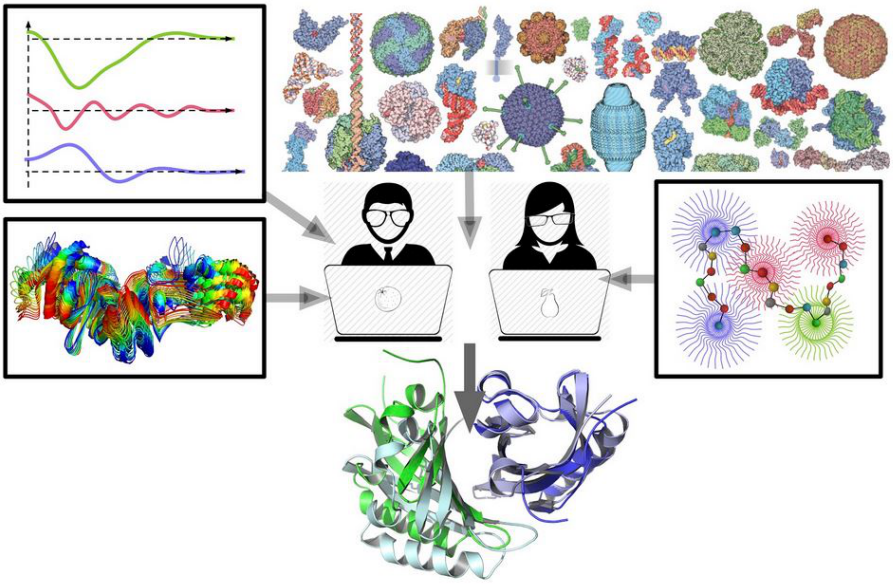
CANCELLED : Algorithms for integrative structural biology



Report of Contributions

https://workshops.ill.fr/e/261

Contribution ID: **1** Type: **Poster**

Studying Protein Structure through Hydrogen Exchange and Coarse-grained Conformational Sampling

Experimentally observing and/or computationally modeling large proteins and macromolecular complexes remain critical challenges for structural biology. Hydrogen-exchange monitoring is cheap and easy to carry out, but cannot produce structural models because of its low resolution. To mitigate this issue, one side of our coupled approach consists of developing computational methods to complement such low-resolution experimental techniques. As these computational methods suffer from the curse of dimensionality when applied to large molecular systems, the other side of our coupled approach consists of guiding them with experimental data.
Here, we present three applications of our coupled approach combining hydrogen exchange on the experimental side and a robotics-inspired method for coarse-grained conformational sampling on the computational side. First, we argue that using coarse-grained conformational sampling of protein structure improves the fit between computationally-generated conformations and experimental hydrogen-exchange data. Second, we show that our approach allows analyzing the variability of a protein's native state described by crystallographic and hydrogen-exchange data. Finally, we explain how to obtain an atomic-resolution structural model of a protein state for which only hydrogen-exchange data is available.

Primary author: DEVAURS, Didier (INRIA)

Presenter: DEVAURS, Didier (INRIA)

Contribution ID: **2** Type: **Oral**

Computer simulations of intrinsically disordered proteins - What are we missing?

Intrinsically disordered proteins (IDPs) lack well-defined three-dimensional structure in solution under physiological conditions. Despite this they are functional and participate in the regulation of many biological processes, in which disorder can enable interactions of high specificity coupled with low affinity. Phosphorylation is one of the most abundant types of post-translational modifications of IDPs. The addition of a phosphoryl group can act as a regulatory mechanism, for example by inducing changes in secondary structure or the association state. Computer simulations and modelling in combination with small angle X-ray scattering (SAXS) is fruitful approach to achieve a molecular understanding of the underlying physics of a system. In this talk I will present how atomistic and coarse-grained modelling in combination with Monte Carlo and molecular dynamics simulations can be used to understand and predict the structure and solution behavior of IDPs, but more importantly, focus will be on "what are we missing". After setting the stage and introducing the models used for the monomeric IDPs under dilute conditions, four different cases will be discussed:

1. Temperature effects
2. Phosphorylations
3. Self-association
4. Crowding

Throughout the talk, comparisons will be made between models with different degrees of details, and SAXS-experiments. The talk is based on the experience from research performed in Skepö research group during the last ten years.

Primary author: Dr SKEPÖ, Marie (Div. of Theoretical Chemistry, Dept. of Chemistry, Lund University, Sweden)

Presenter: Dr SKEPÖ, Marie (Div. of Theoretical Chemistry, Dept. of Chemistry, Lund University, Sweden)

Session Classification: Session 8

Contribution ID: **3** Type: **Oral**

Fragment-based docking to tackle the ssRNA flexibility

Protein-RNA interactions are involved in many biological processes like the traduction of messenger RNA to protein, and their modelisation is important to understand them. In particular, non-paired regions of the RNA, i.e., single-stranded RNA (ssRNA), are involved in most of these interactions and are essential for their specificity.

However, ssRNAs are highly flexible and their conformational space can not be sampled exhaustively enough for docking by using standard algorithms such as harmonic analysis, template-based modeling or MD simulations.

Here, we present a fragment-based approach for ssRNA/protein docking, developed to address the limitations described above. It includes:

- the building of a conformer library of ssRNA 3-mers;
- the docking of these ssRNA 3-mers on the protein;
- the assembling of the docked solutions on geometric criteria;
- in case of ssRNA loops, the new implementation of distance constraints in order to assemble them in a more efficient way;

We will present the results on several cases and discuss the advantages and the limitations of such an algorithm. Indeed, such a method has the advantage to sample ssRNA flexibility for the docking [Chauvot de Beauchene Isaure et al., PloS 2016] but further optimization is required due to the challenge to sample the enormous size of the ensemble of possible fragment-chains.

Primary authors: MONIOT, Antoine; ROY, Rohit; Dr CHAUVOT DE BEAUCHENE, Isaure (CNRS)

Presenter: MONIOT, Antoine

Session Classification: Session 10

Contribution ID: **4** Type: **Oral**

Building a model of an homodimeric multi-domain protein, interacting with ssDNA and dsDNA : The challenges of RelSt3 relaxase modeling

Spread of antibiotic resistance genes by conjugative mechanisms represent a major public health issue. Among the elements involved in these horizontal genes transfers, mechanisms involving integrative and conjugative elements (ICEs) are the most common. The RelSt3 relaxase is a protein that binds to a sequence-specific dsDNA, and to the ssDNA region of a hairpin to cut it prior to its transfer. RelSt3 is active in the form of an homodimer, each monomer formed by 3 domains linked by highly flexible long loops

This complex presents several challenges for its modeling. For example, while the locations of the ssDNA nick site and the RelSt3 catalytic site are both known, the hairpin loop is highly flexible due to its single stranded nature. Thus its structure is not known, and moreover can change upon binding with the relaxase. Additionally, each RelSt3 monomer is made of three domains : the n-terminal "HtH" domain binding a sequence-specific double-stranded DNA, with a conserved fold; the central domain containing the catalytic site cutting the single-stranded DNA "nick site", and for which an homologous structure is available; and the c-terminal domain, with an unknown fold.

Here, I want to expose the processes set up to build a model of this protein and of its interactions with ss- and dsDNA. The steps includes :

• Modeling of the protein domains; Of the three domains of RelSt3, two could be built by homology modeling. However the third, while consensually described as primarily organized in alpha helices, resist to our modeling attempts.

