Complementary techniques to NMR for structure determination of biological macromolecular complexes: From Atoms to the Cell - From NMR to cryoEM

A workshop organized in the frame of the EU- funded Coordination action
Focusing NMR on the machinery of Life NMR-Life

Lucia Banci, CERM, University of Florence and Michael Sattler, EMBL, Heidelberg

Florence, May, 5-6, 2006

Participants and contributors:

Lucia Banci, CERM, University of Florence, Italy
Wolfgang Baumeister, Max-Planck-Institute for Biochemistry Munich, Germany
Ivano Bertini, CERM, University of Florence, Italy
Alex Grishaev, National Institutes of Health, Bethesda, USA
Lee Makowski, Argonne National Laboratory, Chicago, USA
Antonio Rosato, CERM, University of Florence, Italy
Michael Sattler, European Molecular Biology Laboratory, Heidelberg
Dmitri Svergun, European Molecular Biology Laboratory, Germany
Peter A. Timmins, Institut Laue-Langevin, Grenoble, France
Matthias Wilmanns, European Molecular Biology Laboratory, Hamburg, Germany

Future challenges in structural genomics and structural biology

The genomic revolution prompted the need for characterizing the products of the human genome and of an ever increasing number of organisms. The knowledge of their structure is a fundamental and necessary step for the understanding of their molecular function and the mechanisms of life of any living organism at the molecular level. In the meantime, the investigation of the whole network of interacting proteins produced by a living organism (the 'interactome') and of how these interactions are being modulated by changing concentrations of metabolites (the 'metabolome') has become an emerging field of research (1,2). Interactome studies provide information on the biochemical processes which occur through protein-protein and protein-DNA/RNA interactions. Such interactions are often weak presumably because they have to be "reversible" in order to repeat the operation whenever and as many times as needed by the organism. Sometimes protein-protein interactions are strong ($K_D < 10^{-6}$ M), but must be triggered by other events e.g. as in the case of EF-hand proteins where the triggering event is calcium binding. Protein-protein and protein-nucleic acid interactions are relevant to several aspects of the physiology of living cells, such as transcription, DNA repair, RNA processing, regulation of gene expression, cell surface recognition and adhesion, signal transduction, to name a few. Malfunctioning of these processes often is linked with, or can lead to, pathological states. The understanding of the molecular bases of diseases

requires also an atomic-level knowledge of the mechanisms of interaction between different biological macromolecules. Therefore, one of the next major challenges for structural biology will be the integration of structural knowledge at the cellular level in the context of systems biology. A major, further step ahead in the comprehension of the processes of Life will be deciphering the human interactome, which will be completed soon. While this knowledge would contribute to advance the comprehension of functional processes, it will also open a large range of new questions and unresolved problems based on the description of the interactions at the atomic level. These will constitute new challenges for scientists in the next years. Within this frame, NMR can play an essential and unique role.

Usefulness of NMR

NMR, together with X-ray diffraction, is one of two techniques that are at the heart of structural biology. Both are used nowadays to provide molecular structures at the atomic level. Despite structure determination of biological macromolecules by NMR began 30 years later than X-ray diffraction, NMR is now a well established, core technique that contributes about 15% of the structures annually deposited in the Protein Data Bank. Although NMR has some limitations with respect to the size of the molecules and the resolution of the structures, it provides considerably more than mere structural information, i.e. it goes beyond a static picture of the three-dimensional structure of genome products. In particular, it can give functionally relevant information on molecular dynamics and can characterize weak and strong interactions with other biomacromolecules as well as small ligands. Transient intermolecular interactions are crucial for phenomena such as regulation of protein expression, enzyme activation/repression, signalling, etc. NMR is particularly well suited to study weak and transient interactions, as it allows researchers to investigate the systems of interest in solution, which is often the physiologically relevant state, or, for membrane systems, in bi-layers or micelles, which simulate the membrane environment. Protein-protein interactions, when weak, cannot be studied with any other technique at the atomic level. In this respect the results of interaction studies by NMR can provide unique information for structural analysis of the interactome maps.

