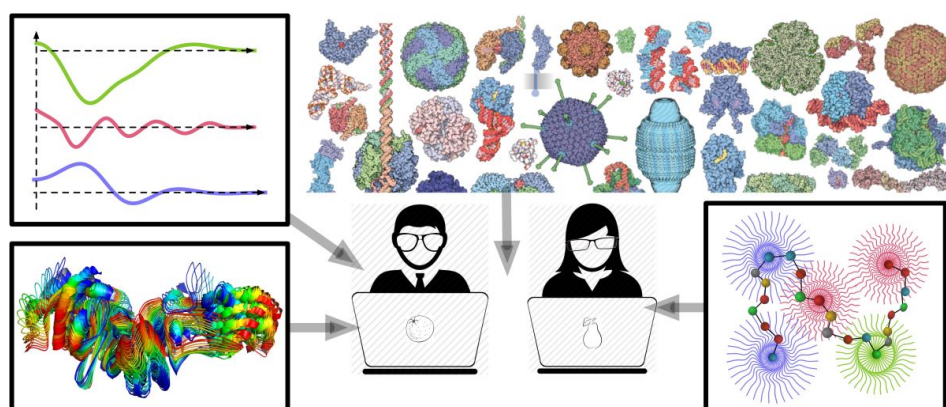# Joint Integrative Computational Biology workshop and CAPRI Meeting

Report of Contributions

Contribution ID: **3**   Type: **Talk**

Hybrid methods for analyzing conformational variability in cryo-EM and cryo-ET data

Monday, 12 February 2024 17:30 (25 minutes)

The elucidation of different conformations of biomolecular complexes is the key to understand the molecular mechanisms behind the biological functions of the complexes and the key to novel drug discovery. Single-particle cryo electron microscopy (cryo-EM) allows 3D reconstruction of multiple conformations of purified biomolecular complexes from their 2D images. Cryo electron tomography (cryo-ET) allows obtaining information on the conformational variability of the complexes in their cellular environment. My group is developing hybrid methods for analyzing continuous conformational changes of biomolecules from cryo-EM and cryo-ET data, which integrate image processing, molecular dynamics simulations, and deep learning approaches. These methods are made available publicly via our open-source, ContinuousFlex software package (a plugin of Scipion, the software largely used in the cryo-EM/ET field). In this talk, I will present our recent work regarding these methodological developments.

Submitting to:

Integrative Computational Biology workshop

Primary author: Dr JONIC, Slavica (CNRS & Sorbonne University, Paris)

Presenter: Dr JONIC, Slavica (CNRS & Sorbonne University, Paris)

Session Classification: Computational Methods and Deep Learning

Contribution ID: **4**   Type: **Talk**

Analysis of interfaces in protein complexes using Voronoi tessellations and graph neural networks

Monday, 12 February 2024 16:15 (25 minutes)

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. 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. In particular, a novel method for assessing inter-subunit interfaces in protein-protein complexes, VoroIF-GNN, will be presented. Given a multimeric protein 3D structural model, the method derives interface contacts from the Voronoi tessellation of atomic balls, constructs a graph of those contacts, and predicts accuracy of every contact using an attention-based graph neural network. The contact-level predictions are then summarized to produce whole interface-level scores. VoroIF-GNN was blindly tested for its ability to estimate accuracy of protein complexes during CASP15 (15th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction) and showed strong performance in selecting the best multimeric model out of many.

Submitting to:

Integrative Computational Biology workshop

Primary author: Dr OLECHNOVIC, Kliment (CNRS Laboratoire Jean Kuntzmann, Grenoble)

Presenter: Dr OLECHNOVIC, Kliment (CNRS Laboratoire Jean Kuntzmann, Grenoble)

Session Classification: Computational Methods and Deep Learning

Contribution ID: **5**   Type: **Talk**

Unsupervised Machine Learning and Phase Space Reduction: A Robust and Generalisable Approach for Concurrently Solving the Protein Complex Conformation Classification and Quantification Problems.

Friday, 16 February 2024 10:30 (30 minutes)

The understanding of biochemical processes and the machinery of life hinges on comprehending the structural aspects of macromolecular interactions. This requires a systematic approach to analysing the vast manifold of macromolecular associations, or complexes, typically determined from crystallographic or electron microscopy experiments. Specific points of interest include similarity measurements, multiple structural alignments and superpositions, conformational analysis, classification, and functional annotation.

While a considerable number of tools have been developed for analysing covalently linked structures, or single chains [1-5], no methods applicable to the analysis of complexes are known to us. This may be explained by the higher diversity and absence of canonical ordering of chains in quaternary structures, leading to higher ambiguity compared to secondary and tertiary structure analyses.

We present FunCLAN, a novel approach and software solution for the analysis of protein complex conformations. FunCLAN combines unsupervised machine learning with physically informed scoring to transform a practicably infinite, continuous, geometric phase space into a tractable and measurable discrete one. This enables the robust classification and quantification of a protein complex's in-sample conformational landscape, canonical chain ordering and optimal multiple complex alignment and superposition.

The approach has been used to classify protein complexes, such as the SARS-CoV-2 spike protein, into distinct conformations, and measure the degree of similarity among them. FunCLAN provides a uniform and consistent approach to the analysis of macromolecular structures. It can be generalised to the case of single chains by applying the algorithm to secondary structures, or highly conserved "rigid" regions.

References:

[1] Krissinel, E. (2012). Enhanced fold recognition using efficient short fragment clustering. Journal of molecular biochemistry, 1(2), 76.

[2] Krissinel, E., & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallographica Section D: Biological Crystallography, 60(12), 2256-2268.

[3] Ellaway, J. I., Anyango, S., Nair, S., Zaki, H. A., Nadzirin, N., Powell, H. R., ... & Velankar, S. (2023). Identifying Protein Conformational States in the PDB and Comparison to AlphaFold2 Predictions. bioRxiv, 2023-07.

[4] Ye, Y., & Godzik, A. (2004). FATCAT: a web server for flexible structure comparison and structure similarity searching. Nucleic acids research, 32(suppl_2), W582-W585.

[5] Li, Z., Natarajan, P., Ye, Y., Hrabe, T., & Godzik, A. (2014). POSA: a user-driven, interactive multiple protein structure alignment server. Nucleic acids research, 42(W1), W240-W245.

Submitting to:

8th CAPRI assessment meeting

Primary author: Dr CELIS GARZA, Daniel (CCP4, Research Complex at Harwell, STFC Rutherford-Appleton Laboratory, UK)

Co-authors: Dr ELLAWAY, Joseph (PDBe, EMBL European Bioinformatics Institute, Genome Campus, UK); Dr VELANKAR, Sameer (PDBe, EMBL European Bioinformatics Institute, Genome Campus, UK); Dr KRISSINEL, Eugene (CCP4, Research Complex at Harwell, STFC Rutherford-Appleton Laboratory, UK)

Presenter: Dr CELIS GARZA, Daniel (CCP4, Research Complex at Harwell, STFC Rutherford-Appleton Laboratory, UK)

Session Classification: CAPRI

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

A new approach for assigning names and persistent, unique identifiers to macromolecular complexes in the Protein Data Bank

Macromolecular complexes are key functional components in nearly all biological processes, and achieving an atomic-level understanding of them is crucial for unravelling their molecular mechanisms. The Protein Data Bank (PDB) is the primary worldwide repository of experimentally determined macromolecular structures, offering valuable insights into the dynamics, conformations, and functional states of biological assemblies. However, the current annotation practices within the PDB lack consistent naming conventions for these assemblies and do not assign persistent, unique identifiers to them, making it challenging in downstream data analysis to identify all the PDB entries that represent the same assembly.

We propose a novel approach that uses external data resources, including the Complex Portal, UniProt, and Gene Ontology, to precisely describe macromolecular assemblies. Using our approach, we successfully assigned standardised names to more than 90% of unique assemblies in the PDB and provided each assembly with persistent, unique identifiers. To further enrich our understanding of these assemblies, we also computed symmetry data for each using the AnAnaS program. This annotation is valuable as symmetry plays a key role in the structural and functional characterisation of macromolecular complexes, offering insights into their biological roles and interactions.

In addition to making the assigned identifiers and assembly names accessible via our FTP area, we also integrated this information to enhance the user experience of the PDBe search system, making it easy to find all the PDB entries for a specific assembly composition.

The naming and standardisation of assembly data enhances the PDBe, makes biological assemblies easier to discover, and facilitates a deeper understanding of macromolecular mechanisms.

Submitting to:

8th CAPRI assessment meeting

Primary authors: APPASAMY, Sri (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)); Dr BERRISFORD, John (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI))

Co-authors: GABOROVA, Romana (CEITEC – Central European Institute of Technology); Mr NAIR, Sreenath (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)); Mr ANYANGO, Stephen (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)); Dr GRUDININ, Sergei (Univ. Grenoble Alpes, CNRS); Mr DESHPANDE, Mandar (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)); Dr ARMSTRONG, David (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)); PIDRUCHNA, Ivanna (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)); Mr ELLAWAY, Joseph (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)); Dr

LEINES, Grisell (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)); Dr GUPTA, Deepti (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)); Dr HARRUS, Deborah (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)); Dr VARADI, Mihaly (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI)); Dr VELANKAR, Sameer (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI))

Presenter: APPASAMY, Sri (Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI))

Contribution ID: 7   Type: **Poster**

Computational Design of Peptides as Detectors, Drugs and Biomaterials

Computational design of peptides that can (a) recognize/bind to specific protein interfaces and (b) self-assemble into nano-scale architectures such as β-sheet-based fibrils (amyloid) have potential applications in healthcare and advanced biomaterial fabrication. Screening sequence space via experiments to discover peptides that bind to proteins and self-assemble is challenging, although some progress has been made in the former category. This presentation focusses on novel computational protocols and their applications to discover peptides that (a) bind to biomolecular protein targets and (b) self-assemble into amyloid structures. The key computational steps involved in both protocols are: (a) a Monte-Carlo based peptide design algorithm (a peptide binding design algorithm, PepBD and a peptide assembly design, PepAD) to identify potential protein-binding peptides and self-assembling peptides, respectively, and (b) atomistic molecular dynamics (MD) simulations to examine the thermo-stability of the designed peptide fibrils and the binding free energy of the peptide–protein complex. An additional step in the design of self-assembling peptides is discontinuous molecular dynamics (DMD) simulations, a coarse-grained approach that reveals the kinetic pathway for fibril formation taken by the in-silico discovered peptides.

First, we focus on our efforts to identify peptides that bind to SARS-CoV-2 Receptor Binding Domain (RBD) and inhibit SARS-CoV-2 cell entry. Our in-silico 23 mer peptide, P4, binds to the Wuhan-Hu-1, Kappa, and Delta strain of the SARS-CoV-2 RBD with micromolar level affinity, as measured using a tryptophan fluorescence quenching assay but cannot outcompete ACE2 in a competitive ELISA assay. Second, we describe a peptide-based drug that we designed to bind to Clostridioides difficile toxin A with the aim of neutralizing Clostridioides difficile toxicity in large intestinal cells. This 8-mer peptide, SA1, binds to the Clostridioides difficile toxin A glucosytransferase domain. The efficacy of peptide SA1 was tested using a trans-epithelial electrical resistance (TEER) assay on monolayers of the human gut epithelial culture model. Peptide SA1 blocks TcdA toxicity in jejunum (small intestine) cells and in colon epithelial cells and exhibits a binding affinity in the nanomolar ranges. Finally, we discuss the design of peptides that form amyloid-like structures. These peptides contribute to a new generation of amyloid-structure inspired peptide-based biomaterials. Overall, the combination of fundamental peptide:protein interaction and protein aggregation studies, and the rational design of proteins and short peptides opens broad opportunities in healthcare and soft-materials.

Submitting to:

8th CAPRI assessment meeting

Primary authors: SARMA, Sudeep (Johns Hopkins University); Prof. HUDALLA, Gregory (University of Florida); Prof. MENEGATTI, Stefano (North Carolina State University); Prof. MAGNESS, Scott (University of North Carolina at Chapel Hill); Prof. CROOK, Nathan (North Carolina State University); Prof. PARAVASTU, Anant (Georgia Institute of Technology); Prof. GRAY, Jeffrey (Johns Hopkins University); Prof. HALL, Carol (North Carolina State University)

Presenter: SARMA, Sudeep (Johns Hopkins University)

Contribution ID: **8**   Type: **Talk**

The Influence of Globular Domains on an Intrinsically Disordered Region of p53: A Molecular Dynamics Investigation

Wednesday, 14 February 2024 08:55 (25 minutes)

Unraveling the mysteries of **Intrinsically Disordered Proteins (IDPs)** and **Regions (IDRs)** is one of the greatest challenges in the 21st century. The flexibility of these regions allow proteins to adopt vastly different conformations, allowing them to facilitate the binding and unbinding of vital activation sites. These regions, even in well-studied proteins such as the tumor suppressor p53 protein, are not well understood. We utilized data from a CryoEM structure of p53, and **terminally locked the regions** to reflect the existence of neighboring globular regions on the behavior. We ran trajectories with the end-to-end distances locked at 0.5, 3.0, 5.0 and 7.0 nm to reflect to observe the changes in the types and quantities of secondary structures. We computed the **average chemical shifts (CS)** for each of these trajectories and found that experimental CS agreement peaks at the distance specified in the CryoEM structure, with root-mean-squared deviation **improved by between 0.12 and 0.26 ppm** (depending on the atom) from the traditional MD trajectory (unrestricted). The results of this investigation are to be published in JCTC entitled *"Exploring the Role of Globular Domain Locations on an Intrinsically Disordered Region of p53: A Molecular Dynamics Investigation"* along with several other articles in the works about other regions within p53. The goal for this investigation is **to give a complete understanding of the multiple conformations of p53**, provide insight into the performance of flexible IDRs in primarily globular proteins, and pave the way for a collaborative SAXS investigation into the inner machinations of these elusive structures.

