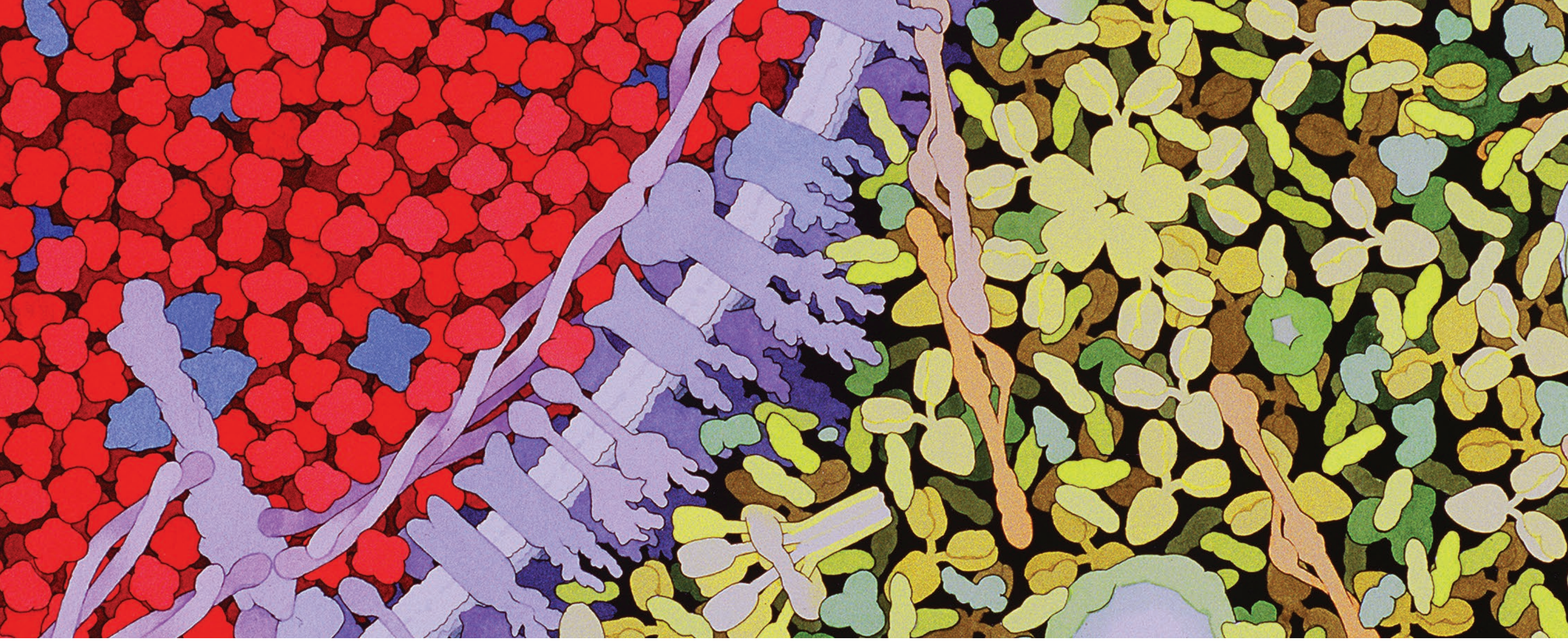

Virtual symposium celebrating the 50th anniversary of the Protein Data Bank

May 4–5, 2021

Day 2 Abstracts

ASBMB

Worldwide PDB Protein Data Bank

Worldwide Protein Data Bank Foundation

Day 2: Wednesday May 5, 2021

POSTER SESSION 4
Wednesday May 5, 2021 1:00pm-1:40pm EDT

Systematic Analysis of Symmetry in Membrane Proteins

Antoniya A. Aleksandrova[1], Edoardo Sarti[1,2], Emily Yaklich[1], Lucy R. Forrest[1]

[1]NIH-NINDS, Bethesda, MD 20892, USA, [2]Sorbonne Université, Institut de Biologie Paris Seine, Paris 75005, France

The burgeoning number of membrane protein structures in the PDB has revealed an abundance of symmetry and pseudo-symmetry, which arose not only by the formation of multi-subunit assemblies, but also by repetition of internal structural elements. In many cases, these symmetry relationships play a crucial role in defining the functional properties of the proteins. Therefore, a systematic study of symmetry should provide a framework for a broader understanding of the mechanistic principles and evolutionary development of membrane proteins. However, available symmetry detection methods have not been tested systematically on this class of proteins because of the lack of an appropriate benchmark set. Hence, we collected membrane protein structures with unique architectures and manually curated their symmetries to create the MemSTATS dataset. Using MemSTATS, we compared the performance of four widely used symmetry detection algorithms (CE-Symm, SymD, AnAnaS, QuatSymm) and pinpointed areas for improvement. To address the identified shortcomings, we developed a robust symmetry detection methodology called MSSD, which takes into consideration the restrictions that the lipid bilayer places on protein structures. MSSD detected symmetries with higher accuracy and lower false positive rate compared to any other tested method. Consequently, we used MSSD to analyze all available membrane protein structures and presented the resultant symmetries in a database called EncoMPASS (encompass.ninds.nih.gov). Using EncoMPASS, we then proceeded to quantify both the extent and diversity of symmetry relationships in known structures of membrane proteins. Ongoing work is addressing diverse questions relating to evolution by analyzing symmetries across species, to the driving forces for oligomerization and fusion by investigating symmetric interfaces, and to function and mechanism by using symmetry to identify and verify ion binding sites.

Local backbone geometry plays a key role in determining the conformational preferences of amino acid residues

Nicole Balasco[1], Luciana Esposito[1], Alfonso De Simone[2], Luigi Vitagliano[1]

[1]Institute of Biostructures and Bioimaging, CNR, 80134 Naples, Italy, [2]Department of Pharmacy, University of Naples "Federico II," 80131 Naples, Italy

The definition of the structural basis of the conformational preferences of the genetically encoded amino acids is crucial to decipher the folding code and would have a huge impact on our understanding of protein structure and function. Although a large number of computational and experimental investigations have highlighted that the different amino acid residues are endowed with distinct conformational propensities, none of the current hypotheses is able to satisfactorily explain these preferences [1-6]. In order to gain insights into this intricate issue, we here determined and compared the amino acid propensity scales for different (φ,ψ) regions of the Ramachandran plot and for different secondary structure elements. These propensities were calculated using the Chou-Fasman approach on a database of 4731 non-redundant protein chains retrieved from the Protein Data Bank. Linear regression analyses were conducted to evaluate similarities between propensity scales of different (φ,ψ) areas. One of the most striking and unexpected findings emerged from this research is that distant regions of the Ramachandran plot occasionally exhibit significantly similar propensity scales. On the other hand, contiguous regions of the Ramachandran plot present anti-correlated propensities. In order to provide an interpretative background to these results, we evaluated the role that the local variability of protein backbone geometry plays in this context. Our analysis indicates that (dis)similarities of propensity scales between different regions of the Ramachandran plot are coupled with (dis)similarities in the local geometry. In recent years, many investigations have highlighted the influence of the conformation on the protein local geometry [7]. The present findings reverse this concept by showing that the local geometric requirements have a significant impact on the preference of individual amino acids for specific conformational states.

References

[1] CB. Anfinsen, HA. Scheraga. Adv Protein Chem 1975,29:205-300. [2] TP. Creamer, GD. Rose. Proc Natl Acad Sci USA 1992,89(13):5937-41. [3] AG. Street, SL. Mayo. Proc Natl Acad Sci USA 1999,96(16):9074-6. [4] F. Avbelj, RL. Baldwin. Proc Natl Acad Sci USA 2003,100(10):5742-7. [5] CL. Towse et al. Biophys J 2016,110(2):348-361. [6] BF. Fisher et al. J Am Chem Soc 2017,139(38):13292-13295. [7] N. Balasco et al. Acta Crystallogr D Struct Biol 2017,73:618–625.

Preparing for Time Resolved Serial Crystallography Studies on Lentinus similis Lytic Polysaccharide MonoOxygenase (LsAA9A)

Sanchari Banerjee[1], Sebastian J. Muderspach[1], Tobias Tandrup[1], Kristian E. H. Frandsen[1], Johan Ø. Ipsen[2], Katja S. Johansen[3], Christina Hernández-Rollán[4], Morten H. H. Nørholm[4], Pedram Mehrabi[5], Eike Schulz[5], David v. Stetten[6], Arwen R. Pearson[7], Leila L. Leggio[1]

[1]Dept. of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100, Denmark, [2]Dept. of Plant & Environmental Sciences, University of Copenhagen, Frederiksberg C, DK-1871, Denmark, [3]Dept. of Geoscience & Natural Resource Management, University of Copenhagen, Frederiksberg, DK-1958, [4]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2800, [5]Max Planck Institute for the Structure and Dynamics of Matter, Luruper Chaussee 149, 22761 Hamburg, [6]EMBL c/o DESY, Notkestrasse 85, 22607 Hamburg, Germany, [7]Hamburg Centre for Ultrafast Imaging, Universität Hamburg, Luruper Chaussee 149, 22761 Hamburg

Lytic Polysaccharide MonoOxygenases (LPMOs) constitute a class of copper metallo-enzymes that act synergistically with other known carbohydrate active enzymes in oxidative degradation of the recalcitrant polysaccharides in plant biomass. Industrially, enzymatic processes are used to convert plant biomass into simple sugars, which can be fermented to bioethanol, an alternative form of renewable fuel. The copper ion in LPMOs is coordinated by two histidine residues forming a His-brace, accompanied by a mostly conserved tyrosine residue. Reduction of Cu(II) to Cu(I) makes LPMOs catalytically active, resulting in oxidative cleavage of the glycosidic bond at either the C1 or C4 position of sugar ring(1,2). LPMOs have been extensively studied industrially, and in our lab particularly the AA9 LPMO LsAA9A from Lentinus similis(3,4). Although the reaction mechanism has been proposed(1,5), the exact sequence of events is still unknown. In order to increase the catalytic efficiency of the enzyme industrially, it is important to understand the correct sequence of the mechanism. We are now preparing for time resolved studies using serial crystallography at synchrotrons. We have optimized crystallization conditions for growing microcrystals by batch method. These results along with the initial diffraction tests will be presented.

1 Quinlan et al., PNAS, 2011
2 Tandrup et al., Biochemical Society Transactions, 2018
3 Frandsen et al., Nature Chemical Biology, 2016
4 Simmons et al., Nature Communications, 2017
5 Vaaje-Kolstad et al., Science, 2010

Structural and dynamic studies of mouse BTNL2, an orphan T cell co-inhibitor

Aditya J. Basak[1], Snigdha Maiti[1], Anita Hansda[2], Gayatri Mukherjee[2], Soumya De[1], Dibyendu Samanta[1]

[1]School of Bioscience, IIT Kharagpur, India, [2]School of Medical Science and Technology, IIT Kharagpur, India

BTNL2 is a T-cell coinhibitory protein that modulates regulatory T-cell differentiation. It is implicated in the pathogenesis of inflammatory diseases such as sarcoidosis, ulcerative colitis and Graft vs Host disease. The potential role of this T cell inhibitory molecule in medicine needs to be explored by investigating the structural basis of its function.

Here we present the NMR structure of the N-terminal IgV domain of mouse BTNL2 (Asp 28 to Ala 143), which has 65% sequence identity with its human counterpart. The protein was expressed in E. coli BL21(λDE3) strain as inclusion bodies, refolded in-vitro by fast-dilution method and purified by size-exclusion chromatography. Functional activity of the refolded protein was demonstrated by CD4+ cell-binding and IL2 suppression assays.
Size-exclusion chromatography (SEC) and analytical ultracentrifugation (AUC) results show that in solution, BTNL2D28-A143 is monomeric. This is further supported by rotational correlation time (τc) calculated using NMR relaxation (R1/R2 ratios) data. Structure was calculated using PONDEROSA C/S. The final ensemble of 10 structures (PDB ID: 6L7Z) has backbone and heavy atom RMSD of 0.56 Å and 0.76 Å, respectively, across all rigid residues.
BTNL2D28-A143 adopts a typical IgV-fold consisting of nine β strands arranged in an antiparallel fashion to form a beta sandwich. A cysteine mediated disulfide bond across strands B and F holds the two sheets together. Fast-time scale NMR dynamics data (15N NOE, R1 and R2) show that the domain is rigid except for loop regions.

In silico generation of holo-like conformations of multi-pocket highly flexible allosteric proteins

Andrea Basciu[1], Giuliano Malloci[1], Andrea Bosin[1], Paolo Ruggerone[1], Alexandre Bonvin[2], Attilio Vargiu[1]

[1]University of Cagliari (Italy), [2]Utrecht University

Quantitative understanding of molecular recognition is crucial for basic research and for structure-based drug design. Key to this goal is the knowledge of the atomic-level structures of the complexes formed by receptors and their ligands. Determining the structure of a protein by experimental techniques remains a quite demanding task in terms of both cost and duration of the experiments. Computational methods have become a valid complement to experiments, although inaccuracy increases with the extent of the conformational changes associated to protein-ligand binding. To address this limitation, we recently introduced “EDES - Ensemble Docking with Enhanced-sampling of pocket Shape”, a computational method based on metadynamics simulations to generate holo-like conformations of proteins by only exploiting their apo structure. Here, we present an improved version of the original protocol enabling to handle multiple - allosteric - binding sites in extremely flexible proteins. We applied our method to a very challenging target, namely the enzyme adenylate kinase (AK), which undergoes very large conformational changes upon ligand binding. Our protocol generated a significant fraction of structures featuring a low RMSD from the experimental geometry of the complex between AK and an inhibitor. These conformations were used in ensemble docking calculations yielding to native-like poses of substrates and inhibitors of adenylate kinase among the top-ranked ones

Using evolutionary covariance to better determine protein quaternary state from a crystal structure

J. Javier Burgos-Mármol[1], Ronan M. Keegan[2], Eugene Krissinel[2], Daniel J. Rigden[1]

[1]Department of Biochemistry & Systems Biology, ISMIB, University of Liverpool, Liverpool L69 7ZB, UK, [2]UKRI-STFC, Rutherford Appleton Laboratory, Research Complex at Harwell, Didcot OX11 0FA, UK

Reliable determination of the quaternary structure of a protein is often crucial to a full understanding of its function. However, for decades, crystallographers have sometimes struggled to distinguish between biologically meaningful interfaces observed in a crystal structure, and the unnatural lattice contacts that allow for crystal formation. In order to solve this problem, Eugene Krissinel developed PISA, and later jsPISA, a CCP4 tool that sorts different candidate quaternary structures according to a likelihood obtained from results on dissociation free energy [1, 2]. Additionally, jsPISA incorporates an "interaction radar" that allows for a rapid visualisation of how likely an interface is according to a number of physico-chemical parameters. In order to provide jsPISA with an additional -- independent -- source of information to determine the probability of a given interface to biologically exist, we propose the use of evolutionary covariance data. This proposal is based upon the observation that pairs of residues whose interaction contributes to a biologically important interface are constrained in their evolution [3, 4]. Consequently, the detection of a pair covariation signal points at the existence of a contact between the two residues contributing to the formation of a biologically important interface. The new extension will enhance the results currently displayed by jsPISA with an additional score and new data based on evolutionary covariance analysis, thereby helping determine the relevant quaternary structure in difficult cases.

[1] Krissinel, E.; Henrick, K. J. Mol Biol., 372 (2007), 774-797, https://doi.org/10.1016/j.jmb.2007.05.022
[2] Krissinel, E. Nucleic Acids Research, 43 (2015), W314–W319, https://doi.org/10.1093/nar/gkv314
[3] Ovchinnikov, S.; Kamisetty, H.; Baker, D. eLife (2014), 3:e02030 https://doi.org/10.7554/eLife.02030
[4] Hopf, T.A. et al. eLife (2014), 3:e03430, https://doi.org/10.7554/eLife.03430

Cryo-EM reveals how hemophore binding to Hb prompts heme removal: a structural and mechanistic insight into the function of IsdB from S. aureus

Omar De Bei[1], Eleonora Gianquinto[2], Dimitri Y. Chirgadze[3], Steven W. Hardwick[3], Lee Cooper[3], Marialaura Marchetti[1], Francesca Spyrakis[2], Stefano Bettati[1, 4], Luca Ronda[4], Ben F. Luisi[3], Barbara Campanini[5]

[1]Interdept. Center Biopharmanet-TEC, Parco Area delle Scienze, Bldg. 33, 43124, Parma, Italy, [2]Dept. of Drug Science and Technology, University of Turin, Via Giuria 9, 10125, Turin, Italy, [3]Dept. of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK, [4]Dept. of Medicine and Surgery, University of Parma, Via Gramsci 14, 43126, Parma, Italy, [5]Dept. of Food and Drug, University of Parma, Parco Area delle Scienze, 27/A, 43124, Parma, Italy

Iron plays a key role in sustaining bacterial metabolism during host colonization, and many pathogens rely on hemoglobin (Hb) as the preferred iron source. IsdB is a surface-exposed receptor of *Staphylococcus aureus* that binds Hb and extracts heme to retrieve iron. The protein is highly expressed during infections and is a recognized virulence factor; hence, its interaction with Hb represents a potential target to develop new antimicrobial agents. Tackling protein-protein interactions requires an in-depth knowledge of the structural details of protein complexes. Cryo-EM was exploited to study the interaction between wild-type IsdB and Hb in two ligation states, carboxyHb (HbCO) and oxidized Hb(metHb). While metHb is the dominant extracellular form upon hemolysis induced by *S. aureus* hemolysins and the natural substrate of IsdB, HbCO forms a complex with IsdB but is resistant to heme extraction. The latter complex thus represents a snapshot of the intermediate step in the heme extraction process.
The stoichiometry of the two complexes and their homogeneity were assessed using size-exclusion chromatography coupled with multi-angle light scattering. Cryo-EM single particle analysis led to the reconstruction of a 2.9 Å map of IsdB:HbCO complex, revealing two IsdB molecules bound to the β-chains of a Hb tetramer, and a dense network of contacts that favour heme extraction and dissociation of Hb into dimers. Reconstruction of the IsdB:metHb complex at 5.8 Å reveals two IsdB molecules bound to one dimer of Hb, with the heme absent from Hb and now present within IsdB. These findings suggest that structural strain induced by IsdB binding favours tetramer dissociation in an intermediate stage of the heme extraction process.

Interplay between hevin, SPARC, and MDGAs: modulators of neurexin-neuroligin transsynaptic bridges

Shanghua Fan[1, 2], Shanti Pal Gangwar[1, 2], Mischa Machius[1, 2], Gabby Rudenko[1, 2]

[1]Department of Pharmacology and Toxicology, [2]Sealy Center for Structural Biology and Molecular Biophysics. University of Texas Medical Branch.

Electronic address: garudenk@utmb.edu.

Hevin (also known as SPARCL1) promotes synapse formation and stabilizes neurexin-neuroligin trans-synaptic bridges in vivo. Hevin is secreted predominantly by astrocytes, and its synaptogenic effects are antagonized by related protein, SPARC. A third protein, the membrane-tethered MDGA, blocks neurexin-neuroligin trans-synaptic bridges. The structure of hevin and how these three proteins impact each other is unknown. Here, we present the crystal structure of the hevin FS-EC tandem and reveal the molecular underpinnings of a regulatory network formed by this trio of proteins. Despite the high sequence homology, the 2.3 Å hevin FS-EC structure reveals profound differences with SPARC, in particular, in the EC domain. The FS domain, however, is structurally conserved, and it houses nanomolar affinity binding sites for neurexin and neuroligin. SPARC also binds neurexin and neuroligin, competing with hevin, so its antagonist action likely is rooted in its shortened N-terminal region. Strikingly, the hevin FS domain competes with MDGA for an overlapping binding site on neuroligin, while the hevin EC domain binds the extracellular matrix protein, collagen, like SPARC. Collectively, these data provide mechanistic insight into how secreted and membrane-tethered synaptic organizers interplay to regulate neurexin-neuroligin trans-synaptic bridges, couple to the extracellular matrix and impact synapse formation, ultimately shaping neural circuits.

Iron necessity meets opportunity: Exploring the molecular mechanism of siderophore recycling in human pathogens

Inês B. Trindade[1], Bruno M. Fonseca[1], Filipe Rollo[1], Pedro M. Matias[1], Elin Moe[1], Ricardo O. Louro[1]

[1]ITQB-NOVA, Oeiras, Portugal

Iron is an essential element for Life although in an oxygen-rich atmosphere its bioavailability is limited. To circumvent this limitation pathogens produce siderophores, small metal chelating compounds with high affinity for ferric iron. Siderophores allow them to proliferate in iron depleted environments such as the human body, where iron availability is tightly regulated. Despite the broad relevance of siderophores for iron metabolism, within the siderophore pathway there is still limited knowledge on the molecular mechanisms for iron release from the ferric-siderophore complexes. Siderophore recycling is the most frequent mechanism for iron release from ferric siderophores and is mediated by Siderophore-Interacting proteins (SIPs). In this work, emphasis will be given to understanding siderophore recycling in opportunistic human pathogenic Gram-negative bacteria from the genus *Shewanella*. Human infections attributed to *Shewanella* are presently increasing, with more than 80% of the known cases being reported since 2000.
Here, we report the structural characterization of a SIP from *Shewanella bicestrii*, isolated from the cerebrospinal fluid of a 1-year-old girl with suspected meningitis, and compare it with its homologue from non-pathogenic *Shewanella frigidimarina*. Preliminary substrate binding studies show mechanistic differences between these two SIPs.
Further biochemical characterization of the SIP from *S. bicestrii* will reveal in more detail its molecular mechanism of action and open the way for the rational design of Trojan-horse conjugates, which are therapeutic agents that combine antibiotics with siderophores to gain entry inside the cells of pathogenic microorganisms. This is essential to determine whether SIPs present feasible targets for the development of novel therapies against pathogens of humans and other hosts.

Acknowledgements: supported by H2020 project TIMB3 contract 810856, Instruct-ERIC PID4509, COST Action CA15133, MOSTMICRO and FCT grant PD/BD/135187/2017

U're the one I want: Cryo-EM Structures of SARS-CoV-2 Nsp15 Reveal Insight into Nuclease Specificity and Dynamics

Meredith N. Frazier[1], Monica C. Pillon[1], Lucas B. Dillard[2], Jason G. Williams[3], Seda Kocaman[1], Juno M. Krahn[2], Lalith Perera[2], Cassandra K. Hayne[1], Jacob Gordon[1], Zachary D. Stewart[1], Mack Sobhany[1], Leesa J. Deterding[3], Allen L. Hsu[2], Venkata P. Dandey[2], Mario J. Borgnia[2], Robin E. Stanley[1]

[1]Signal Transduction Laboratory, National Institute of Environmental Health Sciences, NIH, [2]Genome Integrity and Structural Biology Laboratory, NIEHS, NIH, [3]Epigenetics and Stem Cell Biology Laboratory, NIEHS, NIH

Nsp15, a uridine specific endoribonuclease conserved across coronaviruses, processes viral RNA to evade detection by host defense systems. Crystal structures of Nsp15 from different coronaviruses have shown a common hexameric assembly yet how the enzyme recognizes and processes RNA remains poorly understood. We determined the first series of cryo-EM reconstructions of SARS-CoV-2 Nsp15, in apo, UTP-bound, and pre- and post- cleavage states. The cryo-EM reconstructions, combined with biochemistry, mass spectrometry, and molecular dynamics, expose molecular details of how critical active site residues recognize uridine and facilitate catalysis of the phosphodiester bond. Gel-based and FRET cleavage assays revealed additional sequence preferences beyond the uridine. Mass spectrometry revealed the accumulation of cyclic phosphate cleavage products, while analysis of the apo and UTP-bound datasets revealed conformational dynamics not observed by crystal structures that are likely important to facilitate substrate recognition and regulate nuclease activity. Collectively, these findings advance understanding of how Nsp15 processes viral RNA and provide a structural framework for the development of new therapeutics.

The structural mechanism of a DNA polymerase sliding clamp loader visualized by cryo-EM

Christl Gaubitz[1], Xingchen Liu[1], Nicholas P. Stone[1], Janelle A. Hayes[1], Gabriel Demo[2], Brian A. Kelch[1]

[1]University of Massachusetts Medical School, [2]CEITEC

DNA replication in all life requires the sliding clamp, which is called Proliferating Cell Nuclear Antigen (PCNA) in eukaryotes. PCNA is a mobile, ring-shaped protein that wraps around DNA, where it acts as a scaffold for dozens of proteins to facilitate and coordinate numerous pathways, including DNA replication and repair. PCNA function is governed by its presence or absence on DNA. PCNA is a closed ring in solution that must be opened by clamp loader AAA+ ATPase complexes to be placed on DNA. However, the mechanism by which a clamp loader installs PCNA on DNA is still unknown. Here, we present a comprehensive picture of PCNA opening and DNA loading by the eukaryotic clamp loader Replication Factor C (RFC). Cryo-EM structures of RFC bound to PCNA reveal how RFC latches onto and opens PCNA, while undergoing a large 'crab-claw' like conformational change from an autoinhibited conformation into an open spiral conformation, which is competent to bind DNA. Next, the open RFC:PCNA complex binds to DNA and interrogates the primer-template junction through a limited melting mechanism. The structures of RFC and PCNA with DNA further reveal that PCNA closure around DNA does not require ATP hydrolysis, but is driven by binding energy alone. ATP hydrolysis, which is necessary for RFC release, is triggered by interactions with both PCNA and DNA, explaining RFC's switch-like ATPase activity. The ensemble of structures presented here illustrates a unique AAA+ mechanism that is distinct from the canonical 'spiral staircase' mechanism of the ringed, hexameric AAA+ motor cousins of RFC. Collectively, this work gives insight into the mechanism by which the clamp loader achieves substrate preference and completes the multistep loading process, whilst preventing futile cycles of ATP.

Assembling The Water-splitting Enzyme

Christopher J. Gisriel[2], Zhou Kaifeng[3], Richard J. Debus[1], Yong Xiong[3], Gary W. Brudvig[2]

[1]Department of Biochemistry, University of California, Riverside, CA, USA., [2]Department of Chemistry, Yale University, New Haven, CT 06511, USA, [3]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511, USA.

The photosystem II (PSII) enzyme catalyzes water oxidation to derive reducing equivalents that fuel the organism's metabolism. The mature enzyme is dimeric and harbors an inorganic metal cluster, termed the "oxygen-evolving complex" (OEC), in its active site. Molecular oxygen is produced as a byproduct of water oxidation that can damage to the mature enzyme. To cope with this, the cell has developed a complex mechanism of assembly and repair. Thus, PSII is found in various intermediate states in vivo, and while the mature PSII structure is well known, an understanding of how this repair process occurs, and especially how the active site cofactor is assembled, is lacking. To investigate this, we solved the cryo-EM structure of monomeric PSII lacking extrinsic and peripheral subunits, and the OEC. The apo-PSII's lumenal surface is negatively charged to attract cations, increasingly-so toward the OEC-binding pocket. The configuration of the OEC-binding site is found to be quite different than that observed in mature PSII. Furthermore, nearby looping regions, especially those from the CP43 subunit, are shifted away from the OEC-binding site, resulting in a more open configuration for cation delivery. We use these observations to suggest important characteristics of apo-PSII prior to a structural rearrangement that occurs during OEC assembly that has previously remained elusive, giving insight into its mechanism. This structure provides a platform for future PSII studies that aim to understand OEC assembly by solving molecular structures using a bottom-up approach.

BMRBj: Regional NMR data repository site of PDBj

Masashi Yokochi[1], Takeshi Iwata[1], Yohei Miyanoiri[1], Chojiro Kojima[2], Genji Kurisu[1], Toshimichi Fujiwara[1]

[1]Institute for Protein Research, Osaka University, [2]Graduate School of Engineering Science, Yokohama National University

Biological Magnetic Resonance Data Bank Japan (BMRBj, formerly known as PDBj-BMRB at Osaka, https://bmrbj.pdbj.org) is a regional processing and archiving site of NMR experimental data in collaboration with BMRB (UConn, https://bmrb.io). We launched a BMRB satellite site in 2002 to cope with the rapid growth of NMR structural data at that time. The first BMRB entry processed at our site was released in 2005, since then 10% of BMRB entries have been processed.

We maintain annotation service that is equivalent to the BMRB-UConn for both OneDep and BMRBdep depositions, individually offer SMSDep deposition site to accept NMR-derived structures of small, biologically interesting molecules that are ineligible for wwPDB deposition. To facilitate reusing NMR experimental data, we are strongly supporting recent OneDep developments that enable both NMR deposition and validation using either NEF (NMR exchange format) or combined NMR-STAR format. As for data-out activities, we provide our own search service with a wealth of annotations to help users reuse the NMR data.

Scooped! Estimating Rewards for Priority in Science

Ryan Hill[1], Carolyn Stein[2]

[1]Northwestern University, [2]Massachusetts Institute of Technology

The scientific community assigns credit or “priority” to individuals who publish an important discovery first. We examine the impact of losing a priority race (colloquially known as getting “scooped”) on subsequent publication and career outcomes. To do so, we take advantage of data from structural biology where the nature of the scientific process together with the Protein Data Bank — a repository of standardized research discoveries — enables us to identify priority races and their outcomes. We find that race winners receive more attention than losers, but that these contests are not winner-take-all. Scooped teams are 2.5 percent less likely to publish, are 18 percent less likely to appear in a top-10 journal, and receive 20 percent fewer citations. As a share of total citations, we estimate that scooped papers receive a credit share of 45 percent. This is larger than the theoretical benchmark of zero percent suggested by classic models of innovation races. We conduct a survey of structural biologists which suggests that active scientists are more pessimistic about the cost of getting scooped than can be justified by the data. Much of the citation effect can be explained by journal placement, suggesting editors and reviewers are key arbiters of academic priority. Getting scooped has only modest effects on academic careers. Finally, we present a simple model of statistical discrimination in academic attention to explain how the priority reward system reinforces inequality in science, and document empirical evidence consistent with our model. On the whole, these estimates inform both theoretical models of innovation races and suggest opportunities to re-evaluate the policies and institutions that affect credit allocation in science.
link to working paper: https://economics.mit.edu/files/20197

Structural characterization of DUF2419: a queuine salvage protein

Shr-Hau Hung[1], Sana Alkuzweny[1], Manal A. Swairjo[1, 2]

[1]Department of Chemistry and Biochemistry, San Diego State University, San Diego CA 92182,
[2]The Viral Information Institute, San Diego State University, San Diego CA 92182

Queuine is a modified nucleobase that is biosynthesized *de novo* via 7-deazaguanine pathway in many bacteria and archaea. It is salvaged by eukaryotes where it acts as a precursor of the modified nucleoside Queuosine (Q), found at the anticodon wobble base of cytoplasmic and mitochondrial tRNAs that incorporate Tyr, Asp, Asn and His amino acids. In human, queuine is exclusively retrieved from the diet and microbiome and is subsequently inserted in human tRNA by the human tRNA-guanine transglycosylase enzyme. Queuine is the only micronutrient that directly affects efficiency and accuracy of translation of mammalian proteins, with roles in metabolism and brain health in aging. It is required for the biosynthesis of BH4, a vital cofactor for numerous enzymes in the body, including those involved in the formation of the amino acid tyrosine, and the key neurotransmitters dopamine and serotonin. Recently, DUF2419 (encoded by *C9orf649* in human) was identified as a protein family involved in queuine salvage in plants, animals, fungi, and even in some bacteria lacking the 7-deazaguanine *de novo* biosynthesis. Nevertheless, the exact biochemical function of DUF2419 remains unknown. DUF2419 is deleted in cases in Acute Myeloid Leukemia, and Q-modified tRNA levels correlate with tumor aggression, implying a link between Q metabolism and cancer in humans. Here we report the x-ray crystal structure of DUF2419 from *Sphaerobacter thermophilus* (~37 kDa), determined at 2.1-Å resolution by selenium-based multi-wavelength anomalous dispersion methods. The structure reveals a monomeric protein belonging to the orthogonal helical bundle fold. A DALI search for similar structures returned several 8-oxoguanine DNA glycosylases, which cleave oxidized guanines in DNA. Structure-based alignment of DUF2419 with bacterial, archaeal, and human DNA glycosylases reveals 22% sequence similarity and only two invariant residues: the catalytic Asp residue and the Lys essential for recognition and positioning of the substrate during the base hydrolysis reaction. Moreover, DUF2419 specific helical deletions render its putative active site large enough to accommodate the cyclopentenyl ring of queuine. The results suggest that the role of DUF2419 in q salvage is to function as a ribonucleos(t)ide hydrolase that liberates the queuine base from its nucleos(t)ide precursor.