From a more technical point of view, if NMR chemical shift assignments are available and the structure has been determined by either NMR, X-ray or modeling, then simple NMR chemical shift variations provide information on the protein surfaces which are interacting. Programs are already available for proposing models of the complexes on this basis, making NMR an excellent tool for the characterization of protein-protein as well as protein-DNA/RNA adducts. NMR can be used to determine the structure of the complex in solution, both in the case of strong interactions, when therefore a stable complex is formed, and of weak interactions, when the complex is exchanging with the isolated proteins, thus obtaining an average structure. NMR can characterize internal motions as well as the overall tumbling of macromolecules. Therefore, it can provide also insight into the dynamics of protein-protein adducts, as well as of other intermolecular adducts. The dynamics of adducts can be important for molecular recognition processes (e.g. selectivity or induced fit binding). Changes in the internal motion regime occurring in the partners upon complex formation are also important to understand functional features, and can be fully characterized only by means of NMR techniques.

NMR is also particularly suitable for the study of intrinsically unstructured proteins, as required by their function, and of large protein aggregates, which are the results of pathogenic misfolded states. The role of NMR in drug design is well established. A frontier is represented by the screening of molecules interacting with membrane protein receptors. Finally, the impact of NMR in metabolomics, i.e. the analysis of living organism small molecules, is growing steadily.

The outcomes of Structural Genomics ultimately need to be integrated with cell and systems biology, to be able to describe how a cell functions. This would have a great impact and constitute the background knowledge for understanding pathological states originating from alterations of the necessary cell functional processes and/or the presence of SNPs and mutations. So the outcomes of these studies would be beneficial for human health as they are functional for developing therapeutic treatments and new, more efficient drugs.

For the above reasons, we can anticipate that the role of NMR will become increasingly important. In parallel, the evolution of NMR technical and methodological aspects will contribute to widening the range of systems amenable to NMR studies, in particular by relieving the current limitations in terms of molecular size. NMR can be applied to investigate protein structures and protein-protein interactions, as well as nucleic acid structures and protein-DNA/RNA interactions. These scientific aspects have significant direct implications for the understanding of the information derived from genome sequences, as they directly address not only the question of the 3D structure of gene products but also of their mechanism of function, in particular of how the dynamic interactions between proteins, RNA and DNA work together to define biological pathways.

Scope of the workshop

The workshop was organized with the aim of discussing the potentials of combining NMR spectroscopy with other techniques, especially those relevant for samples in the liquid state, in order to advance the scope and applicability of biophysical methods for structure determination. NMR is now a well established technique for structure determination in solution, i.e. in conditions as close as possible to the physiological ones and in the most recent years it is developing into a relatively HTP method for structure determination. One of the next steps in SG is the structure determination of protein-protein or macromolecular complexes. While efforts in this direction are underway (see for example: http://www.3drepertoire.org/ and http://www.3dem-noe.org), they focus on relatively sturdy complexes involving high-affinity interactions. On the other hand, weak and transient protein-protein interactions are key to most of the functional processes in the cell, but these complexes in most of the cases escape crystallization. In these conditions NMR can be quite powerful as it can characterize weak or transient molecular interactions, where a protein binds in fast exchange with the free form. In this respect, NMR will have a relevant impact in the EC-funded integrated project (IP) SPINE2-COMPLEXES (http://www.spineurope.org/). However, despite the huge advancements seen in the last years, NMR is still limited in the molecular size of the systems that can be studied. Presently, macromolecules and complexes larger than 100 kDa can be hardly characterized by NMR. Therefore, there is the need of integration with other techniques. NMR together with crystallography can provide high resolution structures of individual proteins and protein domains. Complementary techniques such as SAXS, WAXS and SANS (described in more detail in the next sections) allow the study of protein complexes in solution, while cryo-electron microscopy can be employed to provide additional low resolution information on large macromolecular assemblies. Cryo-electron tomography can locate multi-protein complexes within the cell, at a resolution superior to light microscopy. Additional long-range distance information can be obtained from FRET data. Thus, while interactome studies provide coarse and somewhat abstract information about the overall genome, NMR, x-ray and neutron scattering, and cryo-electron microscopy give atomic to molecular level structural information describing cellular processes.