Submitting to:

Integrative Computational Biology workshop

Primary author: BAKKER, Michael (Charles University)

Co-authors: Mr SØRENSEN, Henrik (Lund University); SKEPO, Marie

Presenter: BAKKER, Michael (Charles University)

Session Classification: Protein conformational flexibility and solution experiments

Contribution ID: **9**
Type: **Talk**

Protein Dynamics at extreme Temperatures

Tuesday, 13 February 2024 16:20 (25 minutes)

Life has adapted to extreme conditions on Earth. One of the most striking evidences of adaptation to extreme environments are bacteria that are capable of thriving in a vast temperature range, from below 0° C in glacial waters to above 100° C in deep-sea hydrothermal vents. It is known that the individual molecular components of these organisms, exhibit enhanced stability and resistance to the temperature stress. A research focus lies on the proteins, which are the most abundant and less stable macromolecules in the cell. However, the link among the individual protein stability, and the process of cell death caused by the raise of temperature is not yet clear. Understanding the biophysical determinants of cellular thermostability would be fundamental from a theoretical, biotechnological and clinical prospective [2], [3].

In a theoretical work by Dill and his co-workers, [1], it has been proposed that the temperature induced cell death follows a collective unfolding of the proteome, with the entire set of proteins approaching their individual melting in a narrow temperature range.

However, this picture has been recently challenged by further experimental and simulation work, [4], [5]. Experimentally, it was possible to monitor the amount of proteins actually unfolding at the cell death temperature. These recent studies showed that it is only a small fraction of proteins that unfold at the cell death temperature. This was further confirmed by Sterpone and his co-workers, who investigated the dynamical profile of the proteins in the E. Coli, a mesophilic bacterium. Using a combination of Neutron-scattering experiments and Molecular dynamics simulations they showed that a dynamical catastrophe occurs at the cell death temperature which is caused by only around 10% of the proteins unfolding [6].

The goal of this project is to combine all-atom Molecular Dynamics simulations and Neutron Scattering experiments to further investigate the dynamics of the proteins in the case of extremophilic bacterium and to reveal whether it is a small set of proteins that unfold as it has been shown for the mesophile. The bacteria selected for this investigation are the P. Arcticus, a psychrophile, and the A. Aeolicus, a hyperthermophile.

In this occasion, it will be presented how MD simulations and Neutron scattering experiments can be combined to model the dynamics of the proteins and to quantify the number of unfolded proteins at the cell death temperature. Different dynamical aspects that have been revealed so far using MD simulations and NS experiments, such as the global diffusion coefficient and the mean square displacement of the proteins of these two bacteria, will be demonstrated.

1. Dill, K. A. , Ghosh, K. , Schmit, J. D. (2011) 'Physical limits of cells and proteomes' PNAS 108,17876

2. Coffey, D. S. , Getzenberg, R. H. , DeWeese, T. L. (2006) 'Hyperthermic biology and cancer therapies: A hypothesis for the Lance Armstrong effect' JAMA 296, 445.

3. Polizzi, K. M., Bommarius, A. S. , Broering, J. M. , Chaparro-Riggers, J. F. (2007) 'Stability of Biocatalysts' Curr. Opin. Chem. Biol. 11, 220

4. Leuenberger, P. et al. (2017) 'Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability', Science, 355, 6327

5. Jarazab, A. et al. (2020) 'Meltome atlas—thermal proteome stability across the tree of life', Nature Methods, 17, 495.

6. Daniele Di Bari, Stepan Timr, Marianne Guiral, Marie-Thérèse Giudici-Orticoni, Tilo Seydel, Christian Beck, Caterina Petrillo, Philippe Derreumaux, Simone Melchionna, Fabio Sterpone, Judith Peters, and Alessandro Paciaroni (2023) 'Diffusive Dynamics of Bacterial Proteome as a Proxy of Cell Death', ACS Central Science 9 (1), 93-102

Submitting to:

Integrative Computational Biology workshop

Primary authors: Mr PACIARONI, Alessandro; CAVIGLIA, Beatrice (Universite' Paris Cite', University of Perugia); Mr STERPONE, Fabio; PETERS, Judith

Presenter: CAVIGLIA, Beatrice (Universite' Paris Cite', University of Perugia)

Session Classification: Protein conformational flexibility and solution experiments

Contribution ID: **10**   Type: **Talk**

Ensemble based calculations of spin-spin couplings in proteins: from ordered to disordered peptides

Tuesday, 13 February 2024 18:05 (25 minutes)

To acquire a better understanding of how proteins work at the molecular level, it is crucial to understand their structural characteristics. Various techniques have been developed to achieve this, such as computer-based methods for calculating and predicting NMR measurements including the spin-spin coupling constants (SSCCs). Multi-scale calculations combining molecular dynamics simulations with density functional theory (DFT) calculations have become particularly feasible upon the emergence of fragmentation techniques and many successful studies for structured structured proteins appeared since then. On the contrary, examples of applications for intrinsically disordered proteins (IDPs) remain virtually non-existent due to prohibitive computational demands caused by the use of extensive sequential structural ensembles. To alleviate the problem, we pursue the design of smaller size ensembles through the dimensionality reduction of the IDP conformational landscape and clustering of similar conformations. In our contribution, we will show the performance of the workflow employing the t-distributed Stochastic Neighbor Embedding (t-SNE) and hierarchical clustering for the structured proteins GB3 and Ubiquitine and compare it to the disordered protein fragment Tau(210-240).Sequential and dimensionally reduced/cluster-based ensembles will be validated through correlation of SSCCs predicted with empirically parametrized Karplus equations and experimental NMR data.

Submitting to:

Integrative Computational Biology workshop

Primary author: GAFFOUR, Amina

Co-author: Mr BAKKER, Micheal

Presenter: GAFFOUR, Amina

Session Classification: Protein conformational flexibility and solution experiments

Contribution ID: **11**   Type: **Talk**

VTX: High-performance molecular structure and dynamics visualization

Monday, 12 February 2024 14:55 (25 minutes)

Molecular visualization is a critical task usually performed by structural biologists and bioinformaticians to aid three processes that are essential in science and fundamental to understand structural molecular biology: synthesis, analysis and communication [1].
Here we present VTX, a molecular visualization software that includes a real-time high-performance molecular graphics engine dedicated to the visualization of the structure and dynamics of massive molecular systems. VTX disposes of an interactive camera system controllable via the keyboard and/or mouse that includes different modes: 1. a classical trackball mode where the camera revolves around a fixed focus point and 2. a first-person free-fly navigation mode where the user fully controls the movement of the camera. VTX includes an intuitive and highly usable graphical user interface and tools designed for expert and non-expert users. It is free for non-commercial use at http://vtx.drugdesign.fr

[1] Olson, AJ. Perspectives on structural molecular biology visualization: from past to present. J Mol Biol (2018) ; 430(21): 3997–4012.

Submitting to:

Integrative Computational Biology workshop

Primary authors: Dr MARIA, Maxime (Université de Limoges); Mr GUIONNIÈRE, Simon (Conservatoire National des Arts et Métiers); Mr DACQUAY, Nicolas (Conservatoire National des Arts et Métiers); Mr PLATEAU-HOLLEVILLE, Cyprien (Université de Limoges); Mr NAIMI, Yassine (Qubit Pharmaceuticals SAS); Prof. PIQUEMAL, Jean-Philip (Sorbonne Université); Dr LEVIEUX, Guillaume (Conservatoire National des Arts et Métiers); MONTES, Matthieu (Conservatoire National des Arts et Métiers)

Presenter: MONTES, Matthieu (Conservatoire National des Arts et Métiers)

Session Classification: Computational Methods and Deep Learning

Contribution ID: **12**   Type: **Talk**

Impact of AI-Based Modeling on the Accuracy of Protein Assembly Prediction: The CASP Perspective

Thursday, 15 February 2024 09:00 (30 minutes)

In CASP15, 87 predictors submitted around 11,000 models on 41 assembly targets. The community demonstrated exceptional performance in overall fold and interface contact prediction, achieving an impressive success rate of 90% (compared to 31% in CASP14). This remarkable accomplishment is largely due to the incorporation of DeepMind's AF2-Multimer approach into custom-built prediction pipelines. To evaluate the added value of participating methods, we compared the community models to the baseline AF2-Multimer predictor. In over 1/3 of cases, the community models were superior to the baseline predictor. The main reasons for this improved performance were the use of custom-built multiple sequence alignments, optimized AF2-Multimer sampling, and the manual assembly of AF2-Multimer-built subcomplexes. The best three groups, in order, are Zheng, Venclovas, and Wallner. Zheng and Venclovas reached a 73.2% success rate over all (41) cases, while Wallner attained a 69.4% success rate over 36 cases. Nonetheless, challenges remain in predicting structures with weak evolutionary signals, such as nanobody-antigen, antibody-antigen, and viral complexes. Expectedly, modeling large complexes remains challenging due to their high memory compute demands.

In addition to the assembly category, we assessed the accuracy of modeling interdomain interfaces in the tertiary structure prediction targets. Models on seven targets featuring 17 unique interfaces were analyzed. Best predictors achieved a 76.5% success rate, with the UM-TBM group being the leader. In the interdomain category, we observed that the predictors faced challenges, as in the case of the assembly category, when the evolutionary signal for a given domain pair was weak, or the structure was large. Overall, CASP15 witnessed unprecedented improvement in interface modeling, reflecting the AI revolution in CASP14.

Submitting to:

8th CAPRI assessment meeting

Primary authors: Mr KRYSHTAFOVYCH, Andriy (Protein Structure Prediction Center, UC Davis); Ms OZDEN, Burcu (Izmir Biomedicine and Genome Center, DEU); KARACA, Ezgi (Izmir Biomedicine and Genome Center, DEU)

Presenter: KARACA, Ezgi (Izmir Biomedicine and Genome Center, DEU)

Session Classification: CAPRI

Contribution ID: **13**   Type: **Talk**

Shape retrieval methods for the classification of protein surfaces

Friday, 16 February 2024 09:00 (30 minutes)

Macromolecular complexes play a crucial role in almost all biological functions in living cells. Elucidating structural features in proteins is essential to understanding their underlying functions and binding activity. With the rapid accumulation of protein-structure data through AI methods and experimental techniques, such as cryo-electron microscopy, there is a growing demand for efficient approaches to detect protein-structure similarity in real-time database searches. Recently, high-speed computer vision methods have been adapted to perform efficient protein surface comparisons by computing geometric invariant descriptors of 3D features. Here, we present a pipeline that maps 3D surface shape descriptors of proteins using matching and clustering algorithms to identify structure similarity across the Protein Data Bank. We represent the surface shapes of proteins with moment-based 3D Zernike descriptors and spectral signatures. Using matching algorithms based on functional maps and distances can retrieve local similarities between protein surfaces. This pipeline will be integrated into the PDBe and PDBe Knowledge Base infrastructure. Its primary goal is to improve the classification of macromolecular structures and interfaces, which can significantly aid in identifying similarities in protein conformations and facilitate the use of AI methods for predicting macromolecular complexes.

Submitting to:

8th CAPRI assessment meeting

Primary author: DIAZ LEINES, Grisell (EMBL-EBI)

Co-authors: Prof. KIHARA, Daisuke (Purdue University); Dr UHLMANN, Virginie (EMBL-EBI); VELANKAR, Sameer (EMBL-EBI)

Presenter: DIAZ LEINES, Grisell (EMBL-EBI)

Session Classification: CAPRI

Contribution ID: **14**   Type: **Poster**

Deep Learning in Protein Docking: Discriminative vs. Generative Models

Traditional protein-protein docking algorithms often rely on extensive candidate sampling and subsequent re-ranking. However, these steps are time-consuming, limiting their applicability in scenarios requiring high-throughput complex structure prediction, such as structure-based virtual screening. The advent of deep learning methods has brought acceleration to the docking process, leveraging the power of modern GPUs. Despite this acceleration, a challenge persists in the limited generalization of docking accuracy beyond the training data. Deep learning approaches in protein-protein docking can be broadly classified as either discriminative or generative. Discriminative models establish a functional mapping from input (sequences and structures of docking partners) to output (docked protein complex structure). In contrast, generative models learn the inherent distribution of a given dataset (docked structures) and generate novel and realistic samples that exhibit similarities to the original data. This study compares two representative models from our lab – GeoDock, a discriminative-type model, and DockDiffusion, a generative-type model – for protein docking. Utilizing identical training and testing datasets, along with similar neural network architectures, we compare the generalizability, docking accuracy, and inference speed of both methods. The results suggest that while discriminative models like GeoDock can accurately map input features to output structures within the training data, their performance may decline when faced with unseen protein interactions. On the other hand, generative models like DockDiffusion, by learning the underlying distribution of docked structures, exhibit a higher potential for generalization to diverse protein interfaces.

Submitting to:

Integrative Computational Biology workshop

Primary author: Mr CHU, Lee-Shin (Johns Hopkins University)

Co-author: Dr GRAY, Jeffrey (Johns Hopkins University)

Presenter: Mr CHU, Lee-Shin (Johns Hopkins University)

Contribution ID: **16**   Type: **Talk**

Deciphering protein conformational dynamics with cryo-EM and MD simulations

Tuesday, 13 February 2024 15:00 (25 minutes)

Cryo-Electron Microscopy (cryo-EM) allows conformational studies of macromolecular complexes in their close-to-native state, essential for understanding their working mechanisms and for structure-based drug development. However, deciphering continuous conformational transitions of macromolecules, through the main cryo-EM processing techniques, Single Particle Analysis (SPA) and cryo-Electron Tomography (cryo-ET), is challenging partly due to the low signal-to-noise ratio. I will present new image processing methods based on Molecular Dynamics (MD) simulations that allow extracting continuous conformational variability from SPA and cryo-ET data.