Structural and Functional insights of a Novel Multifunctional Endoglucanase from Thermophilic Thermogutta terrifontis

Naveed Hussain[1, 2, 5], Peter Harrison[3, 6], Halina Mikolajek[6], Neil Paterson[6], Muhammad W. Akhtar[5], Saima Sadaf[2, 5], James H. Naismith[1, 4]

[1]School of Biochemistry & Biotechnology, University of the Punjab, Lahore, Pakistan, [2]The Division of Structural Biology, The Nuffield Department of Medicine, University of Oxford, [3]The Membrane Protein Lab, Research Complex at Harwell (R92), Harwell Science & Innovation Campus, [4]Rosalind Franklin Institute, Harwell Campus, Didcot, OX11 0FA, [5]School of Biological Sciences, University of the Punjab, Lahore, Pakistan, [6]Diamond Light Source Ltd, Diamond House Harwell Science & Innovation Campus

Saccharification of polysaccharides require different types of glycoside hydrolases. Incorporated into industrial processes these improve food quality and have application in the biofuels industry. For practical and economic utility, the saccharification enzymes must be thermostable. The discovery of multifunctional enzymes is urgently needed. Exploiting bioinformatics, recombinant DNA technology and structural biology can help to accelerate this discovery. In the current study, the thermostable multifunctional endoglucanase from glycoside hydrolase family 5 (GH5) from Thermogutta terrifontis was cloned, overexpressed, and purified in E. coli. In its native form it was insoluble, however after truncating its 19 amino acids from its N-terminal 90% soluble expression (100-125mg/L purified protein) was seen in BL-21 cells. Besides having high 1, 3-, 1, 4-endoglucanase activities against glucan from barley and carboxymethylcellulose, this enzyme has shown moderate exoglucanase activities against Avicel and regenerated amorphous cellulose. Moreover, side activities of betaglucosidase and 1, 6-endoglunase have also been observed against salicin and laminarin, respectively. All the activities were seen over a broad range of temperature 50-80°C and pH 6. This enzyme is interesting from a structural perspective, as it has less than 29% sequence identities with other known structural homologs at PDB database. Recently, its apo crystals were obtained after a long period of time, and its structure was solved by sulphur-SAD phasing, an elegant and challenging technique for difficult structures. The H280, A312, N328, E329, H391, Y393, and E448 are its conserved amino acid residues which are present in its active site. By site directed mutagenesis, the role of each residue is being investigated in our ongoing experiments. Moreover, ligand protein complex crystals are underway to reveal ligand-protein interactions. Three structural truncated mutants were also created to see roles of different domains of the enzyme. The enzyme's catalytic domain with a linker of 46 amino acids at C-terminal plays a key role in its folding. Based on its structural novelties along with broad range of substrates' specificity and thermostability, it will be an important industrial enzyme for foods' processing, brewing, saccharification of plant biomass to produce biofuels, pharmaceutical and biotechnological applications.

Regulation of RAS function by active site autophosphorylation

Christian W. Johnson[1], Hyuk-Soo Seo[2], Elizabeth M. Terrell[3], Ezekiel A. Geffken[2], Jimit Lakhani[2], Kijun Song[2], Olesja Popow[1], Carla Mattos[4], Deborah K. Morrison[3], Sirano Dhe-Paganon[2], Kevin M. Haigis[1]

[1]Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, [2]Structural and Chemical Biology Center, Dana-Farber Cancer Institute, Boston, MA 02215, [3]Laboratory of Cell and Developmental Signaling, NCI-Frederick, Frederick, MD 21702, [4]Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA 02115

A unifying feature of the RAS superfamily is a functionally conserved cycle that small GTPases use to transition between active and inactive states. Here, we demonstrate that active site autophosphorylation of the oncogenic mutant K-RASA59T is an irreversible and intrinsic regulatory mechanism that alters nucleotide exchange, effector interaction, and downstream signaling. We used x-ray crystallography, enzymology, molecular dynamics and cell assays to explore the mechanism of autophosphorylation and its impact on RAS function. Our crystal structures of H-RASA59T support a zero-order and intramolecular mechanism of autophosphorylation that involves direct attack of threonine 59 on GTP. Further, we find that post-phosphorylation, intrinsic nucleotide exchange is enhanced and favored over GEF mediated nucleotide exchange. Our crystal structures of H-RASA59E and K-RASA59E support this difference in biochemical activity. The impact of active site reorganization extends beyond stabilization of the nucleotide bound state to RAS effector interactions and utilization. Using effector binding assays, as well as classic assays of K-RAS cell transformation, our data demonstrate that K-RASA59T and K-RASA59E do not depend on RAF kinases for cell transformation. Finally, we used structural analysis and molecular dynamics to show that other small GTPases with a conserved threonine or serine at residue 59 likely undergo autophosphorylation. Many of these small GTPases have similar biochemical features to K-RASA59T, including overall stabilization of the GTP versus GDP bound state. We speculate that autophosphorylation is an unexplored means of regulation for these atypical small GTPases.

Neutron and X-ray diffraction produces accurate models of SARS-CoV-2 main protease for structure-guided drug design of antivirals

Daniel W. Kneller[1], Stephanie Galanie[1], Gwyndalyn Phillips[1], Leighton Coates[1], Andrey Kovalevsky[1]

[1]Oak Ridge National Laboratory, Oak Ridge, TN 37830

COVID-19 from SARS-CoV-2 infection brought about a worldwide health and economic catastrophe. The viral main protease (M^{pro}) is a cysteine protease essential for SARS-CoV-2 replication and thus an attractive target for small-molecule inhibitors. Structure-guided drug-design strategies historically base atomic scale understanding of enzymes from cryogenic synchrotron X-ray diffraction. Where conventional protein X-ray crystallography studies are hindered by cryo-artifacts and the inability to measure protonation states, neutron crystallography provides an ideal technique to determine protonation states of ionizable residues at near-physiological temperatures. This series of studies summarizes the call-to-arms journey of generating large-volume protein crystals of M^{pro} required for neutron diffraction experiments.
Pre-requisite X-ray structures produced rapid insights for drug-design and clinical inhibitor repurposing in the early days of the pandemic. The catalytic cysteine can be trapped in the rare peroxysulfenic acid oxidation state at physiological pH while surface cysteines remain reduced signifying high reactivity of the catalytic cysteine. Diffraction data collected at room temperature from ligand-free M^{pro} crystals established an inherent structural plasticity in the active site proposing room-temperature models are a relevant template for in silico docking studies. Clinical HCV protease inhibitors narlaprevir, boceprevir, and telaprevir are low-micromolar inhibitors of SARS-CoV-2 M^{pro} and potential lead compounds for optimization. Structural comparisons indicate significant malleability of active site to enable induced fit of different inhibitor moieties.
The SARS-CoV-2 M^{pro} neutron crystal structure was determined, providing direct observation of protonation states in an active site in any cysteine protease for the first time. The catalytic Cys-His dyad of M^{pro} exists in the reactive zwitterionic state at rest, with charged thiolate and doubly protonated imidazole side chains. A follow-up neutron structure of a covalent inhibitor complex reveals the net +1 positive charge of the active site is maintained by rearrangements of protonation states. Neutron crystallography of M^{pro} showcases the clear requirement of accurate experimental models for *in silico* and drug design applications.
Finally, a lead molecule identified by multimillion compound supercomputer docking was experimentally validated. All structures determined were immediately shared with the scientific community through the Protein Data Bank allowing for real-time discovery of M^{pro} inhibitors.

Different ATP binding states of the essential AAA (ATPases Associated with various Activities)-ATPase Rix7 facilitate substrate translocation in ribosome biogenesis

Seda Kocaman[1], Yu-Hua Lo[1], Venkata Dandey[1], Jason Williams[1], Mario Borgnia[1], Robin E. Stanley[1]

[1]National Institute of Environmental Health Sciences

Ribosome biogenesis is a complex pathway which orchestrates the spatial and temporal regulation of hundreds of assembly factors. Three AAA-ATPases are essential for ribosome production in eukaryotes. By hydrolyzing ATP these AAA-ATPases create a mechanical force to drive the release of ribosomal assembly factors from maturing ribosome particles. Rix7 is the earliest acting AAA-ATPase in the ribosome biogenesis pathway and is responsible for driving the release of the assembly factor Nsa1 from pre-ribosomes, but the mechanism of substrate release is poorly understood. Moreover, mutations within Rix7 have been associated with mental illness, prostate cancer and a motor neuron disease further highlighting the need to understand how this AAA-ATPase works. We study the substrate translocation mechanism of Rix7 via cryo-EM to understand how it releases assembly factors from the pre-ribosomal complex to ensure ribosome maturation. Rix7 contains tandem ATPase domains named as D1 and D2, which both form hexameric rings. Our earlier cryo-EM structure revealed that Rix7 assembles into a D1-D2 stacked hexamer where the six subunits arrange in a staircase like arrangement with a central channel through which substrates pass. We determined cryo-EM structures of Rix7 with a substrate bound in different ATP occupied states at 4.3 and 4.7Å resolution. The first structure contains a total of 11 ATP molecules with a single D2-AAA domain not bound to ATP. ATP hydrolysis in this D2 domain is the power source for processive substrate translocation. However, molecular details of how the substrate is handed over to the next subunit within the hexamer was unclear. Recently, we identified an 8 ATP bound state of Rix7 where two subunits in both the D1 and D2 domains lack ATP and are not in contact with the substrate. We observed conformational changes in these two subunits in the 8 ATP state compared to the 11 ATP state while rest of the subunits remain unchanged. These structures suggest that different ATP states regulate the contact of the individual Rix7 subunits with the substrate in the central channel. ATP hydrolysis results in conformational changes that facilitate the processive substrate translocation which is essential for driving ribosome maturation.

Structural insights into dimeric surface hydrophobics as a functional regulator of essential CYP121A1 in Mycobacterium tuberculosis

Amit Kumar[1], David E. Fernando[1]

[1]Department of Biochemistry, JSMBS, State University of New York, Buffalo, NY-14203

Capturing the novel functional dimeric interface of cytochrome P450 121A1 from Mycobacterium tuberculosis (Mtb) is the main focus of this work. CYP121A1 mediates a phenol coupling reaction of the tyrosine dipeptide cyclo-L-Tyr-L-Tyr (cYY) in Mtb, the causative pathogenic bacterium of tuberculosis and one of the leading causes of death by infection world-wide. Out of twenty cytochrome P450 enzymes encoded by the Mtb genome, cytochrome P450 121A1 (CYP121A1), an essential gene, remains a target of drug design efforts. We capture the solution dimeric interface of cytochrome P450 121A1 from Mtb by X-ray crystallography at 3.8 Å resolution in this work. This investigation showed that CYP121A1 dimers form via intermolecular contacts on the distal surface and are mediated by a network of solvent-exposed hydrophobic residues. Dimerization of CYP121A1 contributes to specific substrate binding in vitro, with 3/4th loss of function when dimers are disrupted. Using the crystallographic evidence and protein docking, we developed a functional CYP121A1 dimer working model. None of the previously reported CYP121A1 structures reflect a distal-to-distal orientation in the crystal lattice. The results obtained suggest that participation of a new homodimer interface in substrate selectivity represents a novel paradigm of substrate binding in CYPs, while also providing crucial mechanistic insight regarding a relevant drug target in the development of novel anti-tuberculosis agents.

Structure-function characterization of the Cj0843c lytic transglycosylase of Campylobacter jejuni

Vijay Kumar[1], Snigdha A. Mature[1], Mijoon Lee[2], Jacob Boorman[1], Ximin Zeng[3], Jun Lin[3], Dusan Hesek[2], Elena Lastochkin[2], Shahriar Mobashery[2], Focco van den Akker[1]

[1]Department of Biochemistry, Case Western Reserve University, Cleveland OH 44106, USA, [2]Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame IN 46556, USA, [3]Institute of Agriculture, University of Tennessee, Knoxville TN 37996, USA

The soluble lytic transglycosylase Cj0843c from Campylobacter jejuni breaks down cell-wall peptidoglycan (PG). Its non-hydrolytic activity sustains cell-wall remodeling and repair. We report herein our structure-function studies probing the substrate preferences and recognition by this enzyme. Our studies show that Cj0843c exhibits both exo- and endolytic activities and forms the N-acetyl-1,6-anhydromuramyl (anhMurNAc) peptidoglycan termini, the typical transformation catalyzed by lytic transglycosylase. Cj0843c shows a trend toward a preference of substrates with anhMurNAc ends and those with peptide stems. Mutagenesis revealed that the catalytic E390 is critical for activity. In addition, mutagenesis showed that R388 and K505, located in the positively charged pocket near E390, also serve important roles. Mutation of R326, on the opposite side of this positively charged pocket, enhanced activity. Our data point to different roles for positively charged residues in this pocket for productive binding of the predominantly negatively charged PG. We also show by X-ray crystallography and by MD simulations that the active site of Cj0843c is still capable of binding GlcNAc containing di- and trisaccharides without MurNAc moieties, without peptide stems, and without the anhMurNAc ends.

Molecular basis for favorable receptor orientation reveals spurious signalling avoidance mechanism

Swetha Lankipalli[1, 2], Udupi A. Ramagopal[1]

[1]poornaprajna institute of scientific research, [2]Manipal Academy of Higher Education

Even after decades of research, a comprehensive mechanism that elucidates the underpinnings of signaling through the cell membrane is still elusive. Here, we address a simple question – “how does the ectodomain of a membrane-associated protein consisting of multiple domains and connected by flexible linkers stand ‘upright’ on the membrane?”. Our analysis based on the available structural and functional data, looking for a pattern of association of these molecules in the crystal structures and with the concept that ‘random things seldom repeat’ lead to a surprisingly interesting and consistent observation that the weak cis-interaction mediated symmetric oligomerization and/or supramolecular organization of signaling molecules not only support their ‘upright’ orientation but often bury their ligand-binding surface to avoid spurious signaling. With CD4, pMHCII, CD2 and TNFR1 as examples, we show that the observed cis-association of molecules also correlate well with their functional role. Further, our analysis reconciles the long-standing controversies related to these molecules and appears to be generic enough to be applied to other signaling molecules.

Beyond History and “on a Roll”: The List of the most Well-Studied Human Protein Structures and Overall Trends in the Protein Data Bank

Zhen-lu Li[1], Matthias Buck[1]

[1]Case Western Reserve University, Cleveland, OH 44106

Here, we present a summary of our recent study of protein in the Protein Data Bank, the PDB [Protein Science DOI: 10.1002/pro.4038]. 7,077 out of the 20,394 distinct (canonical) human proteins have at least one PDB entry as of January 20, 2021. About 38% of the 7,077 human proteins have almost their entire amino acid sequence solved, while 36% of the 7,077 human proteins contain less than 50% of their full polypeptide chains in their structures. Remarkably, the200 human proteins with the most entries (2.8% of 7,077) have 20,195 cumulative PDB entries,counting for around 40% of total human PDB entries. We list the 273 top human protein structures (of which 100 are membrane proteins) based on the number of their PDB entries. Many of these soluble and membrane proteins are all well-known (with first structures being deposited 15~25 years ago). This set of proteins represent past trends as well as current status for protein structural biology. We briefly discuss the relationship which some of the prominent protein structures have with protein research as a whole and mention their relevance to human diseases. Knowledge of the listed “most highly represented proteins” may be informative to senior scientists but also inspire researchers who are new to protein science, as the physical-biochemical/functional/biomedical properties of these proteins are relatively well studied. It may also encourage/guide the exploration of a broader field of lesser or not yet represented proteins, some of which may be distant structural and/or functional homologue to a member of this list.

Identification of functionally important dynamically rigid segments in intrinsically disordered regions of proteins by solution NMR spectroscopy

Snigdha Maiti[1], Bidisha Acharya[1], Soumya De[1]

[1]School of Bioscience, IIT Kharagpur, India

Amino acid sequences with intrinsically disordered regions (IDRs) are ubiquitously present in eukaryotic transcription factors (TFs). Although their biological significance is well appreciated, the underlying mechanism explaining how these disordered regions interact with DNA and proteins, recruit cofactors, and thus regulate transcription, is still obscure. Disordered sequences are known to interact with other macromolecules through short linear motifs (SLiMs). It is postulated that ~10^5 SLiMs are present in nature, however only ~3542 SLiMs have been discovered so far. Due to low sequence conservation, identification of new SLiMs in IDRs has remained elusive. Using NMR spectroscopy, we show that SLiMs are rigid segments within flexible IDRs. We further show that these rigid segments have relatively low conformational entropy compared to the rest amino acids in IDRs. Thus, upon binding, the IDRs undergo reduced loss in conformational entropy resulting in highly specific interactions. Overall, we present an NMR method for the determination of rigid segments, which can act as SLiMs, and also provide a physicochemical basis of their function (Maiti *et al.* 2019). Specifically, we have characterized the fast time-scale (ps-ns) dynamics of two paralogous HOX transcription factors from Drosophila, Sex combs reduced (SCR) and Deformed (DFD). We found that the flexibility of residues in the disorder regions vary significantly. IDRs in both proteins have short stretches (~5-7) of relatively rigid amino acids. NMR titration experiments showed that one of the identified rigid segments specifically interacts with a partner transcription factor Extradenticle (EXD), which is known to modulate the function of HOX transcription factors. Thus, the rigid segments of IDRs are functionally important SLiMs. We have designed several mutations in the IDR of SCR and show by NMR dynamics experiments that these mutants have altered flexibility. Interestingly, no change in the global structural properties are observed for these mutants as determined by analytical size-exclusion chromatography (SEC) and analytical ultra-centrifugation (AUC). EMSA-based DNA binding experiments reveal altered affinity for these mutant SCRs. Overall, these studies further establish that the rigid and flexible segments in IDRs play distinct but important functional roles.

Cryo-EM study of the nsp binding and cleavage by the SARS-CoV-2 main protease

MANJU NARWAL[1], JEAN P. ARMACHE[1], KATSUHIKO MURAKAMI[1]

[1]Pennsylvania State University, State College, PA 16801

Members of the Coronaviridae family are viruses with positive sense single-stranded RNA genomes capable of causing range of respiratory distress among human and animal hosts. Currently, SARS-CoV-2 has become a global pandemic affecting population throughout the globe. The SARS-CoV-2 main protease Mpro, which cleaves the viral non-structural polyprotein (nsp) for assembling a replication-transcription complex (RTC), is one of excellent antiviral targets to prevent SARS-CoV-2 replication.

Since the release of first SARS-CoV-2 Mpro crystal structure in PDB (Protein Data Bank) on Feb 5, 2020, researchers around the world have submitted more than 280 X-ray crystal structures of Mpro with substrate and inhibitors. Although X-ray crystallographic study can reveal atomic detail of the interaction between Mpro and substrate/inhibitors, however, crystallization of Mpro, which is a relatively small and globular enzyme, unavoidably introduces crystal packing constraints, making this method difficult to reveal physiological conformation of the active site and the substrate binding pocket of enzyme. Furthermore, the structural basis of nsp recognition by Mpro has not yet been elucidated because the crystallization of Mpro and nsp complex was not feasible due to flexibility of nsp. Significant technological advances in the cryo-electron microscopy (cryo-EM) allow the structural study of Mpro in solution.

We report here the cryo-EM structures of the SARS-CoV-2 Mpro enzyme in apo-form and in complex with nsp710 polyprotein at atomic resolutions. The structures show that Mpro forms a dimer with 2-fold symmetry and both active sites can bind a substate peptide simultaneously. The mechanism of substate binding and cleavage by the active site of Mpro will be discussed. Comparison with X-ray crystal structures of Mpro (apo- and substrate bound forms) suggest that the cryo-EM study can reveal structural information around the active site of enzyme at near physiological condition suitable for performing in-silico drug screening and design experiments. The cryo-EM study of Mpro provides a better template for developing drugs effectively inhibiting virual replication and contributes towards therapeutic treatment of SARS CoV-2.

VoroContacts: a tool for the analysis of intera-tomic contacts in macromolecule structures

Kliment Olechnovic[1], Ceslovas Venclovas[1]

[1]Vilnius University Life Sciences Center

VoroContacts is a versatile tool for computing and analyzing contact surface areas (CSAs) and solvent accessible surface areas (SASAs) for 3D structures of proteins, nucleic acids and their complexes at the atomic resolution. CSAs and SASAs are derived using Voronoi tessellation of 3D structure, represented as a collection of atomic balls. VoroContacs web server features a highly configurable query interface, which enables on-the-fly analysis of contacts for selected set of atoms and allows filtering interatomic contacts by their type, surface areas, distance between contacting atoms and sequence separation between contacting residues. The VoroContacs functionality is also implemented as part of the standalone Voronota package, enabling batch processing. VoroContacts is available at bioinformatics.ibt.lt/wtsam/vorocontacts.

Understanding gene regulation driven by the Salmonella RcsB response regulator

Juanjo Huesa[1], Joaquín Giner-Lamia[2], M. Graciela Pucciarelli[5], Francisco Paredes-Martínez[1], Francisco García-del Portillo[3], Alberto Marina[4], Casino Patricia[1]

[1]Universitat de València-Instituto Biotecmed, Valencia, Spain, [2]Departamento de Biotecnología y Biología Vegetal, Universidad Politécnica de Madrid, Spain, [3]Laboratorio de Patógenos Bacterianos Intracelulares. Centro Nacional de Biotecnología CNB-CSIC,Spain, [4]Department of Genomic and Proteomic, Instituto de Biomedicina de Valencia IBV-CSIC, Spain, [5]Departmento de Biología Molecular. Centro de Biología Molecular Severo Ochoa CBMSO-CSIC, Spain

Phosphorelay systems are a complex version of two-component systems (TCS), the main signal transduction systems present in bacteria developed to detect a diverse type of signals to produce adaptive responses (Stock AM et al. Annu Rev Biochem, 2000). Phosphorelay systems are based on phosphoryl transfer between His-Asp residues between a hybrid histidine kinase (hyHK), a phosphotransferase and a response regulator (RR), a three-step reaction in order to provide regulatory complexity and versatility (Laub MT et al. Annu Rev Genet, 2007). Phosphorelay systems, together with TCS, are absent in mammals being potential targets for the development of new antimicrobials. To study phosphorelay systems in bacteria, we focus on the Rcs system in Salmonella, which is conserved in Enterobacteriaceae, that comprises the RcsCDB phosphorelay system coupled to two membrane proteins RcsF and IgaA (Wall EN et al, Annu Rev Microbiol 2018). The system is activated when RcsF detects envelope stress and interacts with IgaA, thus, releasing the inhibition over the RcsCDB system to induce the phosphoryl-transfer which will result in phosphorylation of RcsB. In this way, phosphorylated RcsB acts as a transcriptional factor regulating many genes, some of them related with exopolysaccharide production and flagella inhibition which produces non-motile cells with a mucoidy phenotype (Cano DA et al Genetics 2002). Our studies in the last few years have allowed us to solve the structure of RcsB in its phosphorylated conformation competent to bind DNA as well as to observe an additional conformation able to bind double stranded DNA of a yet unknown biological function (Casino P et al NAR 2018). More recently, we have solved the structure of phosphorylated RcsB bound to two promoter regions belonging to the small RNA rprA and the flagelar gene flhDC which has confirmed our previous structural and functional results (Huesa J et al NAR 2021). Moreover, the structures have allowed us to analyse the DNA sequence recognition of RcsB and propose a new recognition box that has been mapped in the Salmonella genome detecting 190 sites which could be assigned to ~200 genes.

Refactoring the B-factor: intuitively extracting structural dynamics from macromolecular disorder

Nicholas M. Pearce[1, 2], Piet Gros[2]

[1]Vrije Universiteit Amsterdam, The Netherlands, [2]Universiteit Utrecht, The Netherlands

Displacement parameters (B-factors) play a crucial role in macromolecular structure determination, yet are rarely used for biological interpretation. This is somewhat egregious, since they account for the local flexibility of individual protein states/conformations. We have developed a new approach for dividing the disorder information in a macromolecular model into a hierarchical series of components on different length-scales, which reveals the components of the atomic disorder that result from molecular disorder, domain disorder, or local atomic disorder. This makes both molecular and atomic disorder intuitively understandable in terms of likely domain motions and internal atomic motions. Upon application of the method to the study of SARS-COV-2 structures, we investigate and visualise the flexibility in the binding site of the SARS-COV-2 main protease, and in both the extended and non-extended receptor-binding domains of the SARS-COV-2 spike glycoprotein.

Seizing the opportunity: using NMR spectroscopy to elucidate the binding site of sub-family IV bromodomains.

Margaret Phillips

[1]Albany College of Pharmacy and Health Sciences, Vermont, USA

Solution-state NMR spectroscopy has been routinely used to identify and characterize various protein-protein interactions in-vitro by mapping chemical shift perturbations observed in the backbone amide resonances of labeled protein in the presence of its interacting partner. Since chemical shifts are sensitive to local environments of the residue in the protein backbone, using 2D NHSQC NMR experiments, we monitored labeled bromodomains - BRPF1 and ATAD2B in solution in the absence and the presence of increasing concentrations of various acetylated histone tail peptide. These titration experiments helped describe the acetyllysine histone tail binding pocket on the bromodomains and highlighted key residues involved in mono- versus di-acetyllysine histone tail recognition. The NMR experiments proved useful in identifying crucial residues that differ between these bromodomains despite having high structural similarity and belonging to the same sub-family.

Function without structure: FhuF - a ghost ferric-siderophore reductase from Escherichia coli

Inês B. Trindade[1], Guillem Hernandez[1], Filipe Rollo[1], Estelle Lebègue[2], Frédéric Barrière[3], João B. Vicente[1], Elin Moe[1], Pedro M. Matias[1, 4], Tiago Cordeiro[1], Mario Piccioli[5, 6], Ricardo O. Louro[1]

[1]Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Portugal, [2]Université de Nantes, CNRS, CEISAM UMR 6230, F-44000 Nantes, France, [3]Univ Rennes, CNRS, Institut des Sciences Chimiques de Rennes – UMR 6226, F-35000, Rennes, France, [4]Instituto de Biologia Experimental e Tecnológica (iBET), Apartado 12, 2780-901 Oeiras, Portugal, [5]Magnetic Resonance Center (CERM) and Department of Chemistry, University of Florence, Italy, [6]Consorzio Interuniversitario Risonanze Magnetiche Di Metallo Proteine, Sesto Fiorentino, Italy

Iron is a key element for virtually all forms of life. Yet, in the present oxygen-rich atmosphere, iron precipitates in its ferric form which is not readily bioavailable. To circumvent this problem, microorganisms secrete siderophores, small molecules with high affinity for ferric iron. However, once inside the cell, iron release from these complexes does not occur spontaneously. Instead, this process is mediated by specific proteins that can be grouped in two families: the siderophore-interacting protein (SIP) family that have a flavin cofactor and the ferric reductase family (FSR) that contain a 2Fe-2S cluster. Escherichia coli K-12 can use different siderophores including hydroxamates and catecholates. In the cytoplasm, two distinct ferric-siderophore reductases have been isolated. One is YqjH from the SIP subfamily, which is able to catalyze the release of iron from Fe(III)-triscatecholates. The other is FhuF, from the FSR subfamily that showed specificity for hydroxamate-type siderophores, since iron removal from these is significantly reduced in a FhuF-defective mutant. In Gram-negative bacteria, FhuF is the only FSR ever isolated given the instability of pure FSR proteins. It contains an atypical 2Fe-2S with the motif Cys-Cys-X10-Cys-X2-Cys and although its function as a ferric-siderophore reductase is established, there is presently no knowledge regarding its structure and how the atypical 2Fe-2S cluster operates to mediate ferric-siderophore reduction. In this work, a combination of paramagnetic NMR spectroscopy, electrochemistry, circular dichroism (CD) and small-angle X-ray scattering (SAXS) was used to investigate the structural and functional characteristics of this protein. Unlike the previously characterized SIPs, FhuF reveals a structural relationship with siderophore synthesizing enzymes and unprecedented structural and electronic properties of the 2Fe-2S cluster. Furthermore, the two families of enzymes promote siderophore release in the cells using the same mechanistic principle of proton-coupled electron transfer. This mechanism expands their range of redox activity and allows iron release from a broader range of ferric-siderophore substrates [1].

[1] Trindade I.B., Hernandez G, Lebègue E, et al (2021) Conjuring up a ghost: structural and functional characterization of FhuF , a ferric siderophore reductase from E . coli. JBIC in press.

Recent structural findings on the interaction between human α-thrombin and oligonucleotide aptamers

Romualdo Troisi[1], Filomena Sica[1]

[1]Department of Chemical Sciences, University of Naples Federico II, Naples IT 80126

Oligonucleotide aptamers are short DNA or RNA sequences that bind to their target with high affinity and specificity. They are extensively used to modulate the function of most of the factors involved in the coagulation pathway. In this context, a particular case is that of the anticoagulant aptamers able to recognize the human α-thrombin (thrombin) by binding the two electropositive regions, known as exosites I and II, lying on the opposite poles of its surface. One of the most extensively studied anticoagulant aptamer is TBA (also known as HD1), a 15-mer oligonucleotide recognizing the exosite I by adopting an antiparallel G-quadruplex structure. Recently, the interest of researchers moved on a new class of oligonucleotides in which the addition of a duplex-forming sequences to a G-quadruplex module results in an improvement of the binding properties. Among them, NU172 possesses a high anticoagulant activity and is currently the only anti-thrombin aptamer evaluated in Phase II of clinical trials (*ClinicalTrials.gov* identifier *NCT00808964*) for anticoagulation in heart disease treatments by *ARCA Biopharma, Inc*. Despite its excellent anticoagulant properties, researchers are investigating several strategies to overcome some long-standing limitations of both TBA and NU172, as the short circulating half-life *in vivo*. Another interesting duplex/quadruplex aptamer is HD22_27mer that recognizes thrombin exosite II with very high affinity but without exerting a potent antithrombotic activity.
In the last years, great attention has been paid to the study of the effects of the simultaneous binding of two aptamers on the two exosites of thrombin. In particular, it was found that the aptamer binding to exosite II increases the binding affinity of another aptamer to the exosite I, and *vice versa*, consequently influencing the coagulant activity of the protein.
During the symposium we will describe the most recent crystallographic results achieved on aptamer-thrombin complexes. In particular, the peculiar structural features of NU172 in the one-to-one binary complex with the thrombin and in the ternary complex in which exosite II interacts with HD22_27mer will be discussed. The structural details of thrombin complexes involving new modified aptamers will be also illustrated.

