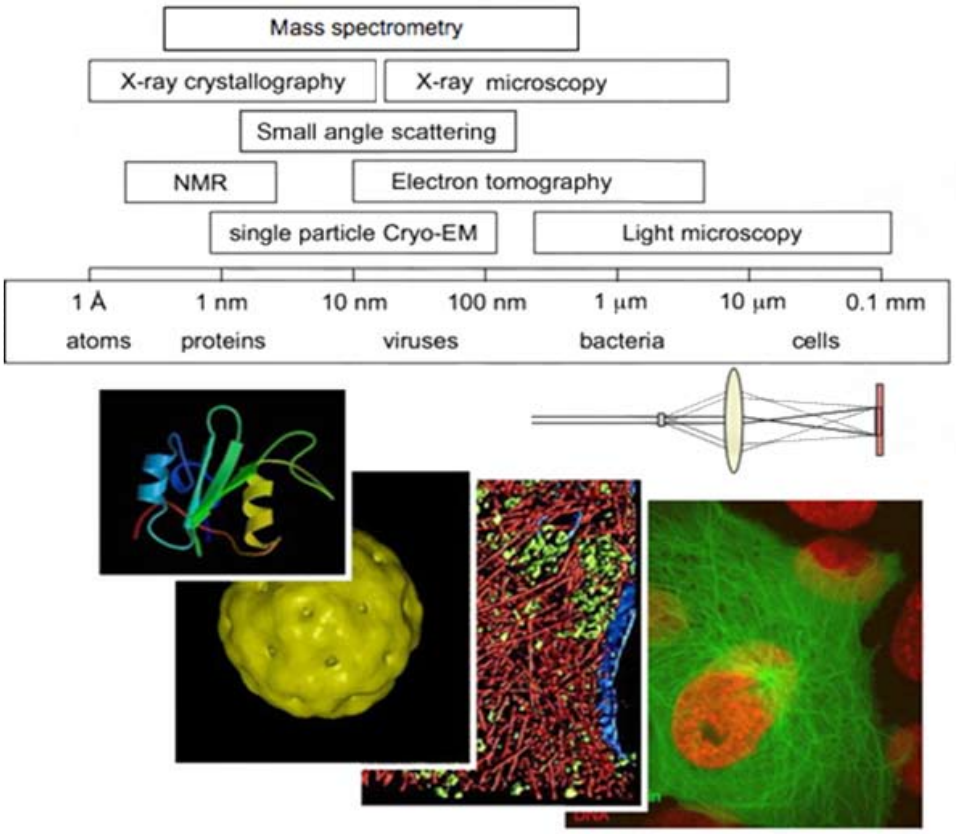


Figure 1. An overview of structural biological techniques showing the range of sizes of molecules studied by each method (Source: Cyttron, Leiden University).

the planning of more ambitious, larger scale and more complex projects.

The preparatory phase

INSTRUCT has just entered the second half of its preparatory phase, which began in April 2008. INSTRUCT will establish a Pan-European infrastructure based on core centres of scientific excellence, reinforced by associate centres with complementary expertise. Each core centre will provide access to cutting-edge equipment such as platforms for protein production, crystallization, mass spectrometry, biophysical methods, NMR, light microscopy and computational methods. Key activities during the preparatory phase have been to develop mechanisms for establishing and maintaining this infrastructure. National user groups have also been established and have met to discuss and compile the requirements of their nations researchers. This input is of great importance in making decisions about the structure of INSTRUCT, especially in terms of establishing guidelines for access and considering national and associate centres. In several countries, unprecedented numbers of scientists have worked together to generate a vital flow of information and ideas.

Identifying bottlenecks

Within the preparatory phase, several of the scientific working groups have been working on feasibility studies to identify bottlenecks in production and to increase efficiency where possible. Several studies have already produced results including workpackage 14.7 which is working on the co-expression of baculovirus systems. In a paper recently published online in Nature methods [1], Acembl, a versatile and automatable system for protein-complex expression in Escherichia coli is presented with the aim of aiding multiprotein complex production for structural and function studies. Other studies include the development of a NMR cryo probe, development in electron microscopy and the automation of eukaryotic expression systems with the development of small volume concentrators. With these techni-

cal developments, the resolution level of structural biology research can be further refined.

The coordination of data management is already an important issue and will continue to be so into the future. With several labs across Europe producing and gathering data, a concise and consistent storage and management of the data will be increasingly vital. Following an extensive feasibility study, researchers have identified PiMS as the management storage of choice and are developing it to make it as user friendly as possible for use across the INSTRUCT community. This will include an online version for labs without the financial means to buy the software, allowing increased data sharing across Europe. Several successful training and dissemination events have already taken place, and the number of participating and interested organizations is rising.

Ties with industry

Another important part of the INSTRUCT vision has been the increased contact and communication with industry. Forging close ties with industry will promote the commercialization of innovative technologies developed within INSTRUCT, and strengthen European industrial competitiveness, in particular Biopharma and Agrochem companies.

Ensuring a smooth transition

To permit a smooth transition to the construction and operational phases, a framework for governance and a mechanism for defining participants has to be clarified. Furthermore, a financial and legal framework is being built to allow major funding bodies to work together to provide a coherent infrastructure with European-wide access. At the time of going to press, the number of affiliated countries is 20, but interest is growing and as the success of the recent annual meeting in Florence shows, researchers from across Europe see and appreciate the need for an integrated infrastructure of structural biology to take biomedical research in Europe to the next level.

References

1. Bieniossek C, Nie Y, Frey D, Olieric N, Schaffitzel C, Collinson I, Romier C, Berger P, Richmond TJ, Steinmetz MO, Berger I (2009) Automated unrestricted multigene recombineering for multiprotein complex production. Nat Methods May 3.

BIOmics Hands-On Workshop

August 30 to September 4, 2009
Weizmann Institute of Science
Rehovot, Israel

http://ispc.weizmann.ac.il/biomics/

The International BIOmics Training & Education Center in Bioinformatics, Proteomics & Functional genomics (BIOmics) announces its first Hands-On Workshop on August 30 to September 4, 2009 at the Weizmann Institute of Science, Rehovot, ISRAEL.

The Workshop is intended for graduate students, postdocs and researchers from academia and industry.

At the Workshop, participants will hear talks from Israeli and international experts and gain first hand experience with:

- high throughput technologies,
- next-generation sequencing,
- viewing molecules in 3D,
- predicting protein–ligand interactions,
- tools for gene ontology;
- approaches for studying whole-genome genetic variation.

Hands-on sessions are limited to 20 participants.

Fellowships are available for attendees from academia, courtesy of the Israel Commission for UNESCO. With your application, let us know whether you will need a fellowship.