Submitting to:

Integrative Computational Biology workshop

Primary author: VUILLEMOT, Rémi (LJK, Université Grenoble Alpes, France)

Co-author: JONIC, Slavica (CNRS & Sorbonne University, Paris)

Presenter: VUILLEMOT, Rémi (LJK, Université Grenoble Alpes, France)

Session Classification: Protein conformational flexibility and solution experiments

Contribution ID: **17**   Type: **Talk**

Studying specificity in protein–glycosaminoglycan recognition with umbrella sampling

Monday, 12 February 2024 14:30 (25 minutes)

Studying specificity in protein–glycosaminoglycan recognition with umbrella sampling

Abstract

Glycosaminoglycans (GAGs) with repeating disaccharide units intricately engage with proteins, playing a crucial role in the spatial organization of the extracellular matrix and the transduction of biological signals in cell to modulate several biochemical processes. Last few decades have shown the importance of GAG research for obtaining insights into various physiological, pathological, and therapeutic aspects mediated by the direct interactions between the GAG molecules and diverse proteins. The structural and functional heterogeneities of GAGs and their ability to bind specific proteins are determined by the sugar composition of the GAG, the size of the GAG chains, and the degree and pattern of sulfation. A deep understanding of the interactions in protein–GAG complexes is essential to explain their biological functions. In this study, we use umbrella sampling (US) to pull away a GAG ligand from the binding site and then pull it back in. We analyze the binding interactions between GAGs of three types (heparin, desulfated heparan sulfate, and chondroitin sulfate) with three different proteins (basic fibroblast growth factor, acidic fibroblast growth factor, and cathepsin K). The focus of our study is to evaluate whether the US approach can reproduce experimentally obtained structures, and how useful it can be for getting a deeper understanding of GAG properties, especially protein recognition specificity and multipose binding. The study shows that the binding free energy landscape in the proximity of the GAG native binding pose is complex and implies the co-existence of several binding poses. The sliding of a GAG chain along a protein surface could be a potential mechanism of GAG particular sequence recognition by proteins. [1]

Reference

1. Marcisz, M., Anila, S., Gaardløs, M., Zacharias, M., & Samsonov, S. A. (2023). Studying specificity in protein–glycosaminoglycan recognition with umbrella sampling. Beilstein Journal of Organic Chemistry, 19(1), 1933-1946.

Submitting to:

Integrative Computational Biology workshop

Primary authors: Dr SEBASTIAN, Anila (Faculty of Chemistry, University of Gdańsk, Gdańsk, Poland.); Mr MARCISZ, Mateusz (Faculty of Chemistry, University of Gdańsk, Gdańsk, Poland.); Dr GAARDLØS, Margrethe (Faculty of Chemistry, University of Gdańsk, Gdańsk, Poland.); Prof. ZACHARIAS, Martin (Physics Department, Technical University of Munich, Garching, Germany.); Dr SAMSONOV, Sergey (Faculty of Chemistry, University of Gdańsk, Gdańsk, Poland.)

Presenter: Dr SEBASTIAN, Anila (Faculty of Chemistry, University of Gdańsk, Gdańsk, Poland.)

Session Classification: Computational Methods and Deep Learning

Contribution ID: **18**   Type: **Talk**

Mechanistic study of MexAB-OprM efflux pump by an integrative approach

Tuesday, 13 February 2024 14:10 (25 minutes)

Bacterial infections remain a major public concern due to the accelerated increase in the appearance of antibiotic resistance. Among the different mechanisms used by bacteria to resist to antibiotics, the active efflux plays a major role. In Gram-negative bacteria, this is achieved by tripartite efflux pumps that form a macromolecular assembly spanning both membranes of the cellular wall, the most studied being AcrAB-TolC in Escherichia coli and MexAB-OprM in Pseudomonas aeruginosa, two pathogens highly involved in patients' death associated to nosocomial diseases.

Along with functional and in silico studies, many structures were solved of the individual components of these pumps and more recently of the whole assembly [1-5]. Nevertheless, a lot of questions concerning the assembly, the mechanism of efflux, and the opening of the whole pump are a matter of active research, as the blockage of these pumps could restore the utility of the actual therapeutic arsenal. The comparison of the structures of the whole MexAB-OprM pump and of MexB we solved by cryo-EM from the same grid, combined with results from different approaches, led us to a better comprehension of the functional mechanism of the pump. These results helped us to clarify the role of the different protein actors, so as to identify the different targets for therapeutic molecule development.

Submitting to:

Integrative Computational Biology workshop

Primary author: BROUTIN, Isabelle (université Paris Cité)

Co-authors: Dr DAURY, Laetitia; Dr LUSTIG, Margaux; Dr LAMBERT, Olivier; Dr NTSOGO-ENGUENE, Yvette; Dr PHAN, Gilles

Presenter: BROUTIN, Isabelle (université Paris Cité)

Session Classification: Protein conformational flexibility and solution experiments

Contribution ID: **19**   Type: **Talk**

MassiveFold: optimized massive sampling with AlphaFold2

Wednesday, 14 February 2024 17:15 (30 minutes)

Massive sampling with AlphaFold-multimer(1,2) showed impressive results for structural prediction of macromolecular assemblies at CASP15-CAPRI(3). Generating a very large number of predictions (>1000) and pushing their diversity by playing with the neural network model versions, the number of recycle steps, the use of templates or not and the activation of the dropout in the Evoformer and in the structure module, ranked this method first for the prediction of complexes(4). Subsequently named AFsample(5), the method allows to run massive sampling with AlphaFold's neural network models v1 and v2. We now created MassiveFold, which is based on AFsample and integrates all these diversity parameters, including all the neural network models provided by all the versions of AlphaFold-multimer (v1 to v3). Our tool is optimized to run on a parallel computing CPU/GPU infrastructure as it automatically performs the multiple sequence alignments on CPU and then sends individual structure prediction runs in batches to GPU servers, afterwards gathering all the prediction results to produce a combined ranking. The final results contain many plots including the well-known plDDT and Predicted Aligned Error plots, but also diagrams and box plots that show the diversity in predictions. MassiveFold allows thus to take full advantage of a CPU/GPU computing infrastructure and to save up to months of calculation with its optimized parallelization feature.

For CAPRI Round 55, we used MassiveFold to compute 6 runs for each target, generating 1005 structures per run. Each run was parameterized with the 3 versions of neural network models, 21 recycles, 0.5 threshold for early stop tolerance on recycling, and of variations in the activation (true/false) of the following parameters: dropout in the Evoformer, dropout in the structure module and/or template activation, totaling 6030 predicted structures for each target. All predictions were ranked following the ipTM+pTM AlphaFold confidence measure. To ensure diversity in the top 5 submitted predictions, the TM-score between the first ranked model and each model was computed, and a K-means clustering was performed on these TM-scores to create 5 clusters. The model with the highest AlphaFold confidence score in each cluster was kept for submission. The last 95 predictions for each target were chosen randomly among the remaining 6025 structures.

1. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021).

2. Evans, R. et al. Protein complex prediction with AlphaFold-Multimer. 2021.10.04.463034 https://www.biorxiv.org/content/10.1101/2021.10.04.463034v1 (2021) doi:10.1101/2021.10.04.463034.

3. Lensink, M. F. et al. Impact of AlphaFold on structure prediction of protein complexes: The CASP15-CAPRI experiment. Proteins 91, 1658–1683 (2023).

4. Wallner, B. Improved multimer prediction using massive sampling with AlphaFold in CASP15. Proteins (2023) doi:10.1002/prot.26562.

5. Wallner, B. AFsample: improving multimer prediction with AlphaFold using massive sampling. Bioinforma. Oxf. Engl. 39, btad573 (2023).

Submitting to:

8th CAPRI assessment meeting

Primary authors: RAOURAOUA, Nessim (CNRS University of Lille); LENSINK, Marc (CNRS); BRYS-BAERT, Guillaume (CNRS University of Lille)

Presenter: BRYSBAERT, Guillaume (CNRS University of Lille)

Session Classification: CAPRI

Contribution ID: **20**   Type: **Talk**

Multistage Docking Approach for Protein-RNA Interactions Prediction in CAPRI and CASP

Thursday, 15 February 2024 14:00 (30 minutes)

Protein-RNA interactions and recognition are essential in gene expression, regulation of transcription, and other biological processes. Prediction of protein and RNA is also a new challenge category for CAPRI and CASP. In this work, we proposed a multistage docking protocol called CoDockPR, which integrates the shape complementarity, knowledge-based scoring functions, and interface similarity evaluation. We trained a knowledge-based scoring function by the iterative method to discriminate the near-native structures of protein-RNA interaction. FFT-based method was used to systematically evaluate shape complementarity, and the retained conformations were evaluated by the knowledge-based scoring function. In addition, we established a protein-RNA interface library and further ranked the conformations based on the interface similarity. By testing on protein-RNA docking benchmark 1.0, CoDockPR remarkably improves the success rate and hit count. The CoDockPR program were applied for the protein-RNA complex prediction of T185 in CAPRI, and T1189 and T1190 in CASP. By compared with the available crystal structures, we analyzed the predicted results and the existing issues. Considering its robust predictive performance, our docking protocol is a good alternative for the protein-RNA interactions prediction.

Submitting to:

8th CAPRI assessment meeting

Primary authors: Dr KONG, Ren (Jiangsu University of Technology); CHANG, Shan (Jiangsu University of Technology)

Presenter: CHANG, Shan (Jiangsu University of Technology)

Session Classification: CAPRI

Contribution ID: **21**   Type: **Talk**

Challenges in visualizing structural flexibility of the SAGA complex

Tuesday, 13 February 2024 14:35 (25 minutes)

Cryogenic electron microscopy (cryo-EM) stands out as a widely utilized technique for elucidating the structures of macromolecular proteins. This method involves capturing projections of protein specimens embedded in thin, amorphous ice, preserving them within the vacuum of an electron microscope. The resulting structure emerges from the averaging of thousands of particle images, revealing only the structurally stable segments of protein complexes, while more flexible regions appear blurred.

Efforts have been undertaken to recover the structure of these flexible regions and characterize their inherent flexibility. In this context, I will discuss such approaches and their associated challenges using the transcription co-activator complex SAGA as a case study. SAGA, or Spt-Ada-Gcn5-acetyltransferase, constitutes a 19-subunit complex that stimulates transcription through two chromatin-modifying enzymatic modules. Additionally, it facilitates the initiation of the pre-initiation complex on DNA by delivering the TATA box binding protein (TBP), a crucial step in the expression of protein-encoding genes.

Submitting to:

Integrative Computational Biology workshop

Primary author: PAPAI, Gabor (INSERM)

Co-authors: Dr BEN-SHEM, Adam (INSERM); Dr SCHULTZ, Patrick (CNRS)

Presenter: PAPAI, Gabor (INSERM)

Session Classification: Protein conformational flexibility and solution experiments

Contribution ID: **22**   Type: **Poster**

Detailed analysis of a thermostable protein-DNA complex: the case of Sac7d as a prototype for protein-DNA interaction

Sac7d is a 7kDa protein belonging to the class of the small chromosomal proteins from archeon Sulfolobus acidocaldarius. Sac7d was discovered in 1974 in Yellowtone National Parks geysers, and studied extensively since then for its remarkable stability at large pH and temperature ranges. Sac7d binds to DNA minor groove to protect DNA from these extreme conditions by increasing its melting temperature. In this study, we analyzed the Sac7d-DNA complex using 1 µs molecular dynamics simulations. The interaction energy ofthe interface was decomposed using Molecular Mechanics Generalized Born Surface Area (MM/GBSA) to determine at the amino acids level which residue contributed most to DNA binding. Twelve amino acids are essential for DNA binding, out of which three were not identified previously by former studies.

One of these new aminoacids, R63, may be involved in a dynamic protein-DNA interaction. Our analysis provides novel insights into how the Sac7d chaperones allows to protect DNA from degradation in extreme conditions.

Submitting to:

8th CAPRI assessment meeting

Primary author: ALVAREZ SANCHEZ, Elena (Nantes Université, US2B, CNRS, UMR6286 & Affilogic, Nantes, France)

Presenter: ALVAREZ SANCHEZ, Elena (Nantes Université, US2B, CNRS, UMR6286 & Affilogic, Nantes, France)

Contribution ID: **23**   Type: **Talk**

Hinge disulfides in human IgG2 CD40 antibodies modulate receptor signaling by regulation of conformation and flexibility

Tuesday, 13 February 2024 18:30 (25 minutes)

The human IgG2 (hIgG2) isotype is unique in its ability to undergo redox based shuffling of its hinge disulfides. Previous work identified hIgG2 as the optimal isotype for mediating activation (agonism) of multiple tumour necrosis factor receptor superfamily (TNFRSF) members, with the hinge disulfides shown to be critical. CD40 is a co-stimulatory TNFRSF receptor that can be targeted using immunostimulatory monoclonal antibodies (mAb) and is of interest for cancer immunotherapy [1].

Using, the clinically relevant anti-CD40 mAb chiLOB7/4, we previously generated a series of cysteine to serine mutation hinge variants that exhibited activities spanning the agonism/antagonism spectrum [2]. To elucidate the mechanisms underlying these activities, a combination of molecular and structural biology along with simulations were utilised [3].

X-ray crystallography was used to determine the structures of the variants of interest, with sulfur-SAD used to confirm the positions of the disulfide bonds in the hinge. It was revealed that disulfide bonds linking opposing heavy chains ('cross-overs') were present in agonistic variants, while a more typical ladder-like topology was observed in the non-agonistic variants. Restrictions imposed by the crystal lattice prevented investigation of the dynamics of the different variants, leading to solution state techniques being used.

Small-angle X-ray scattering (SAXS) experiments revealed that the agonistic variants were conformationally restricted in solution when compared to the non-agonistic variants, which were more flexible and elongated. To bridge the gap between the atomic detail from the crystallographic information and the dynamic information inherent in SAXS, molecular dynamics (MD) and enhanced sampling simulations were used.

