January 23rd 2022
Issue 4

Scientific Annual Review

A collection of scientific advances in the research lines of CIC bioGUNE

www.cicbiogune.es

Advanced Therapies

CIC bioGUNE
MEMBER OF BASQUE RESEARCH & TECHNOLOGY ALLIANCE

Contents

The Center

CIC bioGUNE is a collaborative research center focused on Life Sciences, from Chemistry to Biomedicine, from basic to translational Science. Our cutting-edge scientific activity concentrates on discovering the molecular bases and mechanisms of disease to promote development of advanced therapies. Our activity explores four main biomedical research themes like Cancer, Metabolic Diseases, Rare Diseases, and Infectious Diseases organized in two research programmes "*Metabolism and Cell Signaling in Disease*" and "*Molecular Recognition and Host-Pathogen Interactions*".

With our collaborative philosophy we are deeply engaged in multidisciplinary research collaborations with local, national, and international colleagues and technology experts. The center is impinged in a heterogeneous network of Academic and Clinical Entities, Research and Technology Centers, and is member of the Basque Research and Technology Alliance (BRTA).

The scientific activity is supported by cutting-edge infrastructures and technology platforms, including advanced equipment for nuclear magnetic resonance (NMR), now recognized as ICTS, electron microscopy, a facility for monoclonal antibody production, as well as different core technology platforms where genomes, proteomes and metabolomes can be analyzed.

General View

Direction

CIC bioGUNE activities are strongly related to our specific mission: to build up an EU-referent knowledge pole in biosciences, which should be able to favour the development of the emerging sectors in the bioscience and health fields, and the incorporation of the proper technologies to be able to enhance the competitiveness of the corresponding industrial (biotech, pharma, etc) sectors. Specifically, CIC bioGUNE acts with a strong commitment of collaboration and coordination with the rest of social and scientific agents within BRTA and in the Basque Country to optimize the existing capacities, and jointly conform an integrative scientific and technological offer of excellence. This offer should be able to boost the evolution of the economy by strongly increasing its intrinsically high added value. Our research activities cover from the gene to animal models of cellular processes through the determination of biomolecular structure and assembly and the elucidation the key mechanisms and interaction patterns at the highest resolution. Our scientific objectives are transversal and target the complete characterization of the molecular basis of protein-based processes in human pathophysiology and immune defence, cell proliferation, and development. The final aim is to translate our findings to the clinic, with special interest in precision medicine.

Cancer Immunotherapy in the Post-Pandemic Era

Asís Palazón, *Cancer Immunology and Immunotherapy Laboratory*

After the global hiccup in clinical trial activity suffered by the drug discovery industry during 2020 because of Covid-19 restrictions, 2021 has witnessed several relevant advances in cancer immunotherapy. These include the expansion of checkpoint receptor blockade benefits in more patients and cancer types. Advances in biomarker discovery further help clinicians to predict which patients are going to benefit most from treatment. Combinations remain a favourite approach in clinical testing, and cell therapy is progressing in the challenge of achieving efficacy in solid tumors.

Apart from travel and testing restrictions, cancer patients have suffered from Covid-19 because many cancer therapies, including chemotherapies, can suppress immune responses, resulting in a higher risk of suffering more serious effects after SARS-CoV-2 infection.

Regarding checkpoint receptor blockade, the PD-1/PD-L1 axis remains the most successful in terms of efficacy and development. Additional targets that are being evaluated in the clinic include TIM-3, LAG-3, and many others. Combinatorial approaches are being extensively tested in the clinic. An alternative approach to combinations is the use of bispecific antibodies, which can engage or block two different costimulatory or coinhibitory receptors at the same time. This is the case of an anti-PD-1 and anti-CTLA-4 monovalent bispecific antibody developed by AstraZeneca, which can unlock the potential of CTLA-4 inhibition[1]. This drug candidate is undergoing clinical trials. Other combinations with checkpoint receptor blockade undergoing clinical trials include chemotherapy, bevacizumab, or PARP inhibitors in the case of ovarian cancer.