The registration deadline is **July 31, 2009**

Contact information
Workshop Secretariat:
Bracha Vaknin, ispc@weizmann.ac.il
Dept. of Structural Biology
Weizmann Institute of Science

EMBnet supported workshops

Theme: SYNAPSES-BRIDGING THE GAP BETWEEN BIOLOGISTS AND BIOINFORMATICIANS

6 - 11 July, 2009

CST Conference Room, College of Science & Technology, Covenant University, Ota, Nigeria

http://www.covenantuniversity.com/bioinformaticsIWPDB/iwpdb2.html

COURSE AREAS

1. Bioinformatics for Biologists
2. An Insight to practical Bioinformatics
3. Homology Modelling
4. Biological interpretation of Bioinfornatics results
5. In silico Drug Target Identification with experimental verification
6. Understanding Drug Resistance Mechanism
7. Genetic engineering: Gene Manipulation
8. Software Demonstration

Contacts:

Dr. Ezekiel Adebiyi

Chairman, Workshop Local Organizing Committee
Covenant University, Ota, Nigeria

email: e.adebiyi@daad-alumni.de

Ms. Ijeoma Dike

Secretary, Workshop Local Organizing Committee
Covenant University, Ota, Nigeria

email:ejdike@gmail.com

The Institute of Biochemistry, Molecular Biology and Biotechnology of the University of Colombo is pleased to announce the:

International Conference on Frontiers in Molecular Life Sciences:

Sailing towards New Horizons

http://www.ibmbb.lk/conference2009/index.htm

This will be the first of its kind in Sri Lanka and it expects to attract a wide range of researchers and students working in diverse fields using tools of Molecular Life Sciences.

The main conference will be followed by a Post Conference workshop (19th, 20th and 21st September 2009) on Bioinformatics co-sponsored by the European Molecular Biology Network (EMBnet).

Tentative Conference programme:
http://www.ibmbb.lk/conference2009/tent_pro.htmProgramme

Tentative Workshop Programme:
http://www.ibmbb.lk/conference2009/tent_pro_ws.htm

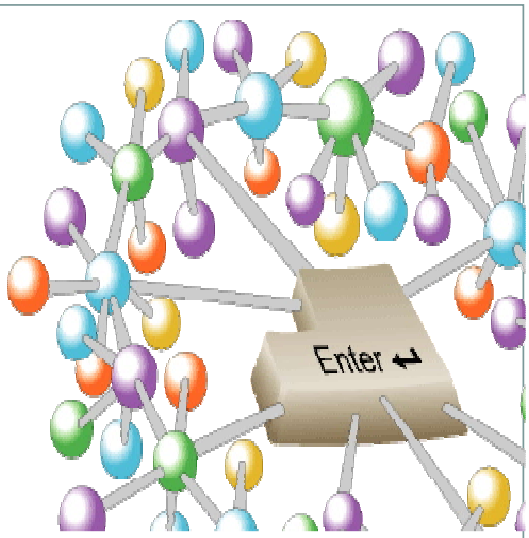
Plenary speakers:
Dr. Marie Allen, University of Uppsala, Sweden
Dr. Erik Bongcam-Rudloff, Linnaeus Centre for Bioinformatics in Uppsala, Sweden
Prof. Ludwig Gortner, University of the Saarland, Homburg/Saar, Germany
Prof. Eric Karunanayake, University of Colombo, SRI Lanka

Deadlines:

Early Bird Registration:	30th June 2009
Registration:	31st July 2009
Accommodation reservation:	31st July 2009

For further details contact us at:
Institute of Biochemistry, Molecular Biology and Biotechnology
University of Colombo, Sri Lanka
Phone: +94 11 2552528 / 34
Fax: +94 11 2552529 / 2553683
e-mail: conference2009@ibmbb.cmb.ac.lk

Joint International *EMBnet-RIBio* Conference 2009



BIOINFORMATICS FOR HIGH THROUGHPUT TECHNOLOGIES AND SYSTEMS BIOLOGY

OCTOBER 25 - 30, 2009
RIVIERA MAYA, CANCUN, MEXICO

CCG Centro de Ciencias Genómicas

Center for Genomic Sciences
Cuernavaca, Mexico
http://www.ccg.unam.mx

Chair:
Julio Collado-Vides

Co-chair:
Cesar Bonavides-Martinez

EMBNETMEXICO

http://embnet.ccg.unam.mx

Teaching the ABCs of Bioinformatics

Jingchu Luo
Centre of Bioinformatics, Peking University, EMBnet China node, China

It has been ten years since we started the semester course Computer Application to Molecular Biology in 2000. We now use Applied Bioinformatics Course, or ABC as a simple acronym for this course. Apparently, as its name indicates, ABC is an entry level introductory course, rather than an advanced one. We run the course in a training room (Fig. 1). Each student has a PC with both Linux and Windows installed.

The course is designed for biological graduate students to solve practical problems. After a brief introduction to the three most popular bioinformatics resources, the National Center for Biotechnology Information (NCBI), the European Bioinformatics Institute (EBI) and the Expert of Protein Analysis System (ExPASy), the students are happy to start with hands-on practice, to retrieve the alpha subunit of human, mouse and rat hemoglobin from Swiss-Prot, then make comparison of the amino acid sequences between each other [Swiss-Prot: HBA _ HUMAN, HBA _ MOUSE, HBA _ RAT]. Couple of online web servers can be found from the links at ExPASy, such as the built-in programs in EBI SRS, and the NCBI tool box. We use the bioinformatics platform WebLab[1] we developed locally.

Figure 1. The students are doing hands-on practice in the training room of the ABC course.

Surprisingly, the output of the Needle global alignment is out of expect, the sequence identity between mouse and rat is less than that of mouse and human. This triggers the interests of the students with enthusiastic discussion. Finally, they find the answer to this question by retrieving and comparing the nucleotide coding sequence (CDS) of the hemoglobin alpha gene of these three mammals. Indeed, the identity of CDS of mouse and rat is higher than that of mouse and human, which tells us that mouse and rat are close relatives based on the analysis of molecular biology data.

A dozen of simple exercises like this are designed. The students can get familiar with various bioinformatics tools by doing these exercises. For example, by comparing the 39kb genomic sequence of a Fugu cosmisd [GenBank: AF164138] to itself using dottup, the dot plot program embedded in EMBOSS, we can easily spot out the tandem duplication of the multi-drug resistance gene in this cosmid. By comparing the predicted results of the trans-membrane helices of a post-floral protein [Swiss-Prot: Q9FY06] identified from a cDNA library of short-day grown G2 pea tissue, using several tools including TMAP, THHMM, TMHMM, TMPred, the students are well aware that different tools may have different output, a common scenario of doing dry lab experiments.

Running BLAST seems an easy job for most of the students. However, running a Good BLAST is not that easy. Literature search reveals that neuroglobin [Swiss-Prot: NGB _ HUMAN] is a member of the human globin family to which nine hemoglobin subunits (alpha, beta, gamma, etc.) as well as myoglobin and cytoglobin belong. Take this as an example for BLAST search, we then ask, "can we find neuroglobin through BLAST search using the alpha subunit [Swiss-Prot: HBA _ HUMAN] as a query"? The answer is "No" if we use the default parameters to run BlastP through the NCBI BLAST server. Nevertheless, we can obtain a good match using PSI-BLAST, by choosing Swiss-Prot as the preferred database, selecting Human as the organism to search, and setting E-value to 0.001. The 12 hits obtained by the above search are ready to retrieve for further analysis for this human globing family, such as making multiple

1 http://weblab.cbi.pku.edu.cn/

sequence alignment, drawing a nice sequence logo, building up a phylogeny tree. By doing so, the students are convinced that it is critical to know the general principles and biological background behind the programs such as BLAST.

