

## Contributions of NMR spectroscopy to the structural characterization of complexes within the 3D-Repertoire project.

The NMR groups at EMBL and CNIO are involved in the characterization of several complexes in collaboration with crystallography groups. Besides this direct participation in specific complexes, the NMR groups can contribute some biochemical and biophysical characterization of the individual components of complexes that could help to modify and optimize crystallization strategies. For this later aspect, interested participants should contact the NMR groups directly. Results of such NMR sample “screens” should then also be documented in the database.

- 1) Structure determination and ligand binding studies of individual components of target complexes using established standard methods. Targets for NMR are more loosely associated subunits and/or those, which resist crystallization.
- 2) Check on the structural integrity of the individual components. A protein could be unstructured in isolation but become structured upon binding to their partners in the complex. If the complex has been unambiguously validated by mass spectrometry and electron microscopy there should be no reasons for concern about that, but if validation is ambiguous and if the protein constructs do not span the whole protein, the investigation of its folded state could help make decisions about how to proceed (new constructs, identification of missing components). In addition, it might be possible to monitor the acquisition of structure (or not) by an isolated unfolded protein upon mixing with their partners in the complex. Checking the folded state and potential oligomerization can be done for proteins of up to 100kDa solubility higher than 50  $\mu$ M and at least 1 mg of protein for natural abundance samples. This is based on 1D experiments and estimation of amide H<sup>N</sup> relaxation times, with the precaution that the larger the mass of the protein the more ambiguous the result will be. Since many proteins will be expressed in *E. coli*, isotopically labelled samples can be prepared (by the NMR groups) that provide more conclusive results and provide a starting point for the characterization of backbone dynamics and ligand binding (see below). Protein samples can be checked for the presence of long unstructured segments that might block crystallization. Again, the unstructured region in the isolated protein could become structured upon binding and not be reason for crystallization failure, but its removal might be a way to move forward in case of being stuck with unsuccessful crystallization trials. Although in the general case we will not be able to identify which region is flexible, just noting its presence, combined with sequence analysis by the participating biocomputing groups, could suggest modifications (i.e. new constructs). Similar experimental requirements than mentioned above apply.
- 3)
- 4) When two or more proteins are known to bind each other but no crystal is obtained, or only a subcomplex crystallizes, NMR can be employed to map the binding surface and the relative orientation of the components. This may be even possible in case of transient, dynamic binding. To be able to do this, it is necessary that the protein sample could be prepared with <sup>15</sup>N(<sup>13</sup>C) and <sup>2</sup>H isotope enrichment, which requires expression in *E. coli*, with a solubility of 0.2 mM or higher, and a molecular weight of the complex <100-150 kDa possibly higher. This strategy has been successfully applied by Wüthrich and coworkers

to map the binding interface in the GroEL/GroES oligomer of 400-900 kDa molecular weight.

- 5) Ligand binding studies using potential small molecule inhibitors can be applied for individual subunits as well as for subunit-selectively labelled complexes (with one subunit  $^2\text{H}$ ,  $^{15}\text{N}$  labelled) based on chemical shift perturbation and other methods. This will allow the identification and mapping of binding sites.

Alternatively, ligand binding experiments can be employed, in which the ligand is detected. These methods require ca. 50  $\mu\text{M}$  unlabelled protein/or complex of without molecular weight limitation (!) and are thus more generally applicable. Both NMR based methods are very efficient and especially useful since they also allow the detection of relatively weak interactions.

---

?

The text below is copied from the technical annex as a reminder.

#### *NMR spectroscopy*

**3D-Repertoire** will also explore strategies of biomolecular NMR in solution, which provides opportunities to characterize the folds and molecular interfaces of proteins and nucleic acids. Due to recent advances in NMR hardware, as well as experimental strategies and isotope labeling it is now becoming possible to study large proteins and protein complexes, i.e. with molecular weights well above 100 kDa. **3D-Repertoire** will use NMR to (1) determine the three-dimensional structure of individual domains and smaller globular recombinant proteins that are constituents of these complexes and (2) characterize the molecular interfaces in reconstituted complexes where subunit-selective isotope labeling is employed. The NMR-based mapping of molecular interfaces and ligand interactions is complementary to crystallographic and EM analysis. It will provide additional information about the molecular interfaces, determine potential conformational changes upon complex formation, characterize the protein dynamics and ligand binding.

#### *NMR spectroscopy and X-ray crystallography for the characterization of small molecule binding.*

NMR and X-ray crystallography will be used to identify small molecules that interact with and thus regulate the molecular functions of these complexes. X-ray crystallography will use classical techniques, such as soaking and co-crystallization, to characterize small molecule binding sites. NMR will be employed to identify binding sites using NMR chemical shift perturbation.