In this workshop we have analysed the above-mentioned techniques in the perspective of their integration with NMR spectroscopy. We discussed their potentialities and limits in this respect, as well as their requirements in terms of experimental conditions and characteristics of the systems investigated in order to define the corresponding ranges of applicability. Finally, we analysed the information that can be obtained from their application in particular to protein-protein

complexes. Focus has been made on possible methods for the assessment and the validation of the results obtained through the above-mentioned techniques.

Description of Complementary Techniques

SAXS

Small-angle X-ray scattering (SAXS) (3) allows one to study native biological macromolecules, from individual proteins to large complexes (1 to 200 nm), in solution under nearly physiological conditions. SAXS is not limited by the molecular size and it is well suited to determine at least the shape of molecules with molecular masses of a few hundred kDa (too large for NMR and too small for cryo- electron microscopy (EM)) in their native conditions. SAXS does not only provide low resolution three-dimensional models of particle shapes. Kinetic SAXS experiments allow one to analyze structural changes in response to variations in external conditions, protein-protein and protein-ligand interactions, and to study kinetics of assembly/dissociation or folding/unfolding. Nowadays, high flux SAXS beamlines at various synchrotrons are essentially intended for time-resolved studies of fast structural transitions in the sub millisecond time region in solution and partly ordered systems with a design goal of SAXS-resolution in the range from 100 to 1-2 nm. The time limit for these measurements is the mixing of the solutions, which is set by the mixing technique used (typically stopped-flow experiments). SAXS requires relatively small amounts of sample (20-50 μl of a solution containing about 1 mg/ml of protein). It should however be taken into account that SAXS may often destroy the sample because of radiation damage. In terms of sample conditions, up to e.g. 4.0 M NaCl is an acceptable ionic strength. Use of denaturants with high electron density such as guanidinum chloride are not recommended because of absorption problems, while urea is generally tolerated.

In the presence of equilibria between species in fast exchange, SAXS cannot discriminate between time averaging and ensemble averaging. Some information can nevertheless be obtained by performing a series of experiments at varying concentrations, e.g. during a titration allowing one to shift the equilibrium and to change the species distribution. SAXS can be performed on the same sample as for NMR. However, the two techniques are affected by sample aggregation at a different extent. Indeed, SAXS is much more sensitive than NMR to this problem, and even a small percentage (5-10%) of aggregated species can be deleterious to data analysis, while it would not be observed by NMR. On the other hand, local small-magnitude structural changes such as those corresponding to conformational exchange can have a significant adverse effect on the NMR spectra, while only marginally affecting SAXS curves.

Initial strategies for the application of SAXS in conjunction to NMR spectroscopy have been already developed. For example, methods for the combination of NMR and SAXS and SANS have been proposed and implemented (4, 5, 6). Here, the utility of NMR residual dipolar couplings (RDCs) is applied to define the relative orientation of structural domains or complex subunits, while scattering data are used to obtain long-range translational restraints. SAXS data thus allow one to reduce the orientational ambiguities of RDC data or to exploit different implementations of restraining potentials. These potentials permit the simultaneous use of SAS and RDC data in structure calculations based on molecular dynamics/simulated annealing methods. Future challenges will be to implement and further improve such protocols for routine structural analysis of macromolecular complexes in solution. For example, it is likely that additional data will be useful and may be required to obtain unambiguous structural models. Other issues are the combined use of SAXS and SANS data, for example, to study protein-lipid or protein-nucleic acid complexes, where contrast matching can be employed to obtain additional information (see below).