Several projects are implemented throughout the course. One of them is the analysis of the sequence, structure and function of the bar-headed goose hemoglobin. As we all know, hemoglobin is one of the most well studied proteins in the last century. More than 700 entries can be found in the Swiss-Prot database. Three-dimensional structures of wild type and mutants from dozens of species have been solved. This provides us a good opportunity to study the relationship among sequence, structure and function of hemoglobin molecules of human and other species. Bar-headed goose is a migration bird. They live in the Qinghai lake during summer time and fly to India all way along over the Tibetan plateau in autumn and come back again in spring. Interestingly, close relative of bar-headed goose, the graylag goose lives in the low land all year around. Sequence alignment of bar-headed goose hemoglobin [PDB: 1A4F] with that of graylag goose [PDB: 1FAW] shows only 4 substitutions. One of them is Pro 119 in the alpha subunit of graylag goose and Ala 119 of bar-headed goose. This residue is located in the surface of the alpha/beta interface. In 1983, Perutz proposed that this substitution reduces the contact between the alpha and beta subunit and increases the oxygen affinity, due to the relation of the tension status in the deoxy form. During the past decade, a research group at Peking University has solved the crystal structure of both deoxy and oxy form of the bar-headed goose, as well as the oxy form of the graylag goose hemoglobin. Using the powerful free software Swiss-PdbViewer[2], the students make a Magic Fit to superpose the alpha/beta heterodimer of the two goose hemoglobins on each other. They are excited to see the conformation difference caused by one amino acid residue substitution Pro119Ala between the two goose hemoglobin molecules (Fig. 2) by measuring the distance of the side chain atoms between Ala/Pro 119 and Leu 55 in the beta subunit. Furthermore, they make a mutation by replacing Leu 55 in the beta subunit with Ala 55 and propose that this mutant may further increase the oxygen affinity.

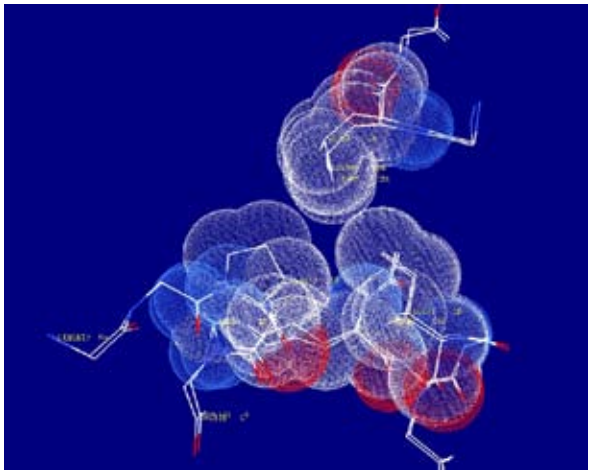


Figure 2. The superposed structure of bar-headed goose [PDB: 1A4F] and graylag goose [PDB: 1FAW] hemoglobin to show the interface of alpha and beta subunit. Pro 119 of alpha subunit (1FAW) has closer contact with Leu 55 of beta subunit than that of Ala 119 (1A4F).

In addition to the pre-defined projects, the students are also encouraged to bring their own projects. They are divided into groups, four students each, and discuss the project outside the class as homework assignments, work together to solve the problems. At the end of the course, a workshop is organized. Speakers chosen by group members make presentation on behalf of each group to summarize what they have learned during the course, and what they plan to continue to learn in the future.

Indeed, the ABC course is aimed to teach ABCs of bioinformatics. We hope that, by learning the course, the students will be well convinced that *"half day on the web, saves you half month in the lab!"*

2 http://spdbv.vital-it.ch/

ELIXIR Project meetings

Pedro Fernandes
Instituto Gulbenkian. Oeiras, Portugal

Another round of meetings of the ELIXIR project, now in the last stretch of the preparatory phase, prior to the presentation of the project proposal to the EU, took place in Copenhagen, on May 18th - 21st 2009.

The meetings, this time organized by Søren Brunak of DTU and his team, were held at the Panum Institute, which hosts the Faculty of Health Sciences of the University of Copenhagen. More than 80 participants attended the meetings.

The main session was the Stakeholders Meeting where the reports of the several workpackages were presented by the respective committee heads. The results of the feasibility studies have also been presented by their respective leaders. The meeting was started and closed by Janet Thornton, the Director of the EBI and coordinator of the ELIXIR Project.

The spirit of mission accomplished was in the air. The material collected allows for the assembly of a very sound and solid proposal, encompassing many aspects of the task of planning an European Infrastructure for Bioinformatics. On the other hand, it was quite evidently shown that the dimension and complexity of the problem have been addressed appropriately. The final proposal is now in preparation.

Several EMBnet members serve in the ELIXIR Workpackage Committees and were involved in the discussions and preparation of reports.

Present in this meeting: Terry Attwood, Pedro Fernandes, Lubos Klucar and Endre Barta.

Current Trends in Bioinformatics:

A Seminar held in Islamabad

Nazim Rahman
Department of Biosciences, COMSATS Institute of Information Technology, Islamabad, Pakistan

The EMBnet Pakistan National Node and the faculty of the Department of Biosciences at the COMSATS Institute of Information Technology (CIIT) together organized a seminar on, "Current Trends in Bioinformatics" on April 4, 2009. Dr. Fauzia Yusuf outlined the accomplishments of the Department of Biosciences as a pioneer in the field of Bioinformatics education and research in Pakistan. Dr. Shahid Nadeem Chohan made a presentation on the Bioinformatics activities in the CIIT emphasizing on the services provided by the EMBnet Pakistan National Node. Guest speakers were invited from COMSATS Institute of Information Technology (CIIT), National Center for Virology and Immunology (NCVI), and National Institute of Biotechnology and Genetic Engineering (NIBGE).

Dr. Syed Habib Bokhari addressing the audience.

Malik Nadeem Akhtar presenting his new Gene Finder Tool.

Bioinfomatics students Zeeshan Fazal and Abbas Bukhari discussing their poster.

Speakers presented their work with the following themes:

- Structural Bioinformatics and Drug Design
- Computational Biology
- Epidemics and Diseases
-

The objectives of the seminar were to:

- Raise awareness about EMBnet and its services
- Invite life scientists to take advantage of the EMBnet Pakistan National Node
- Provide a platform for life science researchers to present bioinformatics tools and techniques applied in their research
- Introduce current trends and developments in bioinformatics

This was a free open access event attended by a large audience comprising of students, researchers and academicians from various universities and research centers both locally and nationally. The audience specially liked the poster session held after the seminar which provided a forum for detailed interaction among the audience and the authors. The event was followed by an award ceremony for the best presentation and the best poster of the event.

Dr. Shahid Chohan, Dr. Raheel Qamar, Dr. Shahzad Muft.

Seminar participants.

Introductory Course in Bioinformatics in Kenya

Etienne de Villiers
EMBnet BecA-ILRI, Nairobi, Kenya

The BecA-ILRI Hub, EMBnet specialist node in Kenya hosted the third "Introductory course in Bioinformatics" between 4-9th May 2009 at ILRI, Nairobi, Kenya. The course team consisted of: Dr. Etienne de Villiers (BecA- ILRI Hub); Dr. Erik Bongcam-Rudloff, Dr. Hans-Henrik Fuxelius and Katharina Truve from Swedish University of Agricultural Sciences (SLU), Uppsala University (UU) and Linnaeus Centre for Bioinformatics (LCB); and George Githinji (KEMRI-Wellcome Trust program).