• Building of the dimeric core domain, based on the dimeric thomologous structure, which also lacks the n-terminal and c-terminal domains.

• Linking the domains; The three domains of RelSt3 are link to each other through long (10 and 13 aa) flexible linkers, leading? to a large conformational space. On the other hand, a precise location of these domain relative to each other is necessary to bind DNA accordingly onto known sites.

• Localizing the HtH domains relative to each other on dsDNA; It was experimentally shown that HtH domains interact with two short (13 nt) repeats, separated by a 9 nt spacer. Due to the symetric nature of the protein, it is suggested that the HtH domains interact with different repeats. This carries out additional constraints to take into account for their positioning on the DNA, relative to each other and relative to the central domain.

• Modeling the ssDNA conformation in contact with the protein; The precise interactions sites are known for both partners. But the fragment-based approach considered here, in which ssDNA sections are fragmented into 3-mers for an efficient docking, was developed for ssRNA/protein docking. It is limited here by the low number of ssDNA/protein complexes in the PDB used to build a fragment library, compared to dsDNA/protein complexes.

The main objective is to open a discussion on some of the challenges encountered during the steps presented above with the people present at the workshop, about ways to overcome them.

Primary author: MIAS-LUCQUIN, DOMINIQUE (LORIA)

Presenter: MIAS-LUCQUIN, DOMINIQUE (LORIA)

Session Classification: Session 4

Contribution ID: **5** Type: **Poster**

Usage of the chemical-crosslink experiment data in a protein structure modeling

Prediction of protein structures is one of the most important problems of current structural bioinformatics. Many proteins in living organisms form multimeric structures or are involved in the complexes with other biomacromolecules, such as other proteins or small ligands. Prediction of protein structure in the monomeric state is already very challenging and therefore the structures of their complexes are even more tricky to determine. There are several experimental techniques whose results can assist in protein structure prediction. For example, chemical crosslinking/mass spectrometry (XL-MS) and small-angle X-ray scattering (SAXS) experiments provide solution data describing approximated shape of molecular system. Such data can be used as a potential term and be included in the energy function in molecular simulations or as a scoring function. Big advantages of SAXS and XL/MS, comparing to the crystallography and NMR measurements, are that they are not that demanding in terms of time, money and sample specificity requirements. This study presents our progress in integrating the XL/MS data in protein structure determination by the probability estimation and rigid-body docking.
In our approach, we integrated information from crosslink experiments to a combination of a statistical potentials derived based on the distance distributions from reference 1 (residue-specific XL, big-X) or based on the calculated actual distance distributions of Cα atoms in experimental structures for restraints provided by previous CASP experiments (CASP11-13, residue-non-specific XL, small-x)2. Depending on the distribution, the data were fitted either using a single Gaussian or a sum of two Gaussians. The function parameters were obtained by a scaled Levenberg-Marquardt fitting algorithm of the specific distance distributions of each XL type. We made a Boltzmann-like hypothesis and considered that there is a pseudo-potential associated with each of the XL constraints, whose value is given by the logarithm of the probability of a certain Cα-Cα distance. Finally, we combined these potentials with a rigid-body docking methods and can conclude that XL-MS data is very useful in the prediction of protein complexes, if the number of false distance restraints is not very high.

Leitner A, Joachimiak LA, Unverdorben P, et al. Chemical cross- linking/mass spectrometry targeting acidic residues in proteins and protein complexes. P Natl Acad Sci U S A. 2014;111(26):9455-9460.
Kryshtafovych, A, Schwede, T, Topf, M, Fidelis, K, Moult, J. Critical assessment of methods of protein structure prediction (CASP)—Round XIII. Proteins. 2019; 87: 1011– 1020.

Primary authors: Dr KARCZYNSKA, Agnieszka (Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France); GRUDININ, Sergei

Presenter: Dr KARCZYNSKA, Agnieszka (Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France)

Contribution ID: **6** Type: **Poster**

An investigation of the interactions between clay nanoplatelets in the presence of monovalent and divalent salts, as well as cationic peptides utilising atomistic molecular dynamics

Clays are negatively charged nanoplatelets with a layered silicate structure. It has been observed from small angle X-ray scattering measurements, and coarse-grained molecular dynamics simulations (CG-MD) that clay nanoplatelets are able to form stacks, known as tactoids, in the presence of divalent or multivalent counterions. In this thesis, the interactions between clay nanoplatelets and counterions of varying valency was studied by the use of atomistic MD, which are of higher resolution than CG-MD. The particular systems studied here are the clay mineral montmorillonite (MMT) with sodium or calcium as counterions, as well as the addition of deca-arginine (Arg_{10}), which is a cationic antimicrobial peptide (CAMP). Experimental studies and CG-MD, have shown that clays are able to sequester CAMPs within tactoids, thus indicating that it can be used as drug delivery vehicles for CAMPs. The aim of this thesis was to investigate whether atomistic MD can be applied to study freely moving clay nanoplatelets and their interactions with CAMPs, to aid in the acquisition of knowledge of such systems.

For this purpose the thesis is comprised of three studies of increasing complexity, simulating MMT systems with atomistic MD using the force field CLAYFF with the SPCE water model. In the first study, the distribution of sodium and calcium ions in the interlayer between MMT surfaces was studied, where a qualitative agreement was found with previous studies. In the second study, two freely moving MMT nanoplatelets were simulated in the presence of either sodium or calcium ions. It was seen that the platelets had attractive edge-to-face interactions regardless of the cation valency, however attractive face-to-face interactions only occurred at high calcium content, above 100 mM, for the platelets that were initially placed in close proximity. Lastly, in the third study, systems of MMT and Arg_{10} were studied. The Arg_{10} chains were modelled with the CHARMM force field, and two water models were used, SPCE and TIP3P. The number of Arg_{10} chains that adsorb to the platelets increased with concentration until saturation occurred, where the total charge of the Arg_{10} chains exceeded the platelet charge, resulting in overcharging, in agreement with experiments and CG-MD. In addition, it is seen that individual Arg_{10} chains are able to interact with multiple platelets simultaneously, which may lead to intercalation within tactoids. However, atomistic MD of multiple platelets with Arg_{10}, did not exhibit tactoidal formation. Therefore, the process behind tactoidal formation is likely more complex than what CG-MD suggests. Although, the simulations here indicate that atomistic MD may capture tactoidal formation given the right circumstances, such as the initial configuration. Further research is required to receive conclusive results of these systems, however, this thesis has shown that atomistic MD is a promising method for studying clay minerals and CAMPs, and may elucidate the complex interactions involved, which is difficult to capture with experiments and not available at atomistic resolution with CG-MD.

Primary authors: KODER HAMID, Mona (Master student); Ms JANSSON, Maria (Doctoral student); SKEPO, Marie

Presenter: KODER HAMID, Mona (Master student)

Contribution ID: **7** Type: **Poster**

SAXS-guided coarse-grained modeling of viral RNA structures through an integrative method

Since the conformation of RiboNucleic Acids (RNA) is a key for its function, experimental techniques at different levels of resolution are implemented to determine its structure. Because of the flexibility and the large size of certain RNAs, their structure cannot be determined through high resolution experiments, such as nuclear magnetic resonance (NMR) and X-Ray crystallography. It is however possible to obtain small angle scattering X-Ray (SAXS) data for large systems in solution that can hint to the structure of the molecule.
The aim of my work is to develop simulation methods to couple SAXS data with molecular modeling tools and high resolution experimental data, allowing the construction of an entire model of RNAs and RNA-protein complexes, similarly to what is done for proteins[1].
We adopt coarse-grained representation of the system with 6-7 beads per nucleotide that allows to study the dynamics of large molecules and for long time scales when used for molecular dynamics simulations[2]. We developed a scheme to compute theoretical SAXS intensity profiles for the coarse-grained systems that gives results of an accuracy similar to CRYSOL[3], without having to resort to an atomistic description that would greatly slow down the dynamics.
The Chi-Squared ($\chi 2$) score (SAXS scoring fonction), giving the agreement between the target experimental intensity curve and the curve computed theoretically on-the-fly in the simulation, is used as a biasing potential generating "SAXS forces" to be added to the force field, and as a Monte Carlo weight for replica exchange simulations. The method is benchmarked on a sample of 4 structures of different sizes and architectures.

