6th ISMB Symposium
17 & 18 June 2014

Report by Dr. Clare Sansom (Department of Biological Sciences, Birkbeck)

The Institute for Structural and Molecular Biology (ISMB) was founded in 2002 to promote close links between researchers in the molecular life sciences based in Birkbeck and UCL. One of its principal activities is its regular programme of symposia and retreats. The ISMB symposium, held biannually since 2004, is a two-day conference in one of the constituent colleges featuring talks by top-ranking researchers in the disciplines it covers. The 2014 symposium was held in Birkbeck's Clore Management Centre on 17 and 18 June with a capacity audience.

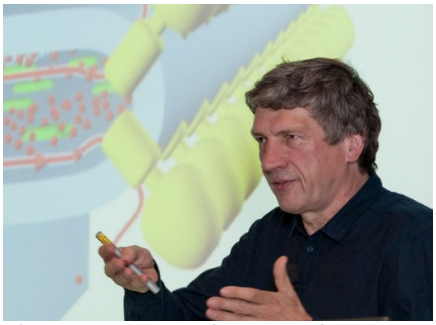
The symposium was introduced by Geraint Rees, director of the Institute of Cognitive Neuroscience at UCL, who is due to take over from Mary Collins as Dean of the Faculty of Life Sciences there in September 2014. Rees mentioned the breadth of the programme, adding that the talk titles as given in the programme seemed more comprehensible to those outside the speakers' disciplines than those of earlier symposia. He praised the vision and energy of the "indefatigable" Gabriel Waksman who in his dozen years as ISMB director has already seen off two deans and one Provost at UCL.

As previously, the programme was split into sections representing the Institute's main research disciplines, with a sixth discipline, proteomics and mass spectrometry, added to those covered in 2012: structural biology, computational biology, biophysics, chemical biology and biochemistry with cell biology. This programme was designed to reflect the breadth of research at the ISMB and of current research in the molecular life sciences more generally. Each section featured two or three talks, with one in each presented by one of the core research staff in the ISMB: there were a total of fifteen talks, two more than in 2012.

The symposium began with the section on structural biology. The first speaker was **Werner Kühlbrandt** of the Max Planck Institute of Biophysics in Frankfurt, Germany, whose talk was one of two to feature electron microscopy. He started by presenting an accessible introduction to the structural biology of membrane proteins. About 30% of all human genes encode proteins that are embedded in cell and organelle membranes, and roughly 50% of all prescription medicines target these proteins. Membrane proteins, however, make up less than 1% of all the structures in the Protein Data Bank. Self-evidently, therefore, these are difficult targets for structural biology. Electron microscopy (EM) is often the most appropriate method for elucidating the structures of membrane-bound proteins, but until very recently it was only possible to obtain low-resolution structures using this technique. Thanks to a new generation of electron microscopes that can detect electrons directly, however, researchers can now solve EM structures to a precision where the position of each non-hydrogen atom can be estimated. Many X-ray crystal structures, particularly of large protein complexes, are no more precise than these.

Kühlbrandt went on to describe some of his recent structural studies of membrane proteins. Cation-proton antiporters are transport proteins found in the membranes

of all organisms. Human antiporters are important drug targets, but their mechanism of action is poorly understood. His group has determined structures of an archaeal sodium-ion antiporter at a range of sodium ion concentrations. They observed that changes in ion concentration lead to changes in the position of some transmembrane helices causing a rapid flip between a form that is open to the outside of the cell and one that is open to the cytoplasm. This has led him to propose a mechanism for rapid



ion exchange across the membrane. He also highlighted studies of the structures of mitochondria using the lower-resolution technique of electron tomography, which have shown the typical organised structure of the mitochondrial membrane to be disrupted during ageing. This is likely to reduce the amount of ATP that mitochondria can supply to cells and thus disrupt cell function.

Andres Ramos from the ISMB then presented his work in understanding how nucleic acid-binding proteins regulate the metabolism of

RNA. There are many multi-functional RNA binding proteins that control RNA biosynthesis and metabolism. These are multi-domain proteins, and many of them use several domains to recognise different RNA targets. A key question in RNA biology is how these proteins are able to recognize a diverse range of RNA targets and therefore to regulate different metabolic processes.