In total 25 applicants from Kenya, Tanzania, Uganda and South Africa were selected from 152 applications. Eight travel fellowships were awarded to participants from Tanzania and Uganda . These fellowships were made possible with generous funding from Prof. Anna Tramontano from University of Rome "La Sapienza", Italy and the BecA-ILRI Hub. The main funding was from the Swedish International Development Cooperation Agency (SIDA) (www.sida.se), Sweden. An additional 22 students from the University of Khartoum, Faculty of Animal Production in Sudan participated online in the course using the EMBnet Internet video conferencing system.

The response from the "online" participants was positive and we will use this medium in future courses. All students received an eBioUSB, a Bioinformatics workbench on a USB memory stick, which contains all the basic bioinformatics tools taught during the course and that enable students to save their data on the same device. The course was favorably evaluated, with 4.4 out of maximum 5 points.

Lots of positive comments were received about the course. Some of these comments were: "a very powerful tool", "We highly appreciate it", "Good course and teachers".

Course participants and teachers.

Constructing and working with protein interaction graphs




Charalampos Moschopoulos[1, 2], Seferina Mavroudi[1], Spiridon Likothanassis[1] and Sophia Kossida[2]

[1] Department of Computer Engineering & Informatics, University of Patras, Rio, Patras, Greece
[2] Bioinformatics & Medical Informatics Team, Biomedical Research Foundation of the Academy of Athens, Athens, Greece

Abstract

Nowadays, the detection of protein – protein interactions is performed mainly from experimental methods that are able to record thousands of them in a single experiment. In this contribution, we present methods that are used to assess the reliability of protein -protein interactions and construct reliable protein interaction graphs. Furthermore, we present the most popular algorithms that are used to detect protein complexes or discover the functionality of unknown proteins through clustering protein interaction graphs. Finally, the most popular visualization software tools are described.

Introduction

Recent development of high-throughput methods produced enormous datasets of protein – protein interaction (PPI) data. Techniques such as yeast two – hybrid and mass spectrometry detect PPIs and give an insight of the cellular organization of an organism. However, these methods suffer from high error rate as they miss an important fraction of protein interactions and yield several protein interactions that do not exist in reality.

Due to the large number of interactions, there is a great need of computational methods and models that would make it easy to extract valuable information from them. Usually, through protein interaction data, information derives about functional modules such as protein complexes (which reveal insights into both the topological properties and functional organization of protein networks) as well as the function of uncategorized proteins. A very efficient way of summarizing these new datasets is by forming protein interaction graphs. These graphs provide a valuable tool that helps the better understanding of the functional organization of the proteome. A graph is represented as $G = (V,E)$, where V is the set of the graph vertices and E is the set of the graph edges. In a protein interaction graph, the vertices represent the proteins and the edges the pairwise interactions between two proteins. Unfortunately, because of the unreliability of protein interaction data, algorithmic methods applied on protein interaction graphs can not produce results of high information value.

In this contribution, we present the most popular methods to assess the reliability of protein-protein interactions. Furthermore, we describe the features of a protein interaction graph and the computational methods which are used to acquire valuable conclusions from them. Additionally, this manuscript presents the most popular open source software tools which visualize PPI data.

Assessing the reliability of protein – protein interactions

Although data sets on the protein interactome, obtained by high-throughput protein interaction assays, are being accumulated rapidly, they usually come at the expense of relatively low quality, containing a high rate of spurious (false positives)

and missing (false negatives) protein-protein interactions [1].

To address the problem of false positives, different confidence scores have been assigned that reflect the reliability and biological significance of each protein interaction pair derived from the experiments [2]. Confidence scores are often computed as single indices, correlating interaction pairs derived from direct experiments (e.g., two-hybrid screens and mass spectrometry) with either indirect biological data sources (e.g., gene expression, protein - DNA binding, biological function, biological process, protein localization, protein class), or sequence based data sources (domain information, gene fusion, etc.). More recently, they are derived from supervised learning methods, which are employed to integrate direct and indirect biological data sources for the prediction task. The training data sets for these methods include known true positives and true negative interactions. For both strategies different approaches have been proposed, where the data sources varied along with the implementations.

Specifically, indices have been based on the sharing of a common cellular localization or a common cellular role [3, 4]. Alternatively, ranking of the reliability of protein interactions have been based on the reproducibility and non-randomness of the observation of an interaction [5-7]. Related to the ideas of functional homogeneity, localization coherence and observational reproducibility are a large number of other approaches based on the use of additional information, such as protein annotation, or the use of information from multiple assays [8-12]. Interaction network topology is a different mean of identifying reliability of interactions relying solely on the topology of the neighborhood of an interacting pair of proteins in the interactome [13, 14].

Bayes classifiers [15] and Bayesian Networks that combine multiple data sources are among the promising machine learning schemes that have been employed to predict true and false protein-protein interactions. Nevertheless, in [16] article, a Bayes classifier has been compared to Random Forest (RF) and Logistic Regression (LR), showing the RF classifier to have the best performance among them. In a more extended comparison including a Random Forest (RF), a RF similarity based k-Nearest-Neighbor classifier, Na□ve Bayes, Decision Tree, Logistic Regression and Support Vector Machine, the superior performance of RF was confirmed along with a satisfying performance of the Suppert Vector Machines (SVMs) [17]. Alternatively, a variant kernel canonical correlation analysis, has been used for predicting pathway protein interactions [18], while in [19] a sum of likelihood ratio scores strategy was explored to predict human PPI confidence.

Even though most of the above approaches are hindered or limited due inherent difficulties (eg., not all model organisms have well annotated genomes, expression of interacting proteins may need not to be correlated over many conditions and conversely protein pairs with correlation patterns do not necessarily physically interact, the number of proteins having known paralogs is limited as well as the number of available structures, etc), nevertheless, most studies suggest that utilizing any of the confidence assignment schemes is always more beneficial than assuming all observed interactions to be true or equally likely.

The problem of false negatives is essentially related to the problem of the ab initio prediction of protein-protein interactions by computational methods. Well known methods rely on gene fusion events [20-22], interacting domains [23, 24], interacting motifs [25-27], co-evolution of proteins or residues [28-30] and the topology of protein–protein interaction networks [31, 32]. Alternatively association rules have been explored [33]. In a different approach some of the confidence scores initially designed for the reliability assignment of observed interactions, are used for the assigning probability scores to putative unobserved interactions pairs.

Protein interaction Graphs

A protein interaction graph can be weighted or unweighted. In a weighted one, each edge connecting two proteins has been characterized by a number that represents the validity of the connection between these two proteins. In an unweighted protein interaction graph, an assumption is made that this number is equal to 1 for all the edges of the graph.

Generally, the protein interaction graphs are undirected and unweighted. Some properties have been identified to be common between the protein interaction graphs of all the organisms. First of all, they are all scale free. Moreover,

it is proved that similar proteins usually interact with each other and that they lie within short distance in the interaction graph. Finally, there are few vertices having many interactions and many that have few interactions. This means that if some proteins are eliminated, the topology of the protein interaction graph does not change which subsequently confirms the robustness of the organisms as they can afford to loose some proteins without jeopardizing the existence or even the normal function of the network.

In protein interaction graphs, the dense subgraphs are valuable since they provide details concerning the functionality of the proteins within the subgraph and the consistency of protein complexes. Given the mathematical representation of a graph, algorithms derived from the graph theory are well suited in order to isolate these dense areas.