MD and metadynamics simulations were used to further characterise the underlying mechanism behind the conformational restriction. SAXS-based reweighting of the conformational pool generated by MD revealed a clear trend of conformational restriction that correlated with increased agonism. Principal component analysis of the MD simulations identified the F(ab)2 hinge bending and torsion angles as being key global motions that differed between agonistic and non-agonistic variants. These motions were then taken into metadynamics simulations as collective variables (CVs) to allow for enhanced sampling of these motions. Reconstructions of the free-energy surfaces for these CVs reveal restrictions in the accessible conformational space that correlate with increasing agonism.

These insights reveal that regulation of mAb conformation and flexibility as a key mechanism by which modulation of agonism can be achieved. Future work aims to investigate if this mechanism of modulation is more broadly applicable to other members of the TNFRSF or other receptor families and if further restriction of flexibility can deliver improved agonism.

References:

[1] Remer M, White A, Glennie M, Al-Shamkhani A, Johnson. P. (2017) The Use of Anti-CD40 mAb in Cancer. Curr Top Microbiol Immunol. 405:165-207.
[2] White, A. L., Chan, H. T. C., French, R. R., Willoughby, J., Mockridge, C. I., Roghanian, A., Penfold, C. A., Booth, S. G., Dodhy, A., Polak, M. E., Potter, E. A., Ardern-Jones, M. R., Verbeek, J. S., Johnson, P. W. M., Al-Shamkhani, A., Cragg, M. S., Beers, S. A., Glennie, M. J. (2015) Conformation of the Human Immunoglobulin G2 Hinge Imparts Superagonistic Properties to Immunostimulatory Anticancer Antibodies. Cancer Cell 27, 138-148.

[3] Orr, C. M., Fisher, H., Yu, X., Chan, C. H., Gao, Y., Duriez, P. J., Booth, S. G., Elliott, I., In-zhelevskaya, T., Mockridge, C. I., Penfold, C. A., Wagner, A., Glennie, M. J., White, A. L., Essex, J. W., Pearson, A. R., Cragg, M. S., Tews, I. (2022) Hinge disulfides in human IgG2 CD40 antibodies modulate receptor signaling by regulation of conformation and flexibility. Sci. Immunol. 7.

Hayden Fisher (1,2,3,4), Christian M Orr (2,3,7), Xiaojie Yu (2), Claude HT Chan (2), Isabel Elliott (2,3,4), Christine A. Penfold (2), Patrick J. Duriez (2,5), Tatyana Inzhelevskaya (2), C. Ian Mockridge (2), Mark D. Tully (1), Jonathan W. Essex (4,6), Mark S. Cragg (2,6), Ivo Tews (3,6).

(1) European Synchrotron Radiation Facility, Grenoble, 38000, FR.
(2) University of Southampton, School of Cancer Sciences, Centre for Cancer Immunology; Southampton, SO16 6YD, UK.
(3) University of Southampton, School of Biological Sciences; Southampton SO17 1BJ, UK.
(4) University of Southampton, School of Chemistry; Southampton SO17 1BJ, UK.
(5) University of Southampton, School of Cancer Sciences, CRUK Protein Core Facility; Southampton, SO16 6YD, UK.
(6) University of Southampton, Institute for Life Sciences; Southampton SO17 1BJ, UK.
(7) Diamond Light Source; Didcot, OX11 0FA, UK.

Submitting to:

Integrative Computational Biology workshop

Primary author: FISHER, Hayden (ESRF)

Presenter: FISHER, Hayden (ESRF)

Session Classification: Protein conformational flexibility and solution experiments

Contribution ID: 24

Type: **Talk**

Asymmetric hydrophobic mismatch in assembly of ATP synthase rotor ring

Tuesday, 13 February 2024 11:35 (25 minutes)

Rotary ATPases are multisubunit enzyme complexes that couple synthesis or hydrolysis of ATP molecules with transport of ions across a membrane. Their transmembrane part includes a homo- or a heterooligomer of *c* subunits called a *c*-ring, with a patch of several lipids confined inside it. Little is known about this patch; it is usually not well resolved in experimental structures, but it is clear that the lipids in it are displaced relative to the surrounding membrane. Here, we show that a protocol involving coarse-grained and atomistic molecular dynamics simulations can be used to obtain a model of the rotor ring protein-lipid assembly that fits well the experimental densities. We then study the mechanism of self-assembly of tetradecameric spinach chloroplast ATP synthase *c*-ring and demonstrate that *c* subunits display unusual asymmetric hydrophobic mismatch. Monomers and partially assembled oligomers cause a deformation of the surrounding membrane, which results in their mutual attraction. Because of asymmetry in the deformation, the subunits assume a relative orientation that favors correct assembly of higher-order oligomers and, eventually, of the whole ring. We estimate the binding energies of different oligomers and build a model for the complete assembly process of a *c*-ring starting from individual protomers. Presented modeling process and biophysical considerations are likely generalizable to assembly of other membrane protein complexes with high-order rotational symmetry.

Submitting to:

Integrative Computational Biology workshop

Primary authors: Mr KUZMIN, Alexander; GUSHCHIN, Ivan (Moscow Institute of Physics and Technology)

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

Session Classification: Computational Methods and Deep Learning

Contribution ID: **25** Type: **Talk**

High-throughput prediction of ATP synthase rotor ring stoichiometries

Friday, 16 February 2024 09:30 (30 minutes)

Rotary ATP synthases are large enzyme complexes present in every living cell. They consist of a transmembrane and a soluble domain, each comprising multiple subunits. The transmembrane part contains an oligomeric rotor ring (*c*-ring), whose stoichiometry defines the ratio between the number of synthesized ATP molecules and the number of ions transported through the membrane. Here, we present an easy-to-use high-throughput computational approach based on AlphaFold that allows us to estimate the stoichiometry of all homooligomeric *c*-rings, whose sequences are present in genomic databases. We validate the approach on the available experimental data, obtaining the correlation of 0.96 for the reference set of *c*-rings with stoichiometry from 8 to 15, and use it to predict the existence of *c*-rings with stoichiometry varying from 8 to 27. We then conduct molecular dynamics simulations of selected *c*-rings to corroborate the machine learning-based predictions. Our work highlights the usability of AlphaFold-based approaches for modeling homooligomeric proteins.

Submitting to:

8th CAPRI assessment meeting

Primary authors: OSIPOV, Stepan D.; ZINOVEV, Egor V.; ANUCHINA, Arina A.; KUZMIN, Alexander S.; MINAEVA, Andronika V.; RYZHYKAU, Yury L.; VLASOV, Alexey V.; GUSHCHIN, Ivan (Moscow Institute of Physics and Technology)

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

Session Classification: CAPRI

Contribution ID: **26**   Type: **Talk**

An Iteratively Derived Knowledge-based Scoring Function at Atomic Level for Protein-DNA Complexes Evaluations

Thursday, 15 February 2024 14:30 (30 minutes)

Protein-DNA interactions play a significant role in biological processes and drug design owing to their prevalence. Computational methods for predicting protein-DNA complex structures serve as a valuable alternative to experimental methods, which, although more accurate, are also time-consuming and resource-intensive. The established framework for predicting protein-protein complex structures can be adapted for protein-DNA complexes, typically involving a Fast Fourier Transform (FFT)-based rigid docking, followed by a scoring function to re-rank the modeled structures. Despite the efficiency and success of this framework, its success rate is influenced by conformational changes induced during the binding process—a common phenomenon in protein-DNA interactions. To address this challenge, we have developed an iterative method, ITScorePD, for training a knowledge-based scoring function on an augmented set. This set includes experimentally resolved crystal structures, and reasonable decoy structures and enriched near-native structures generated through the rotation-translation blocked (RTB) method to account for conformational changes in both proteins and DNAs. Our results indicate that including near-native structures in the training set significantly improves the performance of ITScorePD compared to the scoring function derived from a training set without near-native structures. The detailed test results will be presented.

Submitting to:

8th CAPRI assessment meeting

Primary authors: Dr QIU, Liming (University of Missouri - Columbia); Dr DUAN, Rui (University of Missouri - Columbia); ZOU, XIaoqin (University of Missouri - Columbia)

Presenter: ZOU, XIaoqin (University of Missouri - Columbia)

Session Classification: CAPRI

Contribution ID: **27**   Type: **Talk**

Spectral partitioning into protein structural domains

Tuesday, 13 February 2024 09:00 (25 minutes)

The decomposition of a biomolecular complex into domains is an important step to investigate biological functions, and is also relevant to ease structure determination. A successful approach to do so is the SPECTRUS algorithm, which provides a segmentation based on spectral clustering applied to a graph coding inter-atomic fluctuations derived from an elastic network model. We present a simplification and an extension of SPECTRUS, both straightforward and useful.

For single structures, we show that high quality partitioning can be obtained from a graph Laplacian derived from pairwise interactions, without the use of normal modes. For sets of homologous structures, we introduce a Multiple Sequence Alignment mode, exploiting both the sequence based information (MSA) and the geometric information embodied in experimental structures.

The algorithm compares favorably with the original SPECTRUS as well as state-of-the-art deep approaches such as Chainsaw.

Submitting to:

Integrative Computational Biology workshop

Primary authors: SARTI, Edoardo (Inria); CAZALS, Frederic (Inria); HERRMANN, Jules (CRI Paris)

Presenter: SARTI, Edoardo (Inria)

Session Classification: Computational Methods and Deep Learning

Contribution ID: **28** Type: **Talk**

Analysis of protein-nucleic acid interactions in the PPI3D web server

Monday, 12 February 2024 14:05 (25 minutes)

To perform their functions, proteins frequently interact with other proteins and nucleic acids. Detailed information on these interactions can be obtained from the three-dimensional structures of the corresponding protein-protein or protein-nucleic acid complexes. Since the experimental structure determination is often tedious and expensive, computational structure prediction methods are widely applied. Currently, the structures of protein-protein complexes can be modeled accurately by AlphaFold, but structures protein-nucleic acid complexes can be reliably inferred only based on homology. To facilitate the search and analysis of structural data on protein interactions based on sequence homology, we have developed the PPI3D web server. Here, we present its novel features related to the analysis of protein-nucleic acid interactions. Given the sequences of proteins, PPI3D can now identify not only protein-protein, but also protein-nucleic acid interactions available for homologous proteins in the Protein Data Bank. These structures in the PPI3D database are clustered according to similarity of both protein sequences and interaction interfaces. The identified protein-nucleic acid interfaces can be analyzed in detail at the sequence and structure levels. In addition, homology models of the interactions with nucleic acids can be generated for the query proteins. PPI3D web server is available at https://bioinformatics.lt/ppi3d.

Submitting to:

Integrative Computational Biology workshop

Primary author: DAPKUNAS, Justas (Vilnius University)

Co-authors: Dr TIMINSKAS, Albertas (Vilnius University); OLECHNOVIC, Kliment (Vilnius University); Dr TOMKUVIENE, Migle (Vilnius University); Dr VENCLOVAS, Ceslovas (Vilnius University)

Presenter: DAPKUNAS, Justas (Vilnius University)

Session Classification: Computational Methods and Deep Learning

Contribution ID: **29**   Type: **Talk**

FTDMP: a framework for protein-protein, protein-DNA and protein-RNA docking and scoring

Thursday, 15 February 2024 15:00 (30 minutes)

Understanding the functions of protein-protein and protein-nucleic acid complexes relies on the knowledge of their 3D structures that can either be solved experimentally or predicted computationally. While AlphaFold has revolutionized protein structure prediction, challenges remain, particularly in modeling antibody-antigen interactions, protein-nucleic acid complexes, and proteins lacking close homologs. In these cases, docking can be employed to generate structure models. Consequently, effective methods for selection of the most accurate models are necessary.

Here we present FTDMP, a newly developed framework for protein-protein and protein-nucleic acid docking and scoring. The framework can be used in two ways: to perform docking and subsequent scoring, or to score and rank user provided models coming from different sources (AlphaFold, RoseTTAFold, docking, etc.). The ranking is done by a newly developed method, VoroIF-jury, that is based on the consensus of several scoring functions [1]. VoroIF-jury-based protocol obtained top results in the CASP15-CAPRI scoring experiment [2].

The full FTDMP docking and scoring framework was subsequently tested on protein-protein, protein-DNA, and protein-RNA docking benchmarks [3-5]. All the structures were downloaded directly from the PDB and renumbered, correcting several inconsistencies in the benchmarks' datasets. Compared to currently available docking systems, FTDMP demonstrated improved results of the free unbound-unbound docking when the top-ranked model was considered. Moreover, the success rates were very high for bound-bound docking (up to 83% for the top prediction), which opens new application possibilities when the conformational changes upon binding are negligible. For example, rigid-body docking using FTDMP assisted the identification of high-quality models when AlphaFold and docking predictions were similar for hard CAPRI targets.

FTDMP can be used not only with the built-in, but also with external scoring methods. Thus, the framework can be employed for fast and straightforward evaluation of new scoring functions.

FTDMP, docking benchmarks and docking results are available at https://github.com/kliment-olechnovic/ftdmp.

[1] Olechnovič, K., et al. (2023). Prediction of protein assemblies by structure sampling followed by interface-focused scoring. Proteins, 91(12), 1724-1733.

[2] Lensink, M.F., et al. (2023). Impact of AlphaFold on structure prediction of protein complexes: The CASP15-CAPRI experiment. Proteins, 91(12), 1658–1683.

[3] van Dijk, M., Bonvin, A.M. (2008). A protein-DNA docking benchmark. Nucleic Acids Res, 36, e88.

[4] Guest, J. D., et al. (2021). An expanded benchmark for antibody-antigen docking and affinity prediction reveals insights into antibody recognition determinants. Structure, 29(6), 606–621.e5.