The field of CAR-T has also interesting advances in both haematological cancer and solid tumors. First, idecabtagene vicleucel (Abecma), an anti-BCMA CAR-T, has been approved for the treatment of multiple myeloma in 2021. Approaches under clinical development in hemato-oncology include bispecific agents that are being tested for the treatment of large B cell lymphoma. After treatment with the pioneering anti-CD19 CAR-T therapy, more than 50% of patients experience progressive disease or relapse because of low expression or loss of the target antigen CD19. In order to prevent relapse, Spiegel et al recently tested an antiCD19/anti-CD22 multispecific CAR-T in a phase I clinical trial, demonstrating that low or lack of CD19 expression is a major pathway of resistance after anti-CD19 CAR-T treatment[2]. The clinical success of CAR T-cell therapy in haematological cancers has supported the efforts of further developing this technology in solid tumours. However, solid tumors present certain features that are termed the tumor microenvironment and pose a barrier of resistance to cell therapies that must infiltrate, recognise, and kill the tumor mass. There are several ongoing clinical trials developing CAR-Ts for solid tumors, based on different strategies, either directly targeting tumor cells or different stromal components such as endothelial cells. During 2021, a phase 1 clinical trial (Tmunity Therapeutics) testing anti-PSMA CAR-T cells against prostate cancer was stopped because two patients died from immune effector cell-associated neurotoxicity syndrome (ICANS), a side effect that is not as well characterised as the described cytokine release syndrome (CRS)[3].

In this context, as we enter into 2022, the main challenge for cell therapy is the transfer of CAR-T technology to solid tumors, facing more risks and barriers compared to hemato-oncology indications. Many ongoing clinical trials will clarify the potential of this approach during the coming months.

A Better pair of Glass to Look Cancer

Arkaitz Carracedo, *Cancer Cell Signaling and Metabolism Laboratory*

2021 has left us with outstanding science in the field of cancer research. A common feature of scientific advances this year has been the democratization of state-of-the-art technologies, especially due to a better consensus on the computational and bioinformatics tools that we have at our disposal to squeeze the data emanating from high-throughput approaches. We will highlight 2 areas of research that have seen significant discoveries, namely the relationship between the tumor cell and the cancerous habitat and the cellular states that underscore cancer progression.

The cancer cell and its habitat. The bidirectional relationship of the cancer cell and the tumor microenvironment is unveiling as a critical aspect of the disease, influencing cancer aggressiveness, therapy response and overall patient outcome. We now understand that cancer cells remodel their environment in order to endure, and, conversely, that microenvironmental perturbations reprogram the function of tumor cells.

Cancer cells remodeling the habitat: An illustrative example of the interaction between the tumor cell and the stroma was provided by the Bravo-Cordero lab in a beautiful study[4]. They reported that dormant cancer cells, tumor cells that can remain clinically inactive for years or decades, require a distinct extracellular matrix composition. Interestingly, these cells favor a microenvironment rich in Collagen III, and alterations in the abundance of this collagen type can push the cells out of dormancy into proliferation. Whether other cell types within these target tissues contribute to the production of Collagen III matrix, and if the manipulation of Collagen III production can represent an attractive biomarker or therapeutic target has become a fascinating research question.

More specifically in prostate cancer, single cell analysis has provided some surprising evidence. Years of bulk gene expression analysis has led to a series of assumptions regarding which gene expression alterations in cancer emanate from tumor cells. However, with the advent of single cell gene expression analyses, these results are being reinterpreted. As an example, an exhaustive single cell RNA Seq analysis in prostate cancer specimens revealed that the expression of classical prostate tumor cell-intrinsic genes could be detected in immune cells, including bona fide androgen responsive genes. Beyond the therapeutic implications of these findings, these results are a wakeup call for those of us that build hypothesis based on the transcriptomics analysis bulk cancer specimens[5].

The question remains: what mechanisms do cancer cells employ to remodel their habitat? There are a variety of paracrine communication modes that have been identified in this regard, from extracellular vesicles[6] to metabolites4[7]. The therapeutic actionable nature of this paracrine communication is an emerging field that could yield new modes of therapeutic intervention in the near future.