The amount of data that has been derived from high-throughput approaches, automated text mining techniques, and/or manually from the scientific literature, has been stored in databases called protein-protein interaction databases. These databases are valuable resources for the researchers, where from they can easily retrieve and analyze the stored data [34]. Usually these databases include data of protein interactions obtained from many organisms. The most popular ones are BIND [35], MIPS [36], UniProt [37], IntAct [38].

It must be noted that there is a significant difference in the total number of protein-protein interactions among the various protein-protein interaction databases [39], due to the fact that data for each database were derived using different methods. Apart from the databases, where data obtained from experimental methods are stored, there are some other databases, where protein interactions predicted by computational methods are stored. The most significant one is called STRING database which has integrated known and predicted interactions from a variety of sources as well [40].

Extracting information from protein interaction graphs

Protein interaction graphs are used mainly to detect protein complexes in which individual proteins assemble into functional modules [41] or elucidate the function of uncharacterized proteins.

Various algorithms have been applied for the identification of protein complexes through protein – protein interaction networks. They can be divided in two big categories: those using a local search strategy and those using a hierarchical one. In the first category, the first introduced algorithm in the field was the Molecular Complex Detection (Mcode) [42]. A year before the appearance of Mcode, Enright et al. had introduced an algorithm called TRIBE-MCL [43] based on the Markov cluster algorithm (MCL), a previously developed algorithm for graph clustering. Besides that, King et al. suggested the RNSC algorithm [44] a cost-based local search algorithm. These two algorithms separate the whole protein interaction graph into clusters that represent protein families. This means that not even a single protein is discarded from the final results and several clusters can not be considered as protein complexes. Nevertheless, the RNSC algorithm uses a filtering strategy to achieve the identification of protein complex candidates. Another algorithm of the local search approach is the Local Clique Merging Algorithm (LCMA) [45] which locates local cliques in an interaction graph and subsequently tries to expand them.

On the other hand, all the hierarchical clustering algorithms are based on the concept of dividing the initial graph by removing the minimum set of edges. The Highly Connected Subgraph method (HCS) [46] separates a graph into several subgraphs using minimum cuts and stops when the cut is bigger or equal to the number of the graph vertices divided by 2. Koyutürk suggested the SIdeS algorithm [47] which uses the HCS algorithm philosophy; however the stopping criterion is based on the statistical significance of the derived subgraphs. More specifically, the SideS algorithm uses a framework for analyzing the occurrence of dense patterns in randomly generated graph-structured data with a view to assessing the significance of a pattern based on the statistical relationship between subgraph density and size.

The algorithms applied for the identification of protein complexes can be used to functionally annotate proteins. As it is presented in [48], identifying protein modules helps annotating uncharacterized proteins using the function shared by the majority of the module's proteins. However, these methods are outperformed by more direct "methods" which infer the function of a protein

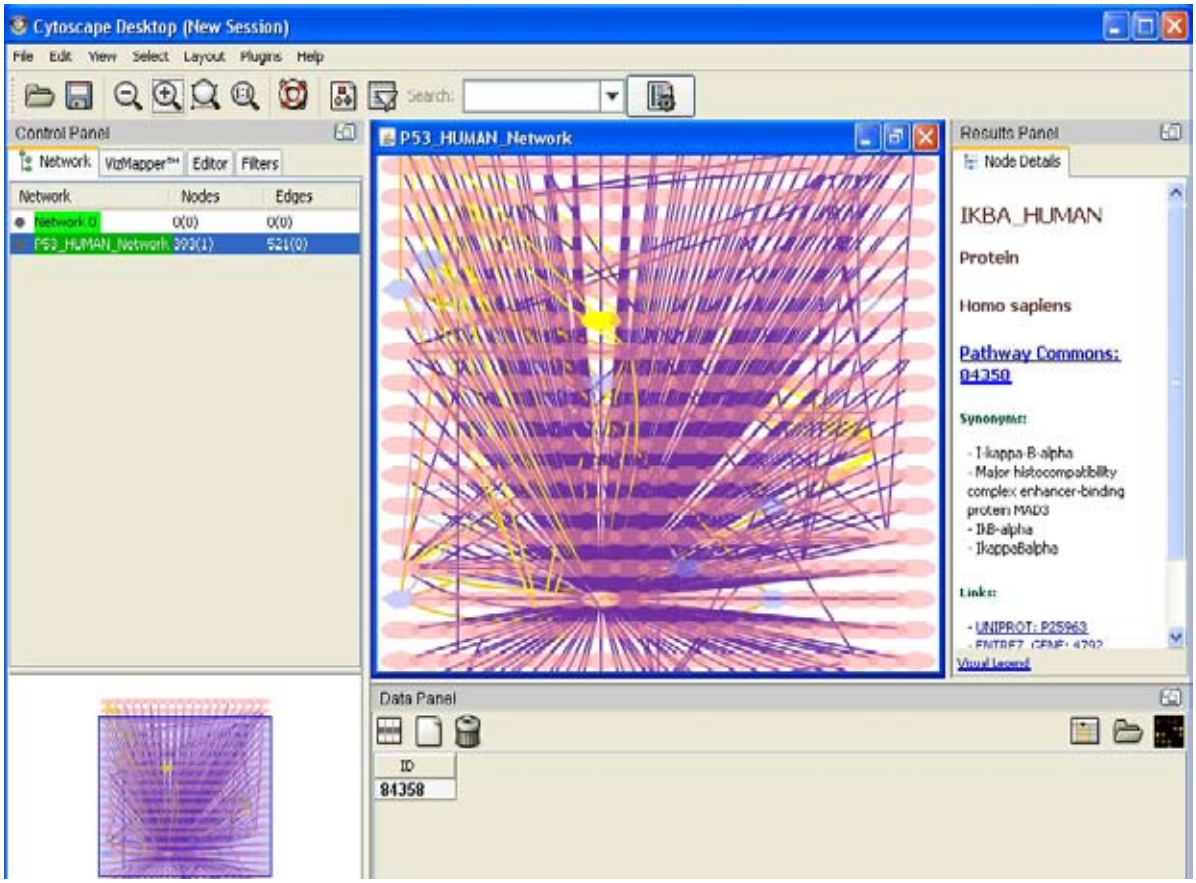


Figure 1. Snapshot of Cytoscape tool.

based on its connections in the network. Based on the principle that proteins that lie closer to one another in the protein – protein interaction graph are more likely to have similar function, they are simpler and more effective from the clustering approaches. They can be divided in those using the neighborhood of a protein [6], the approaches which are graph theoretic [49], the probabilistic ones [50] and those that integrate multiple data sources [51].

Tools for PPI graphs visualization

Availability of large scale experimental data and numerous approaches of extracting information from PPI graphs enable the development of many software tools. The visualization of the vast volume of PPI data allows the observation of the whole proteome of an organism [52].

Among those tools being freely available for academic use, the most popular visualization tool is Cytoscape [53] in which a user can construct his own graph or import PPI graphs from online databases. Additionally, Cytoscape includes a flexible plugin architecture that enables developers to add extra functionality beyond that provided in the core. Another visualization tool is Medusa [54] which is based on the Fruchterman – Reingold algorithm [55]. However, it is less suited for the visualization of big datasets and its own text file format is not compatible with other visualization tools. 2D and 3D representations are offered by BioLayout Express 3D tool [56]. This tool is highly interactive and in the latest version, the MCL algorithm is hosted in this tool. Other visualization tools are VisANT [57] and PIVOT [58] which are best suited for visualizing protein –protein interactions and identifying relationships between them.

