Forum for European Structural Proteomics

Policy Recommendations for Structural Genomics and Structural Proteomics in Europe

Lucia Banci[1], Wolfgang Baumeister[2], Udo Heinemann[3], Gunter Schneider[4], Israel Silman[5] and Joel L. Sussman[6]

[1]Centro Risonanze Magnetiche, University of Florence, Via Luigi Sacconi 6, Sesto Fiorentino, Florence 50019, Italy

[2]Max Planck Institute of Biochemistry, Am Klopferspitz 18a, 82152 Martinsried, Germany

[3]Max Delbrück Center for Molecular Medicine, Robert-Rössle-Str. 10, 13125 Berlin, Germany

[4]Karolinska Institutet, Scheelevägen 2, Stockholm S-171 77, Sweden

[5]Department of Neurobiology, and [6]Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel

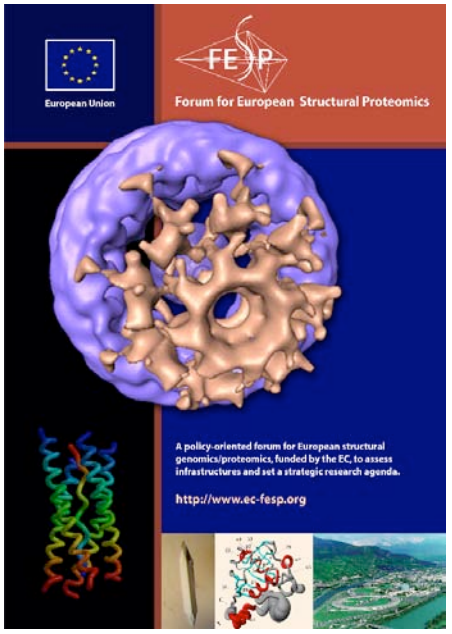

Table of Contents

A Roadmap for Strategic Future of European Directions in Structural Genomics and Structural Proteomics

Executive Summary

The Forum for European Structural Proteomics FESP was delegated to make recommendations to the EC concerning policy in the area of Structural Genomics/Structural Proteomics (SG/SP) over the next 5-10 years. The recommendations fall into two broad categories:

- Specific recommendations relating to infrastructure requirements
- General recommendations concerning research policy

The specific recommendations were largely reached on the basis of questionnaires complete by PIs involved in SG/SP projects and by PIs of individual structural biology (SB) labs, dealing with four topics, synchrotrons, NMR, EM and protein production. The general recommendations arose out of a series of workshops and roundtables organized as satellites of larger SG/SP meetings.

It is the feeling of the panel that many of the recommendations with respect to infrastructure could be implemented within the framework of the INSTRUCT research infrastructure for structural biology recently approved by the European Strategy Forum on Research Infrastructures (ESFRI), together with input from the ELIXIR research infrastructure for bioinformatics which was approved at the same time. But infrastructure is only one issue, however important. There are also requirements for user support and travel, for software and hardware development and for training. Even more important, there is a need for continued funding of integrated projects dedicated to SG/SP topics or to topics at the interface of SG/SP and SB with systems biology, cell biology and clinical research so as to maintain Europe at the 'cutting edge' vis-à-vis Japan and the USA, and to integrate the efforts of structural biologists with those of other biologists and biochemists.

We wish to stress that basic research in structural biology and related fields is vital for maintaining the scientific, technological and entrepreneurial leadership and competitiveness of Europe, not only in these areas but also in broader contexts. Indeed, developments in these basic areas would impact strongly on a broad spectrum of issues related to human quality of life. Accordingly, FESP strongly recommends that the EC should issue regular calls focused on **fundamental research in the life sciences.** This would provide a solid and essential knowledge base, largely at the molecular level, of the functional processes of life, knowledge that would provide the foundation for subsequent research of a more applied character.

Specific recommendations

- Mass spectroscopy – Even high-end equipment is not beyond the means of individual labs, but funding of a training element may be required.
- Bioinformatics – Although it is anticipated that many of the bioinformatics requirements of both the SG/SP consortia and of individual labs will be met by the ELIXIR infrastructure project of ESFRI, certain requirements will need be to directly addressed, including improved software for computational biology and homology modeling, and databases relating to types of information such as expression vectors, gene constructs and

production protocols, which may streamline the production pipeline and eliminate redundant efforts.

- Synchrotrons – It is felt that the number of beamlines already available or anticipated should be adequate for user requirements for the foreseeable future. However, it is recommended that EC support for user access and assistance be continued, and that there should be a substantial investment in automation and in technical advances that will allow the EC to maintain a competitive edge.
- NMR – The five existing EC trans-national BioNMR research infrastructures are deemed to be serving their function and to be adequate for foreseeable user requirements. However there is a need for more powerful instruments, for more solid-state machines and for development of improved hardware and software both for data acquisition and interpretation.
- EM – It is felt that the cost of high-end equipment is fast becoming beyond the means of nearly all individual laboratories. It is thus recommended that the EC should support the establishment and maintenance of trans-national centers very similar to the NMR centers. There is also a serious shortage of trained personnel both for maintenance and utilization. It is recommended that training programmes be established as joint ventures between academia and industry.
- Protein Production – This is a serious bottleneck for many researchers. Nevertheless, it is not recommended that trans-national centers be established, but rather that use is made of non-profit production centers such as exist at many of the SG/SP centers across Europe.

Policy recommendations

- Protein Target Selection – Some of the SG/SP projects worldwide have encountered criticism due to an emphasis on filling 'fold' or 'structure' space, despite their immense achievements in solving thousands of structures and developing powerful high throughput technologies. From the outset, Europe adopted a different approach, emphasizing selection of disease-associated targets, such as relevant human proteins or those of bacterial pathogens. This policy has resulted in the solution of a wealth of protein structures of high clinical importance. It is recommended that such an orientation in target selection be maintained. However, it can be still integrated with a "broad conformation-space coverage" approach as proteomic information becomes available for organisms that cannot presently be cultivated. It should be realized that only 1% of all microorganisms in the environment, as well as in the human body, can be grown. The proteome of these families of organisms may be expected to provide a variety of hitherto unknown conformations, folds and protein architectures, which might prompt us to revisit our knowledge of the principles of protein structure and folding. It is further recommended that the structural biologists involved, even in the framework of an SG/SP project, should be familiar with the underlying science, and liaise closely with the biologist(s) supplying the target
- Systems structural biology is an emergent discipline; two workshops organized by FESP and the EC brought together experts, including structural biologists, other biologists interested in higher order cellular organization, and policy makers, to discuss policy in this area. A consensus was reached that the trajectory of structural biology should be towards the study of higher order structures, eventually producing a 'molecular map' of

the entire cell. Progress in **molecular** systems biology, *i.e.* the ability to **model** systems to predict biological outcome at the molecular level, demands an understanding of the dynamic properties of proteins, the specificities of protein-protein interactions and the resultant properties of molecular machines, pathways and entire networks. We believe that a structural perspective is critical for systems biology to impact significantly on pharmacology and medicine. A structure-based platform for biological systems will provide a particularly effective means for validating and interpreting genetic variation as it relates to disease and will guide more informed and precise therapeutic interventions. Addressing these goals will, of necessity, be a multidisciplinary endeavour, bringing together experts in all experimental areas of structural biology, as well as computer modeling and mass spectroscopy. Obviously, a large investment in state-of-the-art technologies will be required, in all the technologies referred to above, as well as in light microscopy, in EM. Equally important will be effective integration of these resources in core centres of excellence, which may be facilitated within the INSTRUCT infrastructure programme.