A special guest to PDB50: Bacteroides thetaiotaomicron, the human gut microbe that owns the required protein machinery to degrade complex pectic polysaccharides for us

Filipa Trovão[1], Viviana G. Correia[1], Frederico Lourenço[1], Joana L. A. Brás[2], Carlos M. G. A. Fontes[2, 3], Ana Luísa Carvalho[1], Angelina S. Palma[1], Benedita A. Pinheiro[1]

[1]UCIBIO-NOVA, Department of Chemistry, NOVA School of Science and Technology, Portugal, [2]NYTech - genes & enzymes, Campus do Lumiar, Lisbon, Portugal, [3]CIISA, Faculdade de Medicina Veterinária, Universidade de Lisboa, Portugal

The human gut microbiota houses a densely microbial community of commensal and pathogenic bacteria, with a broad capacity to metabolize dietary and host-derived glycans. The variation of the carbohydrate composition in the human diet is known to shape this community, promoting homeostasis or microbial imbalance (dysbiosis) associated with mechanisms of pathology and ultimately, disease [1]. Prominent microbiota strains have multiple enzyme-coding genes in clusters, designated polysaccharide-utilization loci (PULs). The PUL cell-surface expressed proteins include carbohydrate active enzymes with ancillary carbohydrate-binding modules (CBMs). *Bacteroides thetaiotaomicron* (BT) is one of the most common bacteria found in the human gut that can degrade both host and dietary carbohydrates. BT0996 is a modular protein, involved in the early stages of the rhamnogalacturonan II (RG-II) degradation, the most complex plant polysaccharide known. BT0996 protein is composed of an L-arabinofuranosidase and a D-glucuronidase that target chain A and chain B oligosaccharide decorations of RG-II [2]. Besides that, BT0996 has three putative CBMs which role in the RG-II degradation is unknown. Here, we present the characterization of a newly identified CBM located at the C-terminus of BT0996 modular protein – BT0996-C CBM. The high-resolution crystal structure shows an overall β-sandwich fold with a structural calcium and homology to members from CBM CAZY families 6 and 35. However, the putative binding site shows substantial differences in the possible interacting-residues, when compared with CBMs of similar fold. Also, the presence of several charged residues in the BT0996-C surface forms a broad positive patch that comprises the putative binding site. These features may suggest a different binding mode not yet observed in its homologues. Carbohydrate microarray analysis and microscale thermophoresis showed the ability of binding by BT0996-C to the polygalacturonic acid (PGA), revealing a high contribution of electrostatic interactions for the recognition of the anionic polysaccharide. The results support the hypothesis that BT0996-C has evolved to bind to the complex structure of RG-II through a different mechanism than characterised CBMs, approaching BT0996 catalytic modules to their substrate, and potentiating its enzymatic degradation.
1. Martens E. et al., PLoS Biol, 2011, 9, e1001221; 2. Ndeh D. et al., Nature, 2017, 544,65-70

Structural Characterization of the Genetic Switch in the Temperate Bacteriophage TP901-1

Anders K. Varming[1], Kim K. Rasmussen[1], Kristian H. Frandsen[1], Margit Pedersen[1], Karin Hammer[2], Mogens Kilstrup[2], Malene R. Jensen[3], Leila L. Leggio[1]

[1]University of Copenhagen, DK-2100, Denmark, [2]Technical University of Denmark, DK-2800, Denmark, [3]CNRS, Institut de Biologie Structurale Jean-Pierre Ebel, FR-38044, France

Antibiotic resistance is a major threat to the global society, and it is developing rapidly. One of the main reasons for this phenomenon is the ability of temperate bacteriophages (TPs) to perform horizontal gene transfer between bacterial populations. This ability comes from the fact that TPs can implement two different life cycles: *lytic* and *lysogenic*. In the lytic cycle, the TP takes over the machinery of the host and the viral genome is replicated, transcribed and translated producing new TPs, leading to cell lysis and release of phage progeny. In the lysogenic cycle, the viral genome is silenced and incorporated into the host chromosome existing as a prophage. Switching from the lysogenic to the lytic cycle can lead to transfer of host DNA - potentially including resistance genes. Changing between these two alternative life cycles is controlled by a *bistable genetic switch*. Here, we present the work on the genetic switch of the lactococcal phage TP901-1. The genetic switch consists of two divergently oriented promoters, P_R (lysogenic) and P_L (lytic), three operator sites and two transcription factors, the CI repressor and MOR anti-repressor. CI is responsible for initiating the lysogenic cycle by binding the operator sites and repressing the promoters to a varying degree [1]. Switching to the lytic cycle is dependent on the presence of MOR. MOR forms a heterodimer complex with CI, which disrupts the repression of P_L while maintaining repression of P_R. We have solved the structure of the N-terminal domain (alone and in complex with either MOR or DNA) and part of the C-terminal domain of CI using X-ray crystallography and NMR and we have studied the interaction between CI and the operator sites using SAXS [2–5]. All our knowledge gained is now being utilized to study a homologous TP found in the pathogenic *S. aureus*.

1. Pedersen, M. *et al.* J.Mol.Biol. 376, 983–996 (2008)
2. Frandsen, K. H. *et al*. Biochemistry 52, 6892–6904 (2013)
3. Rasmussen, K. K. *et al*. Sci.Rep. 6, 29574 (2016)
4. Rasmussen, K. K. *et al.* FEBS Lett. 592, 1738–1750 (2018)
5. Rasmussen, K. K. *et al.* PNAS. 117, 20576–20585 (2020)

Construction of a multi-epitope vaccine against Mycoplasma pneumoniae applying Immunoinformatic approach.

Thaís Cristina Vilela Rodrigues[1], Arun Kumar Jaiswal[1], Helioswilton Sales-Campos[3], Rodrigo Bentes Kato[1], Vasco Ariston de Carvalho Azevedo[1], Sandeep Tiwari[1], Siomar de Castro Soares[2]

[1]Institute of Biological Sciences, Federal University of Minas Gerais, [2]Department of Immunology, Microbiology and Parasitology, Federal University of Triângulo Mineiro, [3]Institute of Tropical Pathology and Public Health, Federal University of Goias

Pneumonia is a serious health problem with global effects, being the death cause of over one million people annually. Among the main microorganisms responsible for pneumonia, *Mycoplasma pneumoniae* is one of the most common, with a significant increase in the last years along with the development of resistance to some antibiotics. Vaccines are fundamental in disease prevention besides to considerably avoid the need for health services and funding resources. In this way, the proposal of the present study is to construct through immunoinformatic tools, a multi-epitope vaccine against *M. pneumoniae*. Multi-epitope vaccines are constituted by epitopes properly selected to induce targeted immune responses and avoid adverse reactions. In this work, the core proteins from 88 genomes of *M. pneumoniae*, were previously determined through reverse vaccinology, then the search for MHC-I, MHC-II, and B epitopes were performed as well as the check for overlapping epitopes, capable to induce both humoral and cellular responses. Those epitopes were filtered according to their immunogenicity, population coverage, among others. The epitopes with the best features were joined with classical peptide linkers and the heat-labile enterotoxin from *Escherichia coli* as an adjuvant, then, the structure of the vaccine was predicted. The vaccine was considered physically stable, non-toxic, non-allergen, not significantly similar to human proteome and with appropriate antigenic and immunogenic properties. The molecular docking of the vaccine with the Toll-Like Receptor 2 structure, recovered from Protein Data Bank (PDB), was performed and the dynamic simulation was executed to ensure the affinity and stability between this complex. *In silico*a clonagem foi testada em um vetor de expressão com resultados positivos. Além disso, a simulação imune para eficácia da vacina também foi aplicada de diversas maneiras com resultados promissores. Por meio de abordagens de imunoinformática, construímos uma vacina candidata multi-epítopo eficaz que, com testes adicionais, poderia contribuir para a prevenção da pneumonia em grande escala. Além disso, o estudo auxilia no melhor entendimento dos mecanismos imunológicos relacionados às infecções *por M. pneumoniae* e sua interação com o hospedeiro.

Interactive, Flexible Docking to the SARS-CoV-2 Main Protease in Virtual Reality

Rebecca K. Walters[1], Helen M. Deeks[1], Jonathan Barnoud[1], David R. Glowacki[1], Adrian J. Mulholland[1]

[1]Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol, BS8

The main protease (Mpro) of the SARS-CoV-2 virus is a promising target for drug development against COVID-19. Here, we show that interactive molecular dynamics in virtual reality (iMD-VR) is a useful and effective tool for producing complexes of Mpro with a ligand. iMD-VR provides an immersive environment where users can interact with MD simulations, building protein complexes on-the-fly in a fully flexible and physically rigorous way. We have recently shown that iMD-VR is an effective tool for interactive, flexible docking of small molecules to their protein targets.[1]

In this work, we apply similar docking approaches to an Mpro inhibitor and oligopeptide substrate, starting from experimentally determined crystal structures. We test docking to different conformations of Mpro, including an apo (uncomplexed) structure. We also test the use of different docking protocols, i.e. with and without restraints on the protease backbone. As in our previous work, our docked structures are tested with MD (and found to be stable) and then compared to related experimentally observed crystal structures.

We find that, unsurprisingly, iMD-VR docking performs better on cognate protein models i.e., docking to an observed complex, than for docking to apo forms of the protein. However, the results with apo and other protein structures are also in good agreement with related crystal structures, having successfully reformed many key electrostatic interactions. This is because of the flexibility and dynamic response allowed by the physically rigorous, atomically detailed simulation approach of iMD-VR.

Our models are freely available and the software framework we use, Narupa, is open source and uses commodity VR hardware. These should be widely useful in drug development, in education applications, e.g., on viral enzyme structure and function, and in scientific communication more generally.

References: [1] H. M. Deeks, R. K. Walters, S. R. Hare, M. B. O'Connor, D. R. Glowacki, and A. J. Mulholland, PLoS ONE, 2020, DOI: 10.1371/journal.pone.0228461

Half way to hypusine - structural analysis of human deoxyhypusine synthase

Elzbieta Wator[1], Piotr Wilk[1], Przemyslaw Grudnik[1]

[1]Jagiellonian University, Maloposlka Centre of Biotechnology, Krakow, Poland

Deoxyhypusine synthase (DHS) is a transferase catalysing the formation of deoxyhypusine, which is the first, rate-limiting step of unique post-translational modification: hypusination. DHS catalyzes the transfer of 4-aminobutyl moiety of spermidine to a specific lysine of eIF5A precursor in an NAD-dependent manner. This modification occurs exclusively on only one protein: eukaryotic initiation factor 5A (eIF5A) and it is essential for cell proliferation. Malfunctions of the hypusination pathway, including those caused by mutations within DHS encoding gene, are associated with such conditions as cancer or neurodegeneration.
The aim of the presented study was to investigate substrate specificity of the first step of hypusination using macromolecular crystallography as a main tool and additionally to assess the impact of newly recognized pathological mutations in DHS encoding gene on protein stability, activity and structure.
Human DHS wild type and its two mutants were expressed, purified and crystallization attempts were undertaken. Crystallization trials have led to six high-resolution crystal structures determination of DHS wt in apo form and in complexes with natural substrates. Additionally, 2 crystal structures of N173S DHS were determined allowing for detailed analysis and comparison. Based on crystal structures and activity tests it was shown that despite almost identical binding of spermidine, spermine and putrescine, probably only spermidine can serve as a proper substrate of deoxyhypusine formation. Furthermore, it was shown that against the previous studies, no conformational changes occur in DHS structure upon spermidine-binding. Therefore, it is likely that significant conformational change is needed for the further progress of reaction occurring as a result of eIF5A substrate recognition. Based on conducted analysis and comparison, the N173S variant was classified as a “structurally silent”, due to insignificant changes in its stability and structure. The nature of the second mutation leading to the deletion of two amino acids within NAD-binding motif led to significant destabilization of the protein and loss of its function, which is supported by stability and activity tests.
Availability of high-quality structural data will aid the design of novel DHS inhibitors for potential applications in cancer therapy and can significantly advance our understanding of newly recognized genetic DHS deficiency.

Hydration and Dehydration of GPCRs Play Critical Roles in Cellular Signal Transduction

Nipuna Weerasinghe[1], Steven D. Fried[1], Andrey V. Struts[1, 3], Suchithranga M. Perera[1], Michael F. Brown[1, 2]

[1]Department of Chemistry and Biochemistry, University of Arizona, Tucson, AZ 85721, USA, [2]Department of Physics, University of Arizona, Tucson, AZ 85721, USA, [3]Laboratory of Biomolecular NMR, St. Petersburg State University, St. Petersburg 199034, Russia

G-protein-coupled receptors (GPCRs) are a versatile family of membrane proteins that transmit signals across cell membranes and are targeted by more than 30% of pharmaceuticals currently on the market. Crystallography of GPCRs in various conformational states has provided critical breakthroughs to understanding the structural basis for their signaling processes. Moreover, free access to 3-D coordinates of these structures through the PDB has facilitated a deeper understanding of their biological functions. Spectroscopic experiments and computer simulations provide complementary dynamic aspects of GPCR-mediated signaling that cannot be obtained from static structures alone. Here, we used the archetypal GPCR rhodopsin to investigate how hydration is coupled to the protein's structural fluctuations during GPCR activation. We placed rhodopsin under varying levels of controlled dehydration using polyethylene glycol (PEG) solutes and investigated the activation equilibrium using UV-Visible spectroscopy. Our results show a large influx of 80–100 water molecules into rhodopsin during photoactivation, forming a partially disordered, solvent-swollen Meta-II active conformational state [1], a result supported by atomistic molecular dynamics simulations [2]. Effects of osmolytes on rhodopsin activation are size-dependent: large osmolytes back-shift the equilibrium to the inactive Meta-I state, while small osmolytes forward-shift the equilibrium to the Meta-II state. We attribute this size dependence to the degree of osmolyte penetration of the hydrated rhodopsin molecule. Large polymers behave similarly to ideal osmolytes and dehydrate rhodopsin, while smaller polymers wriggle into the protein interior and stabilize the open Meta-II conformation. Application of hydrostatic pressure likewise back shifts the metarhodopsin equilibrium, but for fundamentally different reasons. Integrating the two force-based methods with neutron scattering experiments indicates the active state of rhodopsin is more hydrated yet more locally collapsed [3]. The active GPCR undergoes larger-scale volumetric fluctuations and solvent coupling, which give rise to greater thermal volume. Our results necessitate a new understanding of GPCR activation in which the hydration level of the protein is paramount to governing conformational energy landscapes. [1] U. Chawla *et al.* (2021) *Angew. Chem. Int. Ed.* **60**, 2288–2295. [2] N. Leioatts *et al.* (2014) *Biochemistry* **53**, 376−385. [3] S.M.D.C Perera *et al.* (2018) *J. Phys. Chem. Lett.* **9**, 7064–7071.

Graph-theoretic representation of the topology of large protein complexes

Jan Niclas Wolf[1], Jörg Ackermann[1], Isra Nurhassen[1], Mariella Zunker[1], Ina Koch[1]

[1]Molecular Bioinformatics, Goethe-University, Frankfurt am Main, Germany

The Protein Topology Graph Library (PTGL) [1,2] translates three-dimensional protein structures to graphs. The structures are described on different levels of abstraction. Protein Graphs (PGs) describe the super-secondary structure level and Complex Graphs (CGs) the chain level. In PGs and CGs, each secondary structure element (SSE) and each protein chain is a vertex, respectively. Edges between the vertices denote spatial contacts based on atom-atom distances.
Our program reads a PDB file, either as legacy PDB file or as mmCIF, and a DSSP file. Then, it computes the contacts on different levels of abstraction and the result graphs. We provide the topology graphs as graphic files and as text files for further processing.
PGs describe the topology of one protein chain on SSE level. They can be used to quickly compare the SSE topology of structures and search the database for structural motifs. CGs describe the topology of all protein chains. The edges are weighted with the number of residue-residue contacts, thus describing the interface between the chains. We introduced normalized edge weights that account for the length of the participating protein chains.
We previously extended PTGL by structures of more than 62 chains or 99,999 atoms (large structures) [1]. CGs reduce the high number of atoms in three-dimensional space of such large structures to vertices and edges. CGs can be used, for example, to
- compare complexes across organisms,
- compare different structures of the same complex or
- explore MD simulation dynamics.
Differences can be observed on a topological level, i.e. the presence of edges, and in changes of the edge weights.
We investigated the well-known respiratory complex I as case study. We clustered the CG to find biological functional modules. We hypothesize that protein chains of higher normalized edge weight assemble earlier in the formation of the complex.
Concluding, PTGL uses graph theoretical methods to visualize and analyze protein structures on different levels of abstraction. Currently, PTGL provides in real time pre-computed topology graphs of 151,837 PDB entries, including 921 large structures.

[1] Wolf et al., 2020, Bioinformatics
[2] Koch & Schäfer, 2018, Curr. Opin. Struct. Biol.

Learning in the Lockdown: Free Online CryoEM Study Group

Geoffrey Woollard[1]

[1]University of Toronto, Toronto, Ontario, Canada

During the Great Plague of London, Isaac Newton had to work from home, too. Cambridge sent their students home, and he had an annus mirabilis (year of wonders) on his family estate. While we may not manage to have such insights into calculus, optics and gravity, as this physicist did, we can still make lemonade from lemons because learning the math and theory of cryoEM requires no access to the bench, and can be more fun to do with others!

In the spirit of solidarity during the pandemic, I launched a free online cryoEM study group in November 2020, and have received over 300 registrations from trainees, long term staff, and principal investigators. As one respondent put it “I am interested in being more than a button pusher”.

The learning style is largely a flipped classroom, with pre-lecture readings such as the wonderful free online courses already available provided beforehand, and online meetings spent answering questions, working out problems, and building intuition between data, equations, code, and experimental and data processing choices.

The audience is divided into two streams, from beginners to advanced methods developers. Beginner and intermediate practitioners often have a desire to go deeper into the fundamentals, and understand things under the hood. Some methods developers want to go deeper into the foundations, and interact with other people with strong math and physics backgrounds like themselves.

We have had 13 weekly meetings on topics like complex numbers, Fourier transforms, phase contrast imaging, Nyquist, expectation-maximization, and rotations. Future work includes guest lectures on 3D reconstruction and variational autoencoders, as well as learning from interactive coding notebooks that illustrate commonly encountered problems in data processing and how to diagnose them and troubleshoot with simple examples.

An ongoing question I have is: how much math and theory do cryoEM structural biologists need or want to know, either for solving structures or developing novel methods?

A syllabus, annotated bibliography, and online coding notebooks can be found at https://github.com/geoffwoollard/learn_cryoem_math

Biophysical characterization of SARS-CoV-2 proteins and polyproteins

Ruchi Yadav[1], Sanjay K. Dey[1], Valentine Courouble[2], Jerry J. Harrison[3], Jennifer Timm[1], Francesc X. Ruiz[1], Patrick Griffin[2], Eddy Arnold[1]

[1]Center for Advanced Biotechnology and Medicine, Rutgers University, Piscataway, NJ 08854, USA, [2]Department of Molecular Medicine, The Scripps Research Institute, Jupiter, FL 33458, USA, [3]Department of Chemistry, University of Ghana, Box LG56, Legon-Accra, Ghana

SARS-CoV-2, a member of the *Coronaviridae* family, is responsible for the current COVID-19 pandemic. The non-structural proteins (nsps), formed by nsp3 (PLpro) and nsp5 (Mpro) cleavage of the polyprotein 1ab, assemble to form the replication-transcription complex (RTC) and other viral machinery. A central enzyme of the RTC is nsp12 (RdRp), which is dependent on its co-factors nsp7 and nsp8 for optimal activity[1]. To date, no high-resolution structure of full-length SARS-CoV-2 nsp7:nsp8 complex has been published, with all existing PDBs having truncation (or missing electron density) of their termini. Using solution-based structural proteomic techniques, i.e., hydrogen-deuterium exchange mass spectrometry (HDX-MS) and crosslinking mass spectrometry (XL-MS), we have found that SARS-CoV-2 full-length nsp7 and nsp8 proteins may not assemble into a hexadecameric structure as implied by the SARS-CoV full-length nsp7:nsp8 crystal structure[2]. Instead, our results suggest that the nsp7:nsp8 heterotetramer can dissociate into a stable dimeric unit capable of binding to nsp12 without altering nsp7:nsp8 interactions[3].
Also inspired by novel findings from the HIV-1 polyprotein studies and insights into the maturation of HIV[4], we have initiated similar investigations on SARS-CoV-2 pp1ab. We have found that the nsp7-8 and nsp7-11 polyprotein intermediates are well-behaved and stable upon their expression and purification in bacteria. SAXS and HDX-MS studies of both polyprotein reveal that nsp7-8 is more folded and ordered while nsp7-11 is more flexible. XL-MS studies also reveal that nsp7-11 polyprotein has a high number of crosslinks between the nsp8 and nsp9 domains. From limited proteolysis experiments with Mpro and polyproteins, we observe that nsp7-8 polyprotein is the longest lived intermediate within nsp7-11 with the potential of being specifically targeted with small molecule binders that may inhibit viral maturation[5]. We are also pursuing cryo-EM and crystallographic studies of SARS-CoV-2 nsp7-8 and nsp7-11 polyproteins, which will provide further insights of their role in the viral cycle.

References:

1. Hillen, H.S., et al., Nature, 2020. 584: p. 154–156.
2. Zhai, Y., et al., Nat. Struct. Mol. Bio., 2005. 12: p. 980-986.
3. Courouble, V.V., et al., bioRxiv, 2021, doi: https://doi.org/10.1101/2021.03.06.434214.
4. Harrison, J.J, et al., (in preparation).
5. Dey, S.K., et al., (in preparation).

Cooperativity in the global motions of trimeric SARS-CoV-2 spike: Implications for viral entry

Arangasamy Yazhini[1], Das Swayam Prakash Sidhanta[1], Narayanaswamy Srinivasan[1]

[1]Indian Institute of Science, Bangalore, India -560012

Spike protein is a key player in mediating the primary step of the entry of SARS-CoV-2 virus, causing the ongoing pandemic of COVID-19 infection. Here, we studied the global motions of trimeric SARS-CoV-2 spike protein in distinct functional states to understand conformational states-specific flexibility profiles and the effect of ACE2 binding. We demonstrate using normal mode analysis, residue fluctuations analysis on molecular dynamics simulations trajectories and perturbation-response scanning that intrinsic dynamics of protomers is highly cooperative within the spike protein trimer. We also find that ACE2 binding substantially alters the global motions of not only the receptor-binding domain where it binds but also of other regions in all three protomers. Upon open conformation, coordinated motions between S1 and S2 regions have enhanced compared to the closed state. Furthermore, in the open state, the susceptibility of the S2 region encompassing proteolytic cleavage site and fusion peptide to external perturbations has increased. Gaussian dynamics calculations suggest perturbations at the S1 region may indirectly affect the S1/S2 cleavage sites and S2 regions via two hinge regions (hinge-1: Lys310-Phe318 and hinge-2: Gly593-Val618). Collectively, these observations imply that ACE2 binding at the S1 region in the open state might cause a long-range effect to modulate proteolysis of the S1/S2 cleavage site and the activation of the fusion peptide. Together, our study hints that ACE2 binding allosterically affects proteolytic sites and S2 regions, thereby potentially facilitate the conformational transition from prefusion state to postfusion state.

Torsional Diversity in Nucleosome Core Particle Structures Influences the Folding of Closed DNA Minichromosomes

Robert T. Young, Stefjord Todolli, Wilma K. Olson

Rutgers University, New Brunswick, NJ 08901

Organization of DNA into compact chromosomes hinges on the formation of nucleosome core particles (NCPs), the fundamental packing unit of chromatin where up to 147-bp of DNA wraps around a histone protein complex. Nucleosomal organization is highly dynamic as proteins such as RNA Polymerase interacts with DNA, resulting in localized distortions that must be compensated by neighboring DNA, either in the free linker region or at the ends of the nucleosome. Over two decades since the initial deposit of the 146-bp nucleosome in the Protein Data Bank, several hundred NCP deposits from multiple biophysical methods now exist, providing numerous insights into the packing and processing of genetic information. Recent analysis of NCPs at resolutions at or better than 3.5-Å collected from the PDB has revealed previously unrecognized differences in the end-to-end displacement of nucleosomal DNA and the overall twist of the chain molecule. Here we show how the uptake of twist on different NCPs affects the configuration of closed circular DNA constructs bearing a lone nucleosome. Minichromosomes of lengths 341-bp and 359-bp were constructed from 141 high- resolution NCPs by attaching a Bezier curve of various levels of supercoiling to the nucleosome ends, and then optimizing the elastic energy of the protein-free linker region. Optimization results showed that the clustering of twist found in the NCP structures, at positions ±20 and ±50-bp away from the central dyad base pair, introduce changes in loop topology and overall minichromosome shape. Simulations of nucleosomal breathing, obtained by symmetrically peeling DNA away from the NCP and then optimizing the newly lengthened linker region, lead to an increase in loop opening with greater resistance to structural change nearing an over-twisted NCP. These studies lay the foundation for asymmetric breathing studies at the entry region of the NCP and exploring the influence nucleotide sequence plays in the elastic response of the linker region.

Deciphering the Molecular Mechanism and Role of Fluorination in HCV Protease Inhibitor Potency and Drug Resistance

Jacqueto Zephyr, Desaboini Nageswara Rao, Sang V. Vo, Mina Henes, Klajdi Kosovrasti, Ashley N. Matthew, Elise Chan, Akbar Ali, Nese Kurt Yilmaz, Celia A. Schiffer

University of Massachusetts Medical School, Worcester, Massachusetts 01605, United States

Third generation Hepatitis C virus (HCV) NS3/4A protease inhibitors (PIs), glecaprevir and voxilaprevir, are highly effective across genotypes and against many resistance-associated substitutions (RASs). Hydrogen to fluorine substitution, or fluorination, on the P2-P4 macrocycle and P1 moieties of these PIs are partially responsible for the potency and pan-genotypic efficacy. Fluorination has long been used as a strategy in medicinal chemistry to improve physicochemical properties and binding affinity. However, the molecular basis by which fluorination improves potency and resistance profile of HCV NS3/4A PIs are not well understood. To systematically analyze the contribution of fluorination to potency and resistance profile, we used a multidisciplinary approach involving inhibitor design and synthesis, enzyme inhibition assays, co- crystallography, and structural analysis. A panel of inhibitors in matched pairs were designed with and without fluorine substitutions at the P4 capping groups that reside proximal to major RASs in the S4 pocket; D168A, R155, and A156T. The inhibitors were tested against WT and D168A resistant protease variants, and a total of 23 high-resolution co-crystal structures were determined. While P4 cap fluorination did not significantly alter potency against the WT protease, the PIs with fluorinated P4 capping groups retained much better potency against the D168A protease variant. Detailed analysis of the co-crystal structures revealed that fluorination allowed P4 capping groups to sample alternate binding conformations driven by hydrophobic, orthogonal multipolar, electrostatic, and fluorine-induced intramolecular interactions. Access to the alternate binding conformations enabled the fluorinated P4 caps to adapt to structural changes in the S4 pocket induced by the D168A RAS. Altogether, our results elucidate a novel mechanism of avoiding drug resistance.

THE INVESTIGATION OF BIOPHYSICAL AND BIOLOGICAL FUNCTION OF PRPS FROM NOSTOC PCC 7120

Ruojing Zhang[1], Michael A. Kennedy[1]

[1]Department of Chemistry and Biochemistry, Miami University, Oxford, OH, 45056

Nostoc sp. PCC 7120 are filamentous cyanobacteria capable of both oxygenic photosynthesis and nitrogen fixation, with the latter taking place in specialized cells known as heterocysts that terminally differentiate from vegetative cells under conditions of nitrogen starvation. Pentapeptide repeat proteins (PRPs), which occur most abundantly in cyanobacteria, adopt a right-handed quadrilateral β-helical structure, also referred to as a repeat five residue (Rfr) fold, with four-consecutive pentapeptide repeats constituting a single coil in the β-helical structure. Despite their intriguing structure and importance to understanding ancient cyanobacteria, the biochemical function of PRPs in cyanobacteria remains largely unknown. Here we report the crystal structure of Alr5209 and Alr1298, two PRPs from Nostoc sp. PCC 7120 predicted to be involved in oxidative phosphorylation and response to nitrogen starvation and/or heterocyst differentiation. The Alr5209 structure was analyzed in comparison to all other PRPs to determine how type I β turns can be accommodated in Rfr folds and the consequences of type I β turns on the right-handed quadrilateral β-helical structure. Given that Alr5209 represents the first PRP structure containing type I β turns, the PRP consensus sequence was reevaluated and updated. The structure of Alr1298 displays the typical right-handed quadrilateral β-helical structure and includes a four-α-helix cluster capping the N-terminus and a single α helix capping the C-terminus. Furthermore, we provide the preliminary investigation of the function of Alr5209 and Alr1298 by measurement of phenotype and localization.

Biological Structures, Chemical Features and Machine Learning Predictions of Short Hydrogen Bonds in Proteins

Shengmin Zhou[1, 2], Lu Wang[1, 2]

[1]Institute for Quantitative Biomedicine, Rutgers University, Piscataway NJ 08854, [2]Department of Chemistry and Chemical Biology, Rutgers University, Piscataway NJ 08854

Short hydrogen bonds (SHBs) with donor-acceptor separations below 2.7 Å exhibit unique quantum nature and have been associated with various biological functions, ranging from stabilizing the transition states in enzymatic catalysis and mediating antibiotic resistance. Exact determination of SHBs in proteins requires the biomolecular structures at atomic resolution, creating a challenge to structure determination techniques such as X-ray crystallography, NMR spectroscopy and electron microscopy. The Protein Data Bank (PDB) provides the three-dimensional architecture of over 175,000 biological macromolecules. We statistically analyze the top 1% highest-quality proteins from the PDB and systematically enunciate the factors contributing to the formation of SHBs. We build up a comprehensive idea of SHBs based on the biological structures and chemical features, which points out the preferred enzyme classes, amino acid locations, residue types and charges. All the biological and chemical preferences encode the driving forces together to form SHBs and serve as a reference to validate the presence of SHBs in proteins with low or moderate resolution. We further develop the first machine learning model to effectively predict the formation of SHBs between two amino acids in proteins, which demonstrates both good precision and recall around 80%. This model accounts for the interactions among residues, atoms, charges, locations, secondary structures and the sequence features of both donor and acceptor and categorizes a certain hydrogen bond as a short or normal hydrogen bond. Among all the input variables, our model indicates the major role of donor residue type. Using electronic structure calculations, we further reveal a general dependance of equilibrium proton positions, proton potential energy surfaces and proton sharing energy barriers, on the hydrogen bond length of biological SHBs. Together with the energy decomposition analysis, we elucidate the contributions of electronic quantum effects in stabilizing the SHBs. Our model allows one to efficiently predict and validate the presence of SHBs in proteins and will facilitate the experimental and computational refinement of protein structures.

POSTER SESSION 5
Wednesday May 5, 2021 1:40pm-2:20pm EDT

Biomolecular Visualization of the SARS-CoV-2 Spike Protein through Digital and Physical Media in an Upper Level Undergraduate Biochemistry Laboratory Course

Zahraa Alsilawi[1], Galen Arnold[1], Hunter St. Pierre[1], Benjamin Wheeler[1], Eva Rose M. Balog[1]

[1]University of New England, Biddeford, ME 04005

In Fall 2020, during the COVID-19 pandemic, undergraduate students in the course CHE 350: Biochemistry I (Proteins) at the University of New England (UNE) had the opportunity to learn about the structural biology and biochemistry of the SARS-CoV-2 Spike glycoprotein through weekly journal club-style discussions of the rapidly emerging literature. For our final project, we used knowledge gained from the literature, PDB data, molecular visualization software, and the resources and expertise of UNE's Makerspace laboratory to create multiple representations of Spike protein structure and function, including 3-D printed models, laser engraved art, and a poster display intended for peer and public educational outreach. We learned protein biochemistry in the context of a problem of contemporary, real-world significance; gained experience in searching and critically evaluating scientific literature; practiced multimedia communication of biochemical knowledge; and collaborated on creative works at the intersection of science, technology, and design.

Reconstructed from the PDB: The dynamic multi-enzyme pathway of Aerobic Respiration, from Glucose to ATP.

Drew Berry[1], Justin Muir[1]

[1]WEHI, Melbourne Australia

Combining 100+ PDB structures, cryo-EM, molecular-dynamics simulation and animation, this work presents dynamic visual review of aerobic respiration, from the sugar molecule Glucose to generation of ATP.
Meticulously reconstructed from more than 100 PDB molecular structures and many other forms of data, this animation explores the complete reaction pathway and enzyme mechanisms of Glycolysis, Citric Acid Cycle, Electron Transport Chain, and synthesis of ATP.
The technology pipeline enabling this molecular animation production exploits the extraordinary power of GPU game hardware to generate vast, multi-scale molecular landscapes of mitochondria membranes and dynamic enzyme complex mechanisms.
This work presents the first generation of wehi.tv's new pipeline for cinematic animation production and generation of real-time interactive ‘3D diorama’ molecular visualisation.

27 Years of Structural Classification of Proteins

John-Marc Chandonia[1], Naomi K. Fox[1], Steven E. Brenner[1, 2]

[1]Lawrence Berkeley National Laboratory, [2]University of California, Berkeley

Nearly all proteins have structural similarities with other proteins, and in some of these cases, share a common evolutionary origin. In 1994, the Structural Classification of Proteins (SCOP) knowledgebase became one of the first hierarchical repositories of information on the nascent World Wide Web. For 27 years, SCOP and our successor knowledgebase SCOPe (SCOP–extended) have aimed to provide an accurate, detailed, and comprehensive description of the structural and evolutionary relationships amongst all proteins of known structure, along with resources for analyzing the protein structures and their sequences. The resources are widely used, and have accumulated over 14,000 citations. By providing authoritative information about proteins' relatives, particularly those too ancient to be readily recognized from sequence, SCOP and SCOPe have provided a framework for research and classification.
The classification depends on expert curators to integrate many types of information. To identify ancient homologous relationships, proteins with similar three-dimensional structure and no recognizable sequence similarity are examined to determine whether they possess structural and functional features indicative of homology. If convincing evidence is found of an evolutionary relationship, this is annotated by grouping the homologous domains into a single superfamily; otherwise, the domains of similar structure are annotated as having a common fold but not grouped into a superfamily.
SCOPe (which continues development of the classic SCOP hierarchy after work on the latter ended in 2009) undertakes to provide interfaces and data to support users of protein structure and evolutionary relationships. Applications have included research, education, and policy, at scales ranging from interactive exploration of relationships of proteins of interest, including nuances of their individual structures and variations, to comprehensive studies and methods that draw on the entirety of the protein universe.
A wide variety of predictive algorithms rely on SCOPe as an authoritative standard for annotations of ancient homology, during their development and validation. To meet users' needs, we continue to update SCOPe to classify known structures, including newer PDB structures. We also extend SCOPe to meet new needs; for example, we are building tools to aid the analysis of genetic variants that impact protein structures.