Future perspectives and conlusions

Graph – based model can exploit global and local characteristics of biology and more particularly PPI graphs. Various algorithmic methods and tools have been developed in order to extract information using graph theoretic approaches.

Although the above methods of assessing the protein interactions reliability are useful and some of them exhibit encouraging results, there is still room for improvement. None of the existing methods gain both a high specificity and a good sensitivity at the same time. Data integration usually improves the results. However, different biological sources represent different and apparently biased subsets of the true interactions and simply taking the union may lead to poor performance, while taking the intersection may result in a minimal overlap. New methods, able to cope with partial domain knowledge would be desirable. Supplementary to data integration, model integration could further enhance performance. Besides of accuracy issues, newly designed methods should allow for the interpretability of the results. There are different and often contradicting opinions regarding the biological evidence that should be taken into account for the computation and the evaluation of the reliability of protein-protein interactions. Researchers should be given the means to judge each feature's contribution and to extract new explainable knowledge.

Another future aspect would be the use of heterogeneous source of data to construct weighted graphs. This way, the above mentioned methods can offer better quality of information, while the results of PPI graph analysis would suffer by less error rate. Furthermore, working in the same direction, the variety of information that derives from PPI graph analysis could be retrieved by using web services tools. Web services enable programmers to build complex applications without the need to install and maintain the databases and analysis tools and without having to take on the financial overheads that accompany these. Moreover, Web services provide easier integration and interoperability among applications and the data they require. Finally, it would be interesting to apply these methods in trancsriptomics or metabolomics, sections of research that are still in their infancy.

References

1 von Mering C, Krause R, Snel B, Cornell M, Oliver S G, Fields S,Bork P (2002) Comparative assessment of large-scale data sets of protein-protein interactions. Nature 417: (6887) 399-403.

2 Yu J,Finley R L, Jr. (2009) Combining multiple positive training sets to generate confidence scores for protein-protein interactions. Bioinformatics 25: (1) 105-11.

3 Sprinzak E, Sattath S,Margalit H (2003) How reliable are experimental protein-protein interaction data? J Mol Biol 327: (5) 919-23.

4 Chen J, Chua H N, Hsu W, Lee M L, Ng S K, Saito R, Sung W K,Wong L (2006) Increasing confidence of protein-protein interactomes. Genome Inform 17: (2) 284-97.

5 Nabieva E, Jim K, Agarwal A, Chazelle B,Singh M (2005) Whole-proteome prediction of protein function via graph-theoretic analysis of interaction maps. Bioinformatics 21 Suppl 1: i302-10.

6 Chua H N, Sung W K,Wong L (2006) Exploiting indirect neighbours and topological weight to predict protein function from protein-protein interactions. Bioinformatics 22: (13) 1623-30.

7 Hart G T, Lee I,Marcotte E R (2007) A high-accuracy consensus map of yeast protein complexes reveals modular nature of gene essentiality. BMC Bioinformatics 8: 236.

8 Bader J S, Chaudhuri A, Rothberg J M,Chant J (2004) Gaining confidence in high-throughput protein interaction networks. Nat Biotechnol 22: (1) 78-85.

9 Patil A,Nakamura H (2005) Filtering high-throughput protein-protein interaction data using a combination of genomic features. BMC Bioinformatics 6: 100.

10 Giot L, Bader J S, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao Y L, Ooi C E, Godwin B, Vitols E, Vijayadamodar G, Pochart P, Machineni H, Welsh M, Kong Y, Zerhusen B, Malcolm R, Varrone Z, Collis A, Minto M, Burgess S, McDaniel L, Stimpson E, Spriggs F, Williams J, Neurath K, Ioime N, Agee M, Voss E, Furtak K, Renzulli R, Aanensen N, Carrolla S, Bickelhaupt E, Lazovatsky Y, DaSilva A, Zhong J, Stanyon C A, Finley R L, Jr., White K P, Braverman M, Jarvie T, Gold S, Leach M, Knight J, Shimkets R A, McKenna M P, Chant J,Rothberg J M (2003) A protein interaction map of Drosophila melanogaster. Science 302: (5651) 1727-36.

11 Samanta M P,Liang S (2003) Predicting protein functions from redundancies in large-scale protein interaction networks. Proc Natl Acad Sci U S A 100: (22) 12579-83.

12 Martin S, Roe D,Faulon J L (2005) Predicting protein-protein interactions using signature products. Bioinformatics 21: (2) 218-26.

13 Saito R, Suzuki H,Hayashizaki Y (2003) Construction of reliable protein-protein interaction networks with a new interaction generality measure. Bioinformatics 19: (6) 756-63.

14 Chen J, Hsu W, Lee M L,Ng S K (2006) Increasing confidence of protein interactomes using network

topological metrics. Bioinformatics 22: (16) 1998-2004.

15 Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan N J, Chung S, Emili A, Snyder M, Greenblatt J F,Gerstein M (2003) A Bayesian networks approach for predicting protein-protein interactions from genomic data. Science 302: (5644) 449-53.

16 Lin N, Wu B, Jansen R, Gerstein M,Zhao H (2004) Information assessment on predicting protein-protein interactions. BMC Bioinformatics 5: 154.

17 Qi Y, Bar-Joseph Z,Klein-Seetharaman J (2006) Evaluation of different biological data and computational classification methods for use in protein interaction prediction. Proteins 63: (3) 490-500.

18 Yamanishi Y, Vert J P,Kanehisa M (2004) Protein network inference from multiple genomic data: a supervised approach. Bioinformatics 20 Suppl 1: i363-70.

19 Rhodes D R, Tomlins S A, Varambally S, Mahavisno V, Barrette T, Kalyana-Sundaram S, Ghosh D, Pandey A,Chinnaiyan A M (2005) Probabilistic model of the human protein-protein interaction network. Nat Biotechnol 23: (8) 951-9.

20 Marcotte E M, Pellegrini M, Ng H L, Rice D W, Yeates T O,Eisenberg D (1999) Detecting protein function and protein-protein interactions from genome sequences. Science 285: (5428) 751-3.

21 Enright A J, Iliopoulos I, Kyrpides N C,Ouzounis C A (1999) Protein interaction maps for complete genomes based on gene fusion events. Nature 402: (6757) 86-90.

22 Tsoka S,Ouzounis C A (2000) Prediction of protein interactions: metabolic enzymes are frequently involved in gene fusion. Nat Genet 26: (2) 141-2.

23 Sprinzak E,Margalit H (2001) Correlated sequence-signatures as markers of protein-protein interaction. J Mol Biol 311: (4) 681-92.

24 Han D S, Kim H S, Jang W H, Lee S D,Suh J K (2004) PreSPI: a domain combination based prediction system for protein-protein interaction. Nucleic Acids Res 32: (21) 6312-20.

25 Tong A H, Drees B, Nardelli G, Bader G D, Brannetti B, Castagnoli L, Evangelista M, Ferracuti S, Nelson B, Paoluzi S, Quondam M, Zucconi A, Hogue C W, Fields S, Boone C,Cesareni G (2002) A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. Science 295: (5553) 321-4.

26 Aytuna A S, Gursoy A,Keskin O (2005) Prediction of protein-protein interactions by combining structure and sequence conservation in protein interfaces. Bioinformatics 21: (12) 2850-5.