The habitat re-educates cancer cells. The habitat where a tumor resides represents one of the major natural selection forces to dictate tumor adaptation, evolution, or extinction, and hence cure. When we think about the habitat, we can easily picture the cells in the tumor microenvironment influencing tumor behavior. Indeed, this concept is the foundation of modern immunotherapy, where medicine aims at de-inhibiting immune cells from tumor-induced inactivation to increase therapeutic efficacy. Indeed, resistance mechanism to current therapies, including targeted therapies, can boost an immune-evasive phenotype that reduces the efficacy of immune checkpoint inhibitors[8]. But the span of habitat-dependent reprogramming of the tumor goes beyond this cellular crosstalk. To provide two recent examples, we continue to learn new processes that are

altered in tumor cells when oxygen is scarce, including the epigenetic landscape[6][9,10]. On the other hand, our diet can influence the behavior of the tumor, and how the tumor interacts with stromal cells to acquire aggressive features. Palmitic acid exposure in cancer cells (a lipid heavily increased in western obesogenic diets) triggers a metastasis-promoting cancer cell program that engages profound alterations in the tumor stroma, namely Schwann cells[11].

The persister nature of cancer cells. The mortality of cancer is largely due to its exceptional interindividual and intratumoral heterogeneity, which hampers the development of universal diagnostic and therapeutic strategies. As such, this disease requires tailored and personalized strategies. With better molecular and histological classification of tumors, our therapeutic arsenal has become more selective and efficacious, and the rate of successful cancer eradication has increased. However, many cancer patients exhibit disease recurrence after a long period of response to a given treatment, or present disseminated disease after years in remission. These phenomena have been attributed to the existence of resilient cancer cells that can survive in hostile environments for long periods, until they acquire the capacity to resume active growth. Even with cancer therapies that show high initial efficacy in combating the tumor, a small population of cancer cells often survives to drug treatment that is termed drug tolerant persister cells (DTP). In addition to driving cancer relapse if the treatment is discontinued due to toxic side effects, these DTPs are thought to constitute a reservoir of cells that may eventually acquire further genetic or epigenetic alterations leading to bona-fide drug resistance[12]. While the specific molecular mechanisms for establishing the drug-tolerant state and how it leads to the subsequent development of resistance are not completely understood, cell plasticity is likely to play an essential role. 2021 has seen notorious advances in the comprehension of the regulation of DTPs. Candidate genes and pathways, such as CD74, mitochondrial pyruvate carrier 1 or FAK, have been proposed to promote a DTP state in cancer cells upon therapeutic challenge[13,14,15]. In addition, a physiological parallelism has been proposed for these resilient, low-proliferative cells, which were reported to exhibit molecular programs analogous to diapause, a reversible state of suspended embryonic development triggered by unfavorable environmental conditions[16]. Whether a diapause-like phenotype is a hallmark of all DTP is presented now as an exciting possibility. Last, but not least, although DTPs are frequently described as slow proliferating cells, an ingenuous molecular trick based on label retention and barcoding (termed Watermelon) revealed that DTPs can be classified as cycling and non-cycling, and that cycling DTPs exhibit unique molecular and metabolic alterations. Altogether, 2021 has confirmed that DTP is a yet vague and heterogeneous definition, and that refinement of this phenomenon through computational, molecular and cellular analyses might help tailoring future anticancer therapies.

Tackling Cancer Heterogeneity

Maria dM Vivanco and Robert Kypta, *Cancer Heterogeneity Laboratory*

Cancer cells cannot function in isolation: tumors are populated by epithelial cells, stromal cells and immune cells, as well as small populations of normal stem cells and cancer stem cells (CSCs). In particular, the presence of CSCs proves a challenge to therapy. CSCs are therapy-resistant and their enrichment after treatment leads to tumor recurrence. We previously showed that normal and cancer stem/progenitor cells share some features, such as expression of Sox family transcription factors and increased Wnt signaling activity. The latter Wnt signaling underlies a wide range of pathologies, from bone and metabolic disorders to neurodegenerative diseases and cancer. While several drugs targeting the Wnt pathway are in clinical trials, the challenge is to use them without compromising normal tissue homeostasis. This challenge can be met by developing more specific, patient-tailored inhibitors (personalized medicine) by targeting a specific Wnt ligand or receptor that confers tumor stage selectivity.