[1] S. Yang, S. Park, L. Makowski, B. Roux, Biophys. J. 96, 4449-4463 (2009)
[2] T. Cragnolini, Y. Laurin, P. Derreumaux, S. Pasquali, J. Chem. Theory Comput., 11, 3510 (2015)
[3] D. Svergun, C. Barberato, M. H. J. Koch, J. App. Cryst., 28, 768-773 (1993)

Primary authors: MAZZANTI, Liuba (Université de Paris); ALFERKH, Lina (University of Paris Descartes (University of Paris) - Cibles Thérapeutiques et Concéption de Médicaments (CiTCoM) UMR 8038 CNRS); PASQUALI, samuela (Université de Paris)

Presenter: ALFERKH, Lina (University of Paris Descartes (University of Paris) - Cibles Thérapeutiques et Concéption de Médicaments (CiTCoM) UMR 8038 CNRS)

Contribution ID: **8** Type: **Oral**

Simultaneous determination of protein structure and dynamics using cryo-electron microscopy

Cryo-electron microscopy is rapidly emerging as a powerful technique to determine the structures of complex macromolecular systems elusive to other techniques. Since many of these systems are highly dynamical, characterising also their movements is a crucial step to unravel their biological functions. In this talk, I will present an integrative modelling approach to simultaneously determine structure and dynamics from cryo-electron microscopy density maps. By quantifying the level of noise in the data and dealing with their ensemble-averaged nature, this approach enables the integration of multiple sources of information to model ensembles of structures and infer their populations. I will illustrate the method with two examples: the characterization of structure and dynamics of the ClpP protease by a combined cryo-EM and NMR approach and the study of the effect of acetylation on microtubule structure and stability.

Primary author: Dr BONOMI, Massimiliano (Institut Pasteur - CNRS)

Presenter: Dr BONOMI, Massimiliano (Institut Pasteur - CNRS)

Session Classification: Session 9

Contribution ID: **9** Type: **Oral**

Analyzing macromolecular structures using the Voronoi tessellation of atomic balls

Given a molecular structure, it can be represented as a set of atomic balls, each ball having a van der Waals radius corresponding to the atom type. A ball can be assigned a region of space that contains all the points that are closer (or equally close) to that ball than to any other. Such a region is called a Voronoi cell and the partitioning of space into Voronoi cells is called Voronoi tessellation or Voronoi diagram. Two adjacent Voronoi cells share a set of points that form a surface called a Voronoi face. A Voronoi face can be viewed as a geometric representation of a contact between two atoms. The Voronoi cells of atomic balls may be constrained inside the boundaries defined by the solvent accessible surface of the same balls. The constrained Voronoi cells and their faces are remarkably versatile structural descriptors of atoms and their interactions. Another example of tessellation-derived features are empty tangent spheres corresponding to the vertices of the Voronoi cells. This talk will be focused on some of the protein structural analysis and assessment algorithms that are built upon the aforementioned Voronoi tessellation-derived descriptors.

Primary author: OLECHNOVIC, Kliment (Vilnius University)

Presenter: OLECHNOVIC, Kliment (Vilnius University)

Session Classification: Session 3

Contribution ID: **10** Type: **Oral**

Clustering algorithms for structural studies: insights on novel metrics and cluster stability assessment

Clustering conformations of molecules or molecular assemblies is a central problem faced in most studies, be they concerned with atomic models or coarse grain models. This talk will review recent contributions in this realm.

The first one is concerned with the newly designed molecular distance RMSDcomb [1]. Given the decomposition of a structure into subdomains, RMSDcomb provides a weighted average of the lRMSD observed between these subdomains, stressing the role of local similarities–as opposed to the lRMSD which suffers from the lack of homogeneity inherent to global structural comparisons.

The second one is concerned with clustering methods providing a simple read out for the number of clusters [2]. In this perspective, we will briefly review density based clustering and the associated stability assessment based on topological persistence.

Finally, we shall discuss the first technique finding a many-to-many correspondence between two sets of clusters delivered by two different methods or the same method under two sets of parameters [3], which is particularly useful to consolidate competing clusterings.

All the methods discussed are provided within the Structural Bioinformatics Library, in the following packages:
https://sbl.inria.fr/doc/Molecular_distances_flexible-user-manual.html
https://sbl.inria.fr/doc/Cluster_engines-user-manual.html#sec-Cluster_engines-programs
https://sbl.inria.fr/doc/D_family_matching-user-manual.html

[1] F. Cazals, T. Dreyfus, D. Mazauric, A. Roth and C.H. Robert. Conformational Ensembles and Sampled Energy Landscapes: Analysis and Comparison, J. Comp. Chem., 36, 2015.

[2] F. Cazals and R. Tetley. Characterizing molecular flexibility by combining lRMSD measures, Proteins, 87, 2019.

[3] F. Cazals, D. Mazauric, R. Tetley and R. Watrigant. Comparing two clusterings using matchings between clusters of clusters, ACM J. of Experimental Algorithms, 24, 2019.

Primary author: Dr CAZALS, Frederic (Inria)

Presenter: Dr CAZALS, Frederic (Inria)

Session Classification: Session 3

Contribution ID: **11** Type: **Oral**

Bridging Structural and Systems Biology to investigate the root cause of neurodegeneration

Neurodegenerative diseases are very complex and a holistic understanding of the relation between structure and function requires the integration of molecular, cellular and systems biology. We applied 'omic' technologies to generate of a most complete protein interactome associated with Alzheimer's disease (1). The analysis of this network triggered insights on the mechanisms implicated in the disease such as the formation of amyloid plaques, which surround neurons affected by Alzheimer's, with the dysfunction of mitochondrial respiration (2). We examine the mitochondrial respiratory complex under amyloidogenic conditions by applying an integrative structural and cell biology approach. Deciphering the molecular mechanisms will advance our understanding of the disease etiology by unveiling the causal link between altered mitochondrial respiratory function and early amyloid formation as it may occur in pre-symptomatic stages (3).

References

(1) Soler-López M, Zanzoni A, Lluis R, Stelzl U, Aloy P. Interactome mapping suggests new mechanistic details underlying Alzheimer's disease, Genome Research, 21:364-76, 2011.

(2) Soler-López M, Badiola N, Zanzoni A, Aloy P. Towards Alzheimer's root cause: ECSIT as an integrating hub between oxidative stress, inflammation and mitochondrial dysfunction, BioEssays, 34(7):532-41, 2012.

(3) Giachin G, Bouverot R, Acajjaoui S, Pantalone S, Soler-López M. Dynamics of Human Mitochondrial Complex I Assembly: Implications for Neurodegenerative Diseases, Front Mol Biosci, 3:43, 2016.