Teaching and Learning with the PDB: 28 Years and Counting

Paul A. Craig[1]

[1]Rochester Institute of Technology, Rochester, NY 14624
As a starting assistant professor, I wanted to help my students understand the relationship between structure and function in proteins, starting with hemoglobin and chymotrypsin. At first we just talked about it and looked at textbook images. In year two, a PhD scientist at a local pharmaceutical company demonstrated interactive 3D protein structures on a Silicon Graphics workstation. Shortly thereafter molecular visualization freeware began to appear with instructions that our students used to tell stories about enzymes and proteins. Kinemage and Rasmol were followed by Chime and Jmol, which allowed us to create web pages with molecular graphics. At the same time, the students were learning to probe PDB files for more structural information, and to mine the PDB website to add depth and breadth to their learning. Some of these students also taught me to compare and integrate data from other online databases. Their research led to the BASIL (Biochemistry Authentic Scientific Inquiry Lab; https://basilbiochem.github.io/basil/index.html) project. The focus of this presentation will be on the historical application of the PDB and its considerable resources to biochemistry and molecular biology education, leading to a view of future education opportunities, including coding skills and virtual reality. Support has been provided by Rochester Institute of Technology, NIGMS R15GM078077, and by NSF 0402408, 1503811, and 1709170.

Template-based modeling of protein complexes using the PPI3D web server

Justas Dapkunas[1], Albertas Timinskas[1], Kliment Olechnovic[1], Migle Tomkuviene[1], Ceslovas Venclovas[1]

[1]Institute of Biotechnology, Life Sciences Center, Vilnius University

The PPI3D web server is user-friendly software, focused on searching, analyzing and modeling protein-protein, protein-peptide and protein-nucleic acid interactions in the context of three-dimensional structures. PPI3D uses data from the Protein Data Bank and updates weekly in order to always have the newest structures. Reducing the data redundancy by clustering and analyzing the properties of interaction interfaces using Voronoi tessellation makes this software a highly effective tool for addressing different questions related to protein interactions. In recent CASP and CAPRI experiments, PPI3D also proved to be highly effective in detecting structural templates for modeling protein complexes, thus it may be useful for anyone interested in all types of protein interactions. The web server is available at http://bioinformatics.lt/ppi3d.

Mining the PDB for tractable cases where X-ray crystallography combined with fragment screens can be used to systematically design protein-protein inhibitors.

Gian F. De Nicola[1], Charlie Nichols[1], Joseph Ng[1], Annika Keshu[1], Geoff Kelly[2], Maria R. Conte[1], Michael S. Marber[1], Franca Fraternali[1]

[1]King's College London, [2]The NMR Facility, The Francis Crick Institute

Recent technical advances in automated data collection and interpretation of electron density maps have made it possible to combine X-ray crystallography and fragment screening in a medium throughput fashion. This approach is generally used by academia and the pharmaceutical industry to develop molecules that target the catalytic site of enzymes. Here we show that, contrary to the accepted view, the same approach can be used to chemically probe the surfaces used by proteins to interact and use the outcome of the screens to systematically design protein-protein inhibitors. To prove it we first performed a bioinformatics analysis of the Protein Data Bank (PDB) protein complexes, which revealed over 400 cases where the crystal lattice of the target in the free form is such that large portions of the interacting surfaces are free from lattice contacts and therefore accessible to fragments during soaks. Among the tractable complexes identified, we performed single fragment crystal screens on two particular interesting cases: the Il1β-ILR and p38α-TAB1 complexes. The result of the screens showed that fragments tend to bind in clusters, highlighting the small-molecule hotspots on the surface of the target protein. In most of the cases the hotspots overlapped with the binding sites of the interacting proteins. Finally, for one of the complexes (p38α-TAB1), we carried out a chemical exploration and improved the initial binding affinity of the fragments, bringing it into the micromolar range.

Impact of Protein Flexibility on Enzyme Design

Dennis Della Corte[1]

[1]Brigham Young University, Provo, UT

Our group recently reported on the importance of protein flexibility for the design of enzymes (https://doi.org/10.1142/S0217984921501554) and for protein-ligand docking (https://doi.org/10.1142/S0217984921300027). In another report on the design of a novel biosensor (https://doi.org/10.1038/s41467-020-18400-0), we showed that thermodynamic ensembles are able to provide insights into the effect of single mutations on protein signaling. This poster summarizes our insights from literature search, molecular dynamic studies, and our deep learning methods for protein structure prediction (https://doi.org/10.1101/830273) on the importance of protein dynamics for function prediction. We also present here some new findings, which indicate that sequence encodes not only structure but also full thermodynamic ensemble information, as can be predicted by deep learning networks.

The Future is Now: The RCSB Protein Data Bank, Cryo-EM, and Nanome vs SARS-Cov-2

Carla Gauss[1]

[1]Nanome

In an era of super viruses poised to unleash devastating pandemics, cutting edge science and technology are required for an effective defense. For the past fifty years, the RCSB has been at the forefront of drug discovery research, driving structure-based design of new medicines utilizing state-of-the-art crystallographic methods such as cryo-electron microscopy (Cryo-EM). Nanome is the modern-day leader in collaborative structure-based design through virtual reality (VR) software. Together in VR, Dr. Gregg Crichlow, Biocurator at the RCSB PDB, and the Nanome team explored two SARS-CoV-2 spike protein structural complexes from the PDB: 7A94, spike in complex with ACE2 and published by Benton, Wrobel et al. at the Francis Crick Institute; and 6XCN, spike in complex with a human monoclonal Fab and published by Bjorkman, Nussenzweig, and colleagues at Caltech and Rockefeller University. Detailed investigation of the interfaces of these complexes may aid in the rational design of drugs and biologics to combat SARS-Cov-2.

Mining PDB for insights into protein flexibility

Lukasz Jaroszewski[1], Zhanwen Li[1], Mayya Sedova[1], Mallika Iyer[2], Adam Godzik[1]

[1]University of California Riverside School of Medicine, CA 92521, [2]Graduate School of Biomedical Sciences, Sanford Burnham Prebys Institute , CA 92037

The Protein Data Bank (PDB) is a database collecting information on experimentally solved protein structures and a portal for related analyses. Its entries are files with coordinates of specific conformations of individual proteins (sometimes with alternative positions of some atoms). Therefore, it would seem counterintuitive to use this information to study protein flexibility. But in a quickly growing trend, most PDB entries now provide information on structures of proteins that are identical or closely homologous to already studied proteins, but in different functional states or bound to different inhibitors. This is hardly a surprise, as we know that proteins are highly flexible and exist in a multitude of conformations tied to their specific molecular functions. We also know that homologous proteins have highly similar structures, sometimes with highly conserved, but often significantly modified functions. In both cases subtle differences between the structures, driven by mutations, post-translational modifications or binding to complex partners or substrates/ligands may be important for understanding their functions or designing an optimal inhibitor.Multiple “redundant” coordinate sets for same or homologous proteins provide us with a large dataset to study the more or less subtle differences between protein structures. While most existing programs for protein structure analysis focus on recognizing and classifying similarities between protein structures, our group has developed a series of tools to compare and analyze the differences between similar structures. The servers and algorithms developed and maintained by our group, such as the flexible protein structure alignment program FATCAT, multiple protein structure alignment server POSA, database of structural variations in proteins PDBFlex and our recent server for modeling alternative protein conformations ModFlex, are all publicly available, free servers using information from the PDB and providing users with tools to analyze the next big question in structural biology – how does the structure of a protein of interest change as part of performing its functions or in response to a mutation or inhibitors?On series of examples we show how to analyze groups of homologous proteins or series of alternative conformations of a specific protein to identify and classify function related structural changes in proteins.

Structural and Phylogenetic Relationships Among Beta-Lactamases

Sheuli Zakia[1], Janet Gonzalez[2], Cristina C. Clement[3], Manfred Philipp[4,5]

[1]Department of Natural Sciences, Laredo College, West End Washington St, Laredo, TX 78040, 2Department of Natural Sciences, LaGuardia Community College, 31-10 Thomson Ave, Long Island City, NY, 3Department of Pathology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY, 4Department of Chemistry, Lehman College, 5Biochemistry and Chemistry Doctoral Programs, Graduate Center of CUNY

The Mycobacterium tuberculosis beta-lactamase (BlaC) has been extensively studied in genetic and structural studies. The enzyme, present in M. tuberculosis even prior to the use of beta-lactam antibiotics, is at least partially responsible for the resistance of tuberculosis infection to treatment by various beta-lactams.

This report describes the structure of this enzyme in comparison to proteins that have closely related sequences and structures. This study relies on the previously studied phylogenetic relationships of beta-lactamases and the large number of lactamase structures that have been deposited into the PDB database.

The VAST+ server on the NCBI web site was used to provide lists of protein pdb structures with folding patterns that are closely related to that of BlaC. FASTA protein sequence files for the proteins identified by VAST+ were obtained from the RCSB PDB server. The FASTA protein sequence files were sorted according to the VAST-supplied percent sequence identities and, after removal of His-tag sequences, were submitted to the ngphylogeny.fr server. It was necessary to delete FASTA files of low sequence identity (below 30%) in order to generate phylogenetic trees. After protein 3D-structural alignments were prepared using Swiss Viewer, the phylogenetic trees were used to compare protein structures and the locations of active site residues.

One of the goals of this work was to develop methods that facilitate undergraduate instruction in biochemistry by use of publicly available databases.

Using existing structures to predict M. tuberculosis indole-3-glycerol phosphate synthase ligand interactions

Nina M. Goodey[1], Sarah Cho[1], Katherine Leon H.[1], Huma Booter[1], Ashley Reyes[1], Joseph LaCap[1], Paige Fadden[1], Thomas Oshane[1], David Konas[1]

[1]Montclair State University, Montclair, NJ 07043

Tuberculosis is a lung disease, caused by the pathogen M. tuberculosis that affects millions of people world-wide. Several tuberculosis drugs have lost efficacy due to emerging drug resistance and there is a need to discover new inhibitors and targets. Recent research suggests that indole-3-glycerol phosphate synthase (IGPS) in M. tuberculosis (MtIGPS) could be selectively targeted to treat tuberculosis. Here we examined unpublished MtIGPS structures in the Protein Data Bank to obtain insights into MtIGPS active site residues involved in ligand interactions. We used AutoDock Vina to predict ligand interactions with substrate 1-(o-carboxyphenylamino)-1-deoxyribulose 5'-phosphate (CdRP) and transition state analogs. We examined these predicted structures in light of recent literature findings to obtain insights into structure function relationships. These structural analyses will facilitate the design of inhibitors and the experimental evaluation of MtIGPS as a potential new tuberculosis drug target.

Inspiring Undergraduates Through Problem-Solving Using the PDB

Jennifer E. Grant[1]

[1]University of Wisconsin-Stout, Menomonie, WI 54751

Since the PDB became available on the internet, the PDB has been my muse. As a professor at a Primarily Undergraduate Institution (PUI) I have used the PDB as a resource for teaching undergraduates how to assess protein structure and function as well as develop their problem-solving acumen. Here, I present on how the PDB can be leveraged to promote inquiry among these students and also summarize some of the best vignettes taken from the primary literature. Particularly, how this resource can be used to benefit general education students with the same impact as the benefit to science majors will be addressed.

BINDS: A governmental program supporting research in structural biology and drug discovery in Japan

Tsuyoshi Inoue[1], Haruki Nakamura[2]

[1]Graduate School and School of Pharmaceutical Sciences, Osaka University, [2]Institute for Protein Research, Osaka University

The BINDS program, Basis for Supporting Innovative Drug Discovery and Life Science Research (https://www.binds.jp/en/), has been running since 2017 for five years to promote drug discovery in academia at pre-clinical phase and relating life science research including structural biology by the AMED, Japan Agency for Medical Research and Development. Total 59 research groups were selected in diverse fields, such as Pharmaceutics, Medicine, Chemistry, Genomics, Structural Biology, Informatics and Computer Science, from all over Japan with the basic budget about 3 billion yen every year. The aim is to establish an innovative platform to promote the therapeutic applications of early-stage drug discovery by providing and sharing key technological infrastructure: synchrotron facilities at SPring-8 and the Photon Factory, cryo-Electron Microscopes (EMs), NMR, chemical compound libraries, and next-generation DNA sequencers. The BINDS program has a characteristic feature to support researchers even outside of the program, for drug discovery and general life science. So far, more than 2,500 research supports have been made.

Here, we introduce several recent technological progresses in the field of structural biology and their applications to medical and pharmaceutical science. In particular, the BINDS program strongly promotes development and applications of cryo-EM to high-resolution structural analysis of proteins and their complexes, which were rather poor in Japan before the BINDS started 4 years ago. Several new methods in usage of cryo- EM have been developed for the higher resolution analysis including a new application of chemically modified grids of oxidized graphene for immobilization of target proteins (unpublished), and many structures have been determined at high resolution during the program (e.g. apoferritin: 1.29Å resolution). In order to integrate the cryo-EM usage among many researchers, the BINDS program creates a network, the Cryo-EM network (https://www.cryoemnet.org/), so that most of the cryo-EM machines in Japan are shared with the well- organized schedule by users as many as possible. The BINDS program also supports the EMPIAR at PDBj for management of the cryo-EM raw, 2D-images (https://empiar.pdbj.org/).

We are convinced that various technical and research exchanges as well as human resource development are in progress through BINDS projects.

Neutron crystallography in the fight against COVID-19: Drug Design Targeting SARS-CoV-2 Main Protease

Daniel W. Kneller[1], Leighton Coates[1], Andrey Kovalevsky[1]

[1]Oak Ridge National Laboratory

COVID-19, caused by SARS-CoV-2, is a global health and economic catastrophe. The viral main protease (M^{pro}) is indispensable for SARS-CoV-2 replication and thus is an important target for small-molecule antivirals. Computer-assisted and structure-guided drug design strategies rely on atomic scale understanding of the target biomacromolecule traditionally derived from X-ray crystallographic data collected at cryogenic temperatures. Conventional protein X-ray crystallography is limited by possible cryo-artifacts and its inability to locate the functional hydrogen atoms crucial for understanding chemistry occurring in enzyme active sites. Neutrons are an ideal probe to observe the protonation states of ionizable amino acids at near-physiological temperature, directly determining their electric charges – crucial information for drug design. Our X-ray crystal structures of M^{pro} collected at near-physiological temperatures brought rapid insights into the reactivity of the catalytic cysteine, malleability of the active site, and binding modes with clinical protease inhibitors. The neutron crystal structures of ligand-free and inhibitor-bound M^{pro} were determined allowing the direct observation of protonation states of all residues in a coronavirus protein for the first time. At rest, the catalytic Cys-His dyad exists in the reactive zwitterionic state, with both Cys145 and His41 charged, instead of the anticipated neutral state. Covalent inhibitor binding results in modulation of the protonation states, retaining the overall electric charge of the M^{pro} active site cavity. In addition, high-throughput virtual screening in conjunction with in vitro assays identified a lead non-covalent compound with micromolar affinity, which is being used to design novel M^{pro} inhibitors. Our research is providing real-time data for atomistic design and discovery of M^{pro} inhibitors to combat the COVID-19 pandemic and prepare for future threats from pathogenic coronaviruses.

DeepChain: A platform for protein design

Nicolas Lopez Carranza[1], Marcin Skwark[1], Thomas Pierrot[1], Hippolyte Jacomet[1], Ashraf Guitouni[1], David Sehnal[2], Boulbeba Mallouli[1], Alexander Laterre[1], Zohra Slim[1], Karim Beguir[1]

[1]Instadeep Ltd, [2]National Centre for Biomolecular Research, Faculty of Science and CEITEC-Central European Institute

Here we present a novel, accessible and user-friendly platform for protein design. DeepChain helps biotech companies design and optimize peptide-based macromolecules, and can be directly applied to drug discovery.The user of the platform inputs a PDB file describing the 3D structure of the protein complex and through a simple, visual user interface selects the positions at the interface to apply AI directed mutagenesis to. DeepChain will find a biologically sequence of mutations that improve the binding energy between both parts, and - in more advanced workflows - it will improve any metric designed by the user.Replacement of a single amino acid in a protein leads to a change in the interaction patterns, which may have global effects on the entire complex. Therefore, protein interface optimization is non-factorizable in a general case. We have developed a simulation-based scoring metric, which captures the biological reality well enough, as evidenced by tests on prior, experimental results.This scoring metric, together with state-of-the-art multiple objective optimisation algorithms and pre-trained transformer models allow DeepChain to mutate protein sequences from a PDB file and propose highly plausible designs within a few hundred thousand steps. In order to cross validate the results, we integrated molecular dynamics simulations powered by GROMACS, which can be visualised in the platform.

Capturing the SARS CoV-2 Spike in motion: Visualizing a molecular contortionist

Gael McGill[1, 2], Jonathan Khao[2], Geoffrey Cheung[2]

[1]Harvard Medical School, Boston, MA 02115, [2]Digizyme Inc., Brookline, MA 02446

The rise of SARS CoV-2 has been a rallying cry for the scientific community and we have witnessed an outpouring of research about the virus over the past year. The incredible collaborative effort on the part of the scientific/medical visualization community has paralleled this trend and numerous high-quality models, images and visualizations have been created to help explain the biology of the virus and, in particular, the spike protein. However, few studies (and visualizations based on them) have captured the daunting complexity of the spike's motions as it drives membrane fusion. In this study we share our efforts to accurately model, simulate and visualize the spike-induced membrane fusion process. Our work suggests not only how this molecular contortionist transitions from prefusion, to prehairpin and postfusion intermediates, but also offers potential new insights about the timing of S2' protease cleavage and why inactivated virus may prove to be a poor choice of vaccine immunogen. Finally we share our collaborative work (recently published in Science) visualizing the mechanism of action of a promising class of membrane-tethered peptide fusion inhibitors that block the prehairpin to postfusion transition.

Expanding the Utilities in the ResiRole Server to Provide Automated Assessments of Structure Models Presented in CAMEO and to Enable Interactive Assessments of User-Defined Protein Structure Models

Joshua M. Toth[1], Paul J. DePietro[1], Juergen Haas[2], William A. McLaughlin[1]

[1]Geisinger Commonwealth School of Medicine; 525 Pine Street, Scranton PA, 18509,
[2]Biozentrum, University of Basel and SIB, Klingelbergstrasse 50/70, CH -4056 Basel

The Continuous Automated Model Evaluation (CAMEO) platform presents results of protein structure predictions from hosted structure prediction servers for prelease sequences in the Protein Data Bank (PDB). We have developed a method to evaluate the quality of structural models available through CAMEO based on their abilities to have SeqFEATURE functional site predictions like those at corresponding sites in the reference structures. The ResiRole server, http://protein.som.geisinger.edu/ResiRole/, calculates the average difference scores per structure prediction technique and per structure model. Each difference score is defined as the absolute difference in cumulative probability of functional site prediction in the reference structure versus that at the corresponding site in the structure model. Results are accessible according to defined intervals in which models and reference structures have been made available in CAMEO and the PDB, respectively. Results are further delineated based on target difficulty according to lDDT score ranges. To expand the utilities of the ResiRole server, we are developing automated routine updates that evaluate models in CAMEO. We are also developing means for interactive submissions of structural models and reference structures to further enable the assessments of user-defined models. Support was partly provided by the NIGMS [grant number 5U01 GM093324-02].

A guideline for 3D printing of macromolecular models on the cheap

Marius Mihasan[1]

[1]Department of Biology, Alexandru Ioan Cuza University of Iasi, Romania, RO 700506

The structure and function of biomolecules relationship is the hallmark of biochemistry, molecular biology and life sciences in general. Physical models of macromolecules give students the possibility to manipulate these structures in three dimensions, developing a sense of spatiality and a better understanding of key aspects such as atom size and shape, bond lengths and symmetry. Several molecular model systems were developed specifically to represent particular classes or groups of molecules and hence lack the flexibility of a universal solution. Three dimensional (3D) printing could nevertheless provide such a universal solution, as it can be used to create physical models of biomolecular structures based on the teacher's or demonstrator's needs and requirements. Here, insulin was used as a model molecule and several depictions and printing parameters were tested in order to highlight the technical limitations of the approach. In the end, a set of settings that worked is provided which could serve as a starting point for anyone wishing to print its own custom macromolecular model on the cheap.

Scientists in rapid response to biomedical challenges

Wladek Minor[1], Alexander Wlodawer[2], Zbigniew Dauter[2], Mariusz Jaskolski[3, 4], Bernhard Rupp[5, 6], Aled Edwards[7]

[1]University of Virginia, Charlottesville, VA, USA, [2]Center for Cancer Research, NCI, Frederick, MD, USA, [3]Faculty of Chemistry, A. Mickiewicz University, Poznan, Poland, [4]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland, [5]k.-k. Hofkristallamt, San Diego, CA, USA, [6]Medical University Innsbruck, Austria;, [7]Structural Genomics Consortium, Toronto, Canada

Most research projects take several years to complete, whereas detailed studies and thinking, even if supported by artificial intelligence, are not fast. However, a fast-emerging global crisis, such as the COVID-19 pandemic, usually does not leave a comfortable margin of time for a reaction but requires an immediate and swift response. The amazing development of vaccines for COVID-19 was built on foundational studies on RNA, adenoviruses, lipids, and cellular immunity in the 1980s and the 1990s. The unprecedently rapid response to the deadly threat of COVID-19 has also been possible because of the enthusiasm and vision of a small group of people, who 50 years ago realized that data sharing would be critical to the future of biomedical science. They started an adventure that is now known as the Protein Data Bank, which laid the groundwork for structure validation, archiving and reproducibility, and structure prediction with futuristic applications of artificial intelligence, as illustrated by the spectacular success of the ALPHAFOLD2 system. The success of the COVID-19 vaccines and ALPHAFOLD2 are as much team wins for the scientific community and the scientific method, as they are brilliant accomplishments of their developers. They also provide lessons for the next 50 years.

There are 15,000 Structural Genomics protein structures in the PDB – What is their function?

Suhasini M. Iyengar[1], Kelton Barnsley[1], Penny J. Beuning[1], Mary Jo Ondrechen[1]

[1]Northeastern University

Among the structures in the PDB are 15,000 Structural Genomics proteins. Most of these are of unknown or uncertain biochemical function. Here we present a powerful approach based on computed chemical properties of the individual residues in a protein structure. Local arrays of predicted active residues for sets of proteins of known function are matched with those of SG proteins. For instance, a superfamily consists of proteins with similar structure but multiple different kinds of biochemical function, including different types of reactivity as well as different substrate specificities. Each functional type has a different set of active amino acids. Typically active residues include those in the first layer that make direct contact with the substrate molecule(s) and also some residues in the second and third layers that play supporting roles in the catalytic process. Graph Representation of Active Sites for Prediction of Function (GRASP-Func) establishes local arrays of predicted residues that are common to proteins of the same function. Predicted local arrays for each SG protein are aligned against those of the known members of each functional family. Cases of predicted misannotation, where our prediction differs from the originally assigned function, are especially interesting. Experimental testing is performed by direct biochemical assays to test our predictions. For instance, we show that our predictions are confirmed that RV0760c from Mycobacterium tuberculosis has ketosteroid isomerase activity, whereas NP_103587.1 from Mesorhizobium loti does not. Some cases involving local site matches across different fold types for glycoside hydrolases are also presented. Supported by NSF grant CHE-1905214.

Analysis of Student Mastery of STEM research concepts during a BlendFlex CURE Using a Combination of Self-reported and Empirical Analysis

Shreya Patel[1], Arthur Sikora[1]

[1]NOVA Southeastern University

Due to the ongoing COVID-19 pandemic, institutions across the world have had to make modifications to existing curricula through the university. Course-based Undergraduate Research Experience (CURE) courses are emerging as an effective way to offer research opportunities for students traditionally underrepresented in STEM. The Biochemistry Authentic Student Inquiry Lab (BASIL) operates under the CURE model. Scientific objectives of BASIL focus on characterizing biochemical activities of proteins with known structure and unknown function. BASIL curriculum (freely available on GitHub https://basilbiochem.github.io/basil/) includes modules on computational predictions of protein function, protein purification, and analysis of biochemical activity and kinetics. The BASIL project aims to identify specific functions for proteins of known structure but unknown function, many of which have been deposited in the Protein Data Bank (PDB) as a result of the National Institutes of Health (NIH) protein structure initiative. BASIL students utilize the PDB to access invaluable information and structures that serve as the foundation for their research. This work showcases a two-pronged approach which combines student self-reported mastery with objective evaluation of lab report responses, both aligned with established Anticipated Learning Outcomes (ALOs) for BASIL. Using a pre and post survey, we measured growth in knowledge, experience, and confidence (KEC) as a result of the course. Students reported learning more about bioinformatic experiments and concepts than about their wet-lab counterparts. KEC tied directly to ALOs had an average gain score of 67.0%, while those referring to techniques increased 61.5%. Student lab report responses that aligned to ALOs were analyzed on a scale from one to five. Towards better understanding the impact of the emergency shift online, this model provides a novel method to study the effects on student mastery of ALOs this past year. The student's mastery of wet-lab ALOs coincided with our findings that lab courses need enhanced strategies to teach critical STEM lab-research skills in an online setting. Novel assessment strategies developed based on learning objectives help fill this skills gap and enhance the exposure of undergraduate students to vital STEM research experiences.

DNAmoreDB, a database of DNAzymes

Almudena Ponce-Salvatierra[1], Pietro Boccaletto[1], Janusz M. Bujnicki[1]

[1]International Institute of Molecular and Cell Biology in Warsaw, Warsaw, Poland

Deoxyribozymes, DNA enzymes, or simply DNAzymes are singlestranded DNA molecules that, like proteins and ribozymes, possess the ability to perform catalysis. Although DNAzymes have not yet been found in living organisms, they have been isolated in the laboratory through in vitro selection. The selected DNAzyme sequences have the ability to catalyze a broad range of chemical reactions, utilizing DNA, RNA, peptides, or small organic compounds as substrates. DNAmoreDB is a comprehensive database resource for DNAzymes that collects and organizes the following types of information: sequences, conditions of the selection procedure, catalyzed reactions, kinetic parameters, substrates, cofactors, structural information whenever available, and literature references. Currently, DNAmoreDB contains information about DNAzymes that catalyze 20 different reactions. We included a submission form for new data, a RESTbased API system that allows users to retrieve the database contents in a machine readable format, and keyword and BLASTN search features. The database is publicly available at https://www.genesilico.pl/DNAmoreDB/.

PonceSalvatierra A, Boccaletto P, Bujnicki JM. (2021) NAR 49(D1), D76–D81.

BioMolViz: A Community for Visual Literacy Instruction and Assessment

Kristen Procko[1], Josh T. Beckham[1], Daniel R. Dries[2], Shelly Engelman[3], Margaret A. Franzen[4], Henry Jakubowski[5], Walter Novak[6], Rebecca Roberts[7], Alberto I. Roca[8], Audrey C. Shor[9]

[1]The University of Texas at Austin, [2]Juniata College, [3]Custom EduEval LLC, [4]Milwaukee School of Engineering, [5]College of Saint Benedict/Saint John's University, [6]Wabash College, [7]Ursinus College, [8]DiverseScholar, [9]Saint Leo University

BioMolViz is a community of practice for promoting instruction in visual literacy. Our interactive workshops enhance instructor skillsets through the collaborative development of biomolecular visualization (BMV) assessments for use in the classroom. In these workshops, faculty participants design assessments to probe biomolecular visualization skills, often using Protein Data Bank structures and tools. Employing the BioMolViz Framework (biomolviz.org), a tool that unpacks twelve overarching themes of BMV into a collection of learning goals and objectives, participants use backward design to create high-quality assessment instruments for evaluating visual literacy skills. With inspiration from favorite PDB structures, attendees create assessment questions that probe critical biomolecular visualization skills, for example, molecular interactions and structure-function relationships. The assessment instruments are peer reviewed during the workshops, and then move through a validation process involving review by BioMolViz team members and an expert panel. These tools will be made available to the molecular life sciences community through an online repository, so that the broader visualization community can benefit from the PDB-inspired assessments created through our workshops.

Structural basis of RNA polymerase recycling by ATP-dependent motor enzyme RapA

M Zuhaib Qayyum[1], Vadim Molodtov[2], Rishi K. Vishwakarma[1], Katsuhiko Murakami[1]

[1]Pennsylvania State University, [2]Rugters University

During transcription, cellular RNA polymerases (RNAPs) can become trapped on nucleic acids, threatening genome stability and limiting free enzyme pools for the next round of transcription cycle. It was shown that the E. coli RNAP becomes impounded in an undefined Post-Termination Complex (PTC) after one or few rounds of transcription, and an ATP dependent motor enzyme called RapA, recycles the stalled RNAP during multi-round transcription in vitro. RapA (968aa, ~110kDa) is anRNAP-associated Swi2/Snf2 protein and the ATPase activity of RapA is stimulated upon binding to RNAP. The comprehensive mechanistic understanding of RapA-mediated RNAP recyclingremains elusive. Here, using single particle cryo-electron microscopy, we have determined the structures of E. coli RapA-RNAP complexes along with core RNAP at nominal resolutions. The structures reveal the binding site of RapA on RNAP and the large conformational changes in RNAP and RapA upon their association, implicated in reactivation of seized RNAPs on PTC.Even though RapA binds away from the DNA binding channel (clamp), it is capable to induce structural rearrangements in RNAP allosterically, leading to clamp closure. On the basis of thorough structural analysis and DNA binding assays, we propose that RapA acts as a guardian of RNAP, preventing non-specific association of the post-transcription terminated RNAP with DNA without impeding holoenzyme formation for the next round of transcription cycle, thereby preventing the PTC formation and enhancing RNAP recycling.

Structural Characterization of Diazabicyclooctane (DBO) β-Lactam "Enhancers" in Complex with Penicillin-Binding Proteins PBP2 and PBP3 of Pseudomonas aeruginosa

Malligarjunan Rajavel[1], Vijay Kumar[1], Ha Nguyen[1], Jacob Wyatt[1], Robert A. Bonomo[1], Focco van den Akker[1]

[1]Case Western Reserve University, Cleveland, OH-44120

Multidrug-resistant Gram negative (MDR-GN) pathogenic bacteria are highly resistant to various FDA-approved "front-line" antibiotics and pose a significant public health threat. These MDR-GN pathogens manifest resistance primarily by the degradation of the β-lactam antibiotics by β-lactamases. Novel non-β-lactam penicillin-binding protein (PBP) inhibitors have been shown to bypass such resistance mechanisms through concurrent multiple PBP binding when partnered with appropriate β-lactam antibiotic. Herein, we investigated two DBO compounds [WCK5107 (zidebactam) and WCK 5153, class A and C b-lactamase inhibitors] that function also as β-lactam "enhancers" by binding to penicillin-binding protein 2 (PBP2) of Enterobacterales, *P. aeruginosa* and *A. baumannii*. To this end, we determined the co-crystal structure of P. aeruginosa PBP2 liganded with WCK 5153, which forms covalent bonding interaction with the catalytic S327 of PBP2. We report the first structure for this *P. aeruginosa* PBP. In this complex structure, the 'diacylhydrazide moiety' of WCK 5153 interacts with the aspartate in the classical 'S-X-N/D' signature motif. We next modeled zidebactam in the active site of PBP2, and discerned a similar binding mode. Both DBOs increase the melting temperature Tm of PBP2 through their covalent and structure-stabilizing hydrogen bonding interactions. Also, we crystallographically explored DBOs complexes bound to *P. aeruginosa* PBP3. In contrast to PBP2, these DBOs destabilizes the PBP3 protein upon binding. In conclusion, our findings provide molecular insights into zidebactam and WCK 5153 inhibiting PBP2 compared to PBP3. We also included a structural comparison with their inhibition of the evolutionarily related KPC-2 β-lactamase. These molecular insights into the dual-target DBOs explain zidebactam and WCK 5153's potent anti-*Pseudomonal* enhancer activity and advance our knowledge regarding further DBO optimization efforts to develop novel potent β-lactamase-resistant, non-β-lactam PBP inhibitors.