27 Li H, Li J,Wong L (2006) Discovering motif pairs at interaction sites from protein sequences on a proteome-wide scale. Bioinformatics 22: (8) 989-96.

28 Pazos F,Valencia A (2001) Similarity of phylogenetic trees as indicator of protein-protein interaction. Protein Eng 14: (9) 609-14.

29 Pazos F, Ranea J A, Juan D,Sternberg M J (2005) Assessing protein co-evolution in the context of the tree of life assists in the prediction of the interactome. J Mol Biol 352: (4) 1002-15.

30 Juan D, Pazos F,Valencia A (2008) High-confidence prediction of global interactomes based on genome-wide coevolutionary networks. Proc Natl Acad Sci U S A 105: (3) 934-9.

31 Yu H, Paccanaro A, Trifonov V,Gerstein M (2006) Predicting interactions in protein networks by completing defective cliques. Bioinformatics 22: (7) 823-9.

32 Pei P,Zhang A (2005) A topological measurement for weighted protein interaction network. Proc IEEE Comput Syst Bioinform Conf 268-78.

33 Kotlyar M,Jurisica I (2006) Predicting Protein-Protein Interactions by Association Mining. Information Systems Frontiers 8: (1) 37-47.

34 Suresh S, Sujatha Mohan S, Mishra G, Hanumanthu G R, Suresh M, Reddy R,Pandey A (2005) Proteomic resources: integrating biomedical information in humans. Gene 364: 13-8.

35 Bader G D, Betel D,Hogue C W (2003) BIND: the Biomolecular Interaction Network Database. Nucleic Acids Res 31: (1) 248-50.

36 Mewes H W, Frishman D, Guldener U, Mannhaupt G, Mayer K, Mokrejs M, Morgenstern B, Munsterkotter M, Rudd S,Weil B (2002) MIPS: a database for genomes and protein sequences. Nucleic Acids Res 30: (1) 31-4.

37 Consortium U (2009) The Universal Protein Resource (UniProt) 2009. Nucleic Acids Res 37: (Database issue) D169-74.

38 Kerrien S, Alam-Faruque Y, Aranda B, Bancarz I, Bridge A, Derow C, Dimmer E, Feuermann M, Friedrichsen A, Huntley R, Kohler C, Khadake J, Leroy C, Liban A, Lieftink C, Montecchi-Palazzi L, Orchard S, Risse J, Robbe K, Roechert B, Thorneycroft D, Zhang Y, Apweiler R,Hermjakob H (2007) IntAct--open source resource for molecular interaction data. Nucleic Acids Res 35: (Database issue) D561-5.

39 Mathivanan S, Periaswamy B, Gandhi T K, Kandasamy K, Suresh S, Mohmood R, Ramachandra Y L,Pandey A (2006) An evaluation of human protein-protein interaction data in the public domain. BMC Bioinformatics 7 Suppl 5: S19.

40 von Mering C, Jensen L J, Kuhn M, Chaffron S, Doerks T, Kruger B, Snel B,Bork P (2007) STRING 7--recent developments in the integration and prediction of protein interactions. Nucleic Acids Res 35: (Database issue) D358-62.

41 Hartwell L H, Hopfield J J, Leibler S,Murray A W (1999) From molecular to modular cell biology. Nature 402: (6761 Suppl) C47-52.

42 Bader G D,Hogue C W (2003) An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 4: 2.

43 Enright A J, Van Dongen S,Ouzounis C A (2002) An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res 30: (7) 1575-84.

44 King A D, Przulj N,Jurisica I (2004) Protein complex prediction via cost-based clustering. Bioinformatics 20: (17) 3013-20.

45 Li X-L, Tan S-H, Foo C-S,Ng S-K (2005) Interaction Graph Mining for Protein Complexes Using Local Clique Merging. Genome Informatics 16: (2) 260-269.

46 Hartuv E,Shamir R (2000) A clustering algorithm based on graph connectivity Information Processing Letters 76: (4-6) 175-181.

47 Koyuturk M, Szpankowski W,Grama A (2007) Assessing significance of connectivity and conservation in protein interaction networks. J Comput Biol 14: (6) 747-64.

48 Sharan R, Ulitsky I,Shamir R (2007) Network-based prediction of protein function. Mol Syst Biol 3: 88.

49 Vazquez A, Flammini A, Maritan A,Vespignani A (2003) Global protein function prediction from protein-protein interaction networks. Nat Biotechnol 21: (6) 697-700.

50 Letovsky S,Kasif S (2003) Predicting protein function from protein/protein interaction data: a probabilistic approach. Bioinformatics 19 Suppl 1: i197-204.

51 Tsuda K, Shin H,Scholkopf B (2005) Fast protein classification with multiple networks. Bioinformatics 21 Suppl 2: ii59-65.

52 Pavlopoulos G A G, Wegener A L A,Schneider R R (2008) A survey of visualization tools for biological network analysis. BioData Min 1: (1) 12.

53 Shannon P, Markiel A, Ozier O, Baliga N S, Wang J T, Ramage D, Amin N, Schwikowski B,Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13: (11) 2498-504.

54 Hooper S D,Bork P (2005) Medusa: a simple tool for interaction graph analysis. Bioinformatics 21: (24) 4432-3.

55 Fruchterman T, M. J.,Reingold E, M. (1991) Graph drawing by force-directed placement. Softw. Pract. Exper. 21: (11) 1129-1164.

56 Freeman T C, Goldovsky L, Brosch M, van Dongen S, Maziere P, Grocock R J, Freilich S, Thornton J,Enright A J (2007) Construction, visualisation, and clustering of transcription networks from microarray expression data. PLoS Comput Biol 3: (10) 2032-42.

57 Hu Z, Snitkin E S,DeLisi C (2008) VisANT: an integrative framework for networks in systems biology. Brief Bioinform 9: (4) 317-25.

58 Orlev N, Shamir R,Shiloh Y (2004) PIVOT: protein interacions visualizatiOn tool. Bioinformatics 20: (3) 424-5.

A question of length

Vivienne Baillie Gerritsen

When Charles Darwin accepted the invitation to accompany Captain Fitzroy on HMS Beagle as the ship's naturalist, little did he know that he would bring back with him material that was to haunt him – one way or another - until the end of his days. Amongst the many mineral, plant and animal specimens which were unloaded from the ship on its return in October 1836, there were a number of preserved finches which Darwin had found on the Galapagos Islands. It was the study of these finches, which later became known as 'Darwin's finches', that helped to forge the notion of the transmutation of species. In other words, any given species had the capacity to adapt, evolve and undergo transformations – and it turned out to be in the name of survival. With regards to finches, their beaks were different depending on the kind of diet they had. Charles Darwin had no idea how such changes could occur within a species. Today, we are getting closer and closer to understanding how it happens on the molecular level. And it seems that a protein known as calmodulin has a major role.

'Brown Beakface', by Kaitlin Beckett

Courtesy of the artist

As he set foot on *terra firma* after five years of sailing and as many of nausea, Darwin had no idea that fourteen of the many specimens of birds he brought back to England were in fact all finches. What is more, they seemed to be finches which bore many similarities to a type of finch found along the coast of South America. Darwin had identified them as different birds altogether but when he handed them over to the renowned ornithologist of the time – John Gould – it turned out that these fourteen birds were in fact representatives of twelve different species of finch. Until then, Darwin had believed that there were as many centres of creation as there were of species despite the fact that – within each centre – phenotypical change could occur. With Gould's findings and Darwin's knowledge of the geographical and ecological niches where he had found the birds, he shifted his theory: what if every species of finch on the Galapagos Islands had originated from the one same species on the South American coastline? It marked the very beginnings of his theory on the origin of species.