Ramos described how his group is using NMR spectroscopy in conjunction with other biophysical techniques to probe the structures of a group of these proteins, the FUSE binding proteins (FBPs), and their interactions with their RNA targets. FBPs contain four ssRNA-binding K-homology (KH) domains. Based on their structural and biophysical data, the Ramos group could mutate individual domains to either eliminate RNA binding or change sequence specificity. Studies of the interaction between wild type and mutant forms and their RNA target(s) showed that the KH domains bind RNA in a combinatorial fashion and that individual domains can play a different role when binding to different targets. Changing the nucleic acid sequence specificity of one of the KH domains impairs one of the protein's functions. This implies that it may be possible to target selectively the interaction between the protein and a limited subset of targets, leading to a specific post-transcriptional manipulation of gene expression.

For the first time in its history, the ISMB had the honour of welcoming a Nobel laureate as a symposium speaker. This was **Venki Ramakrishnan** from the MRC Laboratory for Molecular Biology in Cambridge, who was awarded the Nobel Prize for Chemistry in 2009 with Tom Steitz from the US and Ada Yonath from Israel for structural studies of ribosomes. Until very recently, all structures of intact ribosomes were determined using a combination of X-ray crystallography and electron microscopy; now, however, Ramakrishnan's group and others are taking advantage of the new higher-resolution direct electron detectors described by Kühlbrandt to determine near atomic resolution structures using this technique alone.

Ramakrishnan described some of the structural features of the yeast mitochondrial ribosome, the first from a eukaryotic mitochondrion. Mitochondrial ribosomes only synthesise a few proteins, but serious diseases still arise if this synthesis is disrupted. Understanding the structure of this ribosome and how it differs from others may hold clues to understanding and eventually treating these diseases.

Ramakrishnan and his team were able to visualise all the proteins in its large subunit, which is more stable than the small one. They identified two proteins that had not previously been thought to be part of this structure, and traced the path of the exit tunnel through which new protein chains emerge. This is different from that in other ribosomes and may position protein chains so they easily become anchored to the mitochondrial membrane.

The two talks in the biophysics section focused on techniques for visualising single molecules in cells. **Alan Lowe** from the ISMB and the London Centre for Nanotechnology focused on the motion of molecules into the nucleus through the nuclear pore complex. He began with an analogy to help newcomers to the field understand the value of single molecule techniques. Showing real-time data of taxis moving through the streets of San Francisco he explained that fluorescence experiments could now track the path taken by a single molecule through a cell, just as this data is used to track the path of an individual taxi.

Intriguingly, he compared the nuclear pore complex – a large channel with 8-fold radial symmetry - to Maxwell's Demon, a thought

experiment involving a door that opens to let only certain molecules through which produces behaviour that seems to violate the Second Law of Thermodynamics. The similarity lies in that the pore complex treats different types of molecule in different ways. Small molecules simply diffuse through the complex into the nucleus, some larger ones are excluded completely but others can enter the nucleus if bound to transport receptors, perhaps against their concentration gradient. Lowe and his co-workers designed protein-tagged nanocrystals that translocate through the pore and tracked their progress into the nucleus using fluorescence. Most large molecules are rejected at the pore entrance, some can explore the space inside the pore without leaving and only a few percent enter the nucleus. Transport into the nucleus is a multi-step process and several pore and non-pore proteins, including a Ran GTPase, appear to be essential for ensuring that the correct molecules leave the pore.

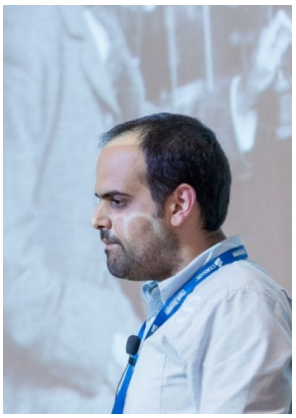
Achilles Kapanidis from the Department of Physics at the University of Oxford presented some of his group's studies of DNA transcription using single-molecule methods. This is a complex process; in prokaryotes, the RNA polymerase binds to the promoter region of DNA before the transcription start site "melts" to form a single-stranded DNA bubble where RNA synthesis can take place. Several of the structures formed during this process are too transient to be observed using traditional structural biology. Kapanidis' group is examining these structures using single-molecule Fŏrster resonance energy transfer (FRET). This technique generates a fluorescent signal that depends on the distance between two probes and can be used as a "molecular ruler" to measure distances of about 2-10 nm. Using real-time measurements of the process as it occurs, they observed a "scrunching" of the DNA conformation at the start of transcription and showed that transcription initiation, like RNA elongation, is a heterogeneous process that can include pausing, back-tracking and recovery in RNA synthesis.