Can 3D-domain-swapping help understand polyQ sequences?

Manjula Ramu[1], Ramaswamy Subramanian[2], Shachi Gosavi[1]

[1]National Institute for Biological Sciences, Bengaluru, India, [2]Purdue University, West Lafayette, IN, USA

Proteins with expanded polyQ repeats can form neurological disease-associated amyloids. The regions surrounding the polyQ sequences are often not resolved in 3D structures (Totzeck, Andrade-Navarro et al. 2017). Thus, understanding the structural details of polyQ stretches even in non-amyloid contexts may provide insight into the mechanism of amyloidogenesis. Previous work on the chymotrypsin inhibitor 2 (CI2) and ribonuclease A (RNase-A) (Sambashivan, 2005) has shown that insertion of polyQ stretches can induce 3D-domain-swapping in proteins. It was hypothesized that the β-strand forming tendency of polyQ and the interprotein β-β interactions similar to those present in amyloids would induce domain-swapping. However, structural information is not available for the RNase-A multimer while the polyQ regions are not resolved in the QQQQ (Q4) inserted CI2 (PDB ID: 1CQ4). We have previously been able to design 3D-domain-swapping in the monomeric sweet protein monellin by inserting the stefin-B derived QVVAG sequence into a tight β-turn. The solvent exposure of the central hydrophobic V of QVVAG in the tight β-turn was expected to destabilize the monellin monomer and promote domain-swapping and dimerization. Consequently, replacing QVVAG with QVNAG reduces dimerization. Based on these observations, it was surprising to see that the QQQQG (Q4G) sequence also induces dimerization in monellin. However, expansion of the insertion from Q4G to Q8G suppresses dimerization and this monellin mutant is entirely monomeric. Thus, polyQ sequences may not be able to form stable tight β-turns and this may be because of the larger bulk of the Q as compared to that of the N. Finally, the structure of dimeric Q4G-monellin was solved using X-ray crystallography and the domain-swapped polyQ hinge region can be clearly seen in this structure. This could pave the way for additional structural characterization of polyQ sequences.

The Protein Data Bank Provides the Foundation for Interdisciplinary Collaboration Through a Course-based Undergraduate Research Experience (CURE)

Rebecca Roberts[1]

[1]Ursinus College, Collegeville, PA 19426

Interdisciplinary collaborations are essential to answer modern, multi-faceted research questions. This approach, however, involves challenges such as understanding the methodologies and norms of a disparate field, as well as achieving effective communication between team members with varying expertise. We must begin to provide young scientists with the skills necessary for success as the call for interdisciplinary science expands. The Biochemistry Authentic Scientific Inquiry Lab (BASIL) can provide a framework for this training. BASIL is a Course-based Undergraduate Research Experience (CURE) in which students attempt to assign biochemical function to proteins with unknown functions that are available in the Protein Data Bank (PDB). At Ursinus College, we have modelled an interdisciplinary collaboration at the undergraduate level using the BASIL CURE. Students in upper-level Biochemistry and Structural Biology courses collaborate over the course of a semester. The Structural Biology students carry out computational, in silico analyses through guided modules on BLAST, Pfam, Dali, the active-site alignment tools ProMol or Moltimate, and perform docking studies to hypothesize protein function. In parallel, the Biochemistry students express and purify the proteins and carry out in vitro activity analyses, informed by the in silico-predicted enzyme function. Students come together in cross-disciplinary teams throughout the project to share progress, explain their discipline-specific methodologies, and discuss next steps. They each present an oral progress report on the entire project, which required that they not only understand their own work but that of their collaborators. Interdisciplinary teams present final posters to the campus community and students prepare future directions documents. A pre-post assessment analyzed the effectiveness of the cross-course collaboration model and assessment of content learning. We see that students can learn the information/methodologies of both their own discipline and that of their collaborators and improve communication skills through this framework.

Disclosing conformational dynamics upon the formation of cysteine synthase complex by hydrogen-deuterium exchange mass spectrometry

Brenda Rosa[1], Eleanor R. Dickinson[2], Marialaura Marchetti[1], Barbara Campanini[3], Barbara Pioselli[4], Kasper D. Rand[2], Stefano Bettati[1, 5, 6]

[1]Biopharmanet-TEC Interdepartmental Center, University di Parma, 43124 Parma, Italy, [2]Protein Analysis Group, Department of Pharmacy, University of Copenhagen, 2100 Copenhagen O, Denmark, [3]Department of Food and Drug, University of Parma, 43124 Parma, Italy, [4]Chiesi Farmaceutici, R & D Department, 43122 Parma, Italy, [5]Department of Medicine and Surgery, University of Parma, 43126 Parma, Italy, [6]Institute of Biophysics, CNR, 56124 Pisa, Italy

Cysteine synthase (CS) is a bienzymatic complex, formed by serine acetyltransferase (CysE) and *O*-acetylserine sulfhydrylase A (CysK), the two enzymes that catalyze the last steps of cysteine biosynthesis in bacteria. CysK is a homodimer, while CysE is a dimer of trimers. Two CysK dimers bind at the two opposite ends of CysE hexamer to form CS. The complex is stabilized by the insertion of the C-terminus of CysE into the active site of CysK. CysE and CysK, and their interaction, have been investigated as potential targets for the development of antibiotics since cysteine is related to many bacterial functions and antibiotic resistance. Considering that many details of the regulatory mechanism and functions of CS complex, together with its three-dimensional structure, are still lacking, HDX-MS was exploited to detect the dynamic movements and conformational changes of CysK and CysE upon CS complex formation. HDX revealed that several regions of CysK show significant changes upon binding to CysE, confirming the engagement of the active site and its participation in complex formation. Importantly, we also identified, for the first time, regions on CysE that display changes in structural flexibility upon complex formation, suggesting allosteric movements induced by CysK binding. HDX data also allowed to highlight plausible pathways for the propagation of allosteric signals, both within the CysE trimer and between the trimers. Following binding of one CysE C-terminus to one CysK subunit, in agreement with the 3:2 binding stoichiometry and recent SAXS data, CysE subunits appear to maintain symmetry in solvent accessibility, supporting the notion that the unbound C-terminal tails fold back into the active site. Inter-trimer communication stabilizes the trimers interface, inducing an allosteric reorganization responsible for the non-random binding of the second dimer of CysK. The new molecular insights into dynamics affecting CysE and CysK upon CS complex formation could be exploited for the design of novel antimicrobial agents.

Playing Bayesian jigsaw with PDB structures: Optimizing rigid body representations for integrative modeling

Tanmoy Sanyal[1, 3], Brian T. Chait[2], Michael O'Donnell[5, 6], Andrej Sali[1, 3, 4]

[1]Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, CA, [2]Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, NY, [3]Quantitative Biosciences Institute, University of California San Francisco, CA, [4]Department of Pharmaceutical Chemistry, University of California San Francisco, CA, [5]DNA Replication Laboratory, The Rockefeller University, NY, [6]Howard Hughes Medical Institute, Chevy Chase, MD

Integrative modeling of macromolecular protein complexes typically involve representing structurally available components of the complex from X-ray crystallography, cryo-electron microscopy or comparative modeling, as rigid bodies, and sampling their relative spatial arrangements to generate an ensemble of models that are consistent with input experimental information. While there exist some sparse guidelines in the literature for optimizing the resolution of flexible regions used to represent missing residues within and between the rigid components, the process of defining rigid regions is entirely ad-hoc, often leading to several tedious rounds of trial and error between the computational modeler and their experimental collaborator, alternating between guessing (based on prior experience and literature) a rigid body definition and running a computationally expensive structural sampling using this definition, until the model achieves both a desirable precision and a quantifiably good fit to input data. Here, we combine concepts from stochastic graph partitioning and on-lattice spin-glass statistical mechanical models, and develop a maximum likelihood algorithm to automatically score and sample through alternate rigid body definitions and provide the modeler with an optimal set of rigid components for subsequent structural assembly. We illustrate our method with the example of a pathological case of structural inconsistency between a cryo-EM structure of the eleven-subunit yeast helicase complex CMG, and a high density dataset of ~1000 chemical crosslinks that describe its binding to the Mcm10 protein. Our algorithm efficiently dissects CMG into an optimal collection of rigid bodies without requiring the modeler to manually enumerate all such cases, and the optimal set of rigid components significantly increases the overall crosslink satisfaction compared to treating the entire CMG as a single rigid body. The method is standalone and could serve as a general tool for pre-processing available structures to ensure greater consistency with available crosslink data, prior to carrying out computationally expensive structural sampling.

Structural and Functional Analysis of the Streptococcus gordonii SspB C123 Domains

Norbert Schormann[1], Sangeetha Purushotham[1], Joshua Mieher[1], Manisha Patel[1], Champion Deivanayagam[1]

[1]University of Alabama at Birmingham

Streptococcus gordonii is a member of viridans streptococci and an early colonizer of the tooth surface, and makes up the base layer of plaque to which other oral microbes including the pathogenic *Streptococcus mutans* adhere through their surface anchored adhesins to colonize the oral cavity. *S. gordonii* SspB belongs to the Antigen I/II(AgI/II)-family, which presents one of the best-studied cell-wall anchored adhesins. The C-terminal region of SspB consists of three tandemly connected individual domains that display the DEv-IgG fold. These domains contain a conserved Ca^{2+} binding site and isopeptide bonds, and bind to Salivary Agglutinin (SAG; also known as Gp340 or SALSA). Here, we report the structural and functional characterization of the SspB_C123 domain resolved at 2.7 Å. Although the individual C-terminal domains of *S. mutans* AgI/II and *S. gordonii* SspB show both a high degree of sequence and structural homology, superposition of these structures highlights substantial differences in their electrostatic surface plots, and this can be attributed to the relative orientation of each individual domain (C1, C2 and C3) with respect to each other. Such differences in surface potential can be attributed to previously reported functional differences in binding of specific extracellular matrix molecules (ECM). Our studies have confirmed that affinity to Gp340 or its scavenger receptor cysteine-rich (SRCR) domains requires two of the three domains of SspB_C123, namely C12 or C23, which is different from AgI/II. We present reasons for this observed functional difference between AgI/II_C123 and SspB_C123 using protein-protein docking, where for AgI/II_C123 only the C12 domains bind to SRCR1 while for SspB_C123 binding is observed for C12 and C23 domains. In addition, SspB_C123 exhibits another unique property compared to AgI/II_C123, since its C2 domain directly interacts with the *Porphyromonas gingivalis* Mfa1 fimbrial antigen receptor while AgI/II_C123 does not.

Structural Basis of Promoter Escape by Bacterial RNA Polymerase

Yeonoh Shin[1], Katsuhiko Murakami[1]

[1]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park

Promoter escape is one of key steps in gene expression and a stage of transcription cycle when RNA polymerase (RNAP) transitions from promoter open complex (RPo) to initially transcribing complex (ITC) to elongation complex (EC). During this process, RNAP is disengaged from a promoter by separating a promoter recognizing σ factor from a catalytic core enzyme by which RNAP extends nascent RNA and also single-stranded DNA region from a transcription start site to downstream DNA. Structures of RPo, ITC and EC were determined by traditional X-ray crystallography and recently by cryo-electron microscopy (cryo-EM); however, structures engaged in the promoter escape could not been solved due to their transient and dynamic states. Therefore, the structural basis of promoter escape remains established. Previously, we established the structural basis of ribosomal RNA (rRNA) transcription and its regulation by global gene regulators DksA and ppGpp by determining four sets of cryo-EM structures of the Escherichia coli σ70 RNAP and rRNA promoter complex (Shin et al., Nature Commun. 2021). In this study, to investigate the mechanism of promoter escape, we added NTPs to RNAP and rRNA promoter complex to initiate RNA synthesis and determined a series of cryo-EM structures of ITCs containing de novo synthesized nascent RNAs from 6-mer to 20-mer in lengths. The structures show that the σ domains 2 and 3 dissociate from their binding sites of core zyme before the σ domain 4 separating from the core enzyme. The ITC containing 12-mer RNA shows that the domain 2 and non-conserved region of σ70 (σNCR) swing toward the B' clamp toe (B'CT) domain to establish an interaction, which allows σ70 binding on the core enzyme during RNA synthesis. This study proposes a new transcription cycle of bacterial RNAP retaining σ70 during entire process of RNA synthesis.

A New Visual Design Language To Create A World In A Cell

Jitin Singla[1,2,] Kylie Burdsall[1,] Brian Cantrell[2,] Jordan R. Halsey[3,] Oscar Li[2,] Alex McDowell[2,] Colleen McGregor[2,] Sanraj Mittal[1,] Raymond C. Stevens[1,] Shaoyu Su[4,] Alexandra Thomopoulos[2,] Theotime Vaillant[2,] Kate L. White[1,] Bryan Zhang[2,] Helen M. Berman[1,2,5,*]

1 Bridge Institute, Michelson Center for Convergent Bioscience, University of Southern California, Los Angeles, California 90089, USA 2 World Building Media Lab, Media Arts + Practice, School of Cinematic Arts, University of Southern California, Los Angeles, California 90089, USA

3 John C. Hench Division of Animation + Digital Arts, School of Cinematic Arts, University of Southern California, Los Angeles, California 90089, USA
4 Immersive Media Lab, School of Cinematic Arts, University of Southern California, Los Angeles, California 90089, USA
5 Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, USA * Corresponding authors: Helen M. Berman (helen.berman@rcsb.org)

Abstract
We have developed a unique platform equipped with novel visual design language for designing an immersive experience of the inside of a cell. Examples of immersive experiences include virtual, augmented or mixed reality. This language expresses the structural features of the cellular components and imparts unique signatures to each molecule. We modeled the insulin secretion pathways of the pancreatic beta cell. We also developed an algorithm whereby the cellular components were directly translated into models composed solely of tetrahedra. Tetrahedra are infinitely scalable and can be precisely connected. The language enables complex animation and user interaction in multiple interactive media. The platform presented here provides tools to bring the immersive experiences of biological systems to a broad audience including K-12 and college students. It can also serve as a collaboration tool for scientific community.

Potential inhibitors of papain-like protease (PLpro) identified by virtual screening

Olena Stasyuk[1], Todd R. Ryder[1]

[1]Southern Connecticut State University

Protease inhibitors are widely used in the treatment of HIV/AIDS and hepatitis C. More recently, inhibitors of the papain-like protease (PLPro) have received attention as antiviral agents for coronaviruses, including SARS-CoV-1 and SARS-CoV-2. We carried out virtual screening using the structure of a known ligand bound to SARS-CoV-2 to identify novel compounds with the potential to inhibit this target. The top hits were then profiled to determine which ones are the most promising for further studies.

Race to the Bottom: Competition and Quality in Science

Ryan Hill[1], Carolyn Stein[2]

[1]Northwestern University, [2]Massachusetts Institute of Technology

This paper investigates how competition to publish first and thereby establish priority impacts the quality of scientific research. We begin by developing a model where scientists decide whether and how long to work on a given project. When deciding how long to let their projects mature, scientists trade off the marginal benefit of higher quality research against the marginal risk of being preempted. The most important (highest potential) projects are the most competitive because they induce the most entry. Therefore, the model predicts these projects are also the most rushed and lowest quality. We test the predictions of this model in the field of structural biology using data from the Protein Data Bank (PDB), a repository for structures of large macromolecules. An important feature of the PDB is that it assigns objective measures of scientific quality to each structure. As suggested by the model, we find that structures with higher ex-ante potential generate more competition, are completed faster, and are lower quality. Consistent with the model, and with a causal interpretation of our empirical results, these relationships are mitigated when we focus on structures deposited by Structural Genomics groups. These are scientists who – by nature of their employment position – are less focused on publication and priority.Link to working paper: https://economics.mit.edu/files/20679

Advanced usage of in-house SAXS for structural biology

Jan Stránský[1], Andreas Keilbach[2], Jan Dohnálek[1]

[1]Institute of Biotechnology, Academy of Sciences of the Czech Republic, Průmyslová 595, Vestec, [2]Anton Paar GmbH, Anton-Paar-Strasse 20, Graz

Small angle X-ray scattering (SAXS) serves as one of the complementary methods in structural biology. SAXS allows both characterization and low resolution structural studies of macromolecules, with the advantage of being able to measure samples in solution. Several complementary methods can be directly coupled with SAXS. Thanks to modern high flux X-ray sources such as MetalJet (Excillum), such applications are now possible not only at synchrotrons, but also in-house. Here, in situ UV-Vis absorption spectroscopy and size exclusion liquid chromatography coupled with SAXS, as available in the Centre of Molecular Structure, will be presented.
The Centre of Molecular Structure is equipped with a SAXSPoint 2.0 (Anton Paar) instrument with a MetalJet C2+ X-ray source and an Eiger R 1M detector. Samples can be loaded to a variety of capillaries using an autosampler. The set of temperature-controlled sample stages includes a low noise cell and a capillary capable of in situ UV/Vis absorption spectrometry. Recently, a dedicated liquid chromatography system AktaGo (GE Healthcare) was introduced.
The Centre of Molecular Structure is supported by: MEYS CR (LM2018127); project Czech Infrastructure for Integrative Structural Biology for Human Health (CZ.02.1.01/0.0/0.0/16_013/0001776) from the ERDF; UP CIISB (CZ.02.1.01/0.0/0.0/18_046/0015974), and ELIBIO (CZ.02.1.01/0.0/0.0/15_003/0000447).

Structural, Biochemical, and Functional Characterization of BcfH, a Novel DsbA-like Protein From Salmonella Typhimurium.

Pramod Subedi[1], Jason J. Paxman[1], Geqing Wang[1], Lilian Hor[1], Yaoqin Hong[2], Anthony D. Verderosa[2], Andrew E. Whitten[3], Santosh Panjikar[4, 5], Carlos F. Santos-Martin[1], Jennifer L. Martin[6, 7], Makrina Totsika[2], Begoña Heras[1]

[1]La Trobe Institute for Molecular Science, La Trobe University, VIC 3086, Australia, [2]School of Biomedical Sciences, Queensland University of Technology, QLD 4059, Australia, [3]Australian Centre for Neutron Scattering, ANSTO, Lucas Heights, NSW 2234, Australia, [4]Macromolecular Crystallography, Australian Synchrotron, ANSTO, Clayton VIC 3168, Australia, [5]Department of Biochemistry and Molecular Biology, Monash University, Clayton VIC 3800, Australia, [6]Griffith Institute for Drug Discovery, Nathan QLD 4111, Australia, [7]Vice-Chancellor's Unit, University of Wollongong, Northfields Ave, Wollongong NSW 2522, Australia

Bacteria use folding enzymes to produce functional virulence factors. These foldases include the Dsb family of proteins, which catalyze a key step in the protein-folding pathway, the introduction of disulfide bonds. The Dsb oxidative system, which includes an oxidative DsbA/DsbB pathway and an isomerase DsbC/DsbD pathway, is present in numerous bacterial species. Conventionally, Dsb proteins have specific redox functions with monomeric and dimeric Dsbs exclusively catalyzing thiol oxidation and disulfide isomerization, respectively. This contrasts with the eukaryotic disulfide forming machinery where the modular thioredoxin protein PDI mediates thiol oxidation and disulfide reshuffling. In this study, we identified and structurally and biochemically characterized a novel Dsb-like protein from Salmonella enterica termed BcfH and defined its role in virulence. Encoded by a highly conserved bcf (bovine colonization factor) fimbrial operon, the Dsb-like enzyme BcfH forms a trimeric structure, exceptionally uncommon among the large and evolutionary conserved thioredoxin superfamily. This protein also displays very unusual catalytic redox centers, including an unwound α-helix holding the redox active site and a trans proline instead of the conserved cis proline active site loop. Remarkably, BcfH displays both thiol oxidase and disulfide isomerase activities contributing to Salmonella fimbrial biogenesis. Typically, oligomerization of bacterial Dsb proteins modulates their redox function, with monomeric and dimeric Dsbs mediating thiol oxidation and disulfide isomerization, respectively. The present study demonstrates a further structural and functional malleability in the thioredoxin-fold protein family. BcfH trimeric architecture and unconventional catalytic sites permit multiple redox functions emulating in bacteria the eukaryotic protein disulfide isomerase dual oxido-reductase activity.

Identification of new integrin ligands by virtual screening of protein data bank using the integrin headpiece as a target.

Yoshikazu Takada, Yoko K. Takada

[1]Department of Dermatology, University of California Davis School of Medicine

Integrins are a family of cell surface heterodimers and act as receptors for extracellular matrix ligands, cell-surface ligands (ICAM-1, VCAM-1), and small soluble ligands. Previous studies showed that antagonists to integrin αvβ3 inhibit fibroblast growth factor-2 (basic FGF, FGF2) signaling suggested that integrin αvβ3 is involved in growth factor signaling through crosstalk and positively regulates growth factor signaling.
For the last 15 years, we identified many new integrin ligands by virtual screening of protein data bank by virtual screening using docking simulation (autodock) using the integrin αvβ3 headpiece as a target. The procedures of docking simulation is briefly described. Docked poses were clustered and usually the first cluster represents the most likely docking poses (We select candidate ligands with docking energy below -20 kcal/mol). We studied the docked models by introducing mutations into the predicted integrin-binding interface of the potential ligands. In vitro binding assays, wild-type ligands bound to integrin αvβ3 as predicted and the binding is suppressed by antagonists to αvβ3, indicating the binding is specific. Notably, most of the ligand mutants defective in integrin binding were defective in signaling (as growth factors) and act as antagonists.
Most of the potential ligands were growth factors, including FGF1, IGF1, neuregulin-1, chemokines (e.g., fractalkine, and SDF-1), indicating that many growth factors bind to integrins. A current model of integrin-growth factor crosstalk is that integrins contribute to growth factor signaling almost exclusively through integrin binding to ECM, and growth factors only bind to their cognate receptors, and two independent signals merge inside the cells. We previously reported that FGF1 and FGF2 directly bind to integrin αvβ3 and subsequent ternary complex formation (integrin-FGF-FGFR) is required for their signaling functions (the ternary complex model). We showed that the ternary complex model can be applied for several other growth factors (e.g., IGF1, IGF2, neuregulin-1, fractalkine, and CD40L). This suggests that direct binding of integrin αvβ3 to growth factors positively regulates signaling by these growth factors.

Virtual screening of protein data bank was useful in elucidating the role of integrins in growth factor signaling. This strategy generated many potential therapeutics (e.g., antagonistic growth factors).

Activity prediction by a Convolutional Deep Learning method using PDB data

Takeshi Tanaka[1], Takao Matsuzaki[1], Ken Ikeda[1], Hirotsugu Komatsu[1]

[1]Interprotein Corporation, Sagamihara, Kanagawa 252-0131, Japan

A Convolutional Deep Learning (CDL) program has been developed to predict Ki or IC_{50} based on the binding 3D-structures of protein and ligand [1]. Among ca. 19,000 complex structures in PDB, we chose 2,600, 600 and 810 structures as a training, validation and test data sets, respectively. Input are the coordinates of the complexes and Ki or IC_{50} values. Output is a predicted class number of activity, which is assigned according to the order of the magnitude of activity. The class-0 compounds has an activity less than 10^{-9} mol/L, the class-1: from 10^{-9} to less than 10^{-8} mol/L, ---, the class^{-7}: from 10^{-3} to less than 10^{-2} mol/L. If the prediction error of +/-1 class number is acceptable in a practical use, success rate was 86% for 2,600 compounds in the training data and 70% for 810 compounds in the test data. The success rate is comparable with that of good Free Energy Perturbation Calculation. The elapsed time is about 3 hours on an Intel Core i7-6700K 4GHz, which means 5 orders of magnitude shorter than FEP calculation. The program, CDL-Predict, was registered in SOFTIC, No. P-10808-1, Feb. 7, 2018. The test application to library screening for Runx1 inhibitor search using docking 3D-coordinates of protein-ligand complexes suggests that CDL-Predict will improve success rate of hit finding. The program has been used successfully to find active compounds and also to enhance the efficiency of hit-to-lead and lead optimization processes.

[1] H. Komatsu, K. Ikeda, T. Tanaka and T. Matsuzaki, Development of Practical Artificial Intelligence System for Drug Discovery and Its Application to Activity Prediction of Small Molecule Protein-Protein Interaction Modulators. *J. Biol. Macromol.*, **19** (1), 5-19, 2019.

Protein Structures and their features in UniProt

Nidhi TYAGI[1], UniProt Consortium [1,2,3]

1 EMBL-European Bioinformatics Institute, Cambridge, UK, 2 SIB Swiss Institute of Bioinformatics, Geneva, Switzerland, 3 Protein Information Resource, Georgetown University, Washington DC & University of Delaware, USA

Annotation of proteins using structure-based analyses is an integral component of the UniProt Knowledgebase (UniProtKB). There are nearly 172,000 experimentally determined 3-dimensional structures of proteins deposited in the Protein Data Bank. UniProt works closely with the Protein Databank in Europe (PDBe) to map these 3D structural entries to the corresponding UniProtKB entries based on comprehensive sequence and structure-based analyses, to ensure that there is a UniProtKB record for each relevant PDB record and to import additional data such as ligand-binding sites from PDB to UniProtKB.

SIFTS (Structure Integration with Function, Taxonomy and Sequences), which is a collaboration between the Protein Data Bank in Europe (PDBe) and UniProt, facilitates the link between the structural and sequence features of proteins by providing correspondence at the level of amino acid residues. A pipeline combining manual and automated processes for maintaining up-to-date cross-reference information has been developed and is run with every weekly PDB release.

Various criteria are considered to cross-reference PDB and UniProtKB entries such as a) the degree of sequence identity (>90%) b) an exact taxonomic match (at the level of species, subspecies and specific strains for lower organisms) (c) preferential mapping to a curated SwissProt entry (if one exists) or (d) mapping to proteins from Reference/Complete proteome or (e) mapping to the longest protein sequence expressed by the gene. Some cases are inspected manually by a UniProt biocurator using a dedicated curation interface to ensure accurate cross-referencing. These cases include short peptides, chimeras, synthetic constructs and de novo designed polymers.

The SIFTS initiative also provides up to date cross referencing of structural entries to literature (PubMed), taxonomy (NCBI), Enzyme database (IntEnz), Gene Ontology annotations (GO) and protein family classification databases (InterPro, Pfam, SCOP and CATH).

Also, a pipeline has been developed to automatically import data from PDB to enhance the unreviewed records in UniProtKB/TrEMBL. This includes details of residues involved in the binding of biologically relevant molecules including substrate, nucleotides, metals, drugs, carbohydrates and post-translational modifications.

UniProt has successfully completed the non-trivial and labour intensive exercise of cross referencing 129,389 PDB entries (365,524 polypeptide chains) to 43,709 UniProtKB entries.

Symmetry and shape-size complementarity for the creation of supramolecular POM-protein assemblies

Laurens Vandebroek[1], Arnout R. Voet[1]

[1]Department of Chemistry, KU Leuven, 3001 Heverlee (Leuven), Belgium

The controlled formation of protein supramolecular assemblies is challenging, but it could provide an important route for the development of hybrid biomaterials. Novel bioinorganic hybrid materials based on proteins and inorganic clusters have enormous potential for the development of hybrid catalysts that synergistically combine properties of both materials. One approach was to combine designer proteins with metal-oxo clusters that have a matching symmetry. Here we report the creation of hybrid assemblies between a computationally designed symmetrical protein Pizza6-S and different polyoxometalates with matching symmetry: the tellurotungstic Anderson-Evans ($Na_6[TeW_6O_{24}]\cdot 22H_2O$) (TEW); Keggin ($H_4[SiW_{12}O_{40}]\cdot 6H_2O$) (STA); and 1 : 2 CeIII-substituted Keggin ($K_{11}[Ce^{III}[PW_{11}O_{39}]_2]\cdot 20H_2O$) (Ce-K). This results in the formation of complexes with clearly defined stoichiometries in solution. Crystal structures validate the complexes as building blocks for the formation of larger assemblies.(1) Another approach explored the role of size and shape complementarity between the designer protein and metal-oxo clusters. In this work, we demonstrate the formation of well-defined complexes between the eightfold symmetrical designer protein Tako8 and soluble metal-oxo clusters from the family of Anderson–Evans, Keggin, and ZrIV-substituted Wells–Dawson polyoxometalates. A combination of X-ray crystallography and solution studies showed that metal-oxo clusters are able to serve as linker nodes for the bottom-up creation of protein-based supramolecular assemblies. Our findings indicate that clusters with larger size and negative charge are capable of modulating the crystal packing of the protein, highlighting the need for a size and shape complementarity with the protein node for optimal alteration of the crystalline self-assembly.(2)

1. L. Vandebroek, H. Noguchi, K. Kamata, J.R.H. Tame, L. Van Meervelt, T.N. Parac-Vogt and A.R.D. Voet, Chem. Commun., 2020,56, 11601-11604
2. L. Vandebroek, H. Noguchi, K. Kamata, J.R.H. Tame, L. Van Meervelt, T.N. Parac-Vogt and A.R.D. Voet, Cryst. Growth Des. 2021, 21, 2, 1307–1313

Elucidation of MEK2 (MAP2K2) 3D Structure Model using Fold Recognition (I-Tasser) and Molecular Dynamics Simulation Based Analysis

Aravind Venkatasubramanian[1, 2], Debjani Dasgupta[1], Pramodkumar P. Gupta[1]

[1]School of Biotechnology and Bioinformatics, D Y Patil Deemed to be University, Navi Mumbai, Maharash, [2]Sunandan Divatia School of Science, NMIMS Deemed to be University, Mumbai, Maharashtra, India

The unavailability of a reliable and accurate protein structure is a strong hindrance to the research that aims to treat diseases that arise due to abnormal protein structure or signaling pathways. Thus, it becomes necessary to elucidate the structure of a protein using techniques like NMR, X-Ray Crystallography, Cryo-Electron Microscopy etc. among others. Alternatively we can also employ Bioinformatics techniques and in-silico modelling and go further with computer aided drug discovery (CADD). MAP2K2 or MEK2 is a protein which participates in the cell cycle via MAPK pathway. Due to its crucial role in regulation of cell division, it becomes a target of cancer research. Modelling of the protein was done since the structure was either partially available or it was of low resolution. In this study we propose a model generated by using Fold Recognition (I-Tasser) and subsequent MD simulation using GROMACS was carried out for 50 ns. The I-Tasser returned a model with C-score of 1.60 placing 98.2% of residues under allowed regions and 1.8% under disallowed regions. The energy minimization reported a potential energy of -1.0748888 e+06 kJ/mol, the residues under allowed regions were 98.8% and 1.2% under disallowed regions. The energy minimized structure yielded an RMSD value of 0.459 when aligned to the template protein (3eqd) and RMSD of 0.841 when aligned to MEK2 3D Structure PDB ID 1S9I. Finally the MD simulated results optimized with potential energy of -9.55204e+05 kJ/mol and total energy -7.95459 e+05 kJ/mol. On aligning the simulated structure to template 3eqd, an RMSD of 2.251 was obtained and when the alignment was done with 1S9I, an RMSD of 2.185 was obtained. The present work designates the full length of MEK2, 400+ amino acid residues. The results show a significant model and will be helpful in further drug discovery projects.

Structural and functional characterization of the glycosyltransferase PglA from Campylobacter concisus

Nemanja Vuksanovic[1], Jozlyn R. Clasman[1], Hannah M. Bernstein[2], Barbara Imperiali[2, 3], Karen N. Allen[1]

[1]Department of Chemistry, Boston University, Boston, MA 02215, [2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, [3]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139

Glycoconjugates represent a class of macromolecules relevant for bacterial viability and virulence, as they play a key role in host adherence. Despite the capability to synthesize a diverse set of glycoconjugates, bacteria share common mechanisms, such as en bloc glycoconjugate synthesis. The biosynthesis involves transfer of glycan moieties from a nucleotide activated sugar to a lipid acceptor undecaprenyl phosphate (Und-P). The glycan-based macromolecules are diversified through the action of glycosyltransferases, which vary in their donor-molecule selectivity. Understanding the factors influencing selectivity of glycosyltransferases could lead to the development of novel highly selective antibiotics, in the light of the emerging threat of antibiotic resistance. In this work we report the first structural characterization of a Type-B glycosyltransferase, PglA, from Campylobacter concisus, involved in the N-linked glycosylation pathway. Using X-ray crystallography, we have determined a 2.5 Å structure of PglA, bound to its native donor substrate, uridine 5′-diphospho-N-acetylgalactosamine (UDP-GalNac). Guided by this structural information we designed a series of site-directed variants in order to gain understanding of the factors influencing PglA substrate specificity.