- Bridging the Gap between Structural Biology and Other Branches of Biology – The various SG/SP endeavours, together with individual SB groups, are supplying large numbers of protein that provide a wealth of valuable information. There is, however, an acute translational problem in ensuring that biologists and clinicians can benefit from it. This is since most of them can neither interpret three-dimensional protein structures, nor take advantage of them to guide their research. This can only be partially overcome by direct collaboration with a trained structural biologist due to the large numbers of structures being generated. It is recommended that the EC should give high priority to training programmes to help non-structural biologists to overcome this obstacle. In parallel, it is recommended that support be given to developing graphic packages that will facilitate extraction of the relevant structural information. Promising efforts in this direction have been made by several groups, but expediting their development, integration and dissemination should be given a high priority.

What is FESP?

The Forum for European Structural Proteomics (FESP) (http://www.ec-fesp.org) is an initiative of structural biologists, supported by the European Commission (EC), under the **"Structuring the European Research Area" specific programme Research Infrastructures Action.** It was established at the beginning of 2006, with a mandate to recommend a policy for structural genomics/proteomics (SG/SP) in the broader context of anticipated developments in biological research. Members of the FESP project are:

- **Professor Lucia Banci**
  Centro Risonanze Magnetiche
  University of Florence, Italy
- **Professor Wolfgang Baumeister**
  Max-Planck-Institute of Biochemistry
  Martinsried, Germany
- **Professor Udo Heinemann**
  Max-Delbruck-Centrum fur Molekulare Medizin
  Berlin, Germany
- **Professor Gunter Schneider**
  Karolinska Institutet
  Stockholm, Sweden
- **Professor Israel Silman**
  Neurobiology Department
  Weizmann Institute of Science
  Rehovot, Israel
- **Professor Joel L. Sussman, Coordinator**
  Department of Structural Biology
  Weizmann Institute of Science
  Rehovot, Israel

FESP has assessed existing European infrastructures and future requirements dedicated to SG/SP. This assessment has resulted in a series of published papers and of documents posted on the FESP web site, which will also be transmitted directly to the EC. The documents make a series of general and specific policy recommendations in the SG/SP area. These recommendations were reached on the basis of a series of surveys, site visits, dedicated workshops, and satellite sessions at conferences, addressing the issues of the experimental and theoretical approaches, technologies and infrastructures which will be required for SG/SP to make an optimal contribution to progress in basic biology and clinical research in Europe over the next 5-10 years (see Appendix-1 for a list of FESP activities, publications and site visits).

Over the next few years, SP will grapple with the problem of visualizing increasingly elaborate and complex structures, from the atomic details of protein structures up to subcellular structures and, eventually, the whole cell. This issue was addressed directly at a seminal workshop organized jointly by the EC and FESP, which took place in Skåvsjöholm, Sweden, on November 15th-17th, 2006. The meeting brought together a body of 40 experts. They included not only structural biologists involved in structural proteomics consortia and other cooperative endeavors in Europe, Japan and the US, but also biochemists, biologists and policy makers. These included three senior representatives of the EC itself, the Director of the Department of Ventures and Initiatives of the Wellcome Trust, and the Director of the US National Institutes of Health Protein Structure Initiative (PSI), as well as a representative

of the pharma/biotech industry. An Associate Editor from *Nature Structural & Molecular Biology* attended the meeting, and a report was subsequently published in that journal (Banci et al. 2007a) (see Appendix-2).

Based on discussion with many scientists throughout Europe, FESP has also proposed a series of nine possible projects in the area of SG/SP for potential calls within FP7 (Appendix-3).

ESFRI & INSTRUCT

An event of great importance that took place during the period of activity of FESP was the decision to fund INSTRUCT, a collaborative project for establishing a trans-national Integrated Structural Biology Infrastructure in Europe. This decision was taken by the European Strategy Forum on Research Infrastructures (ESFRI) (see: http://cordis.europa.eu/esfri). ESFRI, an independent advisory body to the European Commission, was launched in April 2002. It brings together representatives of European Union member states and Associated States, appointed by the ministers in charge of scientific research, and a representative of the European Commission. The role of ESFRI is to support a coherent approach to policy-making on research infrastructures in Europe, and to act as an incubator for international negotiations concerning concrete initiatives. Previously, funding for large-scale infrastructures in Europe had been made on an *ad hoc* basis. Furthermore, it had been confined mostly to the exact sciences, such as physics, computing, and material sciences, being concerned with construction and maintenance of such facilities as synchrotrons, telescopes and super-computer centres. However, when, in 2004, ESFRI initiated a European Roadmap for Research Infrastructures, to recommend appropriate infrastructure strategies, a Biomedical and Life Science Roadmap Working Group (BMS RWG) was set up, with the specific objective of extending support for infrastructures to the Life Science and Clinical Communities. The BMS RWG, like ESFRI itself, includes delegates from virtually all the EC and associate states, as well as a representative of the EC in Brussels. In November 2006, ESFRI selected a set of large-scale research infrastructure projects for funding and, based on the recommendations of the BMS RWG, six of these were biomedical projects, each on a scale of ~300-400 million Euros, to be spent over 10 years, starting in 2008. INSTRUCT was one of the six chosen (see: http://cordis.europa.eu/esfri/large-scale.htm).

INSTRUCT (see: http://www.instruct-fp7.eu), is headed by Prof. David Stuart, of Oxford University. It comprises seven core centres spread across Europe, which will undertake the task of developing infrastructures in the principal areas of X-ray crystallography, NMR spectroscopy, electron microscopy, mass spectroscopy and protein production, while ensuring that the activities of the various nodes are both complementary and tightly integrated. It is envisaged that there will be ample scope for mutual synergies of the core centres with each other, allocation of trans-national access to users in all EU and associate states, and scientific and technical interaction with structural biology laboratories in regions that do not include a core centre, such as the Iberian Peninsula, Eastern Europe, Scandinavia and the Benelux and Baltic states.

It should be noted that although bioinformatics is intimately involved with structural biology and with the corresponding infrastructures, the topic was considered of sufficient importance by the BMS RWG (impinging as it does on such other major areas as epidemiology and medical genetics) for a separate bioinformatics research infrastructure to be approved. This infrastructure, with the acronym ELIXIR, centres on an upgrade to the European Bioinformatics Institute (EBI), but will also involve more than thirty other groups, institutes and government agencies throughout Europe. It, too, will commence its preparatory activities during 2008. Obviously, INSTRUCT will have to work closely with ELIXIR during the implementation phase.

As will be seen in the following, and in the recommendations for individual technologies, a large part of the policy recommendations being made by FESP closely match the stated objectives of INSTRUCT. This is not a coincidence, since many of the major players in INSTRUCT have been involved either in FESP itself, and/or in Structural Proteomics in Europe (SPINE), the first integrated SP project to be established in Europe, during the Vth Framework, and SPINE2-COMPLEXES, a continuation of SPINE, funded within the VIth Framework, which started in early 2007.

It should, however, be borne in mind that infrastructure, important as it may be, is only one issue. It is, therefore, essential that continued funding is provided for research projects in areas related to SG/SP topics or to topics at the interface of SG/SP and SB with systems biology, cell biology and clinical research. This is essential both for maintaining Europe at the 'cutting edge' *vis-à-vis* Japan and the USA, and for integrating the efforts of structural biologists with those of other biologists and biochemists. There are also requirements for user support and travel, for software and hardware development and for training, as will be referred to with respect to each individual category of instrumentation.

Recommendations

1) The policy recommendations made by FESP roughly fall into two broad categories: Specific recommendations relating to infrastructure requirements.
2) General recommendations concerning research policy.

Infrastructure requirements

The recommendations relating to infrastructure requirements are, to a large degree, based on the outcome of the surveys conducted on four topics, *viz.* synchrotrons, NMR and EM facilities, and protein production; they are presented in more detail in the respective individual reports. No survey was conducted with respect to bioinformatics, and mass spectroscopy (MS) was not addressed, even though it is anticipated that it will play a major role in the interdisciplinary effort to integrate proteins and protein complexes into higher order cellular structures (Sali et al. 2003; Nickell et al. 2006; Alber et al. 2007a; Alber et al. 2007b; Robinson et al. 2007), and as a complementary technique in structural studies on individual protein complexes. Overall, it is to be hoped that the establishment of INSTRUCT (see: http://www.instruct-fp7.eu) will facilitate the implementation of the recommendations.