Primary author: Dr SOLER LOPEZ, Montserrat (ESRF)

Presenter: Dr SOLER LOPEZ, Montserrat (ESRF)

Session Classification: Session 2

Contribution ID: **12** Type: **Oral**

Cryo-Electron Tomography for in situ Structural Biology

Short Abstract

Cryo-electron tomography (cryo-ET) is a technique which allows direct visualisation of vitrified biological samples in 3D. In a framework conceptually similar to single particle analysis in cryo-EM, it is possible to achieve high-resolution (subnanometer and beyond!) insights into macromolecules directly within their cellular context by alignment and averaging of multiple copies of the same macromolecule. While possible, obtaining such results requires specialist knowledge of a complex workflow and as such is not routinely achieved in many labs worldwide.

Talk Summary

- Overview of cryo-electron tomography
- Recent results
- Challenges in data processing
- Recent algorithmic developments
- New possibilities?

Primary authors: Mr BURT, Alister (Institut de Biologie Structurale); Dr GUTSCHE, Irina (Institut de Biologie Structurale)

Presenter: Mr BURT, Alister (Institut de Biologie Structurale)

Session Classification: Session 9

Contribution ID: **13** Type: **Oral**

Structural basis for the subversion of MAP kinase signalling by an intrinsically disordered parasite secreted agonist

Pathogens often evolve methods of modulating host cell signalling pathways in order to promote their growth and persistence in the infected cell. One of the most efficient methods to interact with host signalling proteins is the use of intrinsically disordered proteins combined with short linear motifs, as the evolutionary timescales involved are shorter than those required for globular domains. Short linear motifs in particular can be used to allow interactions between a pathogenic protein and that of a host in a simple manner. The causative agent of toxoplasmosis, the obligate intracellular parasite Toxoplasma gondii, has recently been shown to deliver an intrinsically disordered protein, GRA24, into the cells it infects, that directly interacts in vivo with p38α, leading to autophosphorylation and nuclear translocation of the host kinase. The molecular basis of the interaction between p38α and the linear motifs R1 and R2 of GRA24 has been determined via the crystal structure of the interacting regions of the two proteins and characterising the full complex in solution using a combinatorial approach with SAXS, EM and AFM. Through structural and biochemical data, GRA24 is shown to use the classic KIM binding mechanism to form a highly stable complex linking two kinases that promotes p38α autophophosphorylation at the activation loop creating a highly active p38α complex. Additionally, the recombinant complex forms a powerful in vitro tool to evaluate specificity and effectiveness of p38α inhibitors that have advanced to clinical trials, as it provides a hitherto unavailable stable and highly active form of p38α. This is the first structural study of pathogen intervention in a MAP kinase pathway important in the inflammatory response and may provide new lines of investigation for novel modulators of MAP kinase signalling

Primary author: BOWLER, Matthew (EMBL)

Presenter: BOWLER, Matthew (EMBL)

Session Classification: Session 2

Contribution ID: **14** Type: **Oral**

Structural diversity and odd functions in RNA viruses.

In the recent years a massive amount of raw viral genomic data has been produced and released in sequenced databases, leading to the paradoxical situation of generating the Domain of Unknown Function (DUF), the number one domain in knowledge databases. Together with bio-informatics prediction, the knowledge of the three-dimensional structures of DUF proteins is the key to unveil the full potential of viral genomic information. At the dawn of the century, cutting-edge research in structural biology moved in two distinct directions: either tightly focused long-term research in individual laboratories, or large consortia of structural biologists developing strategies to determine new protein structures rapidly. On one hand, the latter succeeded in producing large numbers of homologous proteins, and generated advances in light sources, detectors, new algorithms and methods which benefit the entire community, however failed to unravel accurate structures for most viral DUF. On the other hand, the anticipated results that a small number of viral systems, studied in depth, would provide insights across the field of biology with the aid of genome-based comparative structural analysis, moved slower than expected. This was firstly due to the intrinsic complexity and plasticity of the viral models and the challenging molecular ensembles that need to be generated and stabilized prior to study. Secondly, a large part of the viral genome space falls under the category of the viral Dark Proteome (24% proteins & 25% of domains), for which very few new folds are unraveled and a lack of diversity in the known fold.
In the last decade viral human disease emerging from unforeseen viral family (Nidovirales, Mononegavirales, Bunyavirales, Flaviviridae) has stressed the critical necessity to solve the actual structure, to enlighten structural and functional oddities.

Primary author: Dr FERRON, Francois (Architecture et Fonction des Macromolécules Biologiques, AFMB - CNRS)

Co-authors: Dr PAPAGEORGIOU, Nicolas (Aix-Marseille Université UMR7257); Ms VALLE, Coralie (Aix-Marseille Université, UMR 7257); Dr SHANNON, Ashleigh (Architecture et Fonction des Macromolécules Biologiques, AFMB - CNRS); Dr LABAN, Elsie Y (Aix-Marseille Université, UMR 7257); Ms SAMA, Bhawna (Aix-Marseille Université, UMR 7257); Dr DECROLY, Etienne (Architecture et Fonction des Macromolécules Biologiques, AFMB - CNRS); Dr CANARD, Bruno (Architecture et Fonction des Macromolécules Biologiques, AFMB - CNRS)

Presenter: Dr FERRON, Francois (Architecture et Fonction des Macromolécules Biologiques, AFMB - CNRS)

Session Classification: Session 8

Contribution ID: **15** Type: **Oral**

Deep learning approaches in single particle Cryo-Electron Microscopy

Over the last two decades, single-particle cryo-electron microscopy has emerged as a technique capable of producing routinely near-atomic resolution structures and render to study challenging systems that otherwise defy structural characterization. Recent technical advances in this field have resulted in breakthrough progress in its applicability to various biological macromolecules, resolution attained and throughput. However, the technique still faces challenges in almost every step of its workflow, more significantly on the image analyses, which is still a time-consuming long process including lots of manual interventions. Single Particle Analyses (SPA) benefits recently from deep-learning algorithms to drive the workflow more automatically and efficiently as well. I will discuss few deep-learning approaches available for SPA in various parts of its workflow.

Primary author: Dr KANDIAH, Eaazhisai (ESRF)

Presenter: Dr KANDIAH, Eaazhisai (ESRF)

Session Classification: Session 10

Contribution ID: **16** Type: **Oral**

Approaches to conformational heterogeneity of full antibodies by cryo-EM

Multispecific antibodies are artificially engineered molecules designed to bind simultaneously to several different antigens. Potential advantages of generating viable multispecific antibodies include the identification of malignant cells coupled with the concurrent recruitment of immune cells and the blocking of complex viral escape mechanisms. The cross-over dual-variable immunoglobulin (CODV-Ig) has been proposed as a universal bispecific therapeutic format. Its unique antigen-binding fragment (Fab) architecture provides pM affinities for ligands, no positional effect in target binding and a stable self-supporting structure. However, a disparity between *in vitro* and *in vivo* effects suggests that the three-dimensional arrangement of the constant and antigen-binding fragments in the CODV-Ig format is not ideal. To further understand the structure and function of multispecific antibodies based on the CODV-Ig format, high-resolution structural information is required. Towards this, we use cryo-electron microscopy (cryo-EM).