Validation of SARS-CoV-2-related PDB structures

Alexander Wlodawer[1], Zbigniew Dauter[1], Wladek Minor[2], Mirek Gilski[3, 4], Ivan G. Shabalin[2], Bernhard Rupp[5, 6], Dariusz Brzezinski[4, 7], Marcin Kowiel[4], Mariusz Jaskolski[3, 4]

[1]National Cancer Institute, Frederick, MD, USA, [2]University of Virginia, Charlottesville, VA, USA, [3]A. Mickiewicz University, Poznan, Poland, [4]Institute of Bioorganic Chemistry, Poznan, Poland, [5]k.-k. Hofkristallamt, San Diego, CA, USA, [6]Medical University Innsbruck, Austria, [7]Poznan University of Technology, Poznan, Poland

The response of the scientific community to the COVID-19 pandemic has been remarkable, with over 800 structures related to SARS-CoV-2 submitted to the PDB by mid-January 2021. With structural data obtained very rapidly by many groups using different techniques and varied experimental approaches, such efforts may lead to inevitable mistakes, errors, and interpretational problems. Moreover, extracting the most relevant information might not be straightforward, thus, as a result, a number of dedicated databases have been created. To help biomedical researchers, who may not be experts in structural biology, navigate through and absorb the flood of structural information, we have created an online resource, covid19.bioreproducibility.org, that aggregates expert-validated information about SARS-CoV-2-related macromolecular models. The crystal structures in our database have been standardized and if the quality of electron density fit was found to be suboptimal, such models were carefully re-refined. In particular, a number of protein complexes with inhibitors, which are most vital for drug design, were found to pose significant problems. Some of the issues could be traced to erroneous definition of geometrical restraints for ligands and to a general problem of lack of such information in PDB deposits. Several problems with ligand definition by the PDB itself were also noticed. A large number of deposits resulting from PanDDA projects were not amenable to remediation due to inconsistent contents and description of deposited data. Nevertheless, our critical analysis of a large number of protein structures that are potential or actual targets for drug design, all determined within less than a year using modern experimental tools, is important not only in a current medical emergency but should be also useful as a paradigm study for future projects of high interest to biomedical community.

qPTxM: a Tool for Quantifying [Evidence in a Density Map for] Post-Transcriptional Modifications

Iris D. Young[1], Vanja Stojkovic[2], Alexander Myasnikov[3], Adam Frost[3, 4], Danica G. Fujimori[2, 3, 5], James S. Fraser[1, 4]

[1]Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, [2]Department of Cellular and Molecular Pharmacology, University of California San Francisco, [3]Department of Biochemistry and Biophysics, University of California San Francisco, [4]Quantitative Biosciences Institute, University of California San Francisco, [5]Department of Pharmaceutical Chemistry, University of California San Francisco

Defined as the smallest distance by which two features can be separated and still independently resolvable, the metric of resolution is conceptually water-tight but in practice highly dependent on how it is measured, and whereas resolution is a global (at worst, anisotropic) metric for crystallographic maps, it is a local feature of cryoEM maps that is strongly influenced by processing steps such as masking. As cryoEM maps make up an increasing proportion of near-atomic-resolution macromolecular structures, this makes comparisons across maps quite difficult. Here we present a technique-agnostic computational tool for assessing map quality by the proxy of discernability of known features (github.com/fraser-lab/qptm, Stojkovic et al 2020). The profile of densities composing a feature (say, a C5-methylation) can be predicted from the atomic positions, map type, and resolution, and by comparing a predicted density profile to an experimental one, we may assign some measure of believability to the feature. In a structure with many such features, we may set a threshold for "believable enough" and sort our features into detectable and undetectable groups, generating true positives, true negatives, false positives and false negatives. We are then able to compute sensitivity, selectivity, and overall accuracy, which track with the overall quality of the map.
In addition, cryoEM of biological macromolecules has now breached the atomic resolution barrier, making de novo detection of features tenable, and crystallographic structures are routinely high enough resolution for this exploration. qPTxM can alternatively be used to identify promising sites in a structure where modifications might be present even in the absence of biochemical (e.g. MS) data. We also provide a plugin for the graphics program Coot (Emsley et al 2010) to allow a researcher to jump directly to these sites of possible modifications.
This work is supported by NIH/NIGMS fellowship F32GM133129.
References:
Stojkovic V, Myasnikov AG, Young ID, Frost A, Fraser JS, Fujimori DG. Assessment of the nucleotide modifications in the high-resolution cryo-electron microscopy structure of the Escheria coli 50S subunit. Nucleic Acids Research, 2020, 48: 2723-2732.
Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Cryst., 2010, D66: 486-501.

3D Physical Models of the SARS-CoV-2 Spike Protein: Investigating Mutations Associated with High Infectivity and Transmissibility

Dean Young, Lauren McDonald, Patricia Hill, Betsy M. Martinez-Vaz

Hamline University, Saint Paul MN

The SARS-COV-2 emerged in late 2019, and quickly became a global pandemic, killing millions. The spike glycoprotein plays an important role in the process of viral infection by recognizing and binding to the ACE2 receptor in host cells. Recently, it was discovered that SARS-CoV-2 viruses with a spike mutation at position 614 are more contagious than their wild-type ancestor. The D614G substitution eliminates a hydrogen bond with a threonine residue (T859) in an adjacent protein chain, allowing the spike protein to assume more readily an open conformation.The open state is further stabilized by shortening of a hydrogen bond between alanine 647 and glycine 614 in the mutant protein. This in turn allows for the virus to more readily attach to cell receptors, leading to heightened infectivity. The main goal of this research was to create physical 3D printed models of the wild-type and mutant spike proteins to visualize the molecular changes caused by the D614G mutation.The Protein Data Bank files 6vxx and 6xs6 were used to conduct structural analyses of the wild-type and D614G spike mutants. To visualize the conformational changes caused by the D614G mutation, the protein structure alignments were performed using Pymol. A combined file containing the superimposed structures was imported into Jmol to create the scripts for model construction and 3D printing. The 3D physical model highlights mutations like D614G, as well as the N501Y, E484K, and K417N variants, which were recently discovered in highly transmissible SARS-COV-2 strains isolated in England and South Africa. An additional detailed model was created to illustrate the biochemical interactions at the D614G mutation site. In this model, important amino acids such as Asp614, Ala647, and Thr859 are highlighted due to their role in hydrogen bonding and binding to the ACE2 receptor. To supplement the 3D models and assess students' understanding of the structure and function of the SARS-CoV-2 spike protein, a series of Jmol tutorials were developed.This exercise will enable undergraduate students and faculty to visualize and better understand how single amino acid mutations can lead to changes in viral infectivity and transmissibility.

Integrative modeling of the Smc5/6 complex

You Yu[2], Shibai Li[7], Zheng Ser[8, 9, 10], Tanmoy Sanyal[1, 5], Koyi Choi[7], Bingbing Wan[4, 7], Huihui Kuang[3], Andrej Sali[1, 5, 6], Alex Kentsis[8, 9], Dinshaw J. Patel[2], Xiaolan Zhao[7]

[1]Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, CA, [2]Structural Biology Program, Memorial Sloan Kettering Cancer Center, NY, [3]Simons Electron MIcroscopy Cener, New York Structural Biology Center, NY, [4]Sanghai Center for Systems Biomedicine, Sanghai Jiao Tong University, Sanghai, China, [5]Quantitative Biosciences Institute, University of California San Francisco, CA, [6]Department of Pharmaceutical Chemistry, University of California San Francisco, CA, [7]Molecular Biology Program, Memorial Sloan Kettering Cancer Center, NY, [8]Molecular Pharmacology Program, Memorial Sloan Kettering Cancer Center, NY, [9]Tow Center for Developmental Oncology, Dept of Pediatrics, Memorial Sloan Kettering Cancer Center,NY, [10]Tri-Institutional PhD Program in Chemical Biology, NY

Structural Maintenance of Chromosomes (SMC) complexes are important chromatin modulators. The Smc5/6 complex is directly involved in the regulation and repair of DNA in eukaryotes. Despite overarching topological similarities with other members of the Smc family like cohesin and condensin, Smc5/6 has distinct functions whose molecular bases remain poorly elucidated. Here, we develop an integrative model of the budding yeast Smc5/6 in complex with the SUMO ligase Nse2, and the Nse5 and Nse6 proteins. Specifically, we use crystal structures of Nse2 available from the PDB, our cryo-electron-microscopy derived 3 Å resolution structure of the Nse5/6 subcomplex, comparative models of the globular head and hinge regions of Smc5/6 and molecular mechanics models of their coiled-coil arm regions. A multiscale model representation consisting of the aforementioned structurally covered components as rigid bodies, and all other regions as (coarse-grained) flexible beads is scored using spatial restraints designed from our dataset of ~337 unique inter-residue chemical crosslinks, in addition to fundamental physical interactions like chain connectivity and excluded volume. Structural degrees of freedom of alternate models that satisfy the input information are sampled, and a representative ensemble of good scoring models are filtered out to describe the pentamer at a precision of 30 Å. The integrative model qualitatively and quantitatively corroborates the shape of the Smc5/6 arm region from negative staining 2D electron microscopy class averages, which were not used as input data for modeling. Unlike several examples of other SMC structures in the past literature, as well as contemporary structural investigations of the Smc5/6 complex, we provide a rigorous 3D model of the coiled-coil arm region that is structurally consistent with the input data. We find that in contrast to cohesin and condensin, the Smc5/6 arms do not fold back onto themselves; rather they interact uniquely with the Nse-s 2, 5 and 6, such that the Nse5/6 subcomplex organizes spatially as a linchpin that connects distal parts of the pentameric holocomplex.

A pipeline to find all quadruplex structures in the PDB

Tomasz Zok[1], Mariusz Popenda[2], Natalia Kraszewska[1], Joanna Miskiewicz[1], Paulina Pielacinska[1], Michal Zurkowski[1], Marta Szachniuk[1, 2]

[1]Institute of Computing Science, Poznan University of Technology, [2]Institute of Bioorganic Chemistry, Polish Academy of Sciences

Quadruplexes are intricate, four-stranded motifs found in DNA and RNA structures. They usually form in G-rich regions of genomes. Quadruplexes play multiple roles in cellular processes, and many sources link their abundance or lack thereof with certain diseases.
Structurally, the motif consists of tetrads, i.e., planar arrangements of fours of nucleotides. Tetrads compose a stack which together with loops surrounding it is called the quadruplex. Canonical motifs of this type have three loops of varying length and uninterrupted tracts of guanines. Based on these traits and the glycosidic bond angle values, Webba da Silva proposed a classification of quadruplex types. However, the advances in structural studies of nucleic acids allowed to solve many additional non-canonical quadruplexes which cannot be classified that way. For example, they may consist of another nucleobase other than guanine, or their architecture and tetrad stacking pattern might differ from a canonical quadruplex.
To address these issues, we defined the ONZ classification upon the base-pairing pattern present in a tetrad. The O, N, or Z classes combined with strand directionality allow classifying any quadruplex type. We also developed a computational tool ElTetrado to analyze 3D coordinates of a nucleic acid to detect quadruplex motifs and categorize them accordingly. ElTetrado is part of an automatic pipeline that processes every newly deposited structure in the PDB. The pipeline feeds a database whose content allows computing statistics regarding Webba da Silva's and ONZ classification schemes. Besides, we visualize each quadruplex in several ways:

• A customized VARNA application generates a classical 2D structure diagram of nucleic acid with a modified set of icons for Leontis-Westhof base-pairing types.
• A dedicated R-Chie-based software represents tetrads as arcs according to the ONZ classes.
• Another self-developed tool in the pipeline draws a layer diagram with an automatic layout optimization procedure.

To our knowledge, this is the first pipeline to perform all these operations. It allows the automatic gathering of all quadruplexes present in the PDB-deposited structures. The collected entries are processed to extract classes, parameters, and statistics and visualize them in several ways.

POSTER SESSION 6
Wednesday May 5, 2021 2:20pm-3:00pm EDT

Data and Diversity driven development of a Shotgun crystallisation screen using the Protein Data Bank

Gabriel J. Abrahams[1], Janet Newman[1]

[1]CSIRO Collaborative Crystallisation Centre, Melbourne, Australia

Protein crystallisation has, for decades, been a critical and restrictive step in macromolecular structure determination via X-ray diffraction. Crystallisation typically involves a multi-stage exploration of the available chemical space, beginning with an initial sampling (screening) followed by iterative refinement (optimisation). Effective screening is important for reducing the number of optimisation rounds required, reducing the cost and time required to determine structure. Here, we propose an initial screen (Shotgun II) derived from analysis of the up-to-date Protein Data Bank and compare it with the previously derived (2014) Shotgun I screen. In a further update to that analysis, we introduce the concept of“diversity” as filter to improve screen efficacy across a broad protein space. For this, the C6 metric is evaluated as a diversity proxy against data collected at C3 over thousands of crystallisation experiments. Our data demonstrates that the Shotgun I screen, compared with alternatives, has been remarkably successful over the 6 years it has been in use, indicating that Shotgun II is also likely to be a highly effective screen.

Post-16 protein structure teaching resources for chemistry and biology

David R. Armstrong[1], Peter Hoare[2]

[1]Protein Data Bank in Europe (PDBe), EMBL-EBI, Cambridge, CB10 1SD, [2]STEM Outreach for Schools and Colleges, Newcastle University, NE1 7RU, United Kingdom

We have developed teaching resources for A-Level and early years undergraduate teaching as a collaboration between Protein Data Bank in Europe (PDBe) and Newcastle University School of Chemistry. The resources have been developed to support the teaching and learning of protein structure, function and application for post-16 biology and chemistry. All the available content has been peer-produced by A-level and undergraduate project students and can be used for student self-study, teacher support or for outreach sessions. The resources include theory sheets, worksheets and quizzes across a range of topics relating to protein structure and function. These are augmented by videos explaining some key biochemical concepts mentioned throughout our other resources, along with PyMOL session files to simplify visualisation of these structures and themes in 3D. The resources are available from the PDBe teaching pages at PDBe.org/teaching or directly at tiny.cc/proteinLR

De novo designed proteins in the PDB

Asim Bera[1, 2], Alex Kang[1, 2], Ian Haydon[1, 2], Lauren Carter[1, 2], Lance Stewart[1, 2], David Baker[1, 2, 3]

[1]Institute for Protein Design, University of Washington, Seattle, WA 98105, [2]Department of Biochemistry, University of Washington, Seattle, WA 98105, [3]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98105

The Protein Data Bank (PDB) contains an impressive number of natural protein structures. It also increasingly contains novel structures produced not by evolution but through computational protein design enabling breakthroughs in research and education. In this poster, we present an overview of the structures designed at the Institute for Protein Design, including oligomeric fibers with customized widths, repetitive unbounded sheets, dynamic switch-like proteins, transmembrane beta-barrels, hallucinated sequence-only structures, 3D crystal design, and synthetic mixed-chirality peptides. The development of these structure elucidations is presented on a timeline to illustrate the trajectory of de novo protein design from inactive scaffolds to increasingly functional forms. The atomic coordinates for all of these designs have been or soon will be deposited in the PDB, and we thank the PDB team for their support.

Trends in macromolecular structure data across 50 years of the PDB

wwPDB Biocurators[1]

[1]wwPDB

In 2021, we celebrate 50 years of the Protein Data Bank (PDB) archive - one of the longest-running open access scientific databases - managed collaboratively by the wwPDB (RCSB PDB, PDBe, PDBj, and BMRB). Throughout these 50 years, there has been significant evolution of the data in the PDB archive. This evolution has been driven by a number of factors including development in structural determination techniques, adaptation of biocuration practices, and increase in data capture via updated file formats. Here we will discuss some of the key trends in data across the PDB archive, highlighting how structural biology data has changed over time and how wwPDB biocuration practices have adapted to handle these changes. In addition to long term trends, we will also identify more recent trends and challenges and demonstrate how wwPDB biocuration practices are evolving to meet these challenges in the future.

A fluent python interface for querying structures using the RCSB PDB advanced search API

Spencer E. Bliven[1], Yana Rose[2], Roger M. Benoit[1], Jose M. Duarte[2]

[1]Paul Scherrer Institute, 5232 Villigen PSI, Switzerland, [2]RCSB Protein Data Bank, San Diego Supercomputer Computing Center, University of California San Diego

The RCSB PDB provides the ability to query structures using a rich search interface. Structures can be queried based on textual information, header data, sequence similarity, macromolecular shape, local motifs, and ligand properties. A python interface has recently been developed which allows easy programmatic access to the powerful search API. It provides a pythonic interface to access the search system from bioinformatics tools, scripts, and jupyter notebooks. The recent architectural redesign of rcsb.org provides API access to the full search system (search.rcsb.org)[1]. The *rcsbsearch* package has been developed to ease programmatic access to this API through a pythonic interface. Two syntaxes are supported: an operator-based syntax allows the manipulation of searches using set-like operators, abstracting away the details of the API interactions; and a fluent syntax allows queries to be built up from chains of operations.

The *rcsbsearch* package has been used to identify suitable scaffolds for protein design. Chimeric fusion proteins with rigid connections have been used for epitope display on nanoparticles, structural vaccinology, improved crystal packing, and design for structural biology[2]. For a new design, a series of criteria were established based on oligomeric state, protein size, quaternary symmetry, and other specifications. These could be readily translated into search queries and evaluated to obtain a list of structures. Further analysis was then carried out using the rich ecosystem of python bioinformatics tools to prioritize likely structures for detailed design work and future experimental expression.

The *rcsbsearch* package is documented at https://rcsbsearch.readthedocs.io. Source code is available under the BSD license from https://github.com/sbliven/rcsbsearch.

References

[1] Yana Rose, Jose M. Duarte, Robert Lowe, Joan Segura, Chunxiao Bi, Charmi Bhikadiya, Li Chen, Alexander S. Rose, Sebastian Bittrich, Stephen K. Burley, John D. Westbrook. RCSB Protein Data Bank: Architectural Advances Towards Integrated Searching and Efficient Access to Macromolecular Structure Data from the PDB Archive, Journal of Molecular Biology, 2020. DOI: 10.1016/j.jmb.2020.11.003

[2] Gabriella Collu, Tobias Bierig, Anna-Sophia Krebs, Sylvain Engilberge, Niveditha Varma, Ramon Guixà-González, Xavier Deupi, Vincent Olieric, Emiliya Poghosyan, Roger M. Benoit. Chimeric single α-helical domains as rigid fusion protein connections for protein nanotechnology and structural biology. bioRxiv 2020.09.29.318410; DOI: 10.1101/2020.09.29.318410

Seeing is believing: glycosylation in the crystal structure of human myeloperoxidase

Lucas Krawczyk[1], Shubham Semwal[1], Goedele Roos[1], Pierre Van Antwerpen[2], Julie Bouckaert[1]

[1]Unité de Glycobiologie Structurale et Fonctionnelle, UMR8576 of CNRS and University of Lille, France, [2]Faculty of Pharmacy, Université libre de Bruxelles, Brussels, Belgium

Human myeloperoxidase (MPO) was first isolated in 1941 from purulent pleuritis fluid from tuberculosis patients. When neutrophilic polymorphonuclear leukocytes (neutrophils) entrap microbial or other invasive particulates, they release MPO during degranulation. In a respiratory burst of highly reactive oxygen species, MPO catalyzes the production of hypohalous acids, primarily hypochlorous acid in physiologic situations, from hydrogen peroxide. Mammalian MPO crystal structures were progressively acquired and entered in the Protein Data Bank (PDB) with partial glycosylation identification. Actually, the N-glycan composition of native MPO had been thoroughly investigated with mass spectrometry and shows 5 N-glycans at positions 323, 355, 391, 483 and 729 (1). Recently, MPO's enzymatic activity was reported to depend on the hyper-truncation of 2 out of 5 N-glycosylation sites (2).
In our obtained crystal structure at 2.6 Å resolution containing 4 disulfide-linked homodimers of MPO, an interesting collection of glycans have been characterized. We compared those with the glycans from proteomics studies (1-3) and from 18 human MPO structures in the PDB. We made use of the Symbol Nomenclature for Glycans (SNFG) to illustrate congruence in the experimental data. We found each of the 5 glycosylation sites either non-glycosylated or glycosylated with paucimannosidic, high-mannose and complex N-glycans, with the hyper-truncated N-acetyl-β-D-glucosamine core-type asparagine-linked glycans on Asn355 or Asn391 sites gate keeping the funnel towards the ROS-activated heme group.
In conclusion, more experimental data on the glycosylation of MPO is hereby provided, perfectly illustrating the improvement by crystallographic resolution and its capacity to resolve molecular structure including glycosylation.

1. Van Antwerpen, P., Slomianny, M. C., Boudjeltia, K. Z. et al. and Michalski, J. C. (2010) Glycosylation pattern of mature dimeric leukocyte and recombinant monomeric myeloperoxidase: glycosylation is required for optimal enzymatic activity. J Biol Chem 285, 16351-16359
2. Tjondro, H. C., Ugonotti, J., Kawahara, R., et al. and Thaysen-Andersen, M. (2020) Hyper-truncated Asn355- and Asn391-glycans modulate the activity of neutrophil granule myeloperoxidase. J Biol Chem
3. Reiding, K. R., Franc, V., Huitema, M. G., et al. and Heck, A. J. R. (2019) Neutrophil myeloperoxidase harbors distinct site-specific peculiarities in its glycosylation. J Biol Chem 294, 20233-20245

Structure of a bacterial full-length type 2 isoleucyl-tRNA synthetase reveals the C-terminal tRNA-binding domain

Alojzije Brkic[1], Marc Leibundgut[2], Ita Gruic Sovulj[1], Nenad Ban[2]

[1]Faculty of Science, University of Zagreb, 10000 Zagreb, Croatia, [2]Institute of Molecular Biology and Biophysics, ETH Zürich, 8093 Zürich, Switzerland

Isoleucyl-tRNA synthetases (IleRS) are universally conserved enzymes that covalently couple isoleucine to its cognate $tRNA^{Ile}$ in a two-step aminoacylation reaction. These multi-domain proteins consist of an aminoacylation domain, a proofreading domain and a C-terminal anticodon-binding domain involved in $tRNA^{Ile}$ recognition. IleRSs cluster into two clades, IleRS1 and IleRS2, which differ in antibiotic resistance and the architecture of their C-terminal domain. The structure of the C-terminal anticodon-binding domain of IleRS1 is already known and entails three subdomains (SD): SD1 with helical bundle topology, SD2 consisting of four anti-parallel β-sheets and SD3 that is a αβ-fold with a zinc-binding motif. [1] At the same time, the structure of the C-terminal domain of IleRS2 remained unknown as only structures of truncated enzymes were reported. Here, for the first time, we present the structure of full-length *Bacillus megaterium* IleRS2 with a completely resolved C-terminal domain at 2.3 Å resolution. The structure unveils that the C-terminal domain of IleRS2 consists of three subdomains analogously to IleRS1. SD1 and SD2 in IleRS2 align structurally well with the corresponding subdomains in IleRS1. In contrast, SD3 lacks the zinc-binding motif of IleRS1 SD3 and surprisingly, topologically resembles the SD2. Finally, the structure visualized a novel 75 amino acid long SD2 insertion, which is absent in IleRS1. We prepared a *B. megaterium* IleRS2 mutant with the SD2 insertion exchanged to a $[Gly_4Ser]_2$ loop. The mutant has relatively modest 5-fold decrease in aminoacylation rate as compared to the wild type enzyme, which indicates that the SD2 insertion is important, but not essential for IleRS2 aminoacylation. The results thus open an intriguing question whether SD2 of IleRS2 has a role outside of translation.

References:

[1] F. L. Silvian, J. Wang, A. T. Steitz, *Science* **285** (1999) 1074–1077.

Deciphering membrane surface complexes through integrated structural biology: ADP-ribosylation factor (Arf) and GTPase activating proteins (ArfGAP) dance at the membrane surface to regulate signaling

R. Andrew Byrd[1], Olivier Soubias[1], Yue Zhang[1], Jess Li[1]

[1]Center for Structural Biology, National Cancer Institute

ADP-ribosylation factors (Arfs) are small Ras superfamily GTPases that coordinate membrane traffic and actin remodeling. Arf activity depends on guanine nucleotide exchange factors (GEFs) and GTPase-activating proteins (GAPs). Arf GEFs and Arf GAPs, including ASAP1, control integrin trafficking and focal adhesions, leading to effects on cell migration, proliferation/differentiation switching, and cell survival. ASAP1 also controls cancer progression, invasion and metastasis. We are exploring the membrane surface recognition and binding complexes formed between Arf1 and ASAP1. Surface recognition and complexes present considerably different challenges from the study of integral membrane proteins. The proteins need to recognize and bind to a membrane surface, yet they can (and do) diffuse across the surface. Complex formation becomes a combination of separate protein-membrane recognition and protein-protein recognition. The order and affinity of such interactions controls biological function. In each step, allostery is at play, as well as dynamics within each component. These properties pose significant challenges to crystallography and cyroEM; hence, we have explored the use of NMR spectroscopy, Neutron reflectometry (NR), and molecular dynamics (MD) simulations to gain insight into the recognition and mechansims. The NMR approach uses nanodiscs formulated to reflect appropriate membrane composition, to examine Arf1 and the functional components of ASAP1 bound to the membrane. Binding of the Pleckstrin homology (PH) domain of the ArfGAP ASAP1 to membranes containing the phosphatidylinositol phosphate PIP(4,5)P2 is key for maximum GTP hydrolysis, but not fully understood. By combining NMR, NR and MD simulation, we have shown (*Sci Adv* 2020) that binding of multiple PI(4,5)P2 molecules to the ASAP1 PH domain (i) triggers a functionally relevant allosteric conformational switch involving regions distant from the membrane interface, and (ii) maintains the PH domain in a well-defined orientation, allowing critical contacts between the newly exposed segments and an Arf1 mimic to occur. Using both NMR and MD, we characterize the G-domain of myristoylated-Arf1 bound to the nanodisc. The G-domain is highly dynamic and provides an interaction surface for the ASAP1-PH domain to initiate complex formation. The integration of multiple methods enables development of a functional model and understanding, which may ultimately enable cryoEM investigations.

Mechanism of forkhead transcription factors binding to a novel palindromic DNA site

Yongheng Chen[1], Jun Li[1], Shuyan Dai[1]

[1]NHC Key Laboratory of Cancer Proteomics, Xiangya Hospital, Central South University

Forkhead transcription factors bind a canonical consensus DNA motif, RYAAAYA (R = A/G, Y = C/T), as a monomer. However, the molecular mechanisms by which forkhead transcription factors bind DNA as a dimer are not well understood. In this study, we show that FOXO1 recognizes a palindromic DNA element DIV2, and mediates transcriptional regulation. The crystal structure of FOXO1/DIV2 reveals that the FOXO1 DNA binding domain (DBD) binds the DIV2 site as a homodimer. The wing1 region of FOXO1 mediates the dimerization, which enhances FOXO1 DNA binding affinity and complex stability. Further biochemical assays show that FOXO3, FOXM1 and FOXI1 also bind the DIV2 site as homodimer, while FOXC2 can only bind this site as a monomer. Our structural, biochemical and bioinformatics analyses not only provide a novel mechanism by which FOXO1 binds DNA as a homodimer, but also shed light on the target selection of forkhead transcription factors.

PDBe-KB: a community-driven resource for structural and functional annotations

Preeti Choudhary, Lukas Pravda

EMBL-EBI,Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK

The Protein Data Bank in Europe - Knowledge Base (PDBe-KB, https://pdbe-kb.org) is a community-driven, collaborative data resource. It provides literature-derived, manually curated and computationally predicted structural and functional annotations of macromolecular structure data in the Protein Data Bank (PDB). Such annotations include catalytic sites, ligand binding sites, protein flexibility, post-translational modification sites, and the effect of genetic variability or mutations, to name a few. The goals of PDBe-KB are (i) to increase the visibility and reduce the fragmentation of annotations contributed by specialist data resources, (ii) to make these data more findable, accessible, interoperable and reusable (FAIR) and (iii) to place macromolecular structure data in their biological context, thus facilitating their use by the broader scientific community in fundamental and applied research.PDBe-KB currently collaborates with 31 resources from 10 countries, and their annotations are integrated and interconnected with PDB structural data in a novel and distributable PDBe graph database (https://www.ebi.ac.uk/pdbe/pdbe-kb/graph-download). Users can access these annotations programmatically through a REST API (https://www.ebi.ac.uk/pdbe/graph-api/pdbe_doc). Currently, over 80 endpoints are addressing common scientific questions such as: which small molecules interact with a protein of interest, and what are the ligand binding residues? This API also powers aggregated views web-pages that provide an overview of all the structure data related to a full-length protein (UniProtKB accession) in their biological context instead of the conventional PDBe entry pages that focus on a single PDB entry. The architecture of these pages takes advantage of our publicly available web components (https://www.ebi.ac.uk/pdbe/pdb-component-library/index.html) that users may seamlessly integrate with third- party applications.We have been continuously improving and expanding PDBe-KB since its inception in 2018. Last year we rolled out a significant update to support the global effort to tackle the COVID-19 pandemic by creating dedicated pages gathering all the structural information for the new coronavirus SARS-CoV-2 (https://pdbe.org/covid-19). Additionally, we introduced a process to display clusters of PDB entries related to a protein given their structural similarity, batch downloads and many other improvements, including better visualisation and enhanced search options.

Structure-based drug discovery (SBDD) and biochemical evaluation of D-3,3-Diphenylalanine-based anticoagulant tetrapeptides inhibitors of thrombin-activated platelets aggregation

Cristina C. Clement[1], Janet Gonzalez[2], Sheuli Zakia[3], Morayma Reyes Gil[4], Manfred Philipp[3]

[1]Radiation/Oncology Department, Weill Cornell Medicine, New York, 10065, USA;, [2]Department of Natural Sciences, LaGuardia Community College, New York, 1110, USA;, [3]Department of Chemistry, Lehman College of the City University of New York, Bronx, New York, 10468;, [4]Montefiore Medical Center; Hematology and Coagulation Labs, Bronx, New York, 10061;

Thrombosis-related disorders such as myocardial infarction, stroke, and pulmonary embolism remain a major cause of morbidity and mortality worldwide, a fact that is driving sustained interest in developing novel thrombin inhibitors drugs. Aiming to discover new anticoagulants with lower risk of bleeding, novel peptidic direct thrombin inhibitors (DTI) derived from the d-Phe(P3)-Pro(P2)-d-Arg(P1) tripeptide scaffold were developed. Besides the optimal d-Phe-Pro dipeptide at positions P3-P2, the d-isomer of arginine was selected for position P1 to improve resistance to proteolytic degradation by thrombin. In 2012, Figueiredo, A.C., Clement CC et al. published the biochemical and structural characterization of three noncovalent, DTIs that contain the common sequence D-Phe(P3)-Pro-(P2)-D-Arg(P1)-P1'-CONH2, leading to the deposit of three structures of selected DTIs in complex with human α-thrombin:3U8O.pdb; 3U8R.pdb and 3U8T.pdb.

Herein, we report the optimization of the tetrapeptide scaffold by replacing D-Phe in the P3 position with the un-natural Phe-analog, D-3,3-Diphenylalanine. We performed a structure-based drug design (SBDD) and structure-activity relationship (SAR) at the P1'position by replacing L-amino acids with their D-isomers and other unnatural amino acids analogs using as template our published 3U8O.pdb. All D-3,3-Diphenylalanine-DTI analogs competitively inhibited alpha-thrombin's cleavage of the S2238 chromogenic substrate with K(i) of 500-60 nM. Remarkably, the novel DTIs inhibited the aggregation of human platelets in the "whole blood" thromboelastography (TEG) assay, as well as in the ex-vivo thrombin-activated platelets treatment. In addition, the peptidic DTIs showed potent inhibition of blood clotting monitored by aPTT, PT, and TT assays. The global changes in the protein expression profiles in the thrombin-activated human platelets treated with the lead DTI were further analyzed by label free quantification (LFQ) and nanoLC-MS/MS. Ingenuity Pathways Analysis (IPA) highlighted that many proteins involved in the actin, ILK, rhoA, and integrin signaling pathways were down-regulated in the DTI-treated platelets. These data support the newly designed DTIs as potent inhibitors of thrombin-activated platelet aggregation, and further advocate for the use of these novel compounds as potential biomaterials with improved haemocompatibility.