[5] Zheng, J., et al. (2020). P3DOCK: a protein-RNA docking webserver based on template-based and template-free docking. Bioinformatics, 36(1), 96–103.

Submitting to:

8th CAPRI assessment meeting

Primary authors: BANCIUL, Rita (Vilnius University); OLECHNOVIC, Kliment (Vilnius University); DAPKUNAS, Justas (Vilnius University); Dr VENCLOVAS, Ceslovas (Vilnius University)

Presenters: BANCIUL, Rita (Vilnius University); DAPKUNAS, Justas (Vilnius University)

Session Classification: CAPRI

Contribution ID: **30**   Type: **Talk**

Integrative Biology to tackle mitochondrial respiration in Alzheimer's pathogenesis

Wednesday, 14 February 2024 09:20 (25 minutes)

Respiratory complexes located in the internal membranes of our mitochondria are true macro-molecular batteries: they couple the flow of electrons through clusters of metals and cofactors with a transfer of protons to create a gradient that provides the energy necessary for the ATP production and therefore to the nourishment of essential life processes. The first complex in the respiratory chain, named Complex I (CI), is one of the largest membrane proteins, made up of 45 subunits. The processes of its assembly and its sophisticated regulation are still poorly understood, although it is known that their disruption leads to neurodegenerative diseases such as Alzheimer's.

While exploring the molecular basis for protein recognition in the Mitochondrial CI Assembly (MCIA) complex, using a combination of biochemical, biophysical and structural techniques, we discovered that the assembly of the MCIA complex juggles between two incompatible activities: fatty acid oxidation and CI assembly. Cryo-EM and crystal structures of the partners in complex and alone allowed us further understanding into how they switch from one function to another. Furthermore, our recent mitochondrial analyses in amyloidogenic cells provide insights into the relationship of CI assembly in neurological dysfunction, unveiling whether MCIA components could detect early AD pre-symptomatic stages.

References
Giachin et al. Angew. Chem. Int. Ed. 2021, 60(9):4689.
McGregor & Soler-Lopez. Curr Opin Struct Biol. 2023, 80:102573.
McGregor et al. Nat Commun. 2023, 14(1):8248.

Submitting to:

Integrative Computational Biology workshop

Primary author: SOLER LOPEZ, Montserrat (ESRF)

Presenter: SOLER LOPEZ, Montserrat (ESRF)

Session Classification: Protein conformational flexibility and solution experiments

Contribution ID: **31**   Type: **Talk**

HADDOCK in latest CAPRI rounds

Thursday, 15 February 2024 10:00 (30 minutes)

The HADDOCK team have participated as human and webserver predictor and/or scorer in last CAPRI rounds (47-55), mainly relying on the use of our integrative modelling software HAD-DOCK[1] which is able to use user-provided information to guide the docking process. An important element of HADDOCK is its scoring function, computed from the weighted sum of 4 energetical terms (VdW, Electrostatics, Desolvation and restraints energies), which, despite its simplicity, is still performing competitively.

Besides HADDOCK, a variety of other tools were used related to all kind of predictions regarding biomolecular complexes, ranging from the prediction of interacting residues (ProABC2[2], ARCTIC-3D[3], CPORT[4]) to clustering methods (FCC[5]) to the molecular dynamics-based scoring of docking models[6]. With this set of tools and methods, we participated to both predictions (except the CASP-CAPRI round) and scoring rounds as both human teams and web server.

For rounds 47 and 49, we mainly relied on HADDOCK2.4, available for the community as web service at https://wenmr.science.uu.nl/haddock2.4/, for both prediction and scoring. More recently, for manual predictions and scoring, a new, modular version of the software, HADDOCK3 (https://github.com/haddocking/ha has been used. This new version enables easy setup of different docking and/or scoring workflows. Using it allowed us to predict at least acceptable models for 50% of the targets (at round 55) and pick more than 50% of medium quality complexes (in rounds 54 and 55) at scoring stages.

At the CAPRI assessment meeting, we will present the various approaches and methodologies used for different rounds, including the data introduced in our protocols to guide the docking.

References

[1] Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: A Protein–Protein Docking Approach Based on Biochemical or Biophysical Information. J Am Chem Soc. 2003

[2] Ambrosetti F, Olsen TH, Olimpieri PP, et al. proABC-2: PRediction of AntiBody contacts v2 and its application to information-driven docking. Bioinformatics. 2020

[3] Giulini M, Honorato RV, Rivera JL, Bonvin AMJJ. ARCTIC-3D: automatic retrieval and clustering of interfaces in complexes from 3D structural information. Commun Biol. 2024

[4] de Vries SJ, Bonvin AM. CPORT: a consensus interface predictor and its performance in prediction-driven docking with HADDOCK. PLoS One. 2011

[5] Rodrigues JP, Trellet M, Schmitz C, et al. Clustering biomolecular complexes by residue contacts similarity. Proteins. 2012

[6] Jandova Z, Vargiu AV, Bonvin AMJJ. Native or Non-Native Protein-Protein Docking Models? Molecular Dynamics to the Rescue. J Chem Theory Comput. 2021

Submitting to:

8th CAPRI assessment meeting

Primary authors: Dr GIULINI, Marco (Utrecht University);   Dr REYS, Victor (Utrecht University);   XU, Xiaotong (Utrecht University);   Dr COJOCARU, Vlad (Utrecht University);   ENGEL, Anna (Utrecht University);   Dr HONORATO, Rodrigo (Utrecht University);   VAN NOORT, Charlotte (Utrecht

University); Dr TEIXEIRA, João M.C. (Utrecht University); Prof. BONVIN, A.M.J.J. (Utrecht University)

Presenters: Dr GIULINI, Marco (Utrecht University); Dr REYS, Victor (Utrecht University)

Session Classification: CAPRI

Contribution ID: **32** Type: **Talk**

POU5F1B and GOLPH3, two challenging proteins in structural biology

Tuesday, 13 February 2024 16:45 (25 minutes)

During this seminar, two experimental structural biology projects will be discussed illustrating current challenges and limitations in the field. The first will show how sequence identity may fail in predicting the protein structure and function. The second deals with investigating protein-protein interactions experimentally with the aim of creating small molecule inhibitors that block such interaction. Our results from biophysical experiments (CD and ITC) as well as NMR will be presented and discussed.

Submitting to:

Integrative Computational Biology workshop

Primary authors: Ms THEODOROPOULOU, Anastasia (EPFL); Prof. TRONO, Didier (EPFL); Prof. D'ANGELO, Giovanni (EPFL); Dr SIMO, Laia (EPFL); Dr ABRIATA, Luciano (EPFL); Dr MARCAIDA, Maria (EPFL); Prof. DAL PERARO, Matteo (EPFL)

Presenter: Dr MARCAIDA, Maria (EPFL)

Session Classification: Protein conformational flexibility and solution experiments

Contribution ID: **34**　　　　　　　　　　　　　　　　　　　　　　　　Type: **Talk**

Modeling of immune recognition and protein complexes with deep learning and physics-based docking

Friday, 16 February 2024 11:00 (30 minutes)

The Pierce group has utilized a combination of AlphaFold and traditional docking methods, including Rosetta and ZDOCK, to model protein complexes in recent CAPRI and CASP/CAPRI rounds. This has led to success for several challenging targets, as well as useful lessons learned for prospective modeling efforts. Our group has been particularly focused on the utilization and adaptation of AlphaFold and other deep learning methods for accurate modeling of immune recognition, including antibodies and T cell receptors. We recently reported a comprehensive benchmarking of AlphaFold for its performance on a large set of antibody-protein antigen complexes, and have also been testing its performance for antibody-peptide complexes. Additionally, we adapted AlphaFold to predict T cell receptor complexes with peptide-MHC targets, which is available to the public as the TCRmodel2 web server.

Submitting to:

8th CAPRI assessment meeting

Primary authors:　　PIERCE, Brian (University of Maryland);　YIN, Rui (University of Maryland)

Co-authors:　LEE, Jessica (University of Maryland); GUEST, Johnathan (University of Maryland); CHEUNG, Melyssa (University of Maryland);　PARK, Minjae (University of Maryland)

Presenter:　PIERCE, Brian (University of Maryland)

Session Classification:　CAPRI

Contribution ID: **35**   Type: **Talk**

Prediction and refinement of protein assemblies with ClusPro, Alphafold and Molecular Dynamics.

Wednesday, 14 February 2024 16:45 (30 minutes)

In the latest CAPRI round our group used a combination of Alphafold-Multimer (AFM), the Clus-Pro webserver for docking, and Molecular Dynamics based sampling to refine models for small targets. Assembly prediction was based on a two-stage methodology, in which we first generate an ensemble of initial models using AlphaFold-Multimer using standard protocol for MSA generation. We stop the search if the model is of sufficiently high confidence, otherwise we perform template based search using ClusProTBM and free docking using ClusPro. The resulting structures are transferred to Alphafold-Multimer (AFM) as starting templates for generating "refined" structures of the target complex. For smaller targets such as T231 we have explored an additional refinement step based on constrained MD based sampling, where we constrained the high confidence protein and peptide regions and performed additional sampling in the other regions, followed by template based refinement.

Submitting to:

8th CAPRI assessment meeting

Primary author: KOZAKOV, Dima (Stony Brook University)

Co-authors: SIMMERLING, Carlos; STEPANENKO, Darya; BEGLOV, Dmitri; PADHORNY, Dzmitry; KHAN, Omeir; ASHIZAWA, Ryota; VAJDA, Sandor; GHANI, Usman

Presenter: KOZAKOV, Dima (Stony Brook University)

Session Classification: CAPRI

Contribution ID: **36**   Type: **Talk**

Towards integrative use of pyDock and AI-based modeling in 8th CAPRI

Wednesday, 14 February 2024 16:15 (30 minutes)

We have participated, both as predictors and as scorers, in all proposed targets of the 8th CAPRI edition. Here we will focus on the 11 purely CAPRI targets, that is, excluding the CASP-CAPRI joint rounds 46, 50 and 54, already evaluated and discussed in the respective CASP meetings, and the 4 targets in round 51 related to COVID19, with no available structure for evaluation. These 11 targets involved domain-domain, protein-protein and protein-DNA interactions, with homo- and hetero-meric interfaces.

As predictors, in general, we applied our standard pyDock protocol, based on the generation of rigid-body docking orientations with FTDock 2.0 and ZDock 2.1, and the selection of models by py-Dock 3.0 scoring function. Oligomerization states and other specific issues were carefully treated in a case-by-case basis, applying symmetry restraints in case of homo-dimeric interfaces, and combining docking with homology-based interfaces if templates were available. The multi-domain protein in T160 was built by combining homology modelling with Modeller and domain-domain docking with pyDockTET [1]. In T161 we introduced flexibility in the inter-domain orientations using Normal Mode Analysis with ProDy. Protein-DNA interfaces in T187 and T188 were built by our recent pyDockDNA method [2]. Starting in round 53, interacting subunits were routinely modelled by AlphaFold2. In the last round (targets T231-T234) we combined docking models with those generated by AlphaFold-Multimer (ColabFold versions 1-3), selected on the basis of AF model confidence (0.8 ipTM + 0.2 pTM) [3] and pyDock scores.

We also participated as predictors with our server pyDockWeb [4], using the automatic FTDock and pyDock pipeline with symmetry-based restraints when needed.

As scorers, we mostly used pyDock 3.0 to directly score the provided models, without further minimization, applying the same restraints and additional filters as in predictors.

According to the results so far, our top 5 submitted models were successful in 6 out of the 10 evaluated targets either as predictors or as scorers: T160, T163, T187, T188, T231, and T232. In addition, we got an acceptable model in T161, but had too many clashes. Especially interesting were the results for the protein-DNA interface in T188, where we were the only group with acceptable models as predictors (top 1) and as scorers (top 5). On the other side, the targets where we failed both as predictors and as scorers were actually challenging for all participants (e.g. no successful group in T187 protein-DNA interface or in T234; only one successful group as predictors in T161 and T162).

Overall, the performance of pyDock in this 8th CAPRI edition was in line with that of the best participants. This experiment has shown that the problem of protein-protein docking is far from being solved, and has confirmed the value of energy-based scoring and other approaches in combination with AlphaFold predictions.

[1] Cheng TM, Blundell TL, Fernandez-Recio J (2008) Structural assembly of two-domain proteins by rigid-body docking. BMC Bioinformatics 9:441.

[2] Rodríguez-Lumbreras LA, Jiménez-García B, Giménez-Santamarina S, Fernández-Recio J (2022) pyDockDNA: A new web server for energy-based protein-DNA docking and scoring. Front. Mol. Biosci. 9:988996.

[3] Evans R, O'Neill M, Pritzel A, et al. (2021) Protein complex prediction with AlphaFold-Multimer. BioRxiv.

[4] Jiménez-García B, Pons C, Fernández-Recio J (2013) pyDockWEB: A web server for rigid-body protein-protein docking using electrostatics and desolvation scoring. Bioinformatics, 29:1698-1699.

Submitting to:

8th CAPRI assessment meeting

Primary author: FERNANDEZ-RECIO, Juan (ICVV-CSIC and BSC)

Co-authors: RODRIGUEZ-LUMBRERAS, Luis Ángel (ICVV-CSIC and BSC); ROSELL, Mireia (ICVV-CSIC and BSC); ROMERO-DURANA, Miguel (ICVV-CSIC and BSC); MONTEAGUDO, Víctor (ICVV-CSIC); GLASER, Fabian (Technion, Haifa)

Presenter: FERNANDEZ-RECIO, Juan (ICVV-CSIC and BSC)

Session Classification: CAPRI

Contribution ID: **37**   Type: **Talk**

Integrative structural biology of cell extracts : an application to structurally undercharacterized organisms

Tuesday, 13 February 2024 17:40 (25 minutes)

Integrative structural biology of cell extracts bridges the gap between the high resolution structural characterisation of highly purified, isolated biomolecules and in situ electron tomography. This booming technique combines mass spectrometry (MS)-based proteomics and cryo-electron microscopy (cryo-EM) of fractionated cell extracts to quantitatively and structurally characterise endogenous proteins and complexes. We chose to employ this method on Physarum Polycephalum, which is absent from protein sequence and structure databases (Uniprot, PDB, EMDB) in order to mimic the challenges posed e.g. by emerging pathogens. We will show how structural biology can be applied when only raw genomic data is available, and discuss the difficulties faced to identify and annotate protein and complexes structures when no prior information is available.