In those days, the description of specimens – whichever kingdom they belonged to – depended on a keen eye, a pencil and paper. Today, thanks to novel molecular methods, observation has been magnified by the millions – and scientists are able to see or imagine processes which are going on well beneath the level of feathers and petals. Finding links between a specific gene and the effect it has on an organism is now routine. In this way, scientists discovered that the protein calmodulin – from CALcium MODULated proteIN – has a direct role in the shape of a finch's beak. What is more, they discovered that calmodulin seemed to have an effect only on the length of the birds' beak and not its width, or depth – which are dependent on another gene. From an evolutionary point of view, this is not really surprising since it gives natural selection a form of plasticity. In other words, evolution is fine-tuned.

How can calmodulin affect the length of a finch's beak? It seems difficult to believe that one molecule could have such a massive effect on an organism's appearance. In fact, it doesn't. At least not directly. It happens to be at the very beginning of important molecular processes. Indeed, calmodulin has the power to trigger off a wide variety of biological pathways and, in turn, many activities such as muscle contraction, short-term and long-term memory, intracellular movement, inflammation, nerve growth and the immune response to name a few. It uses calcium ions, which are present in all kinds of tissues both inside the cell and outside it. Calmodulin is just one of the many molecules which use calcium ions to induce a reaction. Nevertheless, without it and calcium, a lot of what goes on inside us would go haywire.

At rest, calmodulin looks a little like a dumbbell. It is composed of two arms attached by a helix hinge. Each arm can hold up to two calcium ions. Once bound, the structural conformation of calmodulin is modified and ready to bind to specific target proteins which it does by wrapping its arms around it in a sort of molecular hug. What is more, depending on the amount of calcium ions bound – up to four – and the kind of post-translational modification calmodulin has undergone, the protein can bind to a great variety of targets ranging from kinases, phosphatases and phosphodiesterases to ion channels, cyclases and cytoskeleton receptors. In turn, each of these target proteins will trigger off cellular processes – from the regulation of metabolism and the cytoskeleton, to ion transport, protein folding and cell proliferation. With regards to the length of finch's beaks, researchers discovered that the long-beaked finches always express a higher level of calmodulin than the shorter and wider-beaked species. And when they upregulated the calmodulin gene in chicken, this had a direct effect on the length of their beaks!

Although a number of anti-calmodulin products had already been described in the 1980s, by the 1990s interest had faded. However, owing to the more recent discoveries of the involvement of calmodulin in so many different physiological processes, there has been a drastic increase in its interest, especially within the world of therapy and drug design. Some synthetic inhibitors are already used clinically as anti-cancer and anti-psychotic agents for example. But scientists have already described over one hundred natural inhibitors, the most potent of which are animal venoms. Such naturally-occurring compounds could be used to develop herbicides or to design drugs for neurodegenerative diseases for example. The future certainly seems bright for calmodulin. HMS Beagle took Darwin around the world; little did the founder of the theory of the origin of species know where his finches would take him.

Cross-references to Swiss-Prot

Calmodulin, *Homo sapiens* (Human) : P62158

References

1. Abzhanov A., Kuo W.P., Hartmann C., Grant B.R., Gran P.R.
The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches
Nature 442:563-567(2006)
PMID: 16885984

2. Patel N.H.
How to build a longer beak
Nature 442:515-516(2006)
PMID: 16885968

National Nodes

Argentina
IBBM, Facultad de Cs. Exactas, Universidad Nacional de La Plata

Australia
RMC Gunn Building B19, University of Sydney, Sydney

Belgium
BEN ULB Campus Plaine CP 257, Brussels

Brazil
Lab. Nacional de Computação Científica, Lab. de Bioinformática, Petrópolis, Rio de Janeiro

Chile
Centre for Biochemical Engineering and Biotechnology (CIByB). University of Chile, Santiago

China
Centre of Bioinformatics, Peking University, Beijing

Colombia
Instituto de Biotecnología, Universidad Nacional de Colombia, Edificio Manuel Ancizar, Bogota

Costa Rica
University of Costa Rica (UCR), School of Medicine, Department of Pharmacology and ClinicToxicology, San Jose

Cuba
Centro de Ingeniería Genética y Biotecnolgía, La Habana

Finland
CSC, Espoo

France
ReNaBi, French bioinformatics platforms network

Greece
Biomedical Research Foundation of the Academy of Athens, Athens

Hungary
Agricultural Biotechnology Center, Godollo

India
Centre for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad

Italy
CNR - Institute for Biomedical Technologies, Bioinformatics and Genomic Group, Bari

Mexico
Nodo Nacional EMBnet, Centro de Investigación sobre Fijación de Nitrógeno, Cuernavaca, Morelos

The Netherlands
Dept. of Genome Informatics, Wageningen UR

Norway
The Norwegian EMBnet Node, The Biotechnology Centre of Oslo

Pakistan
COMSATS Institute of Information Technology, Chak Shahzaad, Islamabad

Poland
Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warszawa

Portugal
Instituto Gulbenkian de Ciencia, Centro Portugues de Bioinformatica, Oeiras

Russia
Biocomputing Group, Belozersky Institute, Moscow

Slovakia
Institute of Molecular Biology, Slovak Academy of Science, Bratislava

South Africa
SANBI, University of the Western Cape, Bellville

Spain
EMBnet/CNB, Centro Nacional de Biotecnología, Madrid

Sri Lanka
Institute of Biochemistry, Molecular Biology and Biotechnology, University of Colombo, Colombo

Sweden
Uppsala Biomedical Centre, Computing Department, Uppsala

Switzerland
Swiss Institute of Bioinformatics, Lausanne

Specialist Nodes

CASPUR
Rome, Italy

EBI
EBI Embl Outstation, Hinxton, Cambridge, UK

ETI
Amsterdam, The Netherlands

ICGEB
International Centre for Genetic Engineering and Biotechnology, Trieste, Italy

IHCP
Institute of Health and Consumer Protection, Ispra. Italy

ILRI/BECA
International Livestock Research Institute, Nairobi, Kenya

MIPS
Muenchen, Germany

UMBER
School of Biological Sciences, The University of Manchester,, UK

for more information visit our Web site
www.embnet.org

EMBnet.news
ISSN 1023-4144

Dear reader,

If you have any comments or suggestions regarding this newsletter we would be very glad to hear from you. If you have a tip you feel we can print then please let us know. Please send your contributions to one of the editors. You may also submit material by e-mail.

Past issues of EMBnet.news are available as PostScript or PDF files. You can get them from the EMBnet organization Web site:
http://www.embnet.org/download/embnetnews

Publisher:
EMBnet Executive Board
c/o Erik Bongcam-Rudloff
Uppsala Biomedical Centre
The Linnaeus Centre for Bioinformatics, SLU/UU
Box 570 S-751 23 Uppsala, Sweden
Email: erik.bongcam@bmc.uu.se
Tel: +46-18-4716696

Submission deadline for the next issue:
August 20, 2009