Bioinformatics infrastructure has a high priority, but it is anticipated that the ELIXIR infrastructure project (http://www.elixir-europe.org), involving an upgrade of the EBI, and integration of its activities with its partners throughout Europe, will be responsible for taking care of the matter. It is recommended, however, that the INSTRUCT team liaises closely with the EBI so as to ensure that the requirements of SG/SP and of Structural Biology are adequately addressed. However, specific aspects may need to be prioritized within the framework of INSTRUCT and of SG/SP consortia in general. These may include investment in improved software for computational biology, homology modeling and computational methods for predicting function on the basis of protein sequences and structures. Another issue that requires a focused effort is making structural information more accessible to biologists and clinicians, as discussed below under Policy Recommendations. Finally, at a more mundane level, it will be necessary to invest in data bases collating such information as purification protocols, expression vectors and gene constructs, so as to streamline the production pipeline and reduce redundancy, on both the trans-European and international levels. Implementation of a standardized trans-European LIMS system would be a worthy objective in this context.

With regard to MS, even high-end equipment is not so expensive as to be beyond the means of individual research groups. There may, however, be a need for funding of a training element as utilization of this technique becomes increasingly integrated into the SG/SP effort.

With respect to synchrotrons, there is a general consensus that the number of beamlines dedicated to structural biology that are already operational, will become operational shortly, or are in advances stages of planning, will suffice for the foreseeable future. However, FESP strongly recommends continued EC support for synchrotron users, including support for user travel, and funding of staff and running costs. Support should also be given to automation, both in order to increase the efficiency of on-site data collection and to permit 'remote' data collection that will eliminate the need for scientists to travel to synchrotrons for routine experiments. Finally, in order for European synchrotrons to maintain a competitive edge, it will be necessary to invest in technical advances, including improved detectors and micro-focus beamlines, and in software for data acquisition and structure solution.

With regard to NMR, overall there is a feeling that the five trans-national BioNMR Research Infrastructures that have already been established, and provide service to users throughout the EU with the support of the EC, are fulfilling their intended role. It is thus recommended that support for both the users and for the facilities should continue to be provided in the same way as for synchrotron beamlines. Although increased development of automation is an essential objective, development of infrastructure permitting 'remote' use does not appear to have a high priority for various technical and logistic reasons. There is strong user demand for more powerful instruments and for increased availability of solid-state NMR instruments. Again, there is a strong recommendation for continued EC support for both software and hardware development, important themes being increased sensitivity, more rapid data collection, improvements in solid-state NMR, and improved software for all aspects of data collection and interpretation.

With regard to EM, the situation is different. Until now, it has been performed in individual laboratories, and true service facilities have not been available in Europe, though such facilities do exist in the USA, an example being the National Center for Macromolecular

Imaging (NCMI) at Baylor University, TX (http://ncmi.bcm.tmc.edu/ncmi). But the high-end 'state-of-the-art' instruments now coming on the market have price tags of ~3 million Euros, and together with ancillary equipment, substantially more; they are thus beyond the means of most individual institutions, and it is anticipated that the next generation of instruments will be significantly more expensive. It is, therefore, recommended that trans-national regional centers be established, much along the lines of the BioNMR Research Infrastructures, with the EC supporting not only the purchase of the equipment, but also maintenance and access. There is also reason for concern with respect to the 'training' element. There is now a serious shortage of trained personnel in the field of electron optics, which affects development, maintenance and utilization of instruments. It is, therefore, recommended that training programmes be established, ideally involving collaboration between academia and industry.

Protein production is often the bottleneck in SG/SP endeavours, as well as in individual structural biology laboratories, and the seriousness of the problem was obvious from the responses to the questionnaires. There are two principal ways to attempt to open up the bottleneck. Without establishing dedicated trans-national centers, it is nevertheless recommended that individual scientists be able to make use of non-profit production centers. Such centers exist, for example, in the framework of many of the SG/SP centers across Europe, and it would be cost-effective to encourage them to fulfill such a service role for clients from outside their own institutions or consortia, and to provide financial support to this end. The ISPC is a good example, as it provides such a service to academia, and also to industry, throughout Israel, and with the expanded facilities that will shortly be at its disposal, would be open to serving other EU research groups. There is also a large measure of redundancy in development of protein production protocols, with respect to both expression and purification. Strong efforts should thus be invested in readily accessible databases for storing such information, in courses and workshops at which techniques can be taught and information exchanged, and in funding of short visits to other laboratories for 'hands-on' experience. Moreover, it would be valuable to provide funding for maintenance of libraries of vectors and/or cell lines at one or more research centers.

Research policy

Considerations of Research Policy by FESP included three major topics:

1) Protein target selection
2) Systems structural biology
3) Bridging the gap between structural biology and other branches of biology

With respect to target selection, ever since large-scale SG projects were established in the USA and Japan, there has been controversy over this issue. In both countries, the initial effort was largely focused on development of high throughput (HTP) methodologies, with an emphasis on either filling in fold space or solving the structures of proteins belonging to a given genome, organelle or other category. These efforts resulted, in both countries, in immense technological achievements, and in the solution, by now, of several thousand structures, a remarkable achievement in its own right. But, again, in both countries 'in-house' criticism surfaced which argued that the emphasis was wrong, with too much effort being expended on 'low-hanging fruit' and on 'boring' proteins, and has resulted in an ongoing heated debate on the value of large-scale SG initiatives (Cyranoski 2006; Banci et al. 2007b; Blundell 2007; Harrison 2007; Janin 2007; McPherson 2007; Moore 2007; Petsko 2007;

Steitz 2007; Yokoyama et al. 2007; Berman 2008; Burley et al. 2008; Chazin 2008; Doublie and Rould 2008; Moult 2008; Vakser 2008). From the outset, Europe adopted a different approach, with SPINE emphasizing selection of disease-associated targets, such as relevant human proteins or those of bacterial pathogens, as reflected in its emphasis on structural proteomics (SP) rather than on SG. A focus on human targets related to human health has also been the stated policy of the SGC. This policy has resulted in the solution of a wealth of protein structures of high importance to clinical research and of direct relevance to the concern of the EC with improving the quality of life within Europe. It is thus recommended that such an orientation in target selection be maintained. Indeed, both in the USA and Japan, the thrust of SG/SP research is taking a similar direction. Furthermore, even important targets should not be solved in a 'vacuum'. It is recommended that the structural biologists involved, even in the framework of an SG/SP facility or consortium, should have a clear knowledge of the underlying science, and liaise closely with the biologist(s) who supplied the target. The EC should be ready to cope with new challenges that may arise in SG/SP. Thus it is anticipated there will be eventual access to the proteomes of microorganisms, which cannot presently be cultivated, which may account for up to 99% of the microorganisms in the environment as well as in the human body (Papamichail et al. 2005). The proteomes of these families of organisms may be expected to provide a wealth of unknown conformations, folds and protein architectures, which may well prompt us to revisit our knowledge of the principles of protein structure and folding.