Submitting to:

Integrative Computational Biology workshop

Primary author: DESFOSSES, Ambroise (CNRS- IBS)

Presenter: DESFOSSES, Ambroise (CNRS- IBS)

Session Classification: Protein conformational flexibility and solution experiments

Contribution ID: **38**   Type: **Talk**

Integrative structural biology with atomic force microscopy data

Wednesday, 14 February 2024 11:05 (25 minutes)

Understanding the molecular and cellular function of biological molecules requires an accurate perception of their functional assemblies. The term "functional" refers to the actual structure of the bioactive molecule: it includes, but is not limited to, oligomerization, molecular partners, and their multiscale dynamics. Achieving such a functional goal requires a combination of techniques that provide details from atomic to macromolecular assembly resolution. Atomic force microscopy (AFM) provides information at nanoscale resolution (10-20 angstroms) with an exceptional signal-to-noise ratio. The output of AFM is, in the best cases, an isolated single molecular topography with nanometer resolution. We have developed tools to use AFM topographic data to assemble macromolecular systems from their constituent units. In addition, the reconstruction of flexible macromolecular systems under the experimental constraints of AFM topography allows us to study molecular dynamics at the structural unit level. Tools and applications will be presented.

Submitting to:

Integrative Computational Biology workshop

Primary author: PELLEQUER, Jean-Luc (IBS)

Presenter: PELLEQUER, Jean-Luc (IBS)

Session Classification: Protein conformational flexibility and solution experiments

Contribution ID: **39**   Type: **Talk**

Integrative ensemble modeling that utilizes distance distribution restraints

Tuesday, 13 February 2024 13:45 (25 minutes)

AlphaFold2 predictions for the human proteome reveal that the great majority of all human proteins contain intrinsically disordered regions (IDRs) or do not cotnain a folded domain at all. The structure of such proteins must be represented by an ensemble of conformers. However, most experimental techniques provide ensemble-average restraints, which do not contain direct information on width of the ensemble. Even single-molecule techniques, such as Förster Resonance Energy Transfer (smFRET) average over many conformers that are visited ny molecular dynamics at time scale of the measurement. Using restraints from such techniques one cannot evaluate the probability for an individual conformer to belong to the ensemble.

In contrast, distance distributions between two spin-labelled sites in a protein, as they can be obtained by pulse electron paramagnetic resonance experiments, such as DEER, directly inform on ensemble width and can be used to compute such probabilities. Therefore, distance distribution restraints can already be used at the stage of sampling conformer space and protect against unrealistic narrowing of the ensemble during ensemble reweighting. To make use of these features in ensemble modelling, we have developed the toolbox MMMx.

MMMx implements the RigiFlex approach that can model flexible multi-domain proteins by first computing a distributed rigid-body arrangement of the foldced domains in the Rigi step. The IDRs that link the folded domains or are terminal are then constructed in the Flex step. The raw ensemble generated in this way is then reweighted by the EnsembleFit step, which can additionally take into account small-angle scattering curves and NMR paramagnetic relaxation enhancements (PREs). Use of smFRET mean-distance restraints will be implemented soon.

We illustrate MMMx ensemble modelling on the examples of (1) the intrinsically disordered N-terminal domain of FUS in the dispersed and condensed state formed by liquid-liquid phase separation, (2) full-length hnRNP A1 that consist of a folded domain and a long C-terminal IDR, and (3) SRSF1$_{\triangle RS}$ in its free form and in complexes with two short single-stranded RNAs. More information on MMMx can be found in the online documentation.

Submitting to:

Integrative Computational Biology workshop

Primary author: JESCHKE, Gunnar (ETH Zürich)

Presenter: JESCHKE, Gunnar (ETH Zürich)

Session Classification: Protein conformational flexibility and solution experiments

Contribution ID: **40**   Type: **Talk**

Generative models and analysis tools for the study of highly-flexible (intrinsically disordered) proteins

Tuesday, 13 February 2024 08:10 (25 minutes)

Proteins can have very different architectures, generally involving a concatenation of relatively rigid domains and flexible regions. Indeed, many proteins in eukaryotes, prokaryotes and viruses are composed of several domains connected by linkers, and flexible tails are also frequently found at the termini of rigid domains. Besides, flexible loops connecting secondary structure elements within domains are omnipresent in proteins. All these types of flexible regions (linkers, tails and loops), in addition to fully flexible, intrinsically disordered proteins, play key functional roles, usually related to inter- or intramolecular interactions.

While the structure of rigid domains can be accurately determined using experimental methods or predictors such as AlphaFold2, the structural study of flexible regions remains a challenge. It requires computational methods for the generation of conformational ensemble models that are fitted or refined on the basis of experimental measurements. In recent years, we have developed several algorithms, based on fragment databases and robotics-inspired techniques, for the conformational sampling of flexible loops and intrinsically disordered regions. Building on this work, we propose a unified approach to sample conformations of proteins with complex architectures composed of rigid and flexible regions. Our approach integrates a multi-agent reinforcement learning technique to improve sampling performance while taking into account the specificities of each flexible/disoriented region of the protein. In addition to these generative models, we have developed statistical tools for the analysis and comparison of conformational ensembles of highly-flexible proteins.

Submitting to:

Integrative Computational Biology workshop

Primary author: CORTÉS, Juan (LAAS-CNRS)

Presenter: CORTÉS, Juan (LAAS-CNRS)

Session Classification: Computational Methods and Deep Learning

Contribution ID: **41**   Type: **Talk**

Structure and function relationships in protein homorepeats

Wednesday, 14 February 2024 10:40 (25 minutes)

Homorepeats, repetitive sequences inserted in disordered proteins, play fundamental roles in biology and mutations in these low-complexity regions are linked to several neurodegenerative and developmental diseases. Despite their relevance, the structural characterization and modelling of these proteins remain challenging. From an experimental perspective, the severe signal overlap of their NMR spectra hampers the site-specific assignment, precluding the high-resolution structural characterization. As a consequence of this lack of structural data, the quality of computational models is difficult to assess.

Our group has developed novel chemical biology tools enabling the site-specific incorporation natural and non-natural amino acids into LCRs to derive residue-specific structural and dynamic information. These tools have been used to study the structural properties of poly-Glutamine, poly-Proline and poly-Alanine segments present in disease-causing proteins. The methodological developments, the structural information obtained from these tailored labeling strategies and the computational integration of these data will be discussed along the presentation.

Submitting to:

Integrative Computational Biology workshop

Primary author: BERNADO, Pau (Centre de Biologie Structural (CBS-Montpellier))

Presenter: BERNADO, Pau (Centre de Biologie Structural (CBS-Montpellier))

Session Classification: Protein conformational flexibility and solution experiments

Contribution ID: **42**   Type: **Talk**

Exploring the conformational space of proteins by enumeration in the frame of the Distance Geometry Problem

Tuesday, 13 February 2024 10:20 (25 minutes)

The continuous development of the methods for the protein structure prediction was taking advantage from the precious experimental information obtained by structural biology as well as by sequencing of multiple organisms. Indeed, the general developed pipeline is based on determining conformations of protein fragments, and then using multiple sequence alignments to obtain long-range distances allowing to get predicted protein conformations. The introduction of the deep learning techniques has recently permitted a important jump in the results obtained in this framework, as illustrated by the success of AlphaFold2, RoseTTAFold and ESMFold. Nevertheless,in some cases, long-range restraints cannot be obtained, as for intrinsically disordered proteins or regions (IDP/IDR), or in the case of orphan proteins for which not enough statistical signal can be obtained by sequence alignment.

Here, we propose to investigate another point of view, in which local information would be mainly used to determine the protein folding, this local conformational information being extracted directly from the primary sequence. Two major obstacles arise from: (i) the variability in protein stereochemistry, inducing a drift of the protein backbone when the residues are successively added; (ii) the size of the protein conformational space described by local conformations, as it would not be reduced by the long-range proximity information. In the framework of the Distance Geometry Problem (DGP), the Branch-and-Prune (BP) approach, based on a graph description of proteins, brings an answer to the problem of the size of the conformational space by performing a systematic enumeration of all protein conformations satisfying a given set of geometric constraints. Applications of the BP approach will thus be presented for the reconstruction of folded structures of proteins, with propositions for dealing with the variability of stereochemistry. The efficiency of the BP approach will be also demonstrated on the cases of disordered or flexible proteins, in particular on the Small EDRK-rich factor 1.

Submitting to:

Integrative Computational Biology workshop

Primary authors: Dr HUANG, Shu-Yu (Academia Sinica); Dr CHANG, Chi-Fon (Academia Sinica); Dr SHIH, Orion (National Synchrotron Radiation Research Center); Dr LIN, Jung-Hsin (Academia Sinica); MALLIAVIN, Therese (Laboratoire de Physique et Chimie Theoriques)

Presenter: MALLIAVIN, Therese (Laboratoire de Physique et Chimie Theoriques)

Session Classification: Computational Methods and Deep Learning

Contribution ID: **43**   Type: **Poster**

Integrative structural biology of cell extracts : a Blob case study

Integrative structural biology of cell extracts bridges the gap between the high resolution structural characterisation of highly purified, isolated biomolecules and in situ electron tomography. This booming technique combines mass spectrometry (MS)-based proteomics and cryo-electron microscopy (cryo-EM) of fractionated cell extracts to quantitatively and structurally characterise endogenous proteins and complexes. We chose to employ this method on Physarum Polycephalum, which is absent from protein sequence and structure databases (Uniprot, PDB, EMDB) in order to mimic the challenges posed e.g. by emerging pathogens. We will show how structural biology can be applied when only raw genomic data is available, and discuss the difficulties faced to identify and annotate proteins and complexes structures when no prior information is available.

Submitting to:

Integrative Computational Biology workshop

Primary authors: DESFOSSES, Ambroise (CNRS-IBS); PAGEOT, Eymeline (UGA-IBS)

Co-authors: GUTSCHE, Irina (CNRS-IBS); GAIFAS, Lorenzo (UGA-IBS); BACIA-VERLOOP, Maria (CNRS-IBS); ADAM, Virgile (CNRS-IBS)

Presenter: DESFOSSES, Ambroise (CNRS-IBS)

Contribution ID: **44**   Type: **Talk**

Dissecting peripheral protein-membrane interfaces

Tuesday, 13 February 2024 08:35 (25 minutes)

Peripheral membrane proteins (PMPs) are soluble proteins that bind transiently to the surface of cell membranes. Having the ability to exist in both a soluble and a membrane-bound form their membrane-binding region is constrained to retain a fine balance of polar and hydrophobic character, which makes it difficult to distinguish it from the rest of their surface. As a result peripheral membrane-binding sites are notoriously difficult to predict.

We collected and curated a dataset containing 2500 structures and compared their membrane-binding sites to the rest of their solvent-accessible surfaces, in order to reveal features of PMPs' membrane-binding sites. To this goal we took advantage of the hydrophobic protrusion model we reported earlier (Fuglebakk and Reuter, PLoS Comp Biol, 2018) but also extended it to consider amino acids neighbouring protrusions. We find that, among positively charged amino acids, lysines are significantly more present than arginines. Protruding hydrophobes are a landmark of the interfacial binding sites of about 2/3 of PMPs, indicating that a majority of PMPs takes advantage of the hydrophobic effect while a non-negligeable minority (1/3) most likely relies on electrostatics interactions or other mechanisms. The IBS of peripheral membrane proteins contain significantly more glycines than the rest of their surface. Furthermore the analysis of 9 superfamilies revealed amino acid distribution patterns in agreement with their known functions and membrane-binding mechanisms. These findings and the collected dataset shed light on properties of protein-membrane interfaces and will be useful for the development of prediction models for membrane-binding sites of PMPs.

Submitting to:

Integrative Computational Biology workshop

Primary author: Prof. REUTER, Nathalie (University of Bergen)

Co-authors: Dr FUGLEBAKK, Edvin; Dr TUBIANA, Thibault

Presenter: Prof. REUTER, Nathalie (University of Bergen)

Session Classification: Computational Methods and Deep Learning

Contribution ID: **45**   Type: **Talk**

Integrative modeling meets deep learning: applications to crosslinking mass spectrometry

Monday, 12 February 2024 13:40 (25 minutes)

Recent progress in protein structure prediction significantly enhanced integrative structure modeling of large macromolecular assemblies by providing improved structural coverage for individual proteins and protein-protein interactions. However, direct prediction of large protein assemblies remains a challenge. I will introduce CombFold, a hierarchical and combinatorial assembly algorithm designed to predict structures of large protein complexes utilizing pairwise interactions between subunits predicted by AlphaFold2. Distance restraints based on cross-linking mass spectrometry can be directly integrated into the assembly algorithm, further improving the accuracy. To optimize the integration of information from cross-linking experiments, we propose a deep-learning model to predict the optimal distance range for crosslinked residue pairs based on their structural neighborhoods. Despite the availability of high-throughput approaches for identifying protein-protein interactions and crosslinks in whole-cell crosslinking experiments, applying integrative modeling to these datasets poses challenges. One major obstacle is the need for methods to decipher temporal and spatial composition and stoichiometry of the cellular assemblies.