An area actively being addressed in our lab is the effect of the surrounding tissue environment on the natural selection of mutation-driven tumor characteristics. Three key studies of intestinal cancer identified a novel role for Wnt signaling in this context[17,18,19], demonstrating that the interaction of malignant cells with neighboring wild-type cells can shape the surrounding environment to the advantage of the tumor. Specifically, these studies revealed that malignant stem cells promote neighboring stem cells to differentiate into less-proliferative cell types, thereby winning a competitive ´proliferation battle´ with normal stem cells. While many researchers are focusing on ways to harness the immune system to target cancer, the studies highlighted here suggest that an alternative approach may be to promote stem cell/tissue contexts that are unfavorable to the evolution of malignancy. Indeed, given that healthy tissues can eliminate malfunctional/malignant cells, boosting normal cell fitness by inhibiting the pro-differentiative influence of malignant cells could be a new therapeutic strategy.

We anticipate that interest in Wnt signaling as a target for therapy in cancer stem cells will continue to grow in 2022. Our development of Wnt function-blocking antibodies, their application in colorectal and prostate cancer[20] and their mechanism of action, as well as our development of Sox family inhibitors[21] put us in an excellent position in this arena.

Mechanisms leading to aging and chronic associated liver disease

Malu Martínez-Chantar, *Liver Disease Laboratory*

Liver disease is a major global health problem, affecting an estimated 1.2 billion people worldwide. Moreover, therapeutic options for these diseases remain limited due to the lack of analyzes that allow defining the molecular mechanism underlying disease progression. Single cell transcriptome technologies are transforming the field by advancing our understanding of cellular heterogeneity, basic liver biology, and cellular mechanisms underlying liver regeneration. The application of these methods also identifies the mechanisms that cause disease states such as liver fibrosis and liver cancer. The application of these technologies not only reveals the existence of transcriptional regulation, but also identifies a post-translational system that controls cell behavior in the liver. In the near future, new techniques such as spatial transcriptomics, proteomics and metabolomics, as well as multi-omics approaches, will further our understanding of disease pathogenesis and enable the identification of new therapeutic targets for the treatment of patients.

Using single cell sequencing, Sharma et al. have identified unexplored oncofetal reprogramming of the tumor ecosystem in liver cancer, supporting novel targets for therapeutic intervention in this pathology[22]. In this context, spatially resolved analysis of FFPE tissue proteomes by quantitative mass spectrometry in hepatocellular carcinoma has been developed, providing clues to protein expression or post-translational modifications in patient samples that could complement information obtained by genomics and standard diagnostic methods[23]. Single-cell omics has also revolutionized the field of aging. Understanding the mechanisms underlying this process and supporting knowledge for the development of interventions for age-associated diseases such as chronic liver disease could be considered[24]. Finally, there is growing evidence linking the gut microbiome to the development of liver disease. Current knowledge of the gut-liver axis in non-alcoholic fatty liver

disease may provide the basis for the development of gut microbiome-based personalized approaches to disease management, including use as a non-invasive biomarker for diagnosis and staging, and as a marker of drug response[25].

Given the new technology being used to develop translational research focused on clinical applications, and considering the clinical networks in which our research team is involved, our efforts will be directed towards a) the identification of new cell populations involved in the development of liver injury, b) intercellular communication, and c) the reprogramming of individual cells in the context of liver zonation throughout the disease. Finally, modulatory mechanisms leading to aging and chronic associated liver pathology and possible modulatory therapeutic interventions will be investigated.

Computational Stem Cell Biology: Designing New Strategies for Tissue Regeneration and Rejuvenation in Ageing and Disease

Antonio del Sol, *Computational Biology Laboratory*

The field of stem cell research has been revolutionized by the discovery of induced pluripotent stem cells almost two decades ago. Although it has created new opportunities for studying human ageing and diseases, the ultimate goal of tissue regeneration and rejuvenation – restoring the function of damaged or aged tissues – still faces a number of challenges. Here, especially the efficient in vitro generation of desired cell types is a major goal of stem cell research and holds great potential for medical applications. While some recent work has been done in this area, there are still no general methods for selecting instructive factors for high conversion efficiency. In addition, although in vivo reprogramming strategies have recently gained momentum in the context of tissue rejuvenation, a holistic rejuvenation that fully restores the function of aged tissues has not yet been achieved. Computational stem cell biology can help addressing these issues by developing computational models at different scales of biological complexity to generate predictions that can guide experimental efforts in the design of therapeutic interventions for tissue regeneration and rejuvenation.