Systems structural biology is an emergent discipline, and a workshop supported partially by FESP, held in Florence, Italy, in October, 2007, brought together a body of experts, both structural biologists and other biologists interested in higher order cellular organization, to discuss policy in this area; a white paper was drafted based on the views expressed at this conference (http://www.ec-fesp.org/FESP/reports/MolecularSystemsBiology.pdf). Some of these issues were also considered at the earlier workshop jointly organized by the EC and FESP, held in November, 2006, in Skåvsjöholm, Sweden, which was referred to above (Banci et al. 2007a). There is a broad consensus that the trajectory of structural biology should be towards the study of higher order structures, starting with protein complexes, proceeding towards development of methodologies for elucidating their location within their host organelles, and, eventually, producing a 'molecular map' of the entire cell. Such a map will obviously serve as an accurate blueprint for understanding how the individual components of the cell interact with each other. This will, of necessity, be a multidisciplinary endeavour, bringing together experts in all experimental areas of structural biology, as well as computer modeling and mass spectroscopy. From the top down, correlational microscopy will be used to map the broader features of the cell, and from the bottom up, high-resolution data on individual proteins and protein complexes will be incorporated, making use of appropriate modeling techniques, into higher order structures visualized by EM tomography. An admirable account of how this may be achieved is presented in a recent review (Robinson et al. 2007). Obviously, a large investment in state-of-the-art technologies will be required, especially in light and electron microscopy and in mass spectroscopy. Equally important will be effective integration of these resources in core centres of excellence, which may be facilitated in the framework of the INSTRUCT infrastructure programme. However, funding of infrastructure is only a means to an end, and cannot replace direct funding of cutting edge research, both fundamental and clinically oriented. It is thus recommended that there will be continued funding of integrated SG/SP projects, along the lines of VIZIER, FSG-V-RNA,

SPINE-2-COMPLEXES, E-MeP, 3D-REPERTOIR, CAMP, 3D-EM, BIOXHIT, Opticryst, IMPS, Extend-NMR, UPMAN, NDDP, NMR-Life, HT3DEM and GENEFUN, as well as of smaller consortia and of individual SB grants. The emphasis should be on fundamental projects aimed at increasingly higher order structures, as well as at the development of technologies for attacking them; on joint projects with cell biologists and systems biologists, so as to integrate the structural, dynamic and regulatory features of cellular of subcellular and cellular organization; and on projects with an anticipated clinical payoff, preferably involving collaboration with clinicians or with investigators associated with the pharma or biotech industries. Possible themes to which priority could be given are listed in Appendix-3.

The various SG/SP endeavours, together with classical SB groups, are providing a steady stream of high-resolution structures of proteins and protein complexes that provide a wealth of information of both basic and applied value. There is, however, an acute translational problem with respect to ensuring that biologists and clinicians can take advantage of this information. Most biologists who have not been trained as structural biologists, and who usually also lack a background in chemistry, have great difficulty in grasping the details of three-dimensional structures, and of taking advantage of them to guide their research and to plan appropriate experiments, *e.g.* to design mutants, trim or remove domains, or seek appropriate ligands with the aim of modulating biological activity in a controlled fashion. One way of overcoming this obstacle is for the biologist or clinician to work very closely with a structural biologist, and this is often the case. But in-depth scrutiny of a protein structure, even by a trained structural biologist, is a protracted process, and the scale of generation of structures by the SG/SP consortia is such that it is imperative for other biologists to have the capacity of extracting the pertinent information independently. It is, therefore, recommended that a high priority be given by the EC to devising curricula for appropriate courses and workshops that will train them to do so. In parallel, it is recommended that support be given to development of graphic packages that will facilitate extraction of the relevant structural information. Efforts in this direction that have been made include:

- Kinemage (Richardson and Richardson 1992)
- ProteinExplorer (http://www.umass.edu/microbio/chime/pe_beta/pe/protexpl)
- FirstGlance (http://molvis.sdsc.edu/fgij)
- iSee (Abagyan et al. 2006)
- PDBSUM (Laskowski 2007)
- TOPSAN (http://www.topsan.org/TOPSAN)
- Preparing Enhanced Figures in IUCr Journals (Einspahr and Guss 2008)
- Proteopedia (http://www.proteopedia.org)

References

Abagyan, R., Lee, W.H., Raush, E., Budagyan, L., Totrov, M., Sundstrom, M., and Marsden, B.D. 2006. Disseminating structural genomics data to the public: from a data dump to an animated story. *TIBS* **31:** 76-78.

Alber, F., Dokudovskaya, S., Veenhoff, L.M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B.T., Rout, M.P., and Sali, A. 2007a. Determining the architectures of macromolecular assemblies. *Nature* **450:** 683-694.

Alber, F., Dokudovskaya, S., Veenhoff, L.M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B.T., Sali, A., and Rout, M.P. 2007b. The molecular architecture of the nuclear pore complex. *Nature* **450:** 695-701.

Banci, L., Baumeister, W., Enfedaque, J., Heinemann, U., Schneider, G., Silman, I., and Sussman, J.L. 2007a. Structural proteomics: from the molecule to the system. *Nat. Struct. Mol. Biol.* **14:** 3-4.

Banci, L., Baumeister, W., Heinemann, U., Schneider, G., Silman, I., Stuart, D.I., and Sussman, J.L. 2007b. An idea whose time has come. *Genome Biol.* **8:** 408.

Berman, H.M. 2008. Harnessing Knowledge from Structural Genomics. *Structure* **16:** 16-18.

Blundell, T. 2007. New Dimensions of Structural Proteomics: Exploring Chemical and Biological Space. *Structure* **15:** 1342-1343.

Burley, S.K., Joachimiak, A., Montelione, G.T., and Wilson, I.A. 2008. Contributions to the NIH-NIGMS Protein Structure Initiative from the PSI Production Centers. *Structure* **16:** 5-11.

Chazin, W.J. 2008. Evolution of the NIGMS Protein Structure Initiative. *Structure* **16:** 12-14.

Cyranoski, D. 2006. 'Big science' protein project under fire. *Nature* **443:** 382-382.

Doublie, S., and Rould, M.A. 2008. The Soul of a New Structure-Function Machine. *Structure* **16:** 3-4.

Einspahr, H., and Guss, M. 2008. A new service for preparing enhanced figures in IUCr journals. *Acta Crystallographica Section F* **64:** 154-155.

Harrison, S.C. 2007. Comments on the NIGMS PSI. *Structure* **15:** 1344-1346.

Janin, J. 2007. Structural Genomics: Winning the Second Half of the Game. *Structure* **15:** 1347-1349.

Laskowski, R.A. 2007. Enhancing the functional annotation of PDB structures in PDBsum using key figures extracted from the literature. *Bioinformatics* **23:** 1824-1827.

McPherson, A. 2007. Some Ill Considered Comments on the Protein Structure Initiative. *Structure* **15:** 1526-1527.

Moore, P.B. 2007. Let's Call the Whole Thing Off: Some Thoughts on the Protein Structure Initiative. *Structure* **15:** 1350-1352.

Moult, J. 2008. Comparative Modeling in Structural Genomics. *Structure* **16:** 14-16.

Nickell, S., Kofler, C., Leis, A.P., and Baumeister, W. 2006. A visual approach to proteomics. *Nat. Rev. Mol. Cell. Biol.* **7:** 225-230.

Papamichail, D., Skiena, S.S., van der Lelie, D., and McCorkle, S.R. 2005. Bacterial population assay via k-mer analysis. In *Proceedings of the 3rd Asia-Pacific Bioinformatics Conference, January 17-21, 2005*. (eds. Y.-P.P. Chen, and L. Wong), pp. 299-308. Imperial College Press, Institute for Infocomm Research, Singapore.

Petsko, G. 2007. An idea whose time has gone. *Genome Biol.* **8:** 107.

Richardson, D.C., and Richardson, J.S. 1992. The kinemage: A tool for scientific communication. *Protein Sci.* **1:** 3-9.

Robinson, C.V., Sali, A., and Baumeister, W. 2007. The molecular sociology of the cell. *Nature* **450:** 973-982.

Sali, A., Glaeser, R., Earnest, T., and Baumeister, W. 2003. From words to literature in structural proteomics. *Nature* **422:** 216-225.

Steitz, T.A. 2007. Collecting Butterflies and the Protein Structure Initiative: The Right Questions? *Structure* **15:** 1523-1524.

Vakser, I.A. 2008. PSI Has to Live and Become PCI: Protein Complex Initiative. *Structure* **16:** 1-3.

Yokoyama, S., Terwilliger, T.C., Kuramitsu, S., Moras, D., and Sussman, J.L. 2007. RIKEN aids international structural genomics efforts. *Nature* **445:** 21.