Submitting to:

Integrative Computational Biology workshop

Primary author: Dr SCHNEIDMAN, Dina (The Hebrew University of Jerusalem)

Presenter: Dr SCHNEIDMAN, Dina (The Hebrew University of Jerusalem)

Session Classification: Computational Methods and Deep Learning

Contribution ID: **46**   Type: **Talk**

Generating backbone conformational changes with seven league boots

Monday, 12 February 2024 15:50 (25 minutes)

Generating large amplitude conformational changes of complex biomolecules remains a challenge. This talk will review recent work on this problem, based on novel insights on loop closure techniques coupling kinematic models in dihedral angle spaces and MCMC sampling techniques of the Hit-and-Run type. Along the way, I will discuss connections with density of states calculations and thermodynamics.

Refs available from http://www-sop.inria.fr/teams/abs/publications/frederic-cazals.html

Enhanced conformational exploration of protein loops using a global parameterization of the backbone geometry
T. O'Donnell, and F. Cazals
J. Comp. Chem., 2023

Geometric constraints within tripeptides and the existence of tripeptide reconstructions
T. O'Donnell, and V. Agashe, and F. Cazals
J. Comp. Chem., 2023

Efficient computation of the the volume of a polytope in high-dimensions using Piecewise Deterministic Markov Processes
A. Chevallier, and F. Cazals, and P. Fearnhead
AISTATS, 2022

Wang-Landau algorithm: an adapted random walk to boost convergence
A. Chevallier, and F. Cazals
J. of Computational Physics, 410 (1), 2020

Submitting to:

Integrative Computational Biology workshop

Primary author: CAZALS, Frederic (Inria)

Presenter: CAZALS, Frederic (Inria)

Session Classification: Computational Methods and Deep Learning

Contribution ID: 47   Type: **Talk**

Mariage of Cryo-EM, Solution Scattering, and Computational Science

Tuesday, 13 February 2024 15:55 (25 minutes)

Over the past five years, structural biology has undergone a significant progress. The Nobel Prize in Chemistry awarded in 2017 to the developers of cryo-electron microscopy (cryo-EM) marked a milestone, and the rapid progress and widespread adoption of this technique have demonstrated that the precise structural analysis of biomacromolecules is not limited to crystallography or solution nuclear magnetic resonance (NMR). Furthermore, the public release of AlphaFold2 in 2021 has enabled high-accuracy structural predictions using computational methods. Each method has its imperfections, and choosing the appropriate one for the purpose is essential. Moreover, the increase in options is giving rise to the potential for combining options—a concept we refer to as "Integrated Structural Biology," which is opening new avenues for exploring complex structures and phenomena that were previously inaccessible.

One of the targets of research in Integrated Structural Biology is the "dynamics" of biomacromolecules. To advance this study, we believe that incorporating a collaborative analysis of X-ray/neutron solution scattering and computational analysis is valuable. Solution scattering allows the observation of biomacromolecules in solution, providing information on unfixed, moving biomacromolecules. However, this information is the temporal and ensemble average of all biomacromolecules in the solution, making analysis challenging. Therefore, it is crucial to use molecular dynamics (MD) simulations to analyze the motion of a single molecule from the scattering data. Furthermore, using detailed structural data obtained from cryo-EM and X-ray crystallography as the initial structures in MD simulations can lead to the elucidation of detailed structural fluctuations and dynamics.

In this presentation, I will introduce two recent examples of our integrated analysis:

1. Integrated analysis of a huge protein complex (24-mer), where full-length analysis was impossible with cryo-EM alone. The analysis combined cryo-EM, X-ray Crystallography, X-ray/Neutron small-angle scattering, computational modeling, and MD simulations.

2. Integration of dynamics analysis using cryo-EM multi-imaging and structural distribution analysis using solution scattering and coarse-grained MD simulation, revealing detailed dynamics of biomacromolecules.

The presentation will demonstrate how Integrated Analysis contributes to gaining insights into dynamic fluctuations of biomolecules in detail.

Submitting to:

Integrative Computational Biology workshop

Primary author: SUGIYAMA, Masaaki (Kyoto University, Institute for Integrated Radiation and Nuclear Science)

Presenter: SUGIYAMA, Masaaki (Kyoto University, Institute for Integrated Radiation and Nuclear Science)

Session Classification: Protein conformational flexibility and solution experiments

Contribution ID: **48**   Type: **Talk**

Integrative spatiotemporal map of nucleocytoplasmic transport

Tuesday, 13 February 2024 10:45 (25 minutes)

The Nuclear Pore Complex (NPC) is one of the larger macromolecular complexes in eukaryotic cells. The NPC facilitates rapid and selective transport between the cytoplasm and the nucleus. Existing models of transport do not provide quantitative mechanistic explanations of how some key emergent properties such as the rapid transport rates of molecules as large as ribosomal units and viral capsids arise from the system components and their interactions. To address this question, we constructed an integrative coarse-grained Brownian dynamics model of transport through a single NPC, followed by coupling it with a kinetic model of Ran-dependent transport in an entire cell. The microscopic model parameters were fitted to reflect experimental data and theoretical information regarding the transport, without making any assumptions about its emergent properties. The resulting reductionist model is validated by reproducing several features of transport not used for its construction, such as the morphology of the central transporter, rates of passive and facilitated diffusion as a function of size and valency, in situ radial distributions of pre-ribosomal subunits, and active transport rates for viral capsids. The model suggests that the NPC functions essentially as a virtual gate whose flexible phenylalanine-glycine (FG) repeat proteins raise an entropy barrier to diffusion through the pore. Importantly, this core functionality is greatly enhanced by several key design features, including 'fuzzy' and transient interactions, multivalency, redundancy in the copy number of FG nucleoporins, exponential coupling of transport kinetics and thermodynamics rationalized by transition state theory, and coupling to the energy-reliant RanGTP concentration gradient. These design features result in the robust and resilient rate and selectivity of transport for a wide array of cargo ranging from a few kilodaltons to megadaltons in size. By dissecting these features, our model provides a quantitative starting point for rationally modulating the transport system and its artificial mimics.

Raveh et al., 2023, bioRxiv 2023.12.31.573409

Submitting to:

Integrative Computational Biology workshop

Primary author: RAVEH, Barak (Hebrew University of Jerusalem)

Presenter: RAVEH, Barak (Hebrew University of Jerusalem)

Session Classification: Computational Methods and Deep Learning

Contribution ID: **49**   Type: **Talk**

Interpretable Affinity Prediction and Generative Design for Protein-Ligand Interactions

Thursday, 15 February 2024 16:15 (30 minutes)

In this talk I will introduce our recent works on machine learning for protein-ligand interactions, including explainable prediction of protein-ligand binding affinity and structure-based *de novo* ligand design.

First, as deep learning methods for modeling protein-ligand interactions are increasingly improving their accuracy, their interpretability is often under-explored. We had previously developed DeepAffinity that, starting with protein sequences and chemical identities, predicts both protein-ligand affinities and underlying intermolecular contacts. DeepAffinity unified recurrent and convolutional neural networks, exploited both labeled and unlabeled data, and adopted attention mechanisms for interpreting affinity predictions.

Our recent advances cover two aspects. (1) As attention mechanisms alone are inadequate, we regularize attentions with predicted 3D structural contexts and supervise attentions with non-bonded intermolecular contacts, which leads to DeepAffinity+. We further design DeepRelations with a physics-inspired, intrinsically explainable architecture. (2) Thanks to the recent breakthroughs in protein structure prediction, we consider protein data as available in both modalities of 1D amino-acid sequences and predicted 2D contact maps; and we introduce cross-modality protein embedding schemes. Moreover, we pre-train the protein embeddings through self-supervision using unlabeled data. Our results indicate that attention supervision, cross-modality, and self-supervision further improve the accuracy, the interpretability, and the generalizability of affinity prediction especially for unseen proteins.

Second, for simultaneous *de novo* design and structure prediction, we formulate the problem as 3D graph generation conditioned on target protein structures; and we solve the problem through diffusion models, a recent class of generative AI methods. By embedding 3D ligand graphs' identity and geometry jointly into a latent space, we diffuse the 3D graph iteratively in the latent space and maintain roto-translational equivariance. Compared to other generative models, our latent diffusion model generates tight and diverse molecules while being an order of magnitude faster to train.

Submitting to:

8th CAPRI assessment meeting

Primary author: SHEN, Yang (Texas A&M University)

Presenter: SHEN, Yang (Texas A&M University)

Session Classification: CAPRI

Contribution ID: **50**   Type: **Talk**

Intrinsically disordered regulators of endocytosis - an integrated NMR/single molecule fluorescence approach

Wednesday, 14 February 2024 08:30 (25 minutes)

Intrinsically disordered proteins (IDPs) lack clearly defined structure and are therefore highly flexible and easily adaptable to different binding partners. This makes them important players in many biological processes, often with vital regulatory functions. Their dynamic features and broad range of interaction modes, however, render them difficult to study and analyzing their complexes often requires integrated approaches. Integrating complementary parameters from of nuclear magnetic resonance (NMR) and single molecule fluorescence approaches allowed us to describe the conformational landscape of IDPs at molecular resolution and promises to shed new light onto various biological processes.

Among those counts clathrin mediated endocytosis. The early phases of clathrin mediated endocytosis are organized through a highly complex interaction network mediated by clathrin associated sorting proteins (CLASPs) that comprise long intrinsically disordered regions (IDRs). We characterize the IDRs of those CLASPs in their entirety and at molecular resolution, uncovering a plethora of interactions of various strengths and dynamic features with their endocytic interaction partners, proposing a rationale for how first interactions and dynamics rearrangement of partners take place during the uptake of a coated vesicle.

Submitting to:

Integrative Computational Biology workshop

Primary author: MILLES, Sigrid (Forschungsverbund Berlin e.V. (FMP))

Presenter: MILLES, Sigrid (Forschungsverbund Berlin e.V. (FMP))

Session Classification: Protein conformational flexibility and solution experiments

Contribution ID: **51**   Type: **Talk**

What have we learned (and memorized) about peptide-mediated interactions in CAPRI and beyond?

Thursday, 15 February 2024 09:30 (30 minutes)

Since the last CAPRI meeting, much has happened in the modeling field. Deep Learning (DL) has revolutionized structure prediction, and we are only beginning to grasp the fruits, as well as the new challenges uncovered in this new era.
I will describe different approaches that we developed and applied to address CAPRI challenges, and how they helped us shape and improve our ability to model, and understand protein-protein interactions.

One of the initial targets, T186 in Round 52, involved the modeling of a long loop within a multi-protein complex. Overlapping with the appearance of the first successful DL modeling approaches (e.g. TrRosetta), this challenge allowed us to experiment with ways to integrate our peptide-protein modeling tools, such as Rosetta FlexPepDock and PatchMAN, with these new approaches.
In turn, one of the latest targets, T231, involved the modeling of a short peptide to its antibody. That challenge came after we had established the use of AlphaFold2 for peptide-protein docking, and allowed us to finally validate its robustness on a truly new structure - and to experiment also with other new protocols, such as RFdiffusion.

In between these targets, we have developed a number of new approaches to dock and design peptides onto proteins, and learned not only about structure prediction, but also basic concepts that govern peptide binding and design - even beyond interactions forming defined, stable structures.

Submitting to:

8th CAPRI assessment meeting

Primary author: SCHUELER-FURMAN, Ora (Hebrew University of Jerusalem)

Co-authors: Ms VARGA, Julia K (Hebrew University of Jerusalem); Dr KHRAMUSHIN, Alisa (EPFL); TRABELSI, Nirit (Hebrew University of Jerusalem); TSABAN, Tomer (Hebrew University of Jerusalem); Dr AVRAHAM, Orly (Hebrew University of Jerusalem)

Presenter: SCHUELER-FURMAN, Ora (Hebrew University of Jerusalem)

Session Classification: CAPRI

Contribution ID: **52**   Type: **Talk**

Complementarity based peptide docking and design

Monday, 12 February 2024 16:40 (25 minutes)

In recent years, computational protein engineering has become instrumental in developing drugs and therapies, especially with the rise of machine learning-based methods, that pushed the performance of such tools to a new level of accuracy. Despite the progress, such approaches still present limitations in terms of robustness and complexity of the design objectives they can accomplish. Two of the most significant hurdles present in the current state-of-the-art binder design approaches (e.g. RFdiffusion, Chroma, MaSIF-seed, etc.) are the lack of diverse and irregular motifs in generated binders and rigidity of the target binding pockets. These hinder the applicability of such methods to more complex targets.

The PatchMAN peptide docking approach can identify structural templates that match local surface patches, from which it extracts fragments that complement the receptor surface, to be used as templates for peptide conformation modeling. Unlike the other methods it allows for generation of diverse binding motifs on the protein surface and is not restricted to most common helical binders. Also, importantly, PatchMAN includes intrinsic receptor flexibility allowing it to approach more difficult targets for which bound structure is not available. PatchMAN is among the top-performing peptide docking tools, and can also provide design solutions for systems that are challenging to target by generating initial binding motifs of high complementarity.

In the new modified version of PatchMAN we show that with additional on-the-fly coarse grained filtering based on buried surface area, up to 90% of the templates can be filtered, significantly improving running times without compromising the performance. Furthermore, reducing the number of initial templates allows intensification of the sampling during the refinement process. Together with the newly implemented locally focused search we present the new PatchMAN as a powerful tool for both peptide docking and PPI design.