We purified CODV-Ig both in an unbound state and in complex and validated sample quality using SDS-PAGE, nanoDSF, mass photometry, Small Angle X-Ray Scattering (SAXS) and negative-stain electron microscopy (NSEM; Tecnai T12 microscope at IBS, Grenoble). NSEM data analyzed using Relion 2.1 and 3.0 resulted in low-resolution structural models and suggested a preferential orientation of CODV-Ig under negative-stain conditions. Close-to-optimal vitrification conditions for CODV-Ig have been identified. Efforts are in progress to reduce the antibody's propensity to aggregation and aversion to conventional cryo-EM supports and conditions. Nevertheless, data of sufficient quality for image analysis was obtained using a Titan Krios microscope (ESRF, Grenoble) equipped with a Quantum LS energy filter and K2 Summit direct electron detector. Image processing of both CODV-Ig alone and in complex with antigens is ongoing. The lack of a conventional center-of-mass and the very high flexibility and conformational heterogeneity of the particles have made every step of the processing pipeline very challenging and has required the usage of a wide-range of software, including Relion, cryoSPARC, cisTEM and EMAN.

Primary author: FERNANDEZ, David (ESRF)

Co-authors: KANDIAH, Eaazhisai (ESRF); Dr MATHIEU, Magali (Sanofi); Dr LEONARD, Gordon (ESRF)

Presenter: FERNANDEZ, David (ESRF)

Session Classification: Session 10

Contribution ID: **17** Type: **Poster**

Structural studies of the interaction between ACAD9 and ECSIT using small-angle scattering methods

The mitochondrial respiratory complex I is the largest of the large mitochondrial respiratory complexes, being roughly 1 MDa in size. Consequently, its assembly process is extremely complicated, requiring multiple assembly factors. Two such factors, acyl-CoA dehydrogenase 9 (ACAD9) and the evolutionarily conserved signalling intermediate in the Toll pathway (ECSIT) appear to act in a complex during complex I assembly. Although pending structural studies and a variety of assays on the ACAD9-ECSIT binding interaction exist, the exact location and stoichiometry of the binding and the associated conformational changes remain unclear. To this end, we attempted to investigate the ACAD9-ECSIT binding interaction using small-angle X-ray and neutron scattering, with the use of isotopic labelling to visualise multiple parts of the binding interface. Initial results showed large changes in the conformation of ECSIT upon binding to ACAD9. Such studies shed further light upon complex I assembly and oxidative phosphorylation in the cell, with valuable implications in neurodegenerative disease.

Primary authors: Dr SOLER-LÓPEZ, Montserrat (European Synchrotron Radiation Facility); Mr HEARD, James (University of Bath); Dr ACAJJAOUI, Samira (European Sychrotron Radiation Facility); Ms SAIDI, Mélissa (European Synchrotron Radiation Facility)

Co-authors: Dr LAUX, Valérie (Institut Laue-Langevin); Dr HAERTLEIN, Michael (Institut Laue-Langevin); Prof. FORSYTH, Trevor (Institut Laue-Langevin)

Presenter: Dr SOLER-LÓPEZ, Montserrat (European Synchrotron Radiation Facility)

Contribution ID: **18** Type: **Poster**

Interaction mapping of ACAD9 and ECSIT interaction by homology sequence conservation analysis

The mitochondrial oxidative phosphorylation system (OXPHOS), which includes five protein complexes (CI-V), is the main source of energy in the cell. CI, the largest and first complex, is composed of several protein subunits that are assembled by different assembly factors, such as Acyl-CoA Dehydrogenase Family Member 9 (ACAD9) and Evolutionarily Conserved Signaling Intermediate In Toll Pathway (ECSIT), which join together to form the Mitochondrial Complex I Assembly (MCIA) complex. However, the way these assembly factors interact to form the MCIA complex is largely unknown.

The ancestor of ACAD9, Very Long-Chain Acyl-CoA Dehydrogenase (VLCAD), might be of crucial importance when mapping the ECSIT:ACAD9 interaction interface. Although ACAD9 and VLCAD show a high sequence similarity, VLCAD does not bind ECSIT and it is consequently not involved in CI assembly, unlike ACAD9. Therefore, the regions where the sequences of ACAD9 and VLCAD differ might correspond to the binding sites between ACAD9 and ECSIT. By comparing the sequences of ACAD9 and VLCAD using multiple sequence alignment programmes, such as CLUSTALX, we have been able to identify these potential critical regions. The next step will consist in designing and generating mutants for ACAD9 by site-directed mutagenesis, which will, in turn, be used in protein-protein interaction tests. Mutations that either create or disrupt functional interactions will be the basis to identify key residues within the ECSIT:ACAD9 interface.

Primary author: Mr MIGUELA BENAVIDES, Albert (European Synchrotron Radiation Facility, Grenoble)

Co-author: Dr SOLER LÓPEZ, Montserrat (European Synchrotron Radiation Facility, Grenoble)

Presenter: Mr MIGUELA BENAVIDES, Albert (European Synchrotron Radiation Facility, Grenoble)

Contribution ID: **19** Type: **Oral**

DECOMPOSITION OF EVOLVING SAXS DATASETS USING CHEMOMETRICS

Many biologically important systems are inherently polydisperse e.g. transient protein-protein complexes, oligomerizing proteins or proteins with different co-existing conformations. Such systems are always in an equilibrium of different species and are therefore heterogenous. Importantly, such systems can't be meaningfully studied in a holistic sense by physically separating the components of the system as they exist in context of each other. Small Angle X-ray Scattering (SAXS) is an excellent tool to study such systems as SAXS data is additive and in principle, can be decomposed into the scattering profiles of the components. We have developed a chemometrics based, model free method to decompose evolving SAXS datasets into the SAXS profiles of the individual components and recover the relative populations at the same time1. Our method can use multiple SAXS datasets and multiple representation of the SAXS profiles to improve the resolution of the curves and reduce the ambiguity of the solutions.

Our tests on simulated SAXS titrations datasets of transient protein-protein complexes demonstrate that we are able to recover the SAXS profile, and therefore the size and shape parameters, of the complex. At the same time, we are also able to extract the fractions of the complex and the dissociated components through the titration series, giving insight into the thermodynamics of the system.

Our approach would be useful to study a variety of evolving SAXS data sets (Oligomerization, Titrations of transient complexes, Time-resolved SAXS, Fibrillation, etc.) and extends the reach of SAXS into the realm of complex, heterogeneous, evolving biological systems.

Primary author: Dr SAGAR, Amin (Centre de Biochimie Structurale)

Co-authors: Dr HERRANZ-TRILLO, Fátima; Prof. TAULER , Romà (Institute of Environment Assessment and Water Diagnostics (IDAEA-CSIC), Barcelona, Spain); Prof. BERNADÓ, Pau (Centre de Biochimie Structurale)

Presenter: Dr SAGAR, Amin (Centre de Biochimie Structurale)

Session Classification: Session 1

Contribution ID: **20** Type: **Oral**

How to combine sequence and structure information in multiple protein alignements?