High-throughput protein X-ray crystallography: from structural genomics to hit generation by library screening in structure- and fragment-based drug discovery

Debanu Das[1], Ashley Deacon[1]

[1]Accelero Biostructures Inc, San Carlos, CA 94070

The past 20 years have witnessed an unprecedented growth in instrumentation and automation for protein X-ray crystallography, covering all stages of the gene-to-structure process. These developments were sparked and primarily driven by high-throughput structural genomics and structural biology efforts in the USA and around the world. Outcomes included methods, processes, protocols, and instrumentation for high-throughput protein expression, purification, crystallization, structure determination, refinement, and analysis. Benefits have now percolated into industry and academia, influencing both structural biology and structure-based drug discovery research and have led to significant gains in efficiency and productivity at reduced cost.Screening molecular compound libraries, such as fragment and scaffold libraries, is widely used to discover novel hits for the development of new medicines. These libraries efficiently sample chemical space, which has important outcomes in early drug discovery and can be especially relevant for challenging “non-druggable” targets. With foundational expertise in structural genomics and high-throughput protein X-ray crystallography, Accelero Biostructures is translating these developments for novel high-throughput structure- and fragment-based drug discovery. Our platform for crystallography-based library screening accelerates the hit-to-lead process. 3D structures of compounds bound to a target of interest are generated, potentially leading to ~5-7x reduction in early drug discovery timelines. Intellectual property is established in the form of structurally validated binding sites and specific protein-ligand interactions. This provides a direct experimental route to high quality, high reliability and high value results: the identification and assessment of target engagement by compound binding sites and poses; target ligandability; and differentiation of orthosteric and allosteric sites and binders. We will present results from applying our platform to different classes of drug targets and how it has also helped to launch a new therapeutics program centered around DNA repair in oncology.

PDB Component Library - a collection of open source web components for interactive visualization of macromolecular structure data and annotations

Mandar Deshpande[1]

[1]Protein Data Bank in Europe (PDBe), EMBL-EBI, Cambridge, UK

Data visualization plays a pivotal role in the research and innovation lifecycle by revealing patterns and correlation between different data types. However, it takes a lot of effort and time to design and develop data visualisations amenable to use by both experts and non-experts. The Protein Data Bank (PDB) Component Library is a collaborative effort to make available a collection of reusable, open-source web components for visualization of macromolecular structure data in the PDB and associated structural and functional annotations that provide the necessary biological context. Currently, there are 8 web components in the PDB Component Library that allow visualizing protein structure data and additional annotations in 1D, 2D and 3D representation. These visualizations are interactive and can be easily embedded into a webpage. The collaborative approach has not only eased the development effort but also allowed to create a consistent experience for users exploring these data across different services utilizing these visualizations.

The next generation RCSB.org

Jose M. Duarte[2], Charmi Bhikadiya[1], Chunxiao Bi[2], Sebastian Bittrich[2], Li Chen[1], Shuchismita Dutta[1], Robert Lowe[1], Alexander Rose[2], Yana Rose[2], Joan Segura[2], John D. Westbrook[1], Stephen K. Burley[1, 2]

[1]RCSB Protein Data Bank, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA, [2]RCSB Protein Data Bank, SDSC, University of California, La Jolla, CA 92093, USA

RCSB Protein Data Bank (PDB) provides tools for analysis and visualization of 3D structures of biological macromolecules stored in the PDB archive. In 2020 RCSB.org experienced a complete redesign of user interfaces and backend services. The new services represent a full overhaul of the software/data management architecture, transforming a monolithic application into a micro-service-oriented and cloud-ready resource.
Some of the highlights of the improved rcsb.org experience are:
New search system that allows for queries with arbitrary Boolean logic. The queries can combine specialised bioinformatic search algorithms (sequence search, structure search, small molecule search, structure motif search, sequence motif search) with text search of any attribute present in annotations from PDB or integrated external data resources. The search results can be presented at any of the granularities of the PDB data hierarchy: entry, polymer entity, assembly and non-polymer entity (selectable with the “Display results as” drop-down menu).
Rapid 3D visualization of macromolecules with the Mol* viewer, developed as a collaboration between RCSB PDB, PDBe and the Central European Institute of Technology.
Next generation Protein Feature View: in-house developed Protein Feature View based on modern web technologies, allowing facile visualization of sequence features from PDB and annotations from external resources, such as domains or PDB validation data.
A new pairwise Structure Alignment application that offers some of the most popular structure alignment algorithms (FatCat, CE, TMalign) with integrated visualization in the Mol* viewer.
Structure motif search with an interface to select motifs integrated in the Mol* viewer.
Ultrafast global protein shape search with the in-house developed BioZernike algorithm. The largest assemblies in the PDB can be searched in milliseconds.
New macromolecular static images rendered by the Mol* library in a unified style. Images are now available for polymer entities as well as for assemblies.

How to optimize the fragment hit rate: take protein crystals first.

Serghei Glinca[1], Janis Mueller[1], Stefan Merkl[1], Lars Ole Haustedt[2], Gerhard Klebe[3]

[1]CrystalsFirst GmbH, Marbacher Weg 6, 35037 Marburg, Germany, [2]AnalytiCon Discovery GmbH, Hermannswerder Haus 17, 14471 Potsdam, Germany, [3]Philipps-University Marburg, Department for Pharmaceutical Chemistry, Marbacher Weg 6, 35037 Marburg

Fragment-based lead discovery (FBDD) has undoubtedly become an invaluable resource for medicinal chemists in the identification of lead candidates. It has resulted in 4 approved drugs and more than 50 candidates in clinical trials.

Since an efficient fragment evolution requires structural data of protein-ligand interactions, a variety of biophysical methods are applied in a cascade approach as a set of prefilter prior to structure determination. Systematic studies suggest that the cascade approach significantly reduces the hit rate compared to a direct crystallographic screening.

The application of X-ray crystallography as a primary fragment screening method has been underutilized due to the limited availability of a robust system for soaking experiments and solubility issues of fragments. In our case studies, demonstrate the strength of direct crystallographic screening to find starting points to access best possible chemical space for FBDD.

One of major bottlenecks in the experimental setup for crystal soaking is the sensitivity of protein crystals. We overcome this limitation using the SmartSoak® technology, which stabilizes protein crystals and enables direct crystallographic screening without trial-and-error optimization of the soaking system.

With hit rate over 30 %, we find fragments with unexpected binding modalities that could be rapidly optimized to inhibitors. These results demonstrate the advantages of putting “structural data first” and the application of crystallographic fragment screening as a primary screening strategy for the hit identification.

The structure of cyclic beta 1-,2-glucan synthase from Streptococcus lutetiensis, a novel diarrheal pathogen.

Dinendra Abeyawardhane, Raquel Godoy-Ruiz, Edvin Pozharskiy, Colin O. Stine, David J. Weber

[1]University of Maryland Baltimore School of Medicine.

Diarrhea continues to be a major public health problem in low income countries causing an estimated 2.4 billion cases and 1,310,000 deaths worldwide. Diarrhea kills more children than HIV and malaria combined. Although many pathogens are known, in many cases even when stools are tested for these pathogens the symptoms cannot be attributed to a known cause. These unattributed cases of diarrhea may result from a previously unrecognized pathogen. An analysis of fecal microbiota from a large case-control study of moderate-to-severe diarrhea in developing countries revealed that 16S sequences from Streptococcus were more abundant in children with diarrhea compared to those without diarrhea. However, neither of the pathogenic species could be identified because 16S sequences do not effectively distinguish among species within the genus Streptococcus. Sequencing of the S. lutetiensis genome did not suggest any obvious toxins or virulence factors that distinguish it from its sibling species S. infantarius and or other Streptococci. We identified ‘cyclic beta 1-,2- glucan synthase’ (CBGS) as the enzyme with a sequence similar to a sequence from Brucella abortus [UniProt: 085166] that is an essential step in the invasion of Brucella into its mammalian host. The resolution of recombinant CBGS structure by single-particle cryoelectron microscopy (CyroEM) confirms, for the first time, the presence of this enzyme in S. lutetiensis infections.

In silico Proteomic Chrestomathy of Ruminantia's Labile Phenylalanine-Methionine Dipeptide in Kappa-Caseins

Thomas E. Gorrell[1]

[1]School of Visual Arts, New York, NY 10010

It has been noted that milk and more specifically bovine milk is one of the most calibrated substrates in

Biological Chemistry with a further focus on sour milk at the root of scientific Microbiology. Peculiarly hydrolysis of a phenylalanine-methionine peptide bond in kappa casein characterizes the classic cheese industry with its emphasis on three domesticated species of ruminants; *Bos taurus* , *Capra hipra* , and *Ovisaries*. The kappa casein of cattle has 190 amino acids that derives from transcriptional processing of eg., CSN3 Allele A 15,365bp and posttranslational processing of the nascent protein. There may be one restriction enzyme (FaiI) that cleaves the hexnucleotide (TTTA^TG). The singular dipeptide at F126M127 is cleaved by chymosan/rennet into two peptides. The carboxy-terminal methionine fragment is part of the soluble whey and a nutritional supplement for those with phenylketonuria. The N-terminal phenylalanine peptide stabilizes three other caseins that more directly form the amorphous calcium-phosphate micelles in cheese. The correlation of the primary structure of the phenylalanine peptide to the secondary, tertiary, and curdled aggregates with additional caseins is difficult to reveal since caseins regarded as implicitly intrinsically disordered and computed by *in silico* methods to be inherently unstable. This study evaluated publically available molecular data of kappa casein in mostly members of the Artiodactylia, Rodentia, and Primates and versions of *Bos taurus*. Initial analysis of the kappa caseins reveal a) hexanucleotides not present in Humans, primate, nor rodent, nor cat, b) dog has hexanucleotide yet not dipeptide phenotype c) rabbit has two dipeptides albeit nearer the N-terminus. Cattle,goat and sheep have duplicate hexanucletide in the 3' terminus after a second stop codon. Of particular interest is discerning relevance of the dipeptide to *in silico* studies of epitopes as guide to *in vitro* and *in vivo* studies of allergenic activity of the protein itself and as aggregates- along with nutritional attributes: especially relative to lipids; and now viewed from the perspectives of a putative circular RNA that may derive from an exon in which the dipeptide resides and samples of milk fermenting over a dense cushion of saccharide(s).

Structure/Function Comparisons of 3’-Phospho Adenosine 5’-Phospho Sulfate (PAPS) Synthase (PAPSS) Isoforms 1, 2a and 2b

Pavel Grinkevich, Dhiraj Sinha, Magda Vojtova, David Reha, Rudi H. Ettrich, Kallidaikurichi V. Venkatachalam

1University of South Bohemia in Ceske Budejovice, Branišovská 1760, 370 05 České Budějovice, Czech Re, 2IHU, Aix Marseille University Marseille, France, 3Institute of Microbiology, Academy of Sciences of the Czech Republic, Zamek 136, 373 33 Nove Hrady,, 4Institute of Microbiology, Academy of Sciences of the Czech Republic, Zamek 136, 373 33 Nove Hrady,, 5College of Biomedical Sciences, Larkin University, Miami, Florida 33169, 6College of Allopathic Medicine, Nova Southeastern University, Ft. Lauderdale, FL-33328
3’-phosphoadenosine 5’-phosphosulfate (PAPS) is the universal sulfuryl donor.

PAPS is synthesized in two sequential steps. First inorganic sulfate reacts with alpha-phosphoryl group of ATP to form adenosine 5’-phosphosulfate (APS) and PPi catalyzed by ATP sulfurylase. Thermodynamically this is an unfavorable reaction since the phospho-sulfate anhydride bond is more energetic relative to phospho-anhydride bond of ATP. However, under physiological conditions the coproduct PPi is hydrolyzed to 2 Pi by ubiquitous pyrophosphatase which drives the whole ATP sulfurylase reaction forward. In the second step the 3’-hydroxyl anion reacts with gamma-phosphoryl of ATP to form PAPS and ADP catalyzed by APS kinase. APS kinase reaction is favorable in the direction of PAPS formation. In mammals ATP sulfurylase and APS kinase activities are fused together into a bifunctional polypeptide termed PAPS synthase (PAPSS). In humans Venkatachalam isolated 3 different isoforms from various tissue samples by RT-PCR, cloned and sequenced. PAPSS1 and 2 are about 70% identical. Whereas PAPSS2b is a splice variant of PAPSS2a that has the extra pentapeptide segment GMALP that is unique to PAPSS2b. Experimental kinetic comparisons of PAPSS1, 2a and 2b has been performed using recombinant proteins expressed in mammalian cancer cell lines and in bacterial over expression systems. From this analysis, it was established that the kinetic parameters Km, Vmax and Kcat/Km were quite different between isoforms 1, 2a and 2b. With these functional differences we set out to look at the structural differences that would relate to the functions. First, ATP sulfurylase-APS kinase domain interactions within the isoforms of 1, 2a and 2b was analyzed. Then what functional advantage could PAPSS2b achieve with the extra pentapeptide GMALP? was pondered. Then how the common substrate ATP binds sequentially so that futility can be avoided during PAPS formation? From initial data we speculate that the linker domain between the sulfurylase-kinase orchestrates coordinated control, whereby one domain forms product which serves as a substrate for second domain activity. The GMALP fine tunes catalysis in 2b by forming special loop structure that is absent in 2a or 1. Thus differences in linker and extra pentapeptide GMALP could be responsible for the functional differences

The PDB Art Project

Deepti Gupta[1]

[1]PDBe, EMBL-EBI

The Protein Data Bank in Europe (PDBe) has an ongoing art-science public engagement project, called the PDB Art project (pdbe.org/art). This collaborative project, now in it's fifth year, brings together art societies (GRANTA and CANTAB), schools' arts and science departments, and scientists to inspire students aged 12-17 to explore 3D biomolecular structures in the Protein Data Bank (PDB) and create artworks depicting themes in structural biology.
The collaboration between schools and the scientists, acts to **INSPIRE** school students to create artistic interpretations of their chosen molecules from the PDB.

Students have the opportunity to **LEARN** about the 3D molecular structures through hands-on training and activities by PDBe staff. In addition, they also engage with scientists at the Wellcome Genome Campus or MRC Laboratory of Molecular Biology, Cambridge.

Students then **EXPLORE, RESEARCH & CREATE** the artwork based on their protein of interest and accompany them by scientific descriptions. This unique learning experience helps the students not only gain interest and confidence in science but also enjoy learning various forms of art in creating their masterpieces'.

To celebrate this collaboration, and the fantastic artworks created by the young artists we SHARE them through a range of public engagement activities. Artworks are exhibited in a number of exhibitions and last year we hosted our very first 'virtual exhibition' (https://www.artsteps.com/view/5f33e61717d8cb633b396232) and a guided tour of the virtual PDB Art exhibition is available at https://youtu.be/9uI5TbuK3Jw. Artworks are also featured in the annual PDBe calendar that is distributed worldwide, bringing the artistic interpretation back to the scientific community.

Evolution of Structural Biology: applying cryo-EM to drug discovery

Sarah G. Hymowitz[1]

[1]Department of Protein Sciences, Genentech, Inc

Structural biology is a critical tool in drug discovery. Structures are the engine of "structure-based drug design" by providing a framework to interpret structure-activity relationships for small molecule projects and generate hypotheses that drive compound design. For antibody or protein-based drug discovery projects, structures can be used to differentiate lead molecules based on their epitope and, more generally, elucidate the molecular mechanism of action of potential clinical candidates. Prior to starting a drug discovery campaign, structures of proteins and protein complexes can lead to a better understanding of biological processes underlying human disease. In addition, the shared need for high quality, well-behaved protein is a key driver of assay development, immunization campaigns and structural studies.

Over the past twenty-five years, crystallography has been the dominant technique for determining structures in support of drug discovery. Nuclear Magnetic Resonance (NMR) is complementary structural technique which has been integrated into drug discovery as a preferred option of determining structures of peptides and monitor dynamic changes in proteins. In addition, NMR is valuable in validating hits from screening assays or as a lead-finding methodology in the context of fragment-based drug discovery.

In recent years, cryo-EM has demonstrated the potential to complement crystallography and NMR a driver of drug discovery due to advances in sample preparation, hardware and data processing by delivering high resolution structures without the need for protein crystallization. Setting up a cryo-EM facility to support a variety of discovery, antibody-based and small molecule projects requires investment in microscopes, personnel, informatics, sample preparation and lab space. Having a strategy to integrate the effort into an existing structural biology and drug discovery community is also critical to ensure uptake of a new technique and maximizing impact on projects. In this talk, I will share our strategy and lessons learned through this process, case studies showing how structures generated using cryo-EM have benefited diverse small molecule, large molecule and discovery projects and potential future directions for cryo-EM and structural biology more generally in drug discovery.

The mystery of the Haff disease - how to find the answer?

Eugeny P. Kalinin[1], Natalya N. Buslaeva[1]

[1]Department of Biochemistry, State Medical University, Tyumen, 625023

Haff disease is an alimentary toxicosis that develops when a person comes into contact with the environment. The main clinical manifestation of the disease is alimentary-toxic paroxysmal myoglobinuria (ICD-10, DiseasesDB). This disease was first described in 1924 and after that numerous cases of the disease were recorded without connection with the region or climatic zone. In 2019-2021, several episodes of Haff disease were also recorded in Tyumen region. Studies devoted to Haff disease in Brazil, China and European countries did not allow obtaining information about the chemical strusture of the toxin, however, these studies made it possible to refute the assumptions that the toxin belongs to the already known groups of toxic compounds (for example, heavy metal salts, low molecular weight toxic compounds and palytoxin). The existing data suggest the lipid nature of the toxin, which based on the fact that the soluble in hexane fraction of substances caused symptoms similar to those of Haff disease in laboratory mice. Also, assumptions about the thiaminase activity inherent in the toxin were put forward. It is generally accepted that cyanobacteria (blue-green algae) are a source of toxic compounds; they generate or accumulate toxin. Fish and crustaceans living in an infected lake are a source of toxin for humans, they only deposit the toxin. Due to the fact that the chemical structure of the toxin has not been established, and there is no way to detect it, patients are treated with symptomatic treatment aimed at accelerating the excretion of rhabdomyolysis products. In our opinion, there is a possibility that the toxin belongs to peptides in which most of the amino acid residues have an aliphatic radical, or to cyclic peptides with aliphatic amino acids. According to our data, such plant peptides can be obtained from silt bottom sediments (peloids), and, despite their low solubility in polar solvents, are capable of forming complexes with proteins, modulating their biological activity. Based on this hypothesis, the laboratory of biochemistry of the Tyumen State Medical University offers its own method for isolating the toxin that causes the Haff disease.

KBase-RCSB PDB Integration Framework

Harmeet Kaur[1], John Westbrook[1]

[1]RCSB Protein Data Bank, Rutgers, The State University of New Jersey

In Biology, function follows form (i.e., function is determined by 3D structure). Atomic-level 3D structural insights obtained from experimental and computational approaches are essential for advancing our understanding of the molecular mechanisms of biological processes. Currently the Department of Energy (DOE) Systems Biology Knowledgebase (KBase) (https://www.kbase.us/) and RCSB Protein Data Bank (PDB) are two separated resources that serve up genomic/omics data and biological macromolecular structure data respectively. Development of efficient and semantically deep integrations between systems biology data in KBase and 3D macromolecular structure data in the PDB archive will amplify the impact of both resources, enabling a deeper understanding of detailed molecular processes at the level of large assemblies operating inside plants, animals, and microbial systems. In this poster, we describe the planned architecture and progress developing new APIs that will enable KBase users to integrate macromolecular structures and associated experimental metadata as an additional source of support for new mechanistic hypotheses. This API development will provide access not only to the experimentally determined 3D structure data available at the RCSB PDB (RCSB.org), but to the Integrative/Hybrid 3D structures in the PDB-Dev repository (https://pdb-dev.wwpdb.org/), in silico 3D structure data in the ModelArchive repository (https://www.modelarchive.org/) and computational models in SwissModel (https://swissmodel.expasy.org/repository/). This will enable advancement of DOE's mission of addressing issues pertaining to sustainable energy and improved environment.

Acknowledgements:RCSB Protein Data Bank is funded by the National Science Foundation (DBI-1832184), the US Department of Energy (DE-SC0019749 and 1F-60284), and the National Cancer Institute, National Institute of Allergy and Infectious Diseases, and National Institute of General Medical Sciences of the National Institutes of Health under grant R01GM133198

Web services in PDBj for electron microscopic data: EM-navigator, gmfit and EMPIAR-PDBj

Takeshi Kawabata[1], Hirofumi Suzuki[2], Kiyo Tsunezumi[1], Kei Yura[2, 3], Genji Kurisu[1]

[1]Institute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan, [2]School of Advanced Science and Engineering, Waseda University, Shinjuku, Tokyo 169-8555, Japan, [3]Graduate School of Humanities and Sciences, Ochanomizu University, Bunkyo, Tokyo 112-8610, Japan

Electron microscopy (EM) is increasingly being used as a standard technique to determine atomic 3D structure. Protein Data Bank Japan (PDBj) provides several WEB databases and services for making use of 3D maps and 2D images of electron microscopy.

First, the database “EM navigator” provides a user-friendly view of 3D density maps stored in EMDB (Electron Microscopy Data Bank). The “Omokage search” service enables us to search 3D maps or atomic models with similar shapes by a query map or model given by the user. Secondly, the “gmfit” service provides fitting calculations between a 3D map and an atomic model through WEB. The calculation is fast due to the technique that approximates the density by Gaussian mixture model (GMM). The gmfit program has been improved to handle high-resolution maps and models by introducing down-sampling algorithms to generate GMM. The algorithms for local and multiple subunits fitting have been implemented, such as grid-layout initialization, masked segmentation & fitting, and combination of single subunit searches (CSS).

Thirdly, the mirror site of EMPIAR (Electron Microscopy Public Archive) has been established in Japan as "EMPIAR-PDBj" (https://empiar.pdbj.org) since 2018. EMPIAR is a public resource for 2D electron microscopy images developed in EMBL-EBI in UK. Raw 2D EM images are necessary to validate 3D map, enhance developments of image processing software, and educate and train EM users. Due to the large size of 2D images, alternative storing and distributing site of EM images in addition to UK site is helpful for researches all over the world. We also encourage EM researchers to deposit their 2D images. Researchers can send storage devices (HDD, SSD or tape) to our institute by postal mail or via courier service.

Protein Data Bank Japan: 20 years of serving as the Asian hub for 3D macromolecular structure data

Genji Kurisu[1], Gert-Jan Bekker[1], Masashi Yokochi[1], Yasuyo Ikegawa[1], Takeshi Iwata[1], Takahiro Kudo[1], Reiko Yamashita[1], Yumiko Kengaku[1], MinYu Chen[1], Junko Sato[1], Ju Yaen Kim[1], Kiyo Tsunezumi[1], Takeshi Kawabata[1], Toshimichi Fujiwara[1]

[1]Institute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan

Protein Data Bank Japan (PDBj, https://pdbj.org), accepts and processes regional 3D structure data of biological macromolecules since 2000. We celebrated our 20th anniversary of our regional Data-in activities last year. Our Data-out service has a much longer history, dating back to before the establishment of PDBj. The first protein structure from Asia was determined at the Institute for Protein Research (IPR) in 1971 at a resolution of 4 Å and a subsequent atomic structure solved at 2.3 Å was deposited to the PDB in 1976 as the 21st entry in the PDB. Based on this early contribution as a depositor, IPR started the PDB data distribution in 1979 and kept serving the PDB data as a regional data center, initially by magnetic tape and later by CD-ROM, until the installation of an official mirror site of Brookhaven PDB in 1998. Since 2001, we have provided our newly developed online Data-out services freely and publicly through our own web site, which we have enhanced through the years, including our molecular graphics viewer, Molmil; a molecular surface database for functional sites, eF-site; and a database of protein dynamics calculated via normal mode analysis, Promode Elastic. With the establishment of the wwPDB in 2003, we have served as one of its founding members, collaborating on Data-in and Data-out activities on a global scale. During the COVID-19 pandemic, we have provided a COVID-19 featured page in three Asian languages (Japanese, Chinese and Korean) and have started a new service archiving raw X-ray image data directly related to deposited PDB entries (XRDa, https://xrda.pdbj.org). Since we already have BMRBj (formerly PDBj-BMRB) and EMPIAR-PDBj on-site, XRDa completes the regional experimental raw data archives of the related PDB, BMRB and EMDB entries from the three major experimental methods; Macromolecular Crystallography, NMR spectroscopy and 3D Electron Microscopy.

Investigating the reaction and substrate preference of indole-3-acetaldehyde dehydrogenase from the plant pathogen Pseudomonas syringae PtoDC3000

Soon Goo Lee[1]

[1]University of North Carolina Wilmington

Aldehyde dehydrogenases are versatile enzymes that serve a range of biochemical functions. Although traditionally considered metabolic housekeeping enzymes because of their ability to detoxify reactive aldehydes, like those generated from lipid peroxidation damage, the contributions of these enzymes to other biological processes are widespread. For example, the plant pathogen Pseudomonas syringae strain PtoDC3000 uses an indole-3-acetaldehyde dehydrogenase to synthesize the phytohormone indole-3-acetic acid to elude host responses. Here we investigate the biochemical function of AldC from PtoDC3000. Analysis of the substrate profile of AldC suggests that this enzyme functions as a long-chain aliphatic aldehyde dehydrogenase. The 2.5 A° resolution X-ray crystal of the AldC C291A mutant in a dead-end complex with octanal and NAD+ reveals an apolar binding site primed for aliphatic aldehyde substrate recognition. Functional characterization of site-directed mutants targeting the substrate- and NAD(H)-binding sites identifies key residues in the active site for ligand interactions, including those in the “aromatic box” that define the aldehyde-binding site. Overall, this study provides molecular insight for understanding the evolution of the prokaryotic aldehyde dehydrogenase superfamily and their diversity of function.

3D architecture and structural flexibility revealed in the subfamily of large glutamate dehydrogenases by a mycobacterial enzyme

Melisa Lázaro[1], Mikel Valle[1], María-Natalia Lisa[2]

[1]Center for Cooperative Research in Biosciences (CIC bioGUNE), 48160 Derio, Spain, [2]Instituto de Biología Molecular y Celular de Rosario (IBR CONICET-UNR), S2002LRK Rosario, Argentina

Glutamate dehydrogenases (GDHs) are widespread metabolic enzymes that play key roles in nitrogen homeostasis. Large glutamate dehydrogenases (L-GDHs) composed of 180 kDa subunits (L-$GDHs_{180}$) contain long N- and C-terminal segments flanking the catalytic core. Despite the relevance of L-$GDHs_{180}$ in the physiology of environmental and pathogenic bacteria, the lack of structural data for these enzymes has limited the progress of functional studies. We report the 3D structure of the mycobacterial L-GDH_{180} isoform from *Mycobacterium smegmatis* (mL-GDH_{180}), obtained through an integrative approach that combined single-particle cryo-EM and X-ray protein crystallography. mL-GDH_{180} adopts a quaternary structure that is radically different from that of related low molecular weight enzymes. Intersubunit contacts in mL-GDH_{180} involve a C-terminal domain that we propose as a new fold and a flexible N-terminal segment comprising ACT-like and PAS-type domains that could act as metabolic sensors. These findings reveal unique characteristics of domain organization and oligomeric assembly in the subfamily of L-GDHs, thus allowing to update the annotation of the Pfam family PF05088 that includes the L-$GDHs_{180}$. Furthermore, our cryo-EM data evidenced fluctuations in the quaternary structure of mL-GDH_{180} that are possibly relevant for the allosteric regulation of the enzyme activity, uncovering unique aspects of the structure-function relationship in the L-GDHs subfamily.

NOE-less protein structures: PioC, a small HiPIP from Rhodopseudomonas palustris TIE-1

Inês B. Trindade[1], Michele Invernici[2], Francesca Cantini[2], Ricardo O. Louro[1], Mario Piccioli[2]

[1]Instituto de Tecnologia Química e Biológica António Xavier (ITQB-NOVA), Universidade Nova de Lisboa,, [2]Magnetic Resonance Center and Department of Chemistry, University of Florence, Via L. Sacconi 6 5001

Metals are essential for life, playing a key role in biological functions through protein interaction. Almost half of all known enzymes require a particular metal to function. In all of these, the metal center(s) are essential for catalysis, electron transfer and/or play a crucial role in ensuring stability and structural properties. NMR spectroscopy is a privileged technique for characterizing proteins and their role in key cellular processes since it can probe both structural and dynamic aspects at atomic resolution in conditions that mimic the physiological environment. However, a significant part of key cellular processes and their enzymes rely on the participation of metals in paramagnetic states. In these conditions, a “blind sphere” is created in the vicinity of these metals where nuclear relaxation is enhanced and signal detection becomes a challenge. Making use of recent developments in novel pulse sequences, here we present the structure determination of a paramagnetic protein PioC. PioC from Rhodopseudomonas palustris TIE-1 is the smallest High Potential Iron-Sulfur Protein (HiPIP) ever isolated and it mediates the electron transfer between the reaction center and the iron-oxidase PioA in the photoferrotrophic metabolism of R. palustris TIE-1. Paramagnetism from the [4Fe-4S] cluster affects 60% of the protein, making it the perfect example of the dual nature of paramagnetic NMR. On the one hand relaxation precludes signal detection, and on the other hand it provides the means for obtaining unique sets of structural information. Paramagnetic relaxation enhancements (PREs) have the same distance dependence as Nuclear Overhauser Enhancements (NOEs) and can be used to define the structure of paramagnetic molecules where NOEs are difficult to measure due to fast signal relaxation. The structure of PioC was determined by NMR using two different sets of restraints, one containing PREs and another containing NOEs. These were used both independently and together revealing that under favorable conditions, PREs can efficiently complement and eventually replace NOES for NMR structural characterization[1].
Acknowledgements: supported by H2020 project TIMB3 contract 810856, Instruct-ERIC PID4509, COST Action CA15133, MOSTMICRO and FCT grant PD/BD/135187/2017
1. Trindade et al (2021) FEBS J in press, https://doi.org/10.1111/febs.15615

Integrative structural biology in molecular parasitology: new strategies for old diseases

Adriana E. Miele[1, 2], Matteo Ardini[3], Andrea Bellelli[2], Giovanna Boumis[2], Francesca Fata[3], Rodolfo Ippoliti[3], Fulvio Saccoccia[4], Ilaria Silvestri[3], David L. Williams[5], Francesco Angelucci[3]

[1]UMR5246 ICBMS-CNRS-UCBL, Université Claude Bernard Lyon 1, France, [2]Dept. Biochemical Sciences, Sapienza University of Rome, Italy., [3]Dept. Life, Health and Environmental Sciences, University of L'Aquila, Italy., [4]CNR - Institute of Biochemistry and Cell Biology, Monterotondo, Italy., [5]Dept Microbial Pathogens and Immunity, Rush University Medical Center, Chicago, USA.

Parasitic diseases are a major threat to human health, especially in countries in which access to sanitation and treatment is extremely limited. Eradication of helminth infections is one of the pillars of the recently launched WHO roadmap 2030 to attain global health goals. Worms have evolved to evade the host immune response while travelling through and residing in the host body. It is important for both diagnostics and therapeutics, to know the molecular players in this survival game.
In particular, for the human parasite *Schistosoma mansoni* the methods of diagnosis are not able to detect the early stages of infection and WHO mass drug administration relies on only one drug. Therefore, the identification and characterization of key macromolecular targets for diagnostic and drug discovery is of utmost importance. Advances in genomics, proteomics and metabolomics have directed this task to more rational and interdisciplinary strategies, where structural techniques, such as macromolecular crystallography, SAXS and electron microscopy are pivotal.
We are integrating structural biology with bioinformatics, molecular biology, biochemistry, biophysics and medicinal chemistry to select protein targets and to functionally and structurally characterise them. In the last fifteen years we have deposited 41 structures in the PDB on two main subjects of research in our laboratories: (a) focussed structural genomics of the thiol-mediated detoxification metabolism; (b) structural characterization of the worm secretome. In the former case we have elucidated the entire pathway and selected 2 candidates to perform cycles of structure-based drug design; in the latter the objective is to understand what is at the interface between the worm and the host and to select which macromolecules might become good diagnostics and/or vaccine candidates. The information thus obtained is currently used to shed light on the challenging biological questions opened by a vector-borne eukaryotic parasite that knows our immune system and its regulation better than we do.