Submitting to:

Integrative Computational Biology workshop

Primary author: KHRAMUSHIN, Alisa (EPFL)

Co-authors: Ms TRABELSI, Nirit (Hebrew University of Jerusalem); VARGA, Julia K (Hebrew University of Jerusalem); SCHUELER-FURMAN, Ora (Hebrew University of Jerusalem)

Presenter: KHRAMUSHIN, Alisa (EPFL)

Session Classification: Computational Methods and Deep Learning

Contribution ID: **53**   Type: **Talk**

Assessing the impact of single-point mutations and alternative splicing-induced variations on protein-protein interactions

Thursday, 15 February 2024 11:00 (30 minutes)

Beyond determining the 3D arrangement of interacting protein partners, assessing the impact of sequence variations on binding affinity and specificity is of utmost importance. I will present two methods we developed to address this question. The first method, DLA-Mutation 1, relies on geometric deep learning for predicting mutation-induced binding affinity changes. It exploits and contrasts information from 3D local environments defined around the wild-type and mutated residues in the protein-protein complex. DLA-Mutation combines self-supervised learning from a large collection of protein-protein complexes with the supervised learning of a small number of experimental binding affinity changes. It reaches a Pearson correlation coefficient of 0.735 on about 400 mutations on unseen complexes and displays better generalization capability than the state-of-the-art methods. Beyond assessing the impact of mutations, I will showcase how to use the learned representations for various downstream tasks. The second method, ASPRING 2, systematically detects repeated protein regions alternatively used in evolution. ASPRING identified about 5000 alternative repeats in the human coding fraction, among which 351 are involved in direct interactions in experimental 3D complex structures. I will highlight some examples with conserved amino acid variations between the repeats pointing to potential interaction specificity-determining sites.

1 Szatkownik A., Zea DJ., H. Richard and E. Laine (2023). Building alternative splicing and evolution-aware sequence-structure maps for protein repeats. J. Struct. Biol. 215:107997 doi: 10.1016/j.jsb.2023.107997

2 Mohseni Behbahani Y., E. Laine and A. Carbone (2023). Deep Local Analysis deconstructs protein-protein interfaces and accurately estimates binding affinity changes upon mutation. Bioinformatics 39:i544–i552 doi: 10.1101/2022.12.04.519031

Submitting to:

8th CAPRI assessment meeting

Primary author: LAINE, Elodie

Presenter: LAINE, Elodie

Session Classification: CAPRI

Contribution ID: **54**   Type: **Talk**

Protein structure evolution and convergence

Monday, 12 February 2024 17:05 (25 minutes)

The number of known folds is limited to a few thousand and this number is surprisingly low, several orders of magnitude lower than the number of sequences in the biosphere. Biological or physical constraints may considerably limit the repertoire of folds. In this case, structural convergence should be frequent. However, several studies showed that distribution in proteomes may be a global proxy to build phylogeny and recent experiments of protein design tend instead to show that the number of observed folds is very small compared to the number of possible stable folds. To address these apparent contradictions, we have mapped SCOP CATH and ECOD folds onto a sample of 210 species across the tree of life (TOL). We have assessed congruence using retention index of each fold for the TOL, and principal component analysis for deeper branches. Among the folds, 20% are universally present in our TOL, while 54% are clade-specific, especially among the Eukaryotic clades.

Reconstructed ancestral states coupled with dating of each node on the tree of life provided fold appearance rates. The rate is on average twice higher within Eukaryota than within Bacteria or Archaea. The highest rates are found in the origins of eukaryotes, holozoans, metazoans, metazoans stricto sensu, and vertebrates: the roots of these clades correspond to bursts of fold evolution. We could correlate the functions of some of the fold synapomorphies within eukaryotes with significant evolutionary events. Among them, we find evidence for the rise of multicellularity, adaptive immune system, or virus folds which could be linked to an ecological shift made by tetrapods.

Submitting to:

Integrative Computational Biology workshop

Primary authors: Prof. LECOINTRE, Guillaume (MNHN); Dr CHOMILIER, Jacques (CNRS); ROMEI, Martin (CNRS); CARPENTIER, Mathilde (SU - MNH - CNRS)

Presenter: CARPENTIER, Mathilde (SU - MNH - CNRS)

Session Classification: Computational Methods and Deep Learning

Contribution ID: **55**   Type: **Talk**

Integrated pipeline for protein docking with extended and enhanced Alphafold-based modeling methods

Wednesday, 14 February 2024 15:00 (30 minutes)

Our group, Kiharalab, participated in both the prediction and scoring stages for complex targets. We combined three components in our pipeline, DistPepFold 1, AFSample 2, and a consensus-based score called ranksum we developed for our LZerD protein docking program [3]. For peptide docking, we used our new approach, DistPepFold. DistPepFold improves protein-peptide complex docking using an Alphafold-Multimer (AFM) based architecture through a privileged knowledge distillation approach. DistPepFold leverages a teacher model that uses native interaction information during training and transfers its knowledge to a student model through a teacher-student distillation process. Benchmark study showed that DistPepFold outperforms AFM on peptide docking. AFSample is a protocol to run Alphafold2 with combinations of several settings, and thereby generating many more models, thousands of models per target. In our pipeline, we ran DistPepFold and an in-house implementation of AFSample. Then generated models were clustered and scored mainly based on the ranksum score. The ranksum uses three scoring functions and identifies models that are ranked consistently high by the three scoring functions. For human submission, we also refer to literature when available.

Reference:
1. Zhang Z, Verburgt J, Kagaya Y, Christoffer C, & Kihara D. Improved Peptide Docking with Privileged Knowledge Distillation using Deep Learning. bioRxiv, DOI: 10.1101/2023.12.01.569671 (2023)
2. Wallner B, AFsample: improving multimer prediction with AlphaFold using massive sampling. Bioinformatics, btad573, (2023)
3. Christoffer C, Terashi G et al., Performance and enhancement of the LZerD protein assembly pipeline in CAPRI 38–46, Proteins, 88: 948-961 (2020)

Submitting to:

8th CAPRI assessment meeting

Primary authors: Dr KAGAYA, Yuki (Purdue University); Mr ZHANG, Zicong (Purdue University); Mr VERBURGT, Jacob (Purdue University); Ms JAIN, Anika (Purdue University); Dr NAKAMURA, Tsukasa (Purdue University); KIHARA, Daisuke (Purdue University)

Presenter: KIHARA, Daisuke (Purdue University)

Session Classification: CAPRI

Contribution ID: **56**   Type: **Talk**

NMR Provides Unique Insight into the Functional Dynamics and Interactions of Intrinsically Disordered Proteins involved in Viral Replication

Wednesday, 14 February 2024 10:15 (25 minutes)

Proteins are inherently dynamic, exhibiting conformational freedom on many timescales,1 implicating structural rearrangements that play a major role in molecular interaction, thermodynamic stability and biological function. Intrinsically disordered proteins (IDPs) represent extreme examples where flexibility defines molecular function. In spite of the ubiquitous presence of IDPs throughout biology, the molecular mechanisms regulating their interactions remain poorly understood. We use NMR spectroscopy, in combination with biophysical tools and molecular modelling, to develop a unified description of the structure and dynamics of IDPs as a function of environmental conditions, from membraneless organelles to in-cell,[2-7] and to map their complex molecular recognition trajectories at atomic resolution, from the highly dynamic free-state equilibrium to the bound state ensemble.[8]

Examples include the replication machinery of Measles virus, where we use NMR to characterize the 92 kDa complex formed between the highly disordered phosphoprotein and the nucleoprotein prior to nucleocapsid assembly – a process that we can also follow in real-time.[9] These proteins undergo liquid-liquid phase separation upon mixing and we can combine NMR and fluorescence to describe the molecular basis and functional advantages of this phenomenon.[10] NMR also sheds new light on the molecular basis of host adaptation of influenza polymerase, via a highly dynamic interaction,[11] and reveals the dynamic assembly of SARS-CoV-2 nucleoprotein with its viral partner nsp3.[12]

1. Lewandowski et al Science 348, 578 (2015)

2. Jensen et al Chem Rev 114, 6632 (2014)

[3]. Abyzov et al JACS 138, 6240 (2016)

[4]. Salvi et al Angew Chem Int Ed. 56, 14020 (2017)

[5]. Salvi et al Science Advances (2019)

[6]. Adamski et al JACS (2019)

[7]. Guseva, Schnapka et al JACS (2023)

[8]. Schneider et al JACS 137,1220 (2015)

[9]. Milles et al Science Advances eaat7778 (2018)

[10]. Guseva et al Science Advances (2020)

[11]. Camacho-Zarco et al Nature Communications (2020), Camacho-Zarco, Lu et al JACS (2023)

[12]. Bessa, Guseva, Camacho-Zarco et al Science Advances (2022)

Submitting to:

Presenter: BLACKLEDGE, martin

Session Classification: Protein conformational flexibility and solution experiments

Contribution ID: **57**   Type: **Talk**

CAPRI 2024: Pre- and post-Alphafold protein docking

Wednesday, 14 February 2024 14:15 (45 minutes)

Submitting to:

Presenter: LENSINK, Marc (CNRS)

Session Classification: CAPRI

Contribution ID: **58** Type: **Talk**

The Expanding Horizon of Protein Interactions

Thursday, 15 February 2024 11:30 (45 minutes)

Submitting to:

Presenter: WODAK, Shoshana (VIB-VUB Center doe Structural Biooogy, Brussels Belgium)

Session Classification: CAPRI

Contribution ID: **59**   Type: **Talk**

Automated protein-to-Structure pipelines for new structures and X-ray based ligand screening at EMBL Grenoble. Experience from our first collaboration with CAPRI

Thursday, 15 February 2024 16:45 (45 minutes)

Submitting to:

Presenter: MÁRQUEZ, José (EMBL Grenoble)

Session Classification: CAPRI

Contribution ID: **61**   Type: **Talk**

Integrative structure of a histone chaperone-histone complex

Tuesday, 13 February 2024 09:25 (25 minutes)

Histone chaperones play a crucial role in regulating the assembly and disassembly of chromatin. Our lab recently reported a novel chaperone binding mode in which histone chaperone APLF single-handedly assembled the histone complexes H2A-H2B and H3-H4 into the histone octamer. The chaperone domain of APLF consists of a short (~60 aa) intrinsically disordered, highly acidic domain (AD). As we could only solve the crystal structure of a peptide fragment of the AD bound to the histone octamer, we used an integrative structural biology approach to define the conformation of the rest of the AD. In this talk I will outline how the recent implementation of shape-based docking in the HADDOCK was used to integrate small angle xray and neutron scattering, cross-linking mass-spectrometry, NMR and the crystal structure to define the integrative structure of this challenging histone chaperone-histone complex.

Submitting to:

Presenter: VAN INGEN, Hugo (Utrecht University)

Session Classification: Computational Methods and Deep Learning

Contribution ID: **62**   Type: **Talk**

Integrative modeling of protein oligomers using Heligeom

Tuesday, 13 February 2024 11:10 (25 minutes)

Protein oligomers can modify their overall architecture in response to changes in the environment, such as ion concentration and composition, the presence of small ligands or mechanic stress. These changes may involve small variations in the subunit-subunit interface which can lead to important changes in the overall shape due to multiplication effect. They may also involve large interface variations and lead to alternative assembly modes. These characteristics pose specific problems for the integrative modeling of oligomeric assemblies. For example, characterizing different oligomerization states in solution from SAXS experiments may be tricky. Theoretical approaches were developed in the past to directly relate peaks of the SAXS profiles to quantities such as pitch or number of monomers per turn. Although these approaches benefit from periodicity, the analytic treatment only holds as long as the repeated subunit has a simple, well-defined geometric form such as a sphere. I will present a strategy to handle more complex protein shapes based on the combination of interface sampling, screw construction (Heligeom 1) and SAXS profile reconstruction (FoxS 2) and I will illustrate this approach in the case of the polymorphic RecA filament of homologous recombination.

In addition to establishing the correspondence between interface geometry at the dimer level and helical properties of the oligomer, our python module Heligeom allows comparative analyses of oligomeric assemblies. It also includes a cyclic adjustment tool which enables generating assemblies with desired ring sizes with interfaces as close as possible to a predicted interface, where the predicted interface may come from docking [3] or from deep-learning calculations. Heligeom will soon be made available as a web server, making its functionalities available to researchers interested in helical or ring protein assemblies.

1. B. Boyer, J. Ezelin, P. Poulain, A. Saladin, M. Zacharias, C. H. Robert and C. Prévost (2015). An integrative approach to the study of filamentous oligomeric assemblies, with application to RecA. PLoS ONE 10 (3): e0116414 — doi: 10.1371/journal.pone.0116414

2. Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A. FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. Nucleic Acids Res. 2016 Jul 8;44(W1):W424-9 — doi: 10.1093/nar/gkw389

3. L. Tran, N. Basdevant, C. Prévost and T. Ha-Duong (2016) Structure of ring-shaped Aβ42 oligomers determined by conformational selection. Scientific Reports 6, 21429 — doi: 10.1038/srep21429

Submitting to:

Presenter: PRÉVOST, Chantal (CNRS et univ Paris Cité)

Session Classification: Computational Methods and Deep Learning

Contribution ID: **63**   Type: **Talk**

From Interaction Prediction to Sequence Design: Unveiling PeSTo's Potential in Structural Biology

Wednesday, 14 February 2024 17:45 (30 minutes)

In the field of structural biology, predicting protein interactions and designing sequences based on backbone scaffolds remain pivotal yet challenging tasks. Built on the same deep learning architectural framework, Protein Structure Transformer (PeSTo) and its derivative, CARBonAra, address these challenges. PeSTo employs geometric transformers to proficiently predict diverse protein binding interfaces, setting a new benchmark in accuracy and computational efficiency. It enables high-throughput analyses and is compatible with the expansive AlphaFold foldome. CARBonAra, adapted from PeSTo, specializes in sequence recovery from backbone scaffolds and uniquely accounts for non-protein entities like nucleic acids and ligands. These methods combine speed, accuracy, and wide applicability, offering promising avenues for advancements in structural biology and biotechnology.

https://www.nature.com/articles/s41467-023-37701-8
https://www.biorxiv.org/content/10.1101/2023.06.19.545381v1

Submitting to:

Presenter: KRAPP, Lucien (Laboratory for Biomolecular Modeling, EPFL)

Session Classification: CAPRI