Multiple protein sequence alignments are used daily in bioinformatics to annotate and predict the characteristics of currently mass produced sequences. The quality of their results have been assessed many times and have recahed a plateau. Proteins fold into stable three-dimensional structures with a topology much more conserved than sequence. Consequently, it should be advantageous to use this other source of information to align the sequences, in order to find the homologous positions. Several programs have been developped to align proteins according to their structure or to their sequence and their structure. In this study, we wanted to assess the added value of structural information in multiple alignments and compared the results of these programs to the results of the sequence alignment programs.
We compared the multiple alignments resulting from 25 programs either based on sequence, structure, or both, to reference align-ments deposited in five databases (BALIBASE 2 and 3, HOMSTRAD, OXBENCH and SISYPHUS). On the whole, the structure-based methods compute more reliable alignments than the sequence-based ones, and even than the sequence+structure-based programs whatever the databases. Two programs lead, MAMMOTH and MATRAS, nevertheless the perfor- mances of MUSTANG, MATT, 3DCOMB, TCOFFEE+TM ALIGN and TCOFFEE+SAP are better for some alignments. The advantage of structure-based methods increases at low levels of sequence identity, or for residues in regular secondary structures or buried ones. Concerning gap management, sequence-based programs set less gaps than structure-based programs. Concerning the databases, the alignments of the manually built databases are more challenging for the programs.

Primary author: CARPENTIER, Mathilde (ISYEB - MNHN - SU - CNRS - EPHE)

Presenter: CARPENTIER, Mathilde (ISYEB - MNHN - SU - CNRS - EPHE)

Session Classification: Session 4

Contribution ID: **21** Type: **Poster**

Spontaneous formation of cushioned model membranes promoted by an intrinsically disordered protein

Histatin 5 (Hst5) is a histidine-rich, 24 amino acid protein, classified as an intrinsically disordered protein (IDP). It contains 7 histidines, an amino acid sensitive to charge regulation. The histidines can titrate and gain a positive charge, making the protein highly cationic. Hst5 is a salivary protein found to play a crucial role in fungicidal activity, and its activity to inhibit the growth and viability of *Candida albicans* has been evaluated using a variety of techniques. We have found that when exposed to a solid supported bilayer, the protein spontaneously forms a cushion below the bilayer and lifts it from the solid support.

The results obtained from neutron reflectometry and QCM-D have shown that the interaction between the peptide and the lipid bilayer is completely governed by electrostatic effects. This was done by changing the charge content in the bilayer and the ionic strength of the buffer. At low ionic strength, the peptide penetrates the bilayer and cumulate close to the solid silica substrate underneath the lipid bilayer. We suggest a mechanism of the formation of the cushion where the histidine residues are charged up by the solid substrate, as well as the charged bilayer. The protein can then penetrate the bilayer without disrupting the lipid-lipid interactions, and adsorb to the solid substrate, which releases counterions bound to the substrate. This counterion release increases the osmotic pressure and lifts the bilayer from the substrate.

Primary authors: Dr GERELLI, Yuri; ERIKSSON SKOG, Amanda; JEPHTHAH, Stéphanie (Lund University); WELBOURN, Rebecca; KLECHIKOV, Alexey; SKEPO, Marie

Presenter: ERIKSSON SKOG, Amanda

Contribution ID: **22** Type: **Oral**

Integrated NMR and cryo-EM atomic-resolution structure determination of a half-megadalton enzyme complex

Atomic-resolution structure determination is the key requirement for understanding protein function. Cryo-EM and NMR spectroscopy both provide structural information, but currently cryo-EM does not routinely give access to atomic-level structural data, and, generally, NMR structure determination is restricted to small (<30 kDa) proteins. We introduce an integrated structure determination approach1 that simultaneously uses NMR and EM data to overcome the limits of each of these methods. The approach enabled determination of the high-resolution structure of the 468 kDa large dodecameric aminopeptidase TET2 to a precision and accuracy below 1 Angstrom by combining secondary-structure information obtained from near-complete magic-angle-spinning NMR assignments of the 39 kDa-large subunits, distance restraints from backbone amides and specifically labelled methyl groups, and a 4.1 Angstrom resolution EM map. The resulting structure exceeds current standards of NMR and EM structure determination in terms of molecular weight and precision. Importantly, the approach is successful even in cases where only medium-resolution (up to 8 Angstroms) cryo-EM data are available, thus paving avenues for the structure determination of challenging biological assemblies.

Primary authors: FAVIER, Adrien (IBS CNRS); Dr GAUTO, Diego F. (Univ. Grenoble Alpes, CEA, CNRS, Institut de Biologie Structurale (IBS), Grenoble, France); FARIAS ESTROZI, Leandro (IBS); Dr SCHWIETERS, Charles D. (Center for Information Technology, National Institutes of Health, Bethesda, USA); Mr EFFANTIN, Gregory (Univ. Grenoble Alpes, CEA, CNRS, Institut de Biologie Structurale (IBS), Grenoble, France); MACEK, Pavel (Univ. Grenoble Alpes, CEA, CNRS, Institut de Biologie Structurale (IBS), Grenoble, France; NMR-Bio, Grenoble, France); Dr SOUNIER, Remy (Univ. Grenoble Alpes, CEA, CNRS, Institut de Biologie Structurale (IBS), Grenoble, France; Institut de Génomique Fonctionnelle, University of Montpellier, Montpellier, France); SIVERTSEN, Astrid C. (Univ. Grenoble Alpes, CEA, CNRS, Institut de Biologie Structurale (IBS), Grenoble, France); SCHMIDT, Elena (Institute of Biophysical Chemistry, Goethe University Frankfurt am Main, Frankfurt, Germany); KERFAH, Rime (Univ. Grenoble Alpes, CEA, CNRS, Institut de Biologie Structurale (IBS), Grenoble, France; Graduate School of Science, Tokyo Metropolitan University, Tokyo, Japan); MAS, Guillaume (Univ. Grenoble Alpes, CEA, CNRS, Institut de Biologie Structurale (IBS), Grenoble, France; Biozentrum University of Basel, Basel, Switzerland); COLLETIER, Jacques-Philippe; GUNTERT, Peter (Institute of Biophysical Chemistry, Goethe University Frankfurt am Main, Frankfurt, Germany; Laboratory of Physical Chemistry, ETH Zurich, Zurich, Switzerland; Graduate School of Science, Tokyo Metropolitan University, Tokyo, Japan); Dr SCHOEHN , Guy (Univ. Grenoble Alpes, CEA, CNRS, Institut de Biologie Structurale (IBS), Grenoble, France); BOISBOUVIER, Jerome (Univ. Grenoble Alpes, CEA, CNRS, Institut de Biologie Structurale (IBS), Grenoble, France); SCHANDA, Paul (IBS Grenoble)

Presenter: FAVIER, Adrien (IBS CNRS)

Session Classification: Session 8

Contribution ID: **23** Type: **Oral**

Integrative methods to study flexible multi-domain enzymes: an experimental approach

Integrative modeling is a powerful tool to study the structure-function relationship in multi-domain flexible systems. Here we will describe the combination of modelling, solution scattering data and flexible fitting as well as biochemical and biophysical analyses that have been used to define the quaternary structure of DDX21, a human DEAD-box helicase with RNA G-quadruplex resolving activity. We discovered that DDX21 is dimeric and propose a model for DDX21 function which suggests that an intact dimerization interface is essential to maintaining dsRNA unwinding and G-quadruplex remodelling activities. We will also discuss the problems that we encounter and questions that remain unsolved.

Primary author: MARCAIDA, Maria (EPFL)

Presenter: MARCAIDA, Maria (EPFL)

Session Classification: Session 1

Contribution ID: **24** Type: **Oral**

General thoughts and Examples for EM fitting: an historical perspective.

In last the decades, significant progress was made by combining cryo-EM data with high-resolution structures. We review our adventure developing computational fitting techniques to interpret EM data in terms of available atomic structures. We will try to illustrate the power and limitations of our fitting tools with key examples.