SAXS experiments are best carried out at dedicated synchrotron beam lines, taking advantage of their high brilliance. These lines are typically equipped with a thermostated flow-through cell to minimize radiation damage, a CCD area detector and a setup allowing for a variable

sample-to-detector distance. In Europe, Grenoble (ESRF), Hamburg (DESY), Daresbury and Trieste (Elettra) provide access to SAXS users, with funds awarded by the EC. Currently, about 10% of the users of the X33 beamilne of the EMBL in Hamburg optimized for biological solution scattering are from the NMR community, and perform SAXS measurements in order to combine their results with NMR data. As the typical SAXS experiments and their analysis require much less time than NMR studies, and the overall structural information from SAXS is highly complementary to more local NMR information, application of SAXS should become a standard supplementary step in the structural NMR studies of proteins and complexes.

An important issue in SAXS method development is validation of the scattering data and the models resulting from its interpretation. In the absence of any prior structural information that could be compared against parameters directly extractable from the data such as approximate radius of gyration or maximum molecular dimension, one has to rely on the interpretation of the zero-angle extrapolated scattering intensity to confirm both the expected oligomerization state and absence of aggregation. For such measurement, acquisition of the scattering data on the samples of standards such as lysozyme or BSA, coupled with an accurate measurement of the macromolecule's concentration, are required. Validation of the structural models produced by the analysis of scattering data is more challenging and ultimately depends on the availability and amount of the prior structural or biochemical information about the molecule. In cases where SAXS data is combined with NMR data, *ab initio* SAXS-only low-resolution density reconstructions could be compared with NMR/SAXS generated high-resolution models using a measure like normalized spatial discrepancy to assess their similarity. An additional way to validate SAXS-only density reconstruction is by performing such reconstructions with several different software packages such as GASBOR, DAMMIN, SASHA, SAXS3D, etc.

A crucial issue for the incorporation of SAXS data into the main-stream structural biology is standardization of the data formats and deposition of the SAXS data into a publicly accessible database. One possible way to approach this is by requiring data deposition upon publication of any work that involves structural interpretation of SAXS data into a database such as RSCB PDB.

SANS

Small angle neutron scattering (SANS) is conceptually very similar to SAXS but uses neutron radiation and exploits important differences between their scattering properties and those of X-rays. The most important factor is the scattering length density, the neutron analogue of electron density for X-rays. Neutrons are scattered by the nucleus of an atom rather than the electronic cloud as is the case for X-rays. The scattering process for neutrons involves the formation of a complex nucleus with subsequent liberation of a neutron and is a complex process which is not related in any simple way to the mass or atomic number of the atom contrary to the X-ray case. Thus the scattering length densities of heavy atoms are not necessarily any greater than those for light atoms and can indeed be different for different isotopes of the same element which, by definition, contain a different number of nuclei and therefore have a different nuclear structure. In practise the scattering length densities for neutrons of the atoms common found in biological molecules vary only from 0.54 for carbon to 0.94 for nitrogen all other atoms, with one exception, being between these extremes. The striking exception is hydrogen which has a negative scattering amplitude and is hence easily distinguished from all other atoms. The other important factor is the scattering power of deuterium, an isotope of hydrogen, which is positive and has a scattering amplitude very similar to that of carbon (0.66). Substitution of hydrogen by deuterium is therefore a powerful labelling technique in all domains of neutron scattering, from crystallography, through SANS to inelastic scattering. Concentrating on SANS, the difference between hydrogen and deuterium scattering leads to the contrast variation method whereby mixtures of H_2O and D_2O have very different scattering density as do proteins, nucleic acids or lipids. Similarly if biological macromolecules can

be perdeuterated then their scattering is very different from their hydrogenated counterparts. Via the contrast variation method it is possible to perform SANS experiments on a macromolecular complex in aqueous solution where one component of the complex can be rendered invisible by choosing an appropriate H_2O/D_2O mixture such that solvent and that component have the same scattering density and hence zero contrast. Hence in this way the structure of different parts of a complex can be determined individually without dissociating the complex. There are numerous examples of this such as the study of the troponin complex[1] where samples having any one or two or all three sub-units could be deuterated and by carrying out SANS experiments in solutions containing 42% D_2O/58% H_2O it was possible to visualize, and hence model, the deuterated sub-units alone. Another potentially important class of molecules which can be investigated using this method is membrane proteins solubilised in lipids or detergent micelles. Lipids or detergents are invisible in water containing no or very little Deuterium and thus a neutron scattering experiment in these condition leads to a structure of the membrane protein alone without its solubilising layer of lipid or detergent.