This year, we have witnessed an upsurge in the development of strategies to counteract ageing and to enhance the regenerative capacity of tissues. In this regard, a computational model has been developed that aims at improving cellular conversion efficiency, a major bottleneck in the generation of transplantable cell grafts[26]. In particular, the computational model is combined with a transposon-based genomic integration system for efficient delivery, which demonstrates the importance of a unified computational and experimental approach. Importantly, the generation of a comprehensive library containing 1,564 transcription factors (TFs) made it possible to generate virtually any desired cell type when combined with a computational modeling framework that guides the experiments[27]. Moreover, TF-based cellular conversions have also been recently employed to rejuvenate cells and to restore tissue function. In this regard, a seminal study has demonstrated that partial reprogramming enhances the regenerative capacity of the heart after myocardial infarction by conferring a rejuvenated cardiomyocyte phenotype[28]. In addition to TF-based cell conversion, a recently developed computational model of stem cell-niche interactions predicted signaling molecules to rejuvenate aged muscle tissue[29]. In particular, this study shows that a fraction of quiescent skeletal muscle stem cells is responsible for regeneration and possess genuine 'stemness' properties that deteriorate during aging. Restoring this population by inhibiting the predicted signaling molecules significantly improved tissue regeneration in aged mice.

In the coming years, we expect increasing efforts directed towards the design of novel tissue regeneration and rejuvenation strategies that can be translated to the clinics. These efforts would greatly benefit from the guidance of computational modeling that can, for instance, generate novel hypotheses to assist experimentalists in the design of stem cell therapies for tissue regeneration. In summary, we expect a closer collaboration between experimental and computational researchers that will significantly accelerate stem cell research, including the translation of regenerative medicine applications to the clinic.

Molecular and Functional Characterization of Extracellular Vesicles

Juan M. Falcón-Perez, *Exosomes Laboratory*

The impact of Extracellular Vesicles is found in different areas going from the biomarker discovery and development of therapeutics products, towards active players in disease establishment and development. During this year, a study about metabolic syndrome associated (MetS) network in circulating extracellular vesicles (EVs), has demonstrated that MetS alters the interactions among the circulating endogenous RNA network components in circulating EVs and that this cargo of circulating EVs may have potential translational ramifications for MetS[30]. The role that metabolic enzymes -like COMT- associated to exosomes play in metabolism is still unknown mostly by technological limitations. In a recent study, a liquid chromatography-fluorescence detection (LC-FD) method is established to quantify COMT activities in human erythrocytes and cell homogenates. This method is based on fluorometric detection of the methylated product of 3-BTD. These results give us another method to determine the native activity of COMT in biological samples, to complete understand the role of COMT both in physiological and pathological conditions[31]. In addition new tools will also help to elucidate the COMT-EV role, and in this context a new inhibitor for COMT has been found named Opicapone which shows higher efficiency and fewer safety problems when compared with entacapone and Tolcapone drugs[32]. Non-alcoholic fatty liver disease (NAFLD) is increasing in prevalence worldwide. NAFLD is associated with excess risk of all-cause mortality, and its progression to non-alcoholic steatohepatitis (NASH) and fibrosis accounts for a growing proportion of cirrhosis and hepatocellular cancer. Non-invasive precise methods to distinguish patients with NASH and NASH with significant disease activity and fibrosis when the disease is still modifiable are crucial. During this year a study has reported the results from five clinical studies (n=543) with participants suspected of NAFLD were used for data meta-analysis, demonstrating that quantitative MRI derived biomarkers cT1 and liver fat are suitable for identifying those with NASH, and cT1 is a better non-invasive technology than liver fat to identify NASH patients at greatest risk of disease progression. MRI cT1 and liver fat therefore have important clinical utility to help guide appropriate use of interventions in NAFLD and NASH clinical care pathways in a non-invasive way. Thus, these parameters prove to be extremely useful to convince volunteers for recruitment in studies related to this liver disease[33].