Primary author: Dr CHACON, Pablo

Presenter: Dr CHACON, Pablo

Session Classification: Session 9

Contribution ID: **25** Type: **Oral**

Toward a refined electrostatic description for biomolecules simulations : introducing an effective pH and titration scheme to a coarse-grained nucleic acids model

Electrostatic interactions play a pivotal role in many (bio)molecular association processes. The molecular
organization and function in biological systems are largely determined by these interactions from pure
Coulombic contributions to more peculiar mesoscopic forces due to ion-ion correlation and proton
fluctuations. By means of constant-pH Monte Carlo simulations based on a fast coarse-grained titration proton scheme,
a new computer molecular model was devised to study protein-RNA interactions and compute local pka values for RNA molecules.
Our results illustrate the importance of the charge regulation mechanism that enhances
the association between biological macromolecules in a similar way as observed for other protein-polyelectrolyte systems typically found in colloidal science. Due to the highly negative charge of RNA,
the effect is more pronounced in this system as predicted by the Kirkwood-Shumaker theory.
These efficient calculations have now been coupled to coarse-grained simulations of nucleic acids to include the role of conformational
changes. We will present our work to obtain pH-dependent coarse-grained nucleic acids simulations.

Primary author: Prof. PASQUALI, Samuela

Presenter: Prof. PASQUALI, Samuela

Session Classification: Session 2

Contribution ID: **26** Type: **Oral**

Integrative structures of protein complexes with intrinsically disordered proteins: a case study and current challenges

Intrinsically disordered proteins (IDPs) are key in wide range of biological processes, from signal-transduction pathways to the protection of genomic integrity. Yet, their unstructure poses a significant challenge for understanding their function at the atomistic level. In our work we aim to decipher the role of the intrinsically disordered acidic domain of APLF (APLFAD) in binding and safeguarding histone proteins, the core protein component of chromatin, during DNA damage repair.

APLF is a scaffold protein for the DNA repair machinery and vital to maintenance of genomic integrity. APLF contains a disordered C-terminal acidic domain (APLFAD) that harbors histone chaperone activity, and deletions of this chaperone domain have been shown to impair cell survival after DNA damage [1]. We recently determined that this acidic domain can bind both H2A-H2B and H3-H4 complexes with similar affinity [2]. Moreover, we found that APLFAD can bind all histones simultaneously and can promote their deposition on DNA, challenging the current paradigm of of step-wise nucleosome assembly.

To understand the APLFAD-histone interaction at the structural level we collected a series of NMR, mass-spectrometry and SAXS data. Yet their integration into a representative atomistic ensemble is challenged by unstructured nature of APLFAD, the presence of disordered tails in the histones, the dynamic oligomeric nature of the histones in vitro as well as the size of the complex (120 kDa). In the presentation I will discuss the information content of the experimental data, our approach in modelling the system using the docking software HADDOCK and the use of SANS data to get new insights in the architecture of the complex. Ultimately, the structure of the APLF-histone complex will give key new insights into the chaperone activity of APLF, in particular how it is able to prevent non-native DNA binding and how it controls histone stoichiometry in nucleosome (dis)assembly.

[1]. Mehrotra, P. V. et al. DNA repair factor APLF is a histone chaperone. Mol. Cell 41, 46–55 (2011). DOI: 10.1016/j.molcel.2010.12.008

[2]. Corbeski, I. et al. DNA repair factor APLF acts as a H2A-H2B histone chaperone through binding its DNA interaction

Primary authors: VAN INGEN, Hugo (Utrecht University); Dr CORBESKI, Ivan (Utrecht University)

Presenter: VAN INGEN, Hugo (Utrecht University)

Session Classification: Session 7

Contribution ID: **27** Type: **Oral**

Assembly of membrane protein complexes with encapsulated lipids

Recent breakthroughs in X-ray crystallography, Cryo-EM and complementary approaches resulted in elucidation of many new structures of membrane proteins (MPs) and their complexes. Several classes of MPs, such as microbial rhodopsins, rotary ATPase subunits c, or light harvesting complexes 2, form ring-like assemblies with several lipid molecules trapped inside. Whereas the proteins are usually well resolved in experimental structures, the lipids are more disordered and cannot be readily modeled; the number and nature of the trapped molecules is also often not clear. This is problematic, because the trapped lipids cannot exchange with the pool of the surrounding membrane lipids in simulations, and the errors in the starting models can not correct themselves, as often happens in normal molecular dynamics simulations of soluble and membrane proteins. To address this problem, we simulated the process of the assembly of the MP-lipid complex explicitly, and then removed the excess lipids to achieve a very good fit to the noisy experimental densities that are available [1]. The presented approach is expected to be generalizable to all MP complexes with encapsulated lipid patches. Besides producing molecular models of MP-lipid complexes, the approach also highlights intriguing adaptations of ring-forming MPs that promote efficient assembly.

1. Novitskaia, O.; Buslaev, P.; Gushchin, I. Assembly of Spinach Chloroplast ATP Synthase Rotor Ring Protein-Lipid Complex. *Front. Mol. Biosci.* 2019, 6.

Primary author: GUSHCHIN, Ivan (Moscow Institute of Physics and Technology)

Co-authors: Dr BUSLAEV, Pavel (Nanoscience Center, Department of Chemistry, University of Jyväskylä, Jyväskylä, Finland); Mr KUZMIN, Alexander (Research Center for Molecular Mechanisms of Aging and Age-Related Diseases, Moscow Institute of Physics and Technology, Dolgoprudny, Russia); Mr SERGEEV, Ilya (Research Center for Molecular Mechanisms of Aging and Age-Related Diseases, Moscow Institute of Physics and Technology, Dolgoprudny, Russia)

Presenter: GUSHCHIN, Ivan (Moscow Institute of Physics and Technology)

Session Classification: Session 4

Contribution ID: **28** Type: **Oral**

Coupling X-ray crystallography and molecular modeling for the structural investigation of flexible antibody paratopes: A case study

By providing a precise snapshot of the structure of the interface between two protein partners, X-ray crystallography helps identifying point mutations that are likely to strengthen the interaction. However, this technique provides very little information about conformational flexibility, in particular in protein loop fragments. In cases where plasticity may be critical for protein function, the flexible fragments must be modeled using alternative methods that complete the partial view offered by X-ray crystallography. As part of a project to produce a strong anti-FGFR4 antibody, several crystal structures were obtained for the antibody, free or bound to FGFR4, after different steps of humanization or maturation. While the structures successfully identified point mutations to improve the affinity between the antibody and its antigen, they also revealed a conformational change happening in the CDR-H3 loop upon antigen binding that was not exploited. Using in silico modeling techniques, structural basis for the flexibility of this loop was clarified. Conformational stability in the absence of the antigen or of crystalline contacts was investigated. The results confirm that the loop presents two stable conformations, one of which is stabilized by the antigen, and by conformational changes in the loop surroundings.

Primary authors: BAROZET, Amélie (LAAS-CNRS, Université de Toulouse); MATHIEU, Magali (Sanofi recherche & développement); PAPIN, David (Sanofi recherche & développement); CAMERON, Béatrice (Sanofi recherche & développement); DABDOUBI, Tarik (Sanofi recherche & développement); SÉVERAC, Anne (Sanofi recherche & développement); FERRARI, Paul (Sanofi recherche & développement); SIMÉON, Thierry (LAAS-CNRS, Université de Toulouse); BIANCIOTTO, Marc (Sanofi recherche & développement); CORTÉS, Juan (LAAS-CNRS, Université de Toulouse)

Presenter: CORTÉS, Juan (LAAS-CNRS, Université de Toulouse)

Session Classification: Session 7

Contribution ID: **29** Type: **Poster**

A new Data Reduction and Analysis group at Synchrotron SOLEIL

The newly formed Data Reduction and Analysis group at Synchrotron SOLEIL is responsible for the development and implementation of data analysis software and methods for users of the facility. Rapid and efficient data analysis adds value to the users' experience and efforts will be made by the group to provide simple and rapid tools for this purpose.