Kapanidis also described his use of fluorescence to localise and monitor RNA polymerase activity in live bacteria. A single *E. coli* cell typically

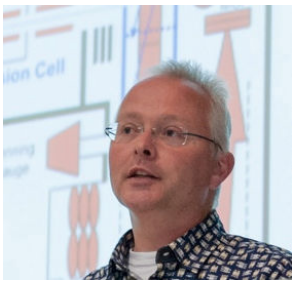
contains about 8,000 molecules of this protein. When the bacteria are grown rapidly in nutrient-rich media, these molecules are organised across each cell in a clear pattern of 2-8 bands that in turn often contain large clusters of particularly tightly packed molecules. In contrast, if the same bacteria are grown in nutrient-poor media, the banding disappears and the tightly-packed clusters of polymerase molecules decrease in size.

The two speakers on proteomics, **Konstantinos Thalassinos** from the ISMB and **Albert Heck** from Utrecht University in the Netherlands were linked in that Heck had been the external examiner for Thalassinos' Ph.D. thesis. Not surprisingly, therefore, the research they presented involved the same technique – native mass spectrometry – and they even started their talks with the same analogy: "electrospray provides wings for molecular elephants". This has an excellent pedigree as it was used by John Fenn, who was awarded a share in the 2002 Nobel Prize for Chemistry for the invention of this technique, in his Nobel lecture. Mass spectrometry has been in use for over 100 years, but it was not until Fenn and others invented ionisation techniques that are gentle enough to leave fragile proteins and protein complexes intact that it found widespread applications in biology. The "molecular elephants" referred to are large protein assemblies such as RNA polymerase and whole viruses that this technique enables to "fly" through a mass spectrometer.

The basic principle of mass spectrometry is the separation of ionised molecules according to their mass-to-charge ratio, allowing the user to determine their molecular mass. Thalassinos described mainly unpublished work using a variant of this technique, ion mobility mass spectrometry (IM-MS). In IM-MS, charged proteins passing through a mobility cell are slowed by collisions with molecules of a buffer gas; the number of these collisions depends on the cross-sectional area of the protein, so differences in cross-section and thus protein shape can be determined from differences in the time the ions spend in the mobility cell. Thalassinos' group has developed a novel software package, Amphitrite, to process IM- MS data, and this is being applied to studies of protein folding, function and aggregation.



Heck described his group's use of mass spectrometry to probe the structure of a protein-RNA complex involved in the defence of bacteria against infection by bacteriophages. This complex, known as Cascade, comprises five types of protein and a single RNA molecule that binds to viral DNA, inhibiting its replication like siRNA does. Adding organic solvent gently dissociated the complex into sub-complexes comprising different combinations of subunits, and electrospray mass spectroscopy then enabled the group to determine the basic subunit structure and stoichiometry of the complex. The protein stoichiometry was found to be ($A_1B_2C_6D_1E_1$) with a single A subunit loosely attached to the rest of the complex, the two B subunits tightly bound together and the six C subunits forming a spiral structure: the intact complex bore a slight resemblance to a seahorse. An EM structure of Cascade determined six months later showed this conformation to be correct apart from the location of the RNA. In another example Heck showed how native mass spectrometry was guiding the design of hexamers of immunoglobulin G, which were shown to enhance complement activation *in vitro* and *in vivo*.



The first day ended with a short but very stimulating "Question Time" session in which PhD students from the ISMB posed questions about research careers and science policy to a panel of five speakers.

The second day started with three talks on chemical biology. The first was given by **Christopher Hunter** from the Department of Chemistry at the University of Sheffield, who described computational methods for predicting DNA packing in the nucleus. Chromosomal DNA is wound tightly around histone proteins to form structures known as nucleosomes. Some regions of DNA are rich in nucleosomes while others, typically close to transcription start sites, are nucleosome-free, and the interplay between nucleosome and transcription factor binding is a determinant of gene expression. It is theoretically possible to calculate nucleosome binding affinities using molecular simulation, but this is time consuming and, with 10^{88} potential nucleosome-binding sequences in the human genome, clearly impossible in practice.