Sample requirements for SANS are similar to SAXS requiring a slightly greater sample volume, 150 – 300ul per sample multiplied by the number of contrasts. There is however, no radiation damage with neutrons which avoids the necessity for multiple or circulating samples. Typical concentrations are 1 – 10 mg/ml. Small molecule additives such as salts or denaturing agents such as urea have no effect on the contrast so long as they do not form nanometre sized aggregates such as micelles. Experiments are usually performed in standard quartz cuvettes which may also be used for UV measurements to determine sample concentration. Exposure times are very variable depending on sample concentration, macromolecular dimensions as for X-rays but also on contrast at signal to noise factors. The highest contrast and hence strongest signals are obtained from deuterated proteins in H_2O or hydrogenated samples in D_2O. The latter conditions provide the lowest background which comes mainly from the incoherent scattering of hydrogen atoms.

Data analysis is in most respects similar to that used for X-rays, particularly in model fitting as exemplified in the methods described above by Svergun and colleagues. The particularity is the increased information content available due to contrast variation leading to several different scattering curves being obtained from the same molecular complex. Interpretation at a simpler level, that of radius of gyration, can also give important information as the contrast variation method allows distances between sub-units of a complex to be determined in a model independent way. Data may also be placed rather easily on an absolute scale leading allowing molecular weight measurements to be made.

The complementarity with NMR lies, as with SAXS, in the possibility of defining an overall molecular envelope delimiting the overall structure of the molecule or molecular complex. Unlike a crystallographic structure it is measured in solution, as with NMR, and the envelope obtained is a space and time average over all solution conformations.

SANS experiments can of course only be carried out at a neutron source which is without exception a central facility – no laboratory neutron sources suitable for scattering exist. Indeed the major difference between neutron and X-ray facilities is the much smaller flux available at even the most powerful sources compared with synchrotron radiation. This is however compensated in part by bigger beams and the flexibilities introduced by the contrast variation method as described above. The most powerful SANS facilities currently available in the world are the D11 and D22 SANS instruments at the Institut Laue-Langevin (ILL) in Grenoble, France. The ILL has the world's most powerful research reactor and is a facility available to all its member countries and even to a limited extent to non member countries. Set up in 1967, the reactor started operation in 1972 and is owned and run by France, Germany and the UK. It has also as scientific members, Austria, Belgium, the Czech Republic, Hungary, Italy, Russia, Spain, Sweden and Switzerland. Beam time is allocated by peer review committees twice a year and experimenters from member

countries who have their experiments selected also have the travel and subsistence paid to carry out the experiments. More details can be obtained from http://www.ill.fr. The ILL also has a laboratory specialising in the production of deuterated macromolecules for neutron scattering experiments. This laboratory is also open to outside users in a way similar to the neutron instruments except that treatment of proposals is entirely by e-mail and application may be made at any time. Further details are available at http://www.ill.fr/deuteration.

SANS facilities are operated also at several other neutron sources in Europe, the USA, Japan and Australia. There are both reactor sources such as the Laboratoire Leon Brillouin at Saclay in France or the FRM-2 reactor in Munich Germany and spallation sources such as ISIS at the Rutherford Laboratory, UK or the SNS in Oak Ridge, USA, the latter currently under construction. All these facilities have external user programs and the list is by no means exhaustive.