Appendix-1 FESP presentations, workshops, & site visits:

Presentations and workshops:

16 Mar 2005: "Future Needs for Research Infrastructures in Biomedical Sciences", Brussels, Belgium - Workshop organized by the EC Research Directorate Generals for Infrastructures and for Health [L. Banci]

31 Aug 2005: "FESP Workshop at SPINE Annual Congress", Montecatini, Italy [L. Banci, U. Heinemann, G. Schneider, I. Silman, J.L. Sussman & J. Enfedaque]

20-25 Aug 2005: "NMR in Molecular Biology", Scania, Sweden - ESF Conference, round table [K. Wuthrich, C. Arrowsmith, H. Oschkinat & L. Banci]

6-8 Dec 2005: "3rd European Conference on Research Infrastructures", Nottingham, UK, organized by EC, Directorate Research Infrastructures [L. Banci]

6-8 Dec 2005: "4th SESAME USERS Meeting", Dead Sea, Jordan [J.L. Sussman]

28 Jan 2006: "EMBL-Hamburg Priorities Committee Meeting", Hamburg, Germany [G. Schneider & J.L. Sussman]

5-6 Apr 2006: "FESP Kick-Off Meeting", Brussels, Belgium [L. Banci, W. Baumeister, U. Heinemann, G. Schneider, I. Silman & J.L. Sussman]

6 Apr 2006: "Infrastructures for EM & NMR", Brussels, Belgium - Participants: A. Engel, C. Griesinger, W. Hax, H. Oschkinat, S. Pittard & G. Wolf [L. Banci & W. Baumeister]

5-6 May 2006: "Complementary Techniques to NMR for Structure Determination of Biological Macromolecular Complexes", Florence, Italy (Joint CA NMR for Life & FESP Activity) [L. Banci] Meeting Report

23 May2006: "SPINE2-Complexes Preparatory Meeting", Budapest, Hungary [L. Banci, U. Heinemann, G . Schneider, I. Silman & J.L. Sussman]

5-6 Jun 2006: "European Research Infrastructures for the Life Sciences", Vienna, Austria - Organized by the European Life Sciences Forum (ELSF) [J.L. Sussman]

23 Oct 2006: "4th ISGO Annual Meeting", Beijing, China - special evening workshop [L. Banci, U. Heinemann, J.L. Sussman & J. Enfedaque]

15-17 Nov 2006: "The Direction of Structural Proteomics - From the Molecule to the System", Skåvsjöholm, Sweden - A Joint EC/FESP Workshop that Addressed Strategic Directions in Structural Biology [L. Banci, W. Baumeister, J. Enfedaque, U. Heinemann, G. Schneider, I. Silman & J.L. Sussman] See: Banci et al. Structural proteomics: from the molecule to the system *Nat. Struct. Mol. Biol.* 14, 3-4 (2007).

18-20 Jan 2007: "Advances and Management of NMR in Life Sciences, Florence, Italy - A Joint Meeting of (I3) EU-NMR and (CA) NMR-LIFE" [I. Bertini (Chair), R. Boelens, L. Emsley, U. Guenther, H. Schwalbe] Presentation by L. Banci.

19 Feb 2007: "Synchrotron Radiation in Structural Biology", G. Schneider, Conference on Structures in Biology - Past, Present and Future, Didcot, UK

2 Aug 2007: “Impact of PSI on Europe & Israel: What’s missing?”, J.L. Sussman, NIGMS, NIH, Bethesda, MD.

18-19 Oct 2007 Joint workshop CA NMR-Life and FESP “New Challenges in the Life Sciences: Prioritizing European Research in Molecular Systems Biology”, Florence, Italy [L. Banci, W. Baumeister, I. Silman & J.L. Sussman]

Site visits

17 Jan 2007: The Centro Risonanze Magnetiche, (CERM), Florence, Italy, by I. Silman

14 Apr 2007: The Photon Factory, Tsukuba, Japan, by I. Silman

15 Apr 2007: Riken Structural Genomics/Proteomics Initiative, Yokohama, Japan, by I. Silman

1-2 Aug 2007: Howard Medical Institute’s Janelia Farm, Ashburn, VA, by W. Baumeister & J.L. Sussman

Appendix-2 List of published papers

Banci L, Baumeister W, Enfedaque J, Heinemann U, Schneider G, Silman I, Sussman JL (2007) Structural proteomics: from the molecule to the system. *Nat Struct Mol Biol* **14:** 3-4

Banci L, Baumeister W, Heinemann U, Schneider G, Silman I, Stuart DI, Sussman JL (2007) An idea whose time has come. *Genome Biology* **8:** 408

Banci L, Baumeister W, Heinemann U, Schneider G, Silman I, Sussman JL (2008) Structural Genomics and Structural Proteomics: A Global Perspective. In *Structural Proteomics and its Impact on the Life Sciences*, Sussman JL, Silman I (eds) (in press). World Scientific Publishing Company

Sussman JL, Silman I (eds) (2008) *Structural Proteomics and its Impact on the Life Sciences*. (in press). World Scientific Publishing Company

Yokoyama, S, Terwilliger, TC, Kuramitsu, S, Moras, D, and Sussman, JL 2007. RIKEN aids international structural genomics efforts. *Nature* **445**: 21

Appendix-3 Recommendation for FP7 projects in the areas of SG/SP

- Multi-Disciplinary Approach to Large macromolecular complexes
- NMR in Structural Genomics and Structural Proteomics
- Structural vaccinology: a new approach to fill the gaps in vaccine design
- High-throughput EM single particle analysis
- Approaches to the Characterization of Intrinsically Disordered Proteins
- Development of novel antibacterials
- Structural Characterization of Functional Protein Complexes in the Normal and Diseased Brain
- Knowledge Data Base
- New strategies for crystallographic data collection

Multi-disciplinary approach to large macromolecular complexes

Rationale for requesting large collaborative project

The area of investigation envisaged here centers on the theme of structural cell biology, and specifically on the large, often transient, protein complexes involved in the interactions of cells with each other and with various pathogens. Although high-resolution structures of key sub-components of such complexes may still need to be determined, it is highly plausible that many complexes of great biological or medical interest will have to be studied at lower resolutions, and will require much more sophisticated methodologies than are currently available to attain even those low resolutions. s

While EM (Electron Microscopy) is making progress towards achieving sub-nanometer resolution, the scope of such work, and the unpredictable difficulties that may be encountered, make it crucial that the capabilities of X-ray work in this resolution range be carefully assessed, and be exploited to the fullest extent possible.

The main theme should therefore be to provide the necessary improvements in instrumentation and methods to tackle those crystallographic studies of large macromolecular complexes, which are still intractable.

A substantial part of these methodological developments would include dealing with non-single crystals by improving the capabilities of micro-diffractometers, of reflection integration software, of data processing software to disentangle the overlapping diffraction patterns of different crystalline domains, and to adequately represent the overlap pattern of reflections, which cannot be separated. Phasing and refinement programs would then have to be modified to handle such overlapped data and to cope with low data resolution.

A related set of developments would be concerned with the limitations that the radiation sensitivity of protein crystals places on the ability of a single diffracting sample to deliver a complete data set to the diffraction limit of a crystal form. Data collection strategies would have to be devised with this limitation in mind, so as to gather all the necessary data from the very limited partial data sets that each small crystalline volume could provide. This would require further work in the exploitation of small goniometers capable of carrying out the re-orientation of very small crystalline samples (with the positional accuracy demanded by a micro-diffractometer) required to fill diffraction data space efficiently. A further source of difficulty will be that the numerous crystals that would be required for implementing such strategies would often be non-isomorphous, again requiring that phasing and refinement methods be reconsidered and adapted to such composite and non-isomorphous datasets.