Hunter described a fast method developed in his lab for calculating the nucleosome-binding affinity of DNA segments based on sequence. X-ray structures of nucleosomes show the DNA coiling round the histone proteins with an average bend of 30^{o} over each 10 base pairs but with a distinct kink each side of the centre. Using a single geometric property of dinucleodide steps known as roll, they were able to calculate the strain energy required to bend any sequence of DNA into this conformation. Twelve sequences selected from a pool of 250,000 random ones as having low strain energies were all found experimentally to have high nucleosome-binding affinities, with one close to the highest-affinity sequence known.

Aletha Tabor from the Department of Chemistry at UCL and a member of ISMB described her work designing and synthesising novel self-assembling nano-particles for use as drug delivery vectors and in medical imaging. Many small-molecule drugs can be successfully delivered to tumours by encapsulating them within lipid vesicles called liposomes. Tabor's work is an extension of this concept to improve selectivity and incorporate nucleic acids as cargo. A lipopolyplex is a self-assembling complex that comprises a cationic lipid, a cationic peptide, and plasmid DNA bound into the centre of the particles. These bind to the target cells via an interaction between the peptide and cell receptors, and the lipids fuse with the cell membrane, releasing the DNA into the cell. They are as effective as viral vectors, have low toxicity and, unlike viral vectors, do not stimulate the immune system.

Polyethylene glycol (PEG) is often added to the surface of liposomes to improve their stability, but this has the disadvantage of reducing target specificity and the efficiency of cargo uptake. Tabor tested the addition of PEG to both the peptide and the lipid components of her nano-particles and found that stability only increases if it is bound to the lipid. She is continuing to improve the properties of these multi-component particles to mimic more of the features of viruses that make them such efficient delivery vectors – without, of course, introducing an immune response – is exploring methods for delivering different types of nucleic acids and designing particles that can be simultaneously used for imaging.

Finally in this section, **Shankar Balasubramanian** from the Department of Chemistry at the University of Cambridge gave an inspiring talk that described how his basic research into DNA polymerase led to the development of a DNA sequencing technique orders of magnitude faster than Sanger's original method. As a young Principal Investigator in the mid 1990s he studied the mechanism of DNA synthesis by incorporating fluorescently labelled nucleotides into the growing strand and realised that labelling each base with a different coloured probe would give the

sequence of the template. This proved to be both an efficient and a precise sequencing method, and as the template strands were immobilised it was possible to sequence many – up to a million – different fragments in parallel. It became clear that it was possible to sequence whole genomes, starting with the tiny genome of a bacteriophage.

Balasubramanian and a colleague, David Klenerman, set up the company Solexa to commercialise this technology in 2007. By the time that Illumina acquired Solexa in 2011 it was possible to sequence 60-100 billion bases in a single long run. Users of current instruments can sequence up to a trillion bases in a run: in other words, about 15 human genomes at 30-fold coverage in three days. This increase in the speed and consequent decrease in the cost of DNA sequencing is transforming many areas of the life sciences and clinical medicine. Balasubramanian, however, is continuing his studies in nucleotide chemistry, incorporating a basic chemical reaction into the sequencing methodology in order to distinguish between chemical modifications of the base cytosine.

The first of two talks in the computational biology programme was given by **Soren Brunak** of the Technical University of Denmark, who is best known as the developer of many widely used tools for protein sequence analysis. In this talk, however, he focused on a completely different type of computational analysis: the data mining of electronic patient records.

Computational biology typically studies molecular data in order to understand disease phenotypes but this analysis essentially turns this around. He is analysing patterns of disease - which diseases are found together and how chronic diseases develop over time – to distinguish between genetic, environmental and treatment-related effects.
Brunak commented that he is fortunate to be working in Denmark, with its relatively small population and very comprehensive patient records. Since 1968, each of six million Danes has been assigned a Personal Identification Number that references, among other things, his or her complete medical history. Once a controlled vocabulary had been defined, it became possible to determine patterns of disease and link back to possible genetic links, finding, for example, that the gene THRA was linked to both male pattern baldness and migraine. Patients with schizophrenia were found to be at low risk of developing some cancers, but it is not yet clear whether it is the disease itself or its treatment that provides the protection. However, many challenges must be overcome before this promising technique can be incorporated into mainstream clinical practice, ranging from legal and ethical concerns to its relative cost.