3D-Beacons: An integrative, distributed platform for FAIR access to experimental and predicted macromolecular structures

Sreenath S. Nair[1], Mihaly Varadi[1], Ian Sillitoe[2], Neeladri Sen[2], Gerardo Tauriello[3], Andrew Waterhouse[3], Stefan Bienert[3], Jun Fan[1], Torsten Swede[3], Maria J. Martin[1], Christine Orengo[2], Sameer Velankar[1]

[1]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), [2]Institute of Structural and Molecular Biology, University College London, [3]SIB Swiss Institute of Bioinformatics, Biozentrum

The overwhelming majority of the known protein sequence space has no structural representations in the Protein Data Bank (PDB), with merely 0.027% of all the known sequences directly mapped to experimentally determined macromolecular structures. Fortunately, rapid advances in template-based and de novo structure modelling methods have resulted in advanced software that can provide high-quality, accurate models. However, the lack of standardisation in data access and quality estimation may hinder the broader scientific user community.
3D-Beacons is an open collaboration between PDB in Europe (PDBe), SWISS-MODEL, Genome3D, the Protein Ensemble Database and other macromolecular model providers that takes advantage of these invaluable structural representations. It is an integrative, distributed platform that aims to provide FAIR access to both experimentally determined and predicted protein structures, combined with their functional annotations. 3D-Beacons supports model providers with tools that simplify making macromolecular models available. The platform provides data access transparently to ensure that data provenance is clear to the end-users, major data providers themselves (UniProt, InterPro and Ensembl) or research groups worldwide.

The Life and Times of the PDB Format - Looking Towards the Future with mmCIF

Ezra Peisach[1], Sanja Abbott[5], Kumaran Baskaran[4], John Berrisford[2], Zukang Feng[1], Yasuyo Ikegawa[3], Catharine Lawson[1], John D. Westbrook[1], Masashi Yokochi[3], Jasmine Y. Young[1], Jeffery Hoch[4], Genji Kurisu[3], Sameer Velankar[2], Stephen K. Burley[1]

[1]RCSB PDB, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA, [2]PDBe, EMBL-EBI, Hinxton, Cambridgeshire, CB10 1SD, UK, [3]PDBj, Institute for Protein Research, Osaka University, Osaka, 565-0871, Japan, [4]BMRB, University of Connecticut, UConn Health, Farmington, CT 06030-3305, USA, [5]EMDB, EMBL-EBI, Hinxton, Cambridgeshire, CB10 1SD, UK

The PDB format file was created when the Protein Data Bank was established In 1971. Originally designed to use 70 columns on punched cards, the format specified fixed column positions and column widths for all data items. As the science and technology evolved, this legacy PDB format can no longer support data representation for large complex structures. This includes large entries that do not fit the PDB file format (those containing >62 chains and/or >99999 ATOM records).

Since the late 1990s, PDB files are also available in mmCIF, an extensible machine readable format that is not limited by fixed column positions and extensible to support new data content. The legacy PDB format has not been modified or extended to support new content since 2012. Since 2014, legacy PDB formatted files have not been produced for entries containing multi-character chain ids or atom serial numbers with greater than five digits.

Working with the community PDBx/mmCIF Working Group, wwPDB is transitioning to sunset the legacy PDB format when the PDB Chemical Component Dictionary (CCD) IDs are extended beyond three alphanumeric characters which cannot be accommodated by the PDB format. We anticipate running out of three-character CCD IDs in the next three to five years. At that time, wwPDB will start issuing four alphanumeric codes for CCD IDs in the OneDep deposition-validation-biocuration system. Entries containing these codes will not have legacy PDB format files. In 2022, wwPDB will begin the implementation of this CCD ID extension. In addition, wwPDB also plans to implement extended PDB IDs that are eight-characters with a PDB prefix, e.g., pdb_00001abc. This extended PDB ID will be included in the PDBx/mmCIF file for new entries issued with four-character PDB IDs. Once the four-character PDB IDs are all consumed, at that point newly deposited PDB entries will only be available in PDBx/mmCIF format.

wwPDB is asking community and user software developers to review their code and remove such limitations for the future.

Insights of unnatural amino acid-containing proteins through X-ray crystallography

Christine M. Phillips-Piro[1]

[1]Department of Chemistry, Franklin and Marshall College, Lancaster, PA 17604

Unnatural amino acids (UAAs) have become a common tool for protein scientists to utilize. These modified amino acids have been used for a variety of reasons to site-specifically study protein structure and function – to add a small vibrational reporter molecule to sense the local environment, to incorporate a novel FRET donor or acceptor, to change the local hydrogen bonding environment of a protein, and to alter the photophysical properties of spectroscopically active proteins, to name a few. All of these uses require that the incorporation of the UAA either does not alter the structure or alters the structure in an expected fashion, which can be assessed by solving the atomic structure of these UAA-containing molecules. In addition, obtaining the structure of these proteins allows for direct structure:function relationships to be drawn and for improved understanding of the utility of various UAAs. Here, the X-ray crystal structures of proteins with a number of phenylalanine-derived UAAs will be presented and the importance of how these structures inform other experimental results will be described.

Connectivity between subunit surface and interior of alpha/beta hydrolase-fold proteins acetylcholinesterase, neuroligin 4 and Candida rugosa lipase, as suggested by analysis of their X-ray structures

Zoran Radic[1]

[1]Skaggs School of Pharmacy and Pharmaceutical Sciences, UC San Diego, La Jolla, CA 92093

Functionally diverse alpha/beta hydrolase-fold proteins combine elements of secondary and tertiary structure similarly in their 500 – 600 amino-acid globular subunits. Capacity of fold to transmit information between protein surface and its interior can have tangible implications for their function. Here, PDB deposited X-ray structures of two enzymes and neuro-adhesive protein of the alpha/beta hydrolase-fold, have been investigated for conformational changes in their alpha carbon backbones upon either binding of large surface ligands or upon environment-triggered conformational change of a surface loop.
Human acetylcholinesterase (hAChE; EC 3.1.1.7) a key enzyme of cholinergic neurotransmission as a near-perfect biological catalyst evolved to accelerate hydrolysis of acetylcholine at diffusion-limited rates. Hundreds of PDB deposited X-ray structures of AChEs reveal largely one conformation, inconsistent with some of functional data that require, at least transient, opening of one or more of its domains. *Candida rugosa* lipase (CRL; EC 3.1.1.3), a serine hydrolase with fold closely related to AChEs was on the other hand crystallized in distinct conformations. In its "active/open" conformation equatorially located "omega" loop (37 amino-acids) flips open to reveal central catalytic cavity to incoming fatty acid substrates. Such, open conformation was never captured, but was suspected to occur transiently, with smaller magnitude in AChE. Finally, human neuroligin 4 (hNL4), one of essential neuroadhesive molecules, participates in formation and stabilization of synaptic architecture. Devoid of catalytic activity it shares, both tertiary and quarternary structure similarities with AChEs. Changes in backbone conformations of hAChE and hNRL4 upon equatorial, reversible association of 61 amino-acid long fasciculin 2 and 179 amino-acid neuro-adhesive peptide, neurexin1β, as well as comparison of open and closed conformations of CRL were analyzed using PACCT (Pairwise Alpha Carbon Comparison Tool) tool (www.ZENODO.org, DOI 10.5281/zenodo. 3992329) revealing multiple points of connectivity including alpha-helices of the C-terminus, suggesting that ligand binding or loop opening, in their equatorial domains could influence homodimerization domain, located apically at their spatially distant C-termini.
Supported by the CounterACT Program, National Institutes of Health Office of the Director (NIH OD), and the National Institute of Neurological Disorders and Stroke (NINDS), [Grant Numbers U01 NS083451 and R21 NS098998].

Hemoglobin binding and heme extraction by Staphylococcus aureus hemophore IsdB investigated with X-ray solution scattering

Omar De Bei[1], Matteo Levantino[2], Marialaura Marchetti[3], Eleonora Gianquinto[4], Francesca Spirakis[4], Barbara Campanini[1, 3], Stefano Bettati[3, 5], Luca Ronda[3, 5]

[1]Dept. of Food and Drug Science, University of Parma, Parma, Italy, [2]ESRF, Grenoble, France, [3]Biopharmanet-TEC, University of Parma, Parma, Italy, [4]Dept. of Drug Science and Technology, University of Turin, Turin, Italy, [5]Dept. of Medicine and Surgery, University of Parma, Parma, Italy

The relationship between bacteria and human host passes through several protein–protein interactions. Iron is a necessary nutrient for bacteria and is acquired from the host that, on the other hand, keeps it mainly sequestered in hemoglobin. Staphylococcus aureus acquires hemic iron from humans by forming a complex between its hemophore IsdB and human hemoglobin (Hb), when the latter protein is released from red blood cells into the plasma upon infection. Once IsdB (or the other bacterial surface-anchored protein IsdH) has extracted heme from the oxidized form of Hb (metHb), other Isd proteins from S. aureus are involved in heme internalization and degradation to make iron available to the bacterium. The presence of an iron containing chromophore (heme) allows to monitor the extraction process by absorption spectroscopy, but the technique, widely used for characterizing heme-containing proteins, offers only local, relative information without clear clues on global structural features or determinants. Static Small (SAXS) and Wide (WAXS) Angle X-ray Scattering measurements were thus performed to characterize the stoichiometry of metHb-IsdB, the dominant species in vivo, as well as oxyHb-IsdB complexes in solution. This latter hemoglobin form is also known to bind IsdB, but heme extraction does not occur, hence providing a valuable intermediate state along the heme extraction process. The obtainment of high quality X-ray scattering spectra, covering SAXS and WAXS range, allowed us to prove that the metHb-IsdB complex in solution is formed by a Hb dimer with two IsdB monomers bound. OxyHb-IsdB showed an Hb tetramer-dimer equilibrium where IsdB is able to bind with only two molecules per tetramer. Time-resolved X-ray scattering techniques were then exploited to better characterize Hb-IsdB interaction and correlated ligand transfer events and to define a sequential model describing the observed kinetics. A novel setup recently developed at the ESRF was optimized and successfully used to follow the structural changes induced by rapid mixing of IsdB with metHb. The time resolved WAXS data revealed that, at 100 ms after mixing, the equilibrium complex is formed. Following changes were attributed to the movement of proteins domains coupled to heme transfer.

RCSB Protein Data Bank: Integrated Searching and Efficient Access to Macromolecular Structure Data from the PDB Archive

Yana Rose[4], Jose M. Duarte[4], Robert Lowe[1, 2], Chunxiao Bi[4], Charmi Bhikadiya[1, 2], Li Chen[1, 2], Alexander S. Rose[4], Sebastian Bittrich[4], Joan Segura[4], Stephen K. Burley[1, 2, 3, 4, 5], John D. Westbrook[1, 2, 3]

[1]RCSB Protein Data Bank, The State University of New Jersey, NJ 08854, [2]Institute for Quantitative Biomedicine, The State University of New Jersey, NJ 08854, [3]Cancer Institute of New Jersey, The State University of New Jersey, NJ 08901, [4]RCSB Protein Data Bank, San Diego Supercomputer Center, University of California, CA 92093, [5]Department of Chemistry and Chemical Biology, The State University of New Jersey, NJ 08854

The US Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) servesmany millions of unique users worldwide by delivering experimentally-determined 3D structures of biomo-lecules integrated with more than 40 external data resources via RCSB.org, application programming interfaces (APIs), and FTP downloads. Herein, we present the architectural redesign of RCSB PDB data delivery services that build on existing PDBx/mmCIF data schemas. New data access APIs (data.rcsb.org) enable efficient delivery of all PDB archive data. A novel GraphQL-based API provides flexible, declarative data retrieval along with a simple-to-use REST API. A powerful new search system (search.rcsb.org) seamlessly integrates heterogeneous types of searches across the PDB archive. Searches may combine text attributes, protein or nucleic acid sequences, small-molecule chemical descriptors, 3D macromolecular shapes, structure and sequence motifs. The new RCSB.org architecture adheres to the FAIR Principles, empowering users to address a wide array of research problems in fundamental biology, biomedicine, biotech-nology, bioengineering, and bioenergy.

NAD(H) linked tetrameric assembly of the cancer target protein, CtBP; insights from early structural investigations into mammalian hemoglobins

Anne M. Jecrois[1], Brendan J. Hilbert[1], Andrew G. Bellesis[1], Celia A. Schiffer[1], William E. Royer[1]

[1]Department of Biochemistry & Mol. Pharm., Univ of Massachusetts Medical School, Worcester, MA 01605

Tetrameric mammalian hemoglobin played a central role in the early history of protein crystallography and set the standard for understanding how tetrameric assembly of subunits can be used to modulate protein function in physiologically useful ways. These early structures provided an illustration of how tightly assembled dimers that lack allosteric function can assemble into dimers of dimers with strong allosteric properties.

C-terminal binding protein (CtBP) is a co-transcriptional factor with pro-oncogenic activity as a result of its regulation of a broad range of genes involved in cell growth and metastasis. It has been known for decades that both paralogs, CtBP1 and CtBP2, are activated by binding of NAD(H) to induce oligomerization. The prevailing view was that NAD(H) triggers assembly of dimers from monomers, as crystal structures revealed an extensive dimeric interface, which buries substantial surface area (~3,000 Å2). Our multi-angle light scattering (MALS) experiments have demonstrated that NAD(H) triggers the assembly of apo dimers into tetramers, which is consistent with the extensive intradimer interface. Analysis of subunit interactions within crystal lattices of both CtBP1 and CtBP2 revealed a similar association of dimeric pairs into tetramers with much more limited interactions; our mutagenesis studies validated that the observed lattice assemblage is maintained in solution, which was further confirmed by our cryoEM structure of CtBP2. Moreover, cellular studies demonstrated that tetramer destabilizing mutants are defective for oncogenic activity, providing strong evidence of the central role of the CtBP tetrameric assembly in transcriptional function. These findings are driving our development of inhibitors that interfere with tetrameric assembly as potential anti-neoplastic agents.

Although key differences exist between the function of heterotetrameric mammalian hemoglobin and homotetrameric CtBP, an important shared theme is the acquisition of beneficial properties from the assembly of very stable dimers into functional tetramers. In both cases, the active sites between subunits within a dimer are much further apart than those across the tetrameric interface. Such arrangements permit the tetrameric interactions to sense ligand binding and functionally respond to environmental conditions.

Understanding the nucleic acid structure in detail: historical reflections and perspectives for the future

Bohdan Schneider[1]

[1]Institute of Biotechnology of the Czech Academy of Sciences

Base pairing has been recognized as the fundamental self-recognition feature of nucleic acids for a long time [1] while the nucleic acid sugar-phosphate backbone has been considered a scaffold-like structure more or less passively accommodating to and enabling base pairing and stacking motifs formed by the four nitrogenous bases. Such a reduction of the NA backbone role is restrictive specifically during the refinement of experimental structures when the lack of tools to position structurally plausible NA fragments is potentially damaging the overall quality of the refined molecular models. Also computer-aided predictions of (especially) RNA models would benefit from the availability of versatile tools able to manipulate with the energetically possible conformers. The lack and in some cases a little use of such tools is in a stark contrast to the emphasis paid to description and analysis of the geometries of the peptidic backbone in both experimental and bioinformatic studies.

It has been obvious for some time that the experimental models archived in the Protein Data Bank allow a comprehensive description of the backbone variability. The structures of DNA oligonucleotides and their complexes with proteins and of RNA architectures, including the breakthrough ribosome structures, allowed empirical studies that started to systematically analyze the NA backbone variability [2-6]. These studies rather naturally concentrated on the RNA structures with visibly rich ensemble of the backbone geometries. In addition, there have been earlier [7] and recent [8] attempts to systematize description of the DNA conformations but only in the last year a fully automated robust classification of geometries of both DNA and RNA molecules was published [9]. I will briefly overview the existing approaches to annotate and analyze nucleic acid structures and discuss the advantages of the tools we have developed.

[1] Spencer Acta Cryst 12:66 (1959). [2] Duarte NAR 31:4755 (2003). [3] Murray PNAS USA 100:13904 (2003). [4] Hershkovitz NAR 31:6249 (2003). [5] Schneider NAR 32:1666 (2004). [6] Richardson RNA 14:465 (2008). [7] Schneider Biopolymers 42:113 (1997). [8] Schneider Acta Cryst D74:52 (2018). [9] Cerný NAR 48:6367 (2020)

Regulation of actin dynamics by phosphoinositides

Yosuke Senju[1]

[1]Research Institute for Interdisciplinary Science (RIIS), Okayama University, Okayama 700-8530 Japan

The actin cytoskeleton powers membrane deformation during many cellular processes, such as migration, morphogenesis, and endocytosis. Phosphoinositides, especially phosphatidylinositol 4,5-bisphosphate [PI(4,5)P2], regulate the activities of many actin-binding proteins (ABPs), including profilin, cofilin, Dia2, N-WASP, ezrin, and moesin; however, the underlying molecular mechanisms have remained elusive. Here, we applied a combination of biochemical assays, photobleaching/activation approaches, and atomistic molecular dynamics simulations to uncover the molecular principles by which ABPs interact with phosphoinositide-containing membranes. We show that, despite using different lipid-binding domains, these proteins associate with membranes through similar multivalent electrostatic interactions, without specific binding pockets or penetration into the lipid bilayer. Strikingly, our experiments reveal that these proteins display enormous differences in the dynamics of membrane interactions and in the ranges of phosphoinositide densities that they sense. Profilin and cofilin display transient, low-affinity interactions with phosphoinositide-rich membranes, whereas F-actin assembly factors Dia2 and N-WASP reside on phosphoinositide-containing membranes for longer periods to perform their functions. Ezrin and moesin, which link the actin cytoskeleton to the plasma membrane, bind membranes with high affinity and slow dissociation dynamics. Unlike profilin, cofilin, Dia2, and N-WASP, they do not require high 'stimulus-responsive' phosphoinositide density for membrane binding. Together, these findings demonstrate that membrane-interaction mechanisms of ABPs evolved to precisely fulfill their specific functions in cytoskeletal dynamics.

Structural insight into Pseudomonas aeruginosa translation Initiation factor 1 and its interaction with the 30S ribosomal subunit

Yonghong Zhang , Nicolette Valdez , James Bullard

The University of Texas Rio Grande Valley, Edinburg, TX 78504

Pseudomonas aeruginosa is a common bacterial pathogen that cause nosocomial infections. The high intrinsic antibiotic resistance makes P.aeruginosa as one of the most difficult organisms to treat, which led to an unmet need for discovery of new antibiotic candidates. Bacterial protein synthesis is an essential metabolic process and a validated target for antibiotic development. Structural information of protein targets in P. aeruginosa protein synthesis (PAPS) is therefore needed for rational design of inhibitors based on structure-activity relationship. This study is focused on P.aeruginosa translation initiation factor 1 (Pa-IF1) and its interaction with the 30S ribosomal subunit for structural insight into initiation of translation in PAPS. Solution NMR techniques were used to determine Pa-IF1 structure and to map out the binding interface of Pa-IF1 to the 30S. The key residues of Pa-IF1 involved in the binding were identified by NMR titration. Pa-IF1 functional assay was performed using the established A/T protein synthesis system from P. aeruginosa. Pa-IF1 consists of a five-stranded β-sheet with an unusual extended β-strand at the C-terminus, and one short α-helix. The structure adopts a β-barrel fold and contains an oligomer-binding motif. A cluster of basic residues (K39, R41, K42, K64, R66,R70, and R72) on the surface near the short α-helix compose the binding interface with the 30S subunit. A structural model of Pa-IF1 and its nteraction with the 30S subunit was built based on the titration result and provides structural information for understanding PAPS machinery.

PDB-Dev: A prototype system for archiving integrative structures

Brinda Vallat[1], Benjamin Webb[2], John D. Westbrook[1], Hongsuda Tangmunarunkit[3], Serban Voinea[3], Carl Kesselman[3], Andrej Sali[2], Helen M. Berman[1, 4]

[1]Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, [2]Department of Bioengineering and Therapeutic Sciences, University of California at San Francisco, [3]Information Sciences Institute, Viterbi School of Engineering, University of Southern California, [4]Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey

Integrative modeling combines information from various experimental and computational techniques to derive the structures of complex macromolecular assemblies that are often elusive to traditional methods of structure determination. Structures determined using integrative modeling present a variety of archiving challenges compared to the structures currently archived in the Protein Data Bank (PDB). To address these challenges, we have created the data standards required for representing integrative models that span multiple spatiotemporal scales and conformational states, as well as the spatial restraints derived from different types of experimental methods. A prototype archiving system called PDB-Dev has been built based on the new data standards. In addition, we are creating a data harvesting system to enable users to put together the information required for archiving integrative models in PDB-Dev. The work is supported by NSF awards DBI-1519158, DBI-1756248 and DBI-1756250 and NIH awards R01GM083960 and P41GM109824.

Outlooks on Claudin Structural Biology and the Basis of Tight Junction Disassembly by a Bacterial Toxin

Alex J. Vecchio[1]

[1]University of Nebraska-Lincoln, Lincoln, NE 68588

Claudins are a family of four-pass integral membrane proteins with 27 subtypes in humans ranging from 23 to 34 kDa. The discrete molecular assemblies of claudins in epithelium and endothelium are linchpins to the supramolecular structure of tight junctions. Claudins impart two critical functions to tight junctions—cell/cell adhesion, and regulation of ions and small molecule transport between intercellular spaces. How claudin structure bestows these and other tissue-specific functionalities to tight junctions remained undetermined until recently. With X-ray crystallography, research teams have now determined nine structures of three mouse and two human claudin subtypes—eight having been enabled by exploiting a claudin/protein interaction refined by nature. The gram-positive bacterium *Clostridium perfringens* causes severe, sometimes lethal foodborne illness in humans by producing an enterotoxin (CpE) that breaks down the gastrointestinal barrier by binding claudins, disrupting their tight junction-forming assemblies. CpE selective recognition of and high affinity binding to receptive claudins is encoded by its 15 kDa C-terminal domain (cCpE). The cCpE adds an essential hydrophilic mass to primarily hydrophobic claudin/detergent complexes and has facilitated crystallization and structural determination of four claudin receptors for CpE. The structures elucidate the basis of cCpE binding and disruption of claudin assemblies. Our laboratory contributed structures of human claudin-4 and -9 in complex with cCpE that provide insights into the molecular and structural element CpE employs for subtype-specific targeting of claudins and the mechanism of CpE-induced claudin disassembly. These structures additionally revealed the molecular bases of distinct claudin functions including selective ion transport, hepatitis C viral entry, and deafness-causing mutants. Here, we describe our cCpE-bound structures, relate them to those of other claudins, and highlight the functional insights they enabled. Further, we present an outlook on what discoveries remain for near atomic-level understanding of claudin assembly, tight junction structure/function, and how structural biology will be central for empowering these breakthroughs.

The analysis of protein fine structure is a valuable tool for quality assessment

Nicole Balasco[1], Luciana Esposito[1], Amarinder S. Thind[2], Mario Guarracino[2], Luigi Vitagliano[1]

[1]1. Istituto di Biostrutture e Bioimmagini, CNR, Napoli, Italy, [2]2. Istituto di Calcolo e Reti ad Alte Prestazioni, CNR, Napoli, Italy

Proteins combine molecular complexity and fine structural regulation. Although protein crystallography has provided an enormous contribution to the development of structural biology, crystallographic data alone are generally not sufficient for effective protein structure refinements. In this framework, the assessment of the validity/quality of protein three-dimensional models represents a fundamental step in the structure determination process. Although over the years impressive advances have been done [1], there are several unmet needs in this field. We and others have shown that backbone geometry and planarity of individual protein residues strongly depend on their local (φ,ψ) values [2-10]. By analyzing the whole structural content of the Protein Data Bank, we here show that the variability of backbone geometry and planarity can be individually detected in the vast majority of protein structures including those refined at low/moderate resolution [Balasco et al. in preparation]. We demonstrate that the detection and the analysis of these fine structural elements are strongly correlated with the overall accuracy of individual structures. These findings clearly demonstrate that the evaluation of these subtle parameters represents an innovative and valuable tool for structure quality assessment. A web server for the automatic evaluation of the local and global quality of protein structures is under development (QuiProQua – QUIck PROtein structure QUAlity assessment at http://study.ibb.cnr.it/quiproqua/index.php).

[1] S. Gore et al. Structure 2017, 25, 1916-1927 [2] L. Esposito et al. J. Mol. Biol. 2005, 347, 483-487 [3] D.S. Berkholz et al. Structure 2009, 17, 1316-1325. [4] D.S. Berkholz et al. PNAS 2012, 109, 449-453. [5] L. Esposito et al. Biomed. Res. Int. 2013, 326914. [6] R. Improta, et al. PLoS One. 2011, 6, e24533. [7] R. Improta et al. Proteins 2015, 83, 1973-86. [8] R. Improta et al. Acta Crystallogr., Sect. D 2015, 71, 1272-83. [9] N. Balasco et al. Acta Crystallogr., Sect. D 2017, 73, 618-625. [10] N. Balasco et al. Biomed. Res. Int. 2017, 2617629.

PDB-101: Molecular Explorations through Biology and Medicine

Maria Voigt[1], Shuchismita Dutta[1], David S. Goodsell[1], Christine Zardecki[1], Stephen K. Burley[1]

[1]RCSB Protein Data Bank, Rutgers University

As of March 2021, the Protein Data Bank (PDB) archive contains ~180,000 structures of proteins and nucleic acids. This information enables many areas of research, leads to discovery of new drugs, and supports new product development.

To make this knowledge accessible beyond the research community, the RCSB PDB maintains the educational portal PDB-101 (**pdb101.rcsb.org**). The portal offers curated learning resources that support exploration of PDB data by students, educators and the curious public . The uniqueness of these resources lies in the fact that the biological concepts can be readily explained with three-dimensional visual evidence of molecular structure, interactions, assembly, and function.

PDB-101 resources include the ongoing Molecule of the Month series, educational materials such as paper models, posters, molecular animations, educational curricula and more. The section “Guide to Understanding PDB Data” is a primer for detailed PDB-specific information: PDB Data, Visualizing Structures, Reading Coordinate Files, scientific methods for structure determination, and more. PDB-101 can be searched by molecule name or keyword. A Browse option displays all available resources, organized by topics such as immune system or renewable energy.

PDB-101 also runs annual Video Challenges for high school students. Participants create short videos that tell molecular stories that connect structural biology and medicine. Previous topics have included HIV/AIDS, diabetes, and antimicrobial resistance. The 2021 challenge will focus on Molecular Mechanisms of Drugs for Mental Disorders.

In 2020, PDB-101 focused on developing COVID-19 related materials. As of March 2021, the PDB contains ~1,000 SARS-CoV-2 related structures. PDB-101 resources communicate this research to wider audiences, explaining the functions of the viral proteins and approaches being taken to address the pandemic. Content includes images, Molecule of the Month articles, a new series on Resources to Fight the COVID-19 Pandemic, and curricular materials. One of the resources, the video Fighting Coronavirus with Soap was an Official Selection of the 2020 American Public Health Association Film Festival.

RCSB PDB is funded by the NSF (DBI-1832184), the US DoE (DE-SC0019749), and the NIH (R01GM133198).

Use of soft X-ray tomography to capture the dynamic biochemical nature of intact pancreatic beta cells

Kate L. White[1]

[1]University of Southern California

Characterizing relationships between cell structures and functions requires quantitative mesoscale mapping of intact cells showing subcellular rearrangements following stimulation; however, current approaches are limited in this regard. Thus, I investigated the application of soft X-ray tomography to generate three-dimensional reconstructions of whole pancreatic β-cells at different timepoints following glucose-stimulated insulin secretion. Reconstructions following stimulation show distinct insulin vesicle distribution patterns reflective of altered vesicle pool sizes as they mature and travel through the secretory pathway. My results show that glucose stimulation causes rapid changes in biochemical composition and/or density of insulin packing, increased mitochondrial volume, and closer proximity of insulin vesicles to mitochondria. Co-stimulation with Exendin-4 (a glucagon-like peptide-1 receptor agonist) prolongs these effects and increased insulin packaging efficiency and vesicle maturation. This approach provides unique perspectives on the coordinated structural reorganization and interactions of organelles that dictate cell responses. To extract as much knowledge as possible from SXT tomograms, we have developed a machine learning auto-segmentation pipeline for a more high-throughput process to quantify the effects of different cellular conditions or drug stimuli on cell structure, function, and biochemical composition. Furthermore, we can use the information from SXT along with electron microscopy and live cell imaging to assemble a data-driven, dynamic model of organelles within pancreatic β-cells. The integration of data from different time and spatial scales is essential for characterizing the interplay between cellular structure and function during complex biological processes, which will open doors to new dimensions of cell biology research.

Structural characteristics, physicochemical properties, and domain architectures of cobra venom cytotoxin

Nurhamimah Misuan[1], Michelle Khai Khun Yap[1, 2]

[1]School of Science, Monash University Malaysia, 47500 Bandar Sunway, Malaysia, [2]Tropical Medicine and Biology Platform, Monash University Malaysia, 47500 Bandar Sunway, Malaysia

A cytotoxin is a three-finger toxin, mainly found only in cobra venom. There are different isoforms of cytotoxin due to species and geographic differences. This study aims to characterize the amino acid compositions, physicochemical and structural properties of a conserved cytotoxin. A cytotoxin sequence (62 amino acids, 7 kDa) was deduced from the multiple sequence alignment of different isoforms. The physicochemical properties of cytotoxin showed its overall hydrophilicity and thermal stability. Based on the amino acid analysis, the cytotoxin was a lysine-rich toxin with a theoretical isoelectric point of 9.38, which supported the basicity of the toxin. The presence of cysteine residues (12.9%) indicated the formation of disulfide bonds within the toxin. On the other hand, the β-pleated sheets constituted the main secondary structure in the cytotoxin, accounting for 71% of the total residues. Furthermore, the domain architecture of cytotoxin displayed the presence of motif sequences ‘PMKSCP’, ‘CIDRCN’, ‘NLCVKM’ which are highly conserved in the three-finger toxin family. The homology model further proved this point, which revealed the existence of these sequences in the functional loop. These functional loops exhibited flanking hydrophobic and polar residues, indicating that the toxin is an amphiphilic protein. The PDB structure of cytotoxin showed the presence of Ser29, thus, it is an S-type cytotoxin. Collectively, the results concluded the conserved physicochemical and structural characteristics of the cytotoxin.

An Examination of Phylogenetically Related Dehydrogenase Active Sites

Mary Kate Boldyrev[1], Song Yu Yang[2, 4], Manfred Philipp[3, 5, 6]

[1]Pelham Memorial High School, 575 Colonial Ave, Pelham, NY 10803, [2]Department of Molecular Biology, NYS Institute for Basic Research in Developmental Disabilities, [3]Biochemistry Doctoral, Program, Graduate Center of CUNY, [4]Biology-Neuroscience Program, Graduate Center of CUNY, [5]Chemistry Doctoral Program, Graduate Center of CUNY, [6]Chemistry Department, Lehman College, Bronx, NY 10468

PDB databases provide tools for the examination of protein structure and protein ligand interactions. These tools were used to examine the interactions of various nicotine-adenine dinucleotide derivatives (abbreviated here as NAD) to proteins that are structurally homologous to HSD10, a hydroxysteroid dehydrogenase that is active in human mitochondria and has been shown to be essential for normal human brain function (Mol Cell Endocrinol 2019, 489, 92-97). HSD10 has also been implicated in Alzheimer’s disease (J Alz Dis 2018, 62, 665-673).
In the current study, the VAST+ system on the NCBI Protein Structure web site was used to identify proteins with the folding patterns of HSD10. The titles of each of the resulting 11027 HSD10-related PDB structures were used to identify those that contain NAD derivatives. PDB identifiers of interest were entered into the Protein Databank in Europe (PDB-E) search window. This site directly identifies the amino acids that are adjacent to relevant ligands, such as the NAD derivatives under study here.

FASTA protein sequence files for selected NAD-containing PDB structures, downloaded from the RCSB PDB database, were provided to the NGPhylogeny.fr site. The resulting phylogenetic trees, generated separately from sequences derived from eukaryotes, bacteria, and archaea, were used to correlate NAD binding interactions with phylogeny.
This study helped determine if web-based analytical tools are suitable for student-based investigations of protein sequence and structure.