The crux of X-ray crystallography is the phase problem - a problem which, when it is solvable by experimental means, turns into an opportunity to determine 3D structures in a totally objective manner, without any pre-conceived model nor its associated possibilities of bias. The capabilities of experimental phasing on large complexes, using different scatterers at different resolution stages, was beautifully illustrated by the solution of such key structures as those of the ribosome and of RNA polymerase II. A focused, systematic effort would be worthwhile so as to consolidate this technique as applied to even larger complexes at lower resolutions. Extensions are also possible towards solution techniques such as NMR and

SAXS (Small-Angle X-ray Scattering), as was demonstrated by Engelman & Moore in their triangulation of the ribosome (with neutrons and deuterated proteins rather than with x-rays and proteins tagged with heavy-atom clusters) in the 1970s, and towards data measured from micro-crystalline powders.

The meshing of low-resolution X-ray crystallographic investigations with EM, through powder and SAXS investigations helping bridge the gap when necessary, would be one of the directions in which fresh support and incentive to collaborate would be extremely beneficial to European Structural Biology as a whole. By bringing about such a wide-ranging consolidation of X-ray techniques appropriate for a gradation of resolution regimes, it would play a decisive role in bringing about a qualitative jump in the complexity of supra-macromolecular systems for which structural information can be obtained in their native state.

NMR in structural genomics and structural proteomics

Rationale for requesting through a NOE

NMR has established itself as an invaluable tool in the investigation of the structure and dynamics of biomolecules. In the post-genomic era, the contributions of NMR to the fields of Structural Genomics and Structural Proteomics will only continue to increase as weak biomolecular interactions are increasingly recognized as central to the biochemical processes of Life, as these weak interactions are best captured with NMR. Decades of work by talented individuals and collaborative teams have allowed Europe to boast both human and infrastructure capitals in NMR that are second to none.

Although NMR has contributed extensively to Structural Biology, it was felt that interactions between top-level laboratories with complementary expertise was limited. For this reason, the EC FP6 funded Coordination Action NMR-Life was created to promote the networking and coordination of NMR research in Structural Genomics. The activities of NMR-Life were developed to achieve five main goals: a) to foster cross-fertilization among the four scientific areas on which the project focuses (i.e. protein-protein interactions, protein-DNA/RNA interactions, protein-ligand interactions, and membrane and immobilized proteins), b) contribute to the development and dissemination of common good practices and methodological approaches, c) foster the transfer of knowledge among different research teams, thereby enhancing the spread of innovative methods and tools, d) implement a common reference point for the European NMR community by maintaining a common virtual on-line laboratory, and e) jointly identify bottlenecks and possible breakthroughs in biological NMR. The efforts to provide a platform for exchange among European researchers spearheaded by the CA have been successful. Given the development of Structural Proteomics towards Systems Biology, it is evident that new experimental approaches such as in-cell NMR, solid-state NMR on native-like membrane samples, dynamic nuclear polarization, and paramagnetic highlighting of transient interactions, among others, will open new avenues for obtaining biological information, in particular of samples or systems that are not yet accessible by other techniques. We believe that a Network of Excellence for NMR in Structural Proteomics is the appropriate next step to further strengthen European NMR research on the international stage.

The funding of an NOE would provide support for the advancement of both technology and methodology through collaborative efforts. NMR is becoming more and more successful in structure determination and in dynamic characterization of biomolecules, but the NMR community, and those working in the field of Fundamental Genomics in general, have stimulating challenges to address in the near future. Namely, the unraveling of biological processes in a complex network of weak transient interactions that must be characterized at an atomic level but within a systems biology approach. Specifically, our efforts should be focused on the investigation of:

a) protein-protein complexes to understand biochemical processes,

b) protein-ligand interactions for drug discovery,

c) protein-DNA/RNA interactions for expression and translation,

d) metabonomics for the investigation of the metabolism in biological fluids,

e) in cell/*in vivo* NMR to monitor *in vivo* processes,

f) solid state NMR for the world of membrane proteins, and

g) molecular imaging.

We want to pursue these objectives through a NOE by i) advancing knowledge in systems biology and investigational methodologies, ii) the exchange of personnel, iii) the exchange of good practices, and iv) the organization of dedicated meetings. The funding of an NOE would therefore not only provide a venue for the transfer of knowledge and ideas, as undertaken by the CA, but would also provide support for the advancement of both technology and methodology through collaborative efforts. The coordination and joint advancement of research, along with enhanced training activities and the encouragement of academic-industrial collaboration are all important roles a Network of Excellence could play, helping Europe to maintain its position as the world leader in NMR.

Our successful experience with NMR-Life leads us to believe that the European Bio-NMR community can successfully take advantage of the opportunities an NOE would provide. We envision an NOE of approximately 12 partners, allowing for some growth in involvement from the current ten NMR-Life partners, while avoiding allowing ourselves to be hindered by an overly large membership. We are hopeful that a call for proposal for an NOE in Structural Genomics or Structural Proteomics will soon be available, and request that the European Commission consider the possibility of such an opportunity.

Structural vaccinology: a new approach to fill the gaps in vaccine design

Rationale for a Large Collaborative Project

For more than a century, vaccines have been developed using the principles of Pasteur, based on use of the whole pathogens or parts of them, also exploiting the continuous technological advancements to make their development more efficient, and resulting in more effective products. This approach has been successful for a large number of diseases, including most of those causing mortality in infants and children.

Still, serious gaps in vaccine development strategies are present at various levels. These gaps mainly reside in two factors, antigenic variability and limited serum antibody induction.

The availability of complete pathogen genomes has allowed major advances in approaches to vaccine development, giving the possibility, within much shorter periods, of screening orders of magnitude larger numbers of potential antigens. Vaccines, developed through a genome-based antigen discovery approach and able to protect against all the strains of a given pathogen, have been already obtained for bacteria for which the "classical" approach led only to vaccines directed against a single specific strain.

To address the challenge in vaccine development arising out of antigenic variability, however, innovative approaches need to be developed. One such approach would be based on individuating the invariant parts of the antigen, which are often poorly visible to the immune system but are still able to elicit an immuno-response, although at a frequency which is not sufficient to give protection. These epitopes are usually called "immunosilent". One example of these epitopes is the one located in the CD4 binding region of HIV gp120 recognized by the monoclonal antibody b12, the structure of which has been recently published in Nature (Kwong, Nature 2007). The key point for developing a strategy for addressing high strain variability is that of understanding the factors, at a molecular level, which discriminate between "immunodominant" and "immunosilent" epitopes. We need to go back to basic research, and understand the molecular nature of immunodominance. One way of addressing this problem is to determine the 3D structure of many immunodominant and silent epitopes in order to understand the basic structural factors involved. Once we understand the rules, we should be in a position to engineer novel immunogens able to induce an immune response against conserved epitopes, and to conquer many diseases that are still beyond our reach.

Accordingly, we need to initiate a project for high throughput determination of the 3D structure of antigens and of antigen-antibody complexes. This approach should become an intrinsic tool of vaccinology.

Structural Vaccinology as an innovative approach to fill the gap in vaccine design for virus with high antigenic variability Projects should focus on the determination of the 3D structures of antigens relevant to vaccines, especially of epitopes covering broadly pathogenic species, universal epitopes and antigen-antibody complexes. X-ray, NMR, cryo-EM methodologies should be applied to elucidating the molecular characteristics of immunodominant and immunosilent epitopes so as to reach the goal of obtaining structure-

based engineered improved immunogens. The idea is to do the structures of all the virus antigens and, based on the immunoresponse of these, design the vaccines.

EC Instrument - Large-scale collaborative project.

High-throughput EM single particle analysis

Rationale for requesting a smaller-scale collaborative project.