Another area of interest in the field of Extracellular vesicles in the context of their application as therapeutics vectors for personalized medicine is to develop better experimental in vitro models to be able to evaluate in an effective and manner in a more physiological condition. In this sense during this year in

the field of hepatology there has been advances in creating protocols to generate those models. Thus, Animizu et al.[34] reported the generation of a hepatobiliary tubular organoid (HBTO) using mouse hepatocyte progenitors and cholangiocytes. Primary hepatocytes form the bile canalicular network and secrete metabolites into the canaliculi, which are then transported into the biliary tubular structure. Moreover, they acquire and maintain metabolic functions including albumin secretion and cytochrome P450 activities over 8 weeks. In conclusion the article establishes a functional liver tissue able to incorporate a bile drainage system ex vivo. HBTO had the capability of reproducing the transport of hepatocyte metabolites in liver tissue, and to investigate the way in which the two types of epithelial cells establish functional connections. The impact of the finding is to have an in vitro tool to study different aspects of the liver functionality. Although the study focused on CYP activity, it could easily be adapted to cancer research and cholangiopathies, metabolic deficiencies and cancer. Since it started from primary hepatocytes, it can be adapted to personalized medicine, starting from patient biopsies. In addition, another work[35] reported during this year describe the generation of liver organoids starting from H1 human embryonic stem cells (WA-01, passage 27-40) and induced pluripotent stem cells (GM23338) with a chemically defined serum-free media to induce formation of posterior foregut cells, later differentiated into hepatic endoderm spheroids (3D conformation) and finally into hepatoblast spheroids. At this stage, spheroids were reseeded in a high-throughput format and induced to form hepatic Organoids. Hepatocytes were functional, based on secretion of albumin and apolipoprotein B and cytochrome P450 activity; cholangiocytes were functional, based on gamma glutamyl transferase and alkaline phosphatase activity and proliferative responses to secretin. To generate disease models, organoids were incubated with oleic and palmitic acid. They compared gene expression profiles of organoids incubated with free fatty acids or without. Organoids incubated with free fatty acids had gene expression signatures similar to those of liver tissues from patients with NASH. Incubation of organoids with free fatty acid–enriched media resulted in structural changes associated with nonalcoholic fatty liver disease, such as decay of bile canaliculi network and ductular reactions. The importance of the article is to generate from a potentially unlimited source of cells, liver models that can be employed for the study of NASH in vitro. The article describes the differences in gene expression, but the application to study metabolism, metabolome profiling, enzyme activity, and secretome associated to NASH makes the tool very valuable.

Finally, an intense area of research in has been the search for biomarkers to improve implantation rates is booming in the field of assisted reproduction techniques (ART). Recently, it has been published an article in which they analyze the RNA content of the extracellular vesicles (EVs) from uterine fluid (UF), in search of a non-invasive biomarker to detect the window of implantation[36]. They show that the transcriptome of UF-EVs correlates with endometrial tissue transcriptome and that the RNA signature of UF-EVs changes with endometrial status and it is different in patients with successful versus failed implantation after the transfer of one euploid blastocyst in the following cycle. They conclude that UF-EVs could serve as reservoir for potential less-invasive collection of receptivity biomarkers and that their work sets the basis for further clinical studies to identify cut-off points of expression levels of genes for the prediction of implantation in ART. Thus, this article represents a step forward in the design of less-invasive approaches for real time monitoring of endometrial status, necessary for advancing the field of reproductive medicine.

Genomics for precision medicine, increasingly a question of finding the context

Urko Martinez Marigorta, *Integrative Genomics Laboratory*

The research at the Integrative Genomics lab revolves around the genetic basis of complex disease. We analyze pan-omic profiles of patients to illuminate our understanding of disease pathogenesis. We focus on two aspects that are key for gearing genomic knowledge towards the goal of achieving precision medicine, namely the need to understand the heterogeneity of mechanisms whereby each patient develops disease, and the development of predictors of prognosis and therapeutic response. These intermediate steps are key for the discovery and characterization of biomarkers useful in the clinic by tracking prognosis and disease status in each patient. As in previous years, we will contextualize one important development in the basic and clinical sides of genomic medicine, respectively.

In regard to the basic side of the coin, we are gaining increasing appreciation of the complexity of genetic effects. Context matters in genetics, and it matters a lot. The genetics of the transcriptome illustrates this well. On the one hand, cis-regulatory effects are ubiquitous. As shown recently by the eQTLgen consortium, expression levels in blood for 90% of genes are influenced by genetic factors[37]. However, causal variant effects can vary a lot depending on the genetic background and/or the socioeconomic status of the person[38]. This can have enormous implications for classical case-control omic studies, since an overwhelming fraction of changes seen in patient cohorts may not be causal but rather induced by disease. Researchers working with any omic datasets should take this into account.