WAXS

Wide-angle x-ray scattering (WAXS) from proteins in solution is an emerging technique which is a direct extension of SAXS derived through the collection and analysis of data to significantly higher scattering angles. Whereas most SAXS studies of proteins in solution are limited to $q < 0.2$ Å^{-1} ($q = 4\pi \sin(\theta)/$and 2θ is the scattering angle), WAXS studies have extended that limit to collect the weak data that extend to $q \sim 3.0$ Å^{-1}. The amount of information in solution scattering patterns is proportional to the resolution limit, so that, in principle, WAXS patterns may contain up to 15 times as much structural information as a SAXS pattern. As outlined above, SAXS data can be used to generate information about the size and shape of a protein. WAXS data provide information about finer details of the structure since scattering at progressively wider angles corresponds to progressively smaller structural features within the protein. X-ray scattering data from proteins in solution correspond roughly to the spherical average of data collected by x-ray crystallography. Thus they preserve information about the relative abundances of interatomic vectors of different lengths, but contain no information about their relative orientations. At very small angles of scattering (i.e. low q), long inter-atomic vectors dominate, limiting the use of small-angle x-ray scattering (SAXS) to determination of protein size and shape. At wider angles (i.e. higher q), shorter inter-molecular vectors contribute proportionately more to the intensity, embedding information about protein secondary and tertiary structure in the weak wide-angle scattering. As such, a WAXS pattern represents a weighted mapping of all the interatomic vector lengths in a protein. Since secondary structural elements such as α-helices or β-sheets have characteristic patterns of interatomic vector length, WAXS data can provide information about the secondary, and potentially, tertiary structure of a protein.

A major advantage of WAXS is that WAXS data can be predicted from atomic coordinates. A number of software packages that carry out this calculation are available, with CRYSOL being the most widely used. This capability makes it possible to test structural models based on atomic coordinates against experimental WAXS data. For instance, if the structure of two proteins are known but the way in which they interact is unknown, WAXS data from a solution of the complex can be used to test models for their interaction by comparing experimental data to that predicted from the atomic coordinates of the individual proteins and their relative positions and orientations. Consequently, WAXS data from complexes that cannot be studied by NMR or x-ray crystallography may provide adequate information for construction of a molecular model of the complex. WAXS is also well suited to the study of conformational changes in macromolecular complexes. Changes in WAXS data reflect changes in quaternary, tertiary or secondary structure, providing them with the potential for being a sensitive, global method for detecting structural changes in proteins induced by ligand binding, changes in environment or interactions with other macromolecules. WAXS is a sensitive tool for detection of the structural changes in proteins that

accompany functional changes, thereby providing an indirect assay for functional interactions with small molecules.

The use of WAXS has, in the past, been limited by the relative weakness of scattering from proteins at relatively wide angles, and by the presence of strong scattering from buffer solutions in this range. More recently, however, third-generation synchrotron sources, such as the European Synchrotron Radiation Facility (ESRF; Grenoble) and the Advanced Photon Source (APS; Chicago) have made possible the collection of wide-angle x-ray scattering (WAXS) of macromolecules in solution to unprecedented resolutions. Although wide-angle x-ray scattering data are significantly weaker than the small angle data collected using SAXS, they can generally be collected using less than 200 μl of a solution with protein concentrations of 5–10 mg/ml in less than 30 s with the high flux available at a third generation synchrotron X-ray source. Higher concentrations result in data with higher signal to noise ratio. In general, WAXS data is less sensitive to the effects of aggregation than SAXS data. That, plus the need to measure much weaker data than in SAXS experiments, makes the use of more concentrated solutions far more favorable than in SAXS experiments. Care must be taken to avoid radiation damage to proteins studied by WAXS. Radiation damage and/or radiation-induced aggregation can be averted by a number of experimental parameters, including flowing the protein sample through the beam during data collection.

WAXS provides data complementary to NMR data in a manner similar to that provided by SAXS data. WAXS data contains information about the relative abundance of interatomic vectors that vary in length from ~ 2Å to the diameter of the scattering particles. As such, it provides information about the global arrangement of material in the protein or protein complex. It is not, in general, possible to obtain information about a specific inter-atomic vector from WAXS data. But the data provide an accurate overall view of the distribution of interatomic vector lengths in the protein which can provide important constraints on the form of the molecular model constructed from NMR data.