We aim to develop cloud-based infrastructure with on-demand remote desktop services, Jupyter notebooks, and micro-services for scientific computation and automatic processing. These services will be filled with scientific applications for synchrotron radiation data treatment, starting with e.g. MX and SAXS data reduction and analysis, tomography and ptychography reconstruction, fluorescence and absorption spectroscopy, photo-emission spectroscopy, and modelling tools (materials, beam-lines). As our workforce is still limited, we shall favour initially workflows gathering existing software solutions.

In the frame of integrated biology, our group will provide in addition a set of data treatment tools to serve the beam-line users for e.g. computational tasks, and data management workflows to merge multi-techniques and multi-scale data sets.

Primary author: Dr FARHI, Emmanuel (Synchrotron Soleil)

Co-authors: ROUDENKO, Olga; PICCA, Frederic-Emmanuel (Synchrotron SOLEIL); BAC, stephane

Presenter: Dr FARHI, Emmanuel (Synchrotron Soleil)

Contribution ID: **30** Type: **Oral**

Current bottlenecks in protein assembly prediction

Protein-protein interactions play a central role in all biological processes. These processes result from the physical interaction of two or more protein molecules, forming a macromolecular assembly.

The CAPRI (Critical Assessment of PRedicted Interactions) experiment has been a proven catalyst of docking algorithms since its inception almost two decades ago. In recent years, the CAPRI experiment has diversified and now includes the prediction of multi-component assemblies, of protein-peptide, protein-nucleic acid and protein-polysaccharide binding, and of binding affinities and high-resolution prediction of interfacial water molecule positioning.

I will give an overview of the experiment and what it can do for you as computational or experimental biologist. In the context of the workshop, I will not only highlight its successes, but very specifically point out the current failures of protein docking and assembly prediction.

Primary author: LENSINK, Marc (CNRS)

Presenter: LENSINK, Marc (CNRS)

Session Classification: Session 3

Contribution ID: **31** Type: **Oral**

Modeling the Solution State through SAXS

Understanding how proteins and complexes move in solution remains a major challenge in structural biology. Biological macromolecules are machines that adopt a variety of structural conformations and efficiently exploring this structural space in a comprehensive and meaningful way can not be achieved solely using the solid-state methods cryo-EM and macromolecular crystallography (MX). Small X-ray scattering (SAXS) measurements of biological macromolecular particles in solution provides a resolution-limited, structural assessment of the thermodynamic ensemble. In a SAXS measurement, all conformations will be sampled albeit with a severe loss resolution and correspondence. Nonetheless„ SAXS observations made at higher resolutions imply a greater detail in the structural measurement. Here, I will present an approach to understanding SAXS data using Information Theory (IT). I will show that the Information theory framework provides a pathway for exploring conformational space when integrated with computational methods such as molecular dynamics. The IT framework can be used to develop structural modeling algorithms for shape determination and docking and I will show the IT approach can be used in antibody-antigen studies to uniquely determine the structure of the complex in the solution state. The IT approach suggests a general method for refining atomistic models derived from homology models or prediction.

Primary author: RAMBO, Robert (Diamond Light Sourcwe)

Presenter: RAMBO, Robert (Diamond Light Sourcwe)

Session Classification: Session 1

Contribution ID: **32** Type: **Poster**

Novel computational tools for integrative structural biology: structure predictions using small-angle scattering, cross-links, symmetry, protein docking, normal mode analysis, and artificial intelligence

While crystallography has been providing atomic-resolution structures of biomolecules for over half a century, the real challenge of today's biophysics is to correlate molecules' structure and dynamics in solution with their function. Owing to the complexity of the problem, the answer can only be found if multiple sources of information are used simultaneously. I will present a set of tools developed for this purpose in our group [1].

Recently, we introduced SAXS and SANS packages called Pepsi-SAXS, and Pepsi-SANS, correspondingly [1,2]. Pepsi-SAXS is a very efficient method that calculates small angle X-ray scattering profiles from atomistic models. It is based on the multipole expansion scheme and is significantly faster than other methods with the same level of precision. One of the challenges in the field is, however, flexible fitting of atomistic models into small-angle scattering profiles. We designed a computational scheme that uses the nonlinear normal modes [1,3] as a low-dimensional representation of the protein motion subspace and optimizes protein structures guided by the SAXS and SANS profiles. For example, this scheme was ranked first in the recent CASP13 structure prediction challenge. Another challenge is data-assisted protein docking. We have designed a scheme for SAXS-assisted rescoring of docking predictions. This was made possible due to the polynomial representation of partial scattering amplitudes for each of the docking partners.

We are also interested in protein symmetry, as many protein complexes are symmetric homooligomers. We have designed a novel free-docking symmetry-assisted method [4] and also a dual analytical technique for identification of symmetry in a protein assembly [5-7]. This allowed us to study principles of protein organizations on the PDB scale.

Finally, artificial intelligence has made a big leap forward and found many applications in structural bioinformatics. On our side, we have been using it for multiple tasks of protein structure prediction, including protein-protein [8] and protein-ligand [9] docking, shape analysis [7] and structure prediction [10-11].

References

[1] - https://team.inria.fr/nano-d/software/.
[2] - S. Grudinin, et al, Acta Cryst D, (2017), D3, 449.
[3] - A. Hoffmann & S. Grudinin, J. Chem .Theory Comput., 2017, 13 (5), 2123.
[4] - Ritchie, D. W. & Grudinin, S (2016). J. Appl. Cryst., 49, 1-10.
[5] - Pages, G., Kinzina, E, & Grudinin, S (2018). J. Struct.Biology, 203 (2), 142-148.
[6] - Pages, G. & Grudinin, S (2018). J. Struct.Biology, 203 (3), 185-194.
[7] - Pages, G. & Grudinin, S (2018). arXiv preprint arXiv:1810.12026.
[8] - Neveu, E., D. Ritchie, P. Popov, & S. Grudinin (2016). Bioinformatics, 32 (7), pp.i693-i701.
[9] - Kadukova, M. & S. Grudinin (2017). J. Comp.-Aid. Mol. Des. 31 (10), pp.943-958.
[10] - Karasikov, M., G. Pagès, & S.Grudinin (2018), Bioinformatics, bty1037.
[11] - Pages, G., B. Charmettant & S. Grudinin (2019). Bioinformatics, btz122.

Primary authors: GRUDININ, Sergei; PAGÈS, Guillaume (Inria); KADUKOVA, Maria (Nano-D, Inria Grenoble / MIPT Moscow); Mr KARASIKOV, Mikhail; MARTEL, Anne (ILL); PREVOST, Sylvain

(Institut Laue-Langevin)

Presenter: GRUDININ, Sergei

Contribution ID: **33** Type: **Oral**

Using NMR and Molecular Modelling to Understand the Function of Highly Dynamics Biomolecules

TBA

Presenter: BLACKLEDGE, martin

Session Classification: Session 7