In regard to the clinical side of the coin, this year has brought solid basis for the generalization of “genomics for the clinic”. For example, implementation of population screening for breast and ovarian cancer in 30-year-old women may lead to 75 fewer cancer cases per 100,000 women screened[39]. In turn, this implementation will benefit massively from targeted use of big data analyses. For instance, electronic health records can be used to detect children at high-risk of suffering a rare disease. Concomitant genetic testing based on rapid whole-genome sequencing leads to extremely quick diagnosis (median of 3 days!) in 40% of acutely ill infants, overall decreasing the length of stay by 90%[40].

Although most approaches implemented at present involve mainly cancer and rare diseases, we expect similar programs attempting generalization to complex disease. The Baylor clinic genetic program, focused on sequencing key genes involved in cardiovascular risk, illustrates well the clinical potential in polygenic disease. Up to one third of individuals tested harbor genetic variants with implications for clinical management, and one in five was diagnosed with either a Mendelian condition or a polygenic risk score value that warrants for changes in diet and lifestyle[41]. Despite the free walk-in nature of the program, the participants presented high rates of dyslipidemia, arrhythmias, and cardiomyopathies. This demonstrates that important fractions of the population need better medical focus, and suggests ways to gear genomic medicine towards improved outcomes in circumstances that at present are not properly attended.

The Role of Glycans in Immune Regulation

Jesús Jiménez-Barbero, *Chemical Glycobiology Laboratory*

Molecular recognition of glycans by glycan-binding-proteins (lectins) are well recognised key events in the modulation of diverse biological phenomena most of them related to host-pathogen interactions, immunity, and cancer. Significant advances in Glycoscience in the past decades have permitted to identify at the oligosaccharide level the preferred glycan binding partners for many human lectins, as well as to characterise their binding complexes. In spite of these developments, our understanding of the rules governing the glyco-code is still incomplete, what severely hampers our possibilities of therapeutically exploit glycosylation, which has been in fact anticipated as a new frontier in cancer and personalised medicine. Important challenges in the field are now to identify glycan-bearing counter-receptors (glycoproteins, lipo-polysaccharides, glycolipids), to understand and control receptor clustering and multivalency, to define the effect of glycan presentation or to characterise in-vivo glycosylation. Recent advances in glycoproteomics[42], mammalian cell glycoengineering[43], and the development of strategies to map glycan binding protein partners in live cells[44] will bring and allow important breakthroughs in the field. Without any doubt, one of the landmarks in Glycosciences in this 2021 has been the identification of glycosylation in RNA[45]. The biological or medical significance of this fact is still unknown, but new research avenues are clearly opened.

In this 2021, we have continued to experience the blows of the pandemic produced by the SARS-CoV-2 virus. It has been demonstrated[46] different C-type lectins (Siglec-1, DC-SIGN, L-SIGN) enhance trans-infection and moreover, influence the role of neutralizing antibodies and provide a lectin-dependent pathway that can enhance infection, even in the lack of ACE-2.

From the molecular recognition perspective, many research groups worldwide, including ours, are interested on deciphering the role of sialic acids in diverse processes related to health and disease. Indeed, these sialic acids are monosaccharides found at the terminal part of N- and O-linked glycans attached to proteins and glycolipids (and now RNA!). Due to the high presence of sialic acid in the surface of cells, these molecules serve as “self” signals for immune cells. Sialic acid-binding immunoglobulin-type lectins (Siglecs) are receptors, expressed on the majority of white blood cells of the immune system, which play critical roles in immune cell signalling through recognition of sialoglycans. Such recognition usually results in blockage of the immune response. Thus, changes in sialylation, such as hypersialylation on cancer cells, can contribute to impaired immune responses via Siglecs. Targeting sialoglycan-Siglec axis is a potential therapeutic approach to suppress the immune system. Different companies are currently developing multiple strategies, including glycan-editing and Siglec-targeting monoclonal antibodies (mAbs) to address this therapeutic area.

Besides Siglec-1 mentioned above, it has been demonstrated that Siglec-9 can also participate as anti-inflammatory and pro-apoptotic checkpoint molecules in COVID-19 disease[47]. Interestingly, engagement of Siglec-9 to lipid-conjugated glycopeptides bearing modified sialic acids can simultaneously inhibit proinflammatory NETotic cell death and induce apoptotic cell death in COVID-19-related inflammation. Our group is now analysing the interaction mode of these modified sialic acids with Siglec-9 at the molecular level, using both structural and biophysical techniques.