FRET

Fluorescence Resonance Energy Transfer (FRET) imaging is a powerful microscopy technique that overcomes some of the usual limitations of light microscopy to allow researchers to visualize and quantify protein associations under physiological conditions in individual cells. Conventional widefield fluorescence microscopy enables localization of fluorescently labelled molecules within the optical spatial resolution limits defined by the Rayleigh criterion, approximately 200 nanometers (0.2 micrometer). FRET, when applied to optical microscopy, permits determination of the approach between two molecules within several nanometers, a distance sufficiently close for molecular interactions to occur. The introduction of the green fluorescent protein (GFP) to FRET-based imaging microscopy allowed the use of this technique as a sensitive probe of protein–protein interactions and protein conformational changes *in vivo*. This was the beginning of real-time *in vivo* imaging of dynamic molecular events, providing researchers with crucial insight into the biological mechanisms as well as the physiological functions of a cell. This technique allows studying interactions that occur in the pico and nanosecond time scale. However, it can also be as a powerful tool to characterize conformational changes, both in vitro and in vivo, thereby providing a unique link to correlate atomic resolution structural information with biological function in a living cell. In many cases qualitative information derived from FRET changes can suffice to provide an understanding of conformational changes.

While examples are available of FRET measurements used in combination with X-ray structures, few examples are available for the combined use of FRET and NMR data applied to structural analysis of macromolecular complexes. NMR is being used to characterize and study the

molecular basis of a newly designed FRET probe that exploits a phosphorylation-induced conformational switch.

FRET data provide local information on protein-protein distances (at longer distances than NMR, typically tens to hundreds of Å). As a consequence, these data are best suited for the validation of a structural model of an adduct or to suggest possible ways of interaction. The application of FRET requires that the protein partner can be chemically modified to introduce on specific residues (generally engineered cysteine residues for proteins) a pair of donor and acceptor fluorophores. This is needed in order to be able to measure a residue-residue distance within the complex in solution. While in some instances this kind of information can be of invaluable importance, it is unlikely that FRET can become a standard technique for the high-throughput structural characterization of complexes.

Cryo EM

Single particle reconstruction methods from EM of frozen-hydrated samples allow the structural analysis of macromolecular objects of molecular weights > 150 kDa. While resolutions between 10 and 15 Å can be achieved routinely, higher (< 10 Å) resolution can only be attained with advanced instrumentation and extensive image classification and averaging. Higher (atomic) resolution structures of components (subunits, domains) derived from crystallography and/or NMR can be fitted into the medium resolution structures of complexes obtained by cryo EM. Image reconstruction relies on averaging of large numbers (up to 100,000) of individual particles. When there are sample inhomogeneities or when different conformational states coexist, image classification must precede averaging. EM of frozen-hydrated samples ensures a close-to-life preservation at the expense of a low signal-to-noise ratio. Negative stain provides high contrast and is therefore less demanding in terms of data processing but it is prone to artefacts and yields only low-resolution structural information (20-30 Å).

Cryoelectron tomography is the only technique that enables the study of large pleomorphic structures such as organelles or whole cells embedded in vitrified ice and, therefore, in a close-to-life state, with a resolution of a few nanometers. With the advent of computer-controlled transmission electron microscopes equipped with eucentric goniometers, and the availability of highly sensitive CCD cameras, it became possible to develop and implement automated image-acquisition procedures, which operate in low-dose conditions, thereby keeping the dose applied to the sample as low as possible. A major limitation in electron tomography is specimen thickness and the limited tilt-range, which allows the recording of only 70% of the necessary data in order to obtain a complete reconstruction. Nevertheless the information provided in the cryoelectron tomograms is sufficient to recognize distinct macromolecular complexes. The macromolecules are embedded in their natural environment and their density distribution is not affected by staining reactions, which tend to produce aggregations, and therefore compromise their molecular interpretation.