Nick Luscombe from the UCL Genetics Institute returned to familiar territory in describing research into the regulation of gene expression. Eukaryotic genes contain protein-coding regions or exons separated by non-coding introns; alternative splicing allows exons to be arranged in different ways to form several mature mRNAs and thus proteins from a single gene. Occasionally, a non-coding sequence can be incorporated into a mature mRNA transcript by mistake, and these "pseudo-exons" produce aberrant proteins. The splicing process is regulated by two ubiquitous proteins that compete with each other to bind to splice sites: U2AF65, which activates splicing, and hnRNP C, which

represses it. A survey of all human mRNA transcripts showed that almost all contain at least one binding site for these proteins.

Working in collaboration with Jernej Ule at the UCL Institute of Neurology, Luscombe used an immunoprecipitation method, iCLIP, to study protein binding to these sites at single-nucleotide resolution. They discovered that over 20% of hnRNP C binding to human DNA is to repetitive sequences known as Alu elements. There are about 650,000 of these in the human genome, and they can be incorporated into mature mRNA giving rise to non-functional proteins. The repressor hnRNP C has been found to play a crucial role in preventing Alu elements being incorporated into mRNA transcripts as "pseudo-exons": Luscombe described it as a "guardian of the transcriptome" reminiscent of the "guardian of the genome", p53.

The final section was devoted to biochemistry and cell biology. **Marino Zerial** of the Max Planck Institute of Molecular Cell Biology and Genomics had given his talk in this section on the first day in a swap with Alan Lowe. He began by describing his studies of endocytosis, the complex process through which molecules that cannot pass through cell membranes are taken up into vesicles to enter cells, at different levels of complexity. Some years ago, he had identified the small GTPase Rab5 as an important regulator of the early parts of this pathway, and demonstrated that it is replaced by Rab7 in a process termed Rab conversion during early-to-late endosome maturation. A quantitative functional genomic screen has allowed his group to identify several pathways regulating endocytosis and discover that growth factor mediated signalling controls the number and size of endosomes and their distribution within the cell. A quantitative analysis of phosphorylated epidermal growth factor receptor internalized in endosomes has revealed that such a mechanism regulates the extent, duration and robustness of the signalling response.

He then moved up a level of complexity to discuss the 3D organisation of liver cells to form tissue. The first models of liver structure, involving almost crystalline sheets of epithelial cells, were formulated over sixty years ago. His group tested this model using new imaging and analysis tools and found the tissue to be more disordered than it assumed. However, each hepatocyte showed an underlying polarity which, at the tissue level, gives rise to an ordered network of sinusoids (capillaries) and bile canaliculi through which bile is drained.

The next talk was given by **Emmanuel Boucrot** of the ISMB, who presented unpublished work characterising a novel pathway for endocytosis that does not require the vesicular protein clathrin but instead is dependent on and controlled by several endophilin isoforms.

Alison Lloyd of the MRC Laboratory for Molecular Cell Biology, also at UCL, gave the final presentation. She described her studies of the complex process through which mammalian peripheral nerves regenerate after injury. In the peripheral nervous system, neurons are surrounded by cells known as Schwann cells. Some of these – myelinating Schwann cells – wrap around axons to form the protective myelin sheath. Lloyd showed that when a nerve is severed the Schwann cells will de-differentiate into a stem cell-like state, proliferate and form a bridge between the stumps that allows the nerve to re-grow.

Neurofibromas are benign nerve cell tumours that are rare except in individuals with mutations in the gene *NF1*. They are derived from non-myelinating Schwann cells and resemble wound tissue but cannot regenerate. Loss of *NF1* leads to hyper-

activation of the Ras/Raf signalling pathway, which in this case drives the de-differentiation of the Schwann cells. Lloyd used a transgenic mouse model to show that Raf activation in Schwann cells caused reversible de-differentiation, an inflammatory response and degradation of the myelin sheath. Further proteins have been shown to mediate the organised cell migration that leads to the re-growth of nerves along their original pathway: these are the "stemness factor" Sox, the tyrosine kinase Ephrin-B and its receptor. And finally, axon re-growth was shown to be mis-directed in mice in which this kinase receptor, EphB2, is knocked out, suggesting that this process is important *in vivo* as well as *in vitro*.

Nicholas Keep, Dean of the Faculty of Science at Birkbeck, concluded the symposium. He mentioned the wide variety of biochemical, biophysical and computational techniques that had been discussed and the elegant results described, and commented on the absence of his own discipline, X-ray crystallography, from the programme. Finally, he paid tribute to Waksman's leadership and highlighted the essential role played by the ISMB administrator, Andrew Service.

The next ISMB symposium will take place in June 2016.