Cryoelectron microscopy of single particles has become a powerful tool for structural studies of macromolecular complexes. The method involves the averaging over large data sets of individual particles, following their alignment and classification. It is a distinct advantage of EM single particle analysis that it can cope with some heterogeneity of the sample; therefore the material needs not to be purified to exhaustion since some 'purification' can be performed *in silico* using smart image classification tactics. Although it has been demonstrated in a few cases, that resolutions can be attained that are good enough to discern secondary structure elements, most studies so far have been at a lower resolution (1-2 nm) level. While this is often sufficient for hybrid approaches, in which high-resolution structures of components obtained by other methods are fitted into the lower resolution structures of large multisubunit complexes, it is desirable to obtain subnanometer resolutions. Key to the attainment of higher resolution is the availability of large, high-quality data sets. The manual recording of large data sets of consistent quality is prohibitively time consuming but entirely feasible when using automated data acquisition methods. Hence there is an urgent need to develop and implement automated procedures allowing the acquisition of EM data in a high-throughput mode.

Detectors for cryo EM

Currently, the performance of cryoelectron microscopy is seriously compromised by the performance of the available detectors, usually CCD cameras. This is particularly true for structural studies performed at intermediate voltages (300-400 keV) as preferred in electron crystallography and in electron tomography. The availability of more sensitive and sufficiently radiation-hard detectors (active pixel sensors or detectors based on CMOS technology) would bring about a great improvement in data quality. Europe has unique potential in adapting these technologies to the specific need of EM.

Approaches to the characterization of intrinsically disordered proteins

Rationale for requesting a smaller-scale collaborative project.

It is now becoming apparent that a significant percentage of the human proteome is constituted by Intrinsically Disordered/Unstructured Proteins (IDPs/IUPs), *i.e.* proteins which, in their native state, are either unfolded or can sample a broad range of conformations. These proteins defy the classical structure-function paradigm, yet they carry out important and varied functions in signal transduction, transcription regulation, and as scaffolding proteins at synapses in the nervous system. They often serve as 'hub' proteins in interactomes. Due to their central roles in regulation, their mutations and/or formation of aggregates are implicated in a variety of severe diseases, such as cancer and neurodegeneration.

Intrinsic disorder is crucial to the function of these proteins and, accordingly, IDPs need to be characterized in terms of conformation, range and time-scale of internal motions, and of their functional interactions with protein or nucleic acid partners. Thus, the eventual goal is to gain a detailed understanding of their properties and mode of action in a cellular context, taking a systems biology approach.

However, due to their dynamical and transient properties, characterization of IDPs is a truly challenging task. They have, thus far, been studied only to a very limited extent as compared to globular proteins with a well-defined 3D structure. To achieve a comprehensive characterization of IDPs, specific biophysical and biochemical approaches will need to be developed, demanding use of a varied repertoire of tools and techniques. An integrated approach will be required to study both the individual IDPs and their functional interactions at various levels of resolution and complexity, and should also include the fourth dimension, i.e. the time frame of processes.

Ideally, we should develop tools which will allow us, starting from the characterization of single IDPs in their multiple conformations, to then define their pattern of interactions, to characterize the individual interactions, and finally, to understand their integration into such complex intra- and intercellular processes as signal transduction, embryonic development and synaptic patterning and plasticity.

An integrated approach to the characterization of intrinsically disordered proteins (IDPs): from single molecules to the cell. – Projects should be highly multidisciplinary and should integrate, coordinate and optimise various methodologies, from bioinformatics and computational approaches to structural tools, with particular emphasis on NMR, EM, X-ray spectroscopy, SAXS, two-dimensional gel electrophoresis (2DE) & mass spectrometry (MS) and cellular imaging technology.

Development of novel antibacterials

Rationale for requesting a large-scale and smaller-scale collaborative project, with significant involvement of SME's.

The fight against infectious diseases is not going well for humans at present. The emergence of extensively drug-resistant tuberculosis (XDR-TB) where the second-line of defense has been broken is indicative of an impending disaster. At the same time, about 20% of all patients undergoing surgery in hospitals in EU countries, such as the UK or Sweden, suffer bacterial infections as a result of medical treatment. Unfortunately, the pathogens responsible for such infections, typically from the geni *Pseudomonas*, *Streptococci* etc also show alarming signs of multi-drug resistance and in particular strains exist that are resistant to vancomycin, the last line of defense.

As a result the WHO and the medical research councils of many countries have declared antibiotic drug resistance as a major threat to human health. The leading journal Nature Biotechnology devoted recently a whole issue to this topic, and Sir David Hopwood, a pioneer of the molecular biology of antibiotic biosynthesis, published a "call to arms" in Nature Reviews of Drug Discovery.

A major requirement for novel antibiotics is that they will be prescribed as restrictively as possible, to postpone the on-set of resistance. This makes the economic return small, and consequently large pharmaceutical industries are currently reluctant to engage in the discovery of new antibacterials. Proposals have thus been put forward both in Europe and the US to enable national (US) or international (EC) academic and governmental laboratories to take a lead in this enterprise. The EC has already supported actions towards this goal, and there is a topic for drug development against gram-negative bacteria in the second call of the 7^{th} framework. Already in the 5^{th} and 6^{th} framework the EC supported projects aimed at drug discovery against tuberculosis. These projects have produced very promising results, but if not continued the developments will not be capitalized. In view of the seriousness of the problems, we propose two topics, both related to the development of novel antibacterials. Structure biology will play a leading role in both these enterprises.

1. Novel antibiotics against multi-drug resistant bacteria

This project targets common pathogens of such geni as *Streptococci*, *Staphylococci* and *Pseudomonas*. Alternatively it could specifically target MRSA (meticillin resistant *Staphylococcus aureus*, a major pathogen causing infections). The aim is the extensive structure determination of proteins of essential biosynthetic or signaling pathways, coupled to genetic target validation. Lead compounds will be identified a combination of in-silico techniques and experimental compound screening. Successful hits will be followed up by organic chemistry and verified *in vivo*. This will involve structural biology groups, microbiologists, SME's providing libraries of natural products and synthetic compounds, and chemists.

EC Instrument - Large-scale collaborative project.

2. Highly drug-resistant TB

This will be a focused project to develop further novel drugs against tuberculosis, both in the active and latent phases of infection. The project involves collaboration between structural biologists, microbiologists and chemists.

EC Instrument - A smaller-scale collaborative project, with significant involvement of SME's.

Structural characterization of functional protein complexes in the normal and diseased brain

Rationale for requesting large collaborative project.

In the brain, specialized sets of proteins and protein complexes provide the scaffolding and the functional modules for the rapid computational processes necessary for analysis of sensory inputs, for execution of motor functions, and for performing the cognitive processes involved in learning and memory. Many of the proteins and of the more complex entities involved, such as the complex structures associated with the synapse and with axonal processes, are unique to the brain, such as the complex network of the postsynaptic density. Furthermore, brain-specific proteins whose functions are still poorly understood, such as the amyloid precursor protein, synuclein and the prion protein, are key players in conformational diseases such as Alzheimer's disease, which impose heavy social and economic burdens on the European Community and elsewhere.

Accordingly, it is most timely to establish an integrated structural genomics/proteomics project targeted at developing tools for high-resolution structure analysis of protein complexes, and of even more complex entities associated with synapses, dendrites and axons.

In such a project an integrated and generally applicable technology platform will be established to permit the identification, preparation and structure analysis of protein complexes differing in size and physical properties. It will identify complexes using proteomic methods starting from unbiased screens or from already known proteins and assess their physiological role by biochemical, biophysical, cell and systems biology methodologies. It will develop high-throughput techniques for recombinant protein complex production in bacterial and eukaryotic host systems, protein crystallization and structure analysis. Since many brain proteins are either intrinsically unfolded or contain long disordered regions, NMR spectroscopy will also be utilized for structure determination, and EM tomography and advanced imaging techniques will be required for investigating the complex supramolecular synaptic structures. It is envisaged that many of the complexes selected for study will be targets for drug design aimed at proteins associated with neurodegenerative disorders or such prevalent conditions as epilepsy and highly malignant brain tumors.