X-ray crystallography

X-ray crystallography provides the most powerful technique for the determination of structures of biological macromolecules. There is virtually no size limit for targets to be investigated, as long as they fold into a defined three-dimensional structure. The method is equally applicable for routine applications, for instance in the context of structural genomics projects, and for challenging multi component complexes. Advanced protocols have been developed for routine structure determination of integral membrane proteins as well as for complexes with partially unfolded protein components.

X-ray crystallography has enormously advanced during recent years because of major efforts to

provide state-of-the-art data acquisition facilities at 3rd generation synchrotron radiation facilities, accompanied by developments and implementations of highly automated pipelines in terms of hardware and software provision. These efforts need to continue to ensure competitiveness. Because of the advances in automation and user-friendliness, translating into remote forms of execution of experiments, we are observing a decrease of a profound knowledge about the underlying theoretical foundations of this technique. It will be essential to keep a critical mass of experts in the field with the ability for further methods and technology developments, as well as to provide advanced methods-oriented training opportunities for young researchers.

At a time where an emerging key objective by the scientific community is to unravel functional/structural relations within entire biological systems, there is an increasing and urgent need to confirm and to validate high resolution molecular data by complementary *in vitro* and *in vivo* imaging methods as well as by functional assays. While the first category (presented by electron microscopy, electron tomography, small angle X-ray scattering, for instance) generally does not require fundamentally different sample preparation facilities, the application of *in vivo* validation methods generally requires access to cell culture facilities and specific know-how on respective applications.

Probably the most serious caveat of X-ray crystallography is its limitation to static structures, which is natural because of the requirement for fixed, diffracting crystal lattices. Therefore, in order to capture the dynamics of a given structure as a function of external parameters of a given biological system, it appears to be essential to combine X-ray crystallography either with experimental or computational structural biology methods to explore dynamic processes. At the molecular level, NMR spectroscopy provides a leading complement, as long as the method is feasible for a given biological system. At the cellular and/or whole-organism level, state-of-the-art complements are provided by life-video imaging methods.

Concluding Remarks

From the analysis of the applications of the techniques addressed during the workshop, and in particular of their possible integration with high-resolution NMR spectroscopy for the structural characterization of macromolecular complexes, a few conclusions can be drawn:

- SAXS and SANS can be integrated with NMR data with relative ease. Indeed, some pioneering protocols and programs have been developed. These techniques have the features and potentialities to attain high-throughput application in the structural characterization of complexes;
- WAXS is stilly in its infancy as far as application to biological macromolecules (and their adducts) is concerned. In principle, it can be well integrated with NMR spectroscopy, and can have a precious role thanks to its ability to provide information on conformational changes and aggregation modes of proteins and nucleic acids. Further methodological developments are needed to bring WAXS to a high-throughput level of application.
- X-ray crystallography and NMR are already employed as complementary methods, where structural data from crystallography are complemented by information concerning intermolecular interactions and dynamics obtained from NMR. Some examples of joint structural refinement against crystallographic and NMR data have been reported. Structures obtained from crystallography, but also from NMR or homology modelling, can be used to obtain structural models of protein complexes when combined with NMR data and computational approaches.
- Cryo EM and cryo electron tomography are complementary to NMR analysis. For example, structures and protein interfaces determined by NMR can be fitted into electron densities derived from EM. However, a direct combined application of the two methods is more difficult given the different size preferences.

- A direct combination of FRET data with NMR is less practical, since very different experimental conditions and, for FRET, extensive sample modifications are required. FRET is of potential interest because it provides long-range residue-residue distance information. Indeed, similar to crystallographic structures, NMR derived structural data may be validated by FRET experiments in vivo. Moreover, NMR can be used to characterize the intermolecular interactions of FRET probes, e.g. also in terms of structural perturbations induced. FRET however will probably remain a system-specific technique for at least some time.

In summary, the currently available techniques for the structural investigation of macromolecular complexes have been experiencing tremendous developments in the last few years. These advances allow scientists to tackle the study of more and more challenging systems, such as weak and/or transient protein interactions. These adducts are often hardly addressed with a single method but they can benefit of the integration of a variety of methods able to characterize different properties, scales and aspects of macromolecular complexes.