It will aim to recruit leading structural biologists currently involved in national and European proteomics and structural genomics consortia, who will act synergistically to utilize preexisting resources such as database systems, protein production facilities, crystallization robotics, NMR centers and synchrotron light sources, for running and managing their research. However, the project will also aim to incorporate neurologists, neuropathologists, neurochemists, neuroanatomists and molecular pharmacologists, who will both guide the choice of targets and interprete and reap the benefits of the novel structures that may be anticipated to emerge from such a project. This, in turn, will provide added value for the EC, and will increase European competitiveness, in the face of competing structural genomics activities elsewhere, by creating new research tools, promoting defragmentation of European research, and bringing basic research to the biotech and pharma industries. By focusing on

protein complexes from the human brain, the project will be addressing research topics that have the potential to provide major medical benefits to an aging European population.

Knowledge data base

Rationale for requesting large collaborative project.

Recently the National Institute of General Medical Sciences (NIGMS) in the NIH issued a call for a **Structural Genomics Knowledgebase** (see: http://grants.nih.gov/grants/guide/rfa-files/RFA-GM-06-004.html#PartII). Below is the synopsis of the project that was proposed; it relates to a topic that is also vital to the European research community. The EC Directorate General for Research should seriously consider sponsoring a programme with similar overall objectives within the area of Genomics and Systems Biology.

Background

With the completion of the Human Genome Project, our understanding of the genomic content of many species, including ourselves, has advanced dramatically. The challenge now is to elucidate the biological functions and interrelationships of all the products encoded in the genome. The three-dimensional (3D) protein structure provides a unique contribution in bridging the gap between our knowledge of a protein's sequence and its molecular function. Recent technological advances in X-ray crystallography and NMR spectroscopy have significantly reduced the time and effort required for protein structure determination, making high-throughput structure solution feasible. As a result, substantial structural genomics efforts have been launched around the world.

The purpose of this RFA is to announce support for a **Structural Genomics Knowledgebase** (hereafter referred as the **Knowledgebase**) in support of the PSI and other related programs. The Knowledgebase will be one component of the PSI Research Network along with large-scale production centers, specialized technology development centers, and a research material repository.

As the PSI is progressing into production mode, a high level of coordination of target selection will be required. To minimize overlap between projects, the NIGMS supported the development of a central target website, the TargetDB (http://targetdb.pdb.org), by the PDB. In the current phase, a single unified target list for all the PSI centers will be developed with the oversight of the Steering Subcommittee for Target Selection of the PSI Research Network. Several considerations in target selection include 1) having the broadest possible representation of sequence families, 2) incorporating plans for improving computational modeling, and 3) maximizing the impact of the structures in biomedical research by inviting community participation in target prioritization. To satisfy these needs, an enhanced target solicitation, posting, and illustration platform needs to be developed to take over the role TargetDB played during PSI-1.

Listed below are the main functions of the Knowledgebase in priority order.

- Structural Genomics Information Repository and Web Portal

  The PSI large-scale production centers and specialized technology development centers are required to make their research resource and data available to the public through a central information repository. This is the foremost function of the Knowledgebase. It must develop capabilities to accommodate heterogeneous types of information and large amounts of data. It must also provide tools for easy deposition of information concerning experimental materials like clones, proteins, etc., and experimental conditions and results. Both positive and negative results must be captured. A separate RFA is being developed to establish a PSI material repository as a physical archive of the clones generated by the PSI centers (link to the material repository RFA). The Knowledgebase should provide links to the clones maintained by the material repository. The Knowledgebase group will be expected to collaborate with researchers in the structural genomics and structural biology community to develop standards and a unified format in information gathering, organization, and presentation. The Knowledgebase will also be a central point-of-entry for accessing all the information and resources related to the PSI. It should develop user-friendly interfaces and tools for information retrieval and analysis.

- Structural Genomics Target Management

  The Knowledgebase should provide detailed information on the targets selected by the PSI centers and other structural genomics projects, with the inclusion of sequence, family size and composition, links to annotations and functions of the family members provided by other databases, status of the targets, and other information proposed by the applicant. Dynamic and schematic illustration should be provided to demonstrate the projected and current progress in covering the sequence space by selected and solved targets respectively. A mechanism should be developed in collaboration with the PSI Research Network to invite community participation in target selection and prioritization. All activities of the TargetDB will be subsumed by the Knowledgebase.

- Structural and Functional Annotation

  The applicant should establish collaborations with the PSI centers to develop a system of structural and functional annotation for the PSI-generated structures. In designing the annotation platform, the applicant should consider the level of heterogeneity of structural genomics targets and structures. While a high level of adaptability is required, standardization is also important. The Knowledgebase will not be responsible for performing annotations. Instead, the Knowledgebase should provide a user-friendly environment for deposition and query of the annotations generated by PSI projects. Links to publications and annotations at other databases should be established. Soliciting community (outside the PSI Centers) participation in annotation is another requirement. The applicant needs to be creative in proposing incentives to community information providers and moderators that will likely be effective in promoting community participation. A quality assurance mechanism should also be proposed.

- Computational Modeling Support

  Computational modeling of protein structures plays an essential role in reaching the goals of structural genomics. The Knowledgebase should provide support in this area by hosting a limited number of high-quality models of proteins that are in the sequence families covered by the structural genomics targets. The applicant should collaborate with the modeling community to develop filters and assessment criteria that only allow the best available models to be selected for posting. These models should have designation of quality/confidence levels and potential utilities. The Knowledgebase should not collect and present all computational models without discretion. The applicant should have plans for continually reevaluating and improving model assessment methods and promoting the most promising models. The Knowledgebase should also provide links to modeling tools and servers developed by other groups, allowing users to generate models using these tools, and provide computational model quality evaluations and uncertainty estimations to the users.

- Training and Education Coordination

  The Knowledgebase will be expected to coordinate with the PSI centers to develop workshops and training courses for active knowledge dissemination.

- Progress Analysis

  The Knowledgebase should develop metrics to analyze periodically the progress of the structural genomics projects and their impact on biomedical research.

New strategies for crystallographic data collection

Rationale for requesting a smaller-scale collaborative project.

At modern synchrotrons, the combination of high photon fluxes, short detector read-out times, multi-circle diffractometry, and high-throughput sample mounting devices creates opportunities for devising new strategies for diffraction data acquisition that will enable:

- On-line monitoring of radiation damage
- Optimized data collection from multiple crystals
- Construction of full data sets by combination from many partial data sets.

New strategies implementing such features would increase the scope of synchrotron-based crystallography by making more difficult systems amenable to high-resolution structural studies.

The very high photon fluxes of modern synchrotron radiation sources in combination with fast-readout detectors allow for extremely fast acquisition of diffraction data from protein crystals. This offers opportunities to introduce new data collection strategies to address the major problem of current synchrotron-based crystallography, radiation damage. The key features of these strategies are (1) online monitoring of radiation damage and (2) construction of complete data sets from partial data sets originating from many samples with limited damage.

Fast data acquisition schemes will allow for repeated acquisition of (reference) data providing a direct measure of progressive radiation damage. However, for very fast acquisition schemes, current technologies are limited due to synchronization problems. By moving towards data collection strategies such as ‘large-bandpass’ or ‘pseudo-Laue’ methods, synchronization problems can be avoided as the sample is kept stationary. Diffraction data can then be acquired stepwise on the milli-second time scale providing the possibility to directly quantify radiation damage and to stop the measurement when the damage becomes unacceptable.

Characterization of protein crystals involves mounting them in the X-ray beam followed by data collection and data evaluation. The time constants of this process are currently dominated by the mounting process requiring minutes. Given that the actual data acquisition at present already takes place in seconds, the technologies for moving crystalline samples into the X-ray beam must be improved. This requires developments in crystal preparation, robotics for crystal mounting and alignment, and in software for evaluation of sample quality.

Finally, when data sets can be collected quickly from many crystals, computational methods will be needed to identify sets of compatible ‘isomorphous’ samples and assemble complete data sets from these.

In addition to providing means to deal with radiation damage, the availability of high-throughput technologies for crystallographic data acquisition would also facilitate large-scale screening of crystals in the context of challenging structural biology projects.