On this side, our group aims to analyze at the molecular and atomic level, the interactions between Siglec receptors and natural and modified sialic acid ligands. This information is helping us to better understand the mechanism of action and the rational generation of molecules that can modulate the natural interactions between Siglecs and its ligands. However, the fine binding specificities of Siglecs for elaborated complex glycan structures and the contribution of the sugar and protein context for recognition of sialoglycans at the cell surface are not fully elucidated. Thanks to the development of a library of isogenic human HEK293 cells selectively modified in glycosylation pathways, it has been possible to dissect Siglec interactions in the natural context of glycoconjugates on the cell surface[48]. Using this novel approach, it was found, for example, that sulfation can affect preferences for binding to O-glycan patterns.

On the other hand, structural biology can identify the molecular details underlying the successful transformation of an epitope of a mAb on the Siglec receptor into an effective immunotherapeutic target. By solving the crystal structure of CD22 in complex with the m971 fragment antigen-binding (Fab), we found that a key aspect of the efficacy of chimeric antigen receptor (CAR)-T cell therapy is based on its ability to target a membrane-proximal epitope on the CD22 extracellular domain[49]. Here, we also report that the antibody paratope contains electrostatic surfaces compatible with interactions with phospholipid head groups.

Additionally, sialoglycans serve as ligands or receptors for several viruses, including the influenza A virus and Middle East respiratory syndrome-related coronavirus (MERS-CoV). Much controversy has been generated around the possibility that this could also be the case for SARS-CoV-2 infection[50],[51]. The detection and characterization of a potential interaction between the SARS-CoV-2 spike glycoprotein and host sialoglycans represents a real challenge, as the binding is expected to be very weak and cis inhibition by the heavily glycosylated spike protein containing terminal sialic acid residues is expected to produce a “masking” effect. In the group we have just devised an NMR-based strategy to unambiguously detect the direct binding between the SARS-CoV-2 spike glycoprotein and common sialoglycans.

Finally, we pass to other family of lectins, galectins. While having long been known and thoroughly studied, new biological functions are continuously been disclosed for most of their members, such as Galectin-3 in bacterial infections[52] or Galectin-9 in B-cell signalling[53]. In fact, a remarkable and disturbing fact about galectins is their involvement in a huge range of functions in many cell types and environments. In our group, we have recently characterised the binding specificity of Galectin-4 for differently presented blood group antigens[54], whose biological relevance seems to be related to its antibacterial functions, in particular to its capacity to bind and destroy self-type-glycan antigens producing bacteria. One of the endogenous ligands for Galectins are the poly-N-acetyl lactosamine chains. Its presence in glycolipids and glycoproteins has been related to immune response regulation. Their interaction with the different members of the galectin family is expected to be different, but its characterization at the molecular level represents a real challenge due to the inherent glycan flexibility and repetitive structure. We have recently developed an NMR method to identify, in a residue-specific manner, the preferred binding epitope/s of each galectin in polyLacNAc chains[55].

Extraordinary Breakthroughs in Protein Folding and Design in the Rise of Artificial Intelligence

Gonzalo Jiménez, Osés *Computational Chemistry Lab*

Proteins are essential to life and understanding their structure paves the way for a mechanistic understanding of their function. Through an enormous experimental effort sustained through decades, the structures of around 100,000 unique

proteins have been determined by X-ray crystallography, cryo-electron microscopy (cryo-EM) and nuclear magnetic resonance (NMR), but this only represents a small fraction of the billions of known protein sequences. Very often, structural coverage is bottlenecked by the months to years of effort required to determine a single protein structure.

The development of computational methods to predict three-dimensional protein structures from protein sequences (i.e. 'the protein folding problem') has advanced through two complementary paths that focus on either physical interactions or evolutionary history. The physical interaction approach heavily integrates our understanding of molecular driving forces into either thermodynamic or kinetic simulation of protein physics or statistical approximations to them. While theoretically very appealing, this approach has proven extremely challenging for even moderate-sized proteins due to the computational intractability of molecular simulation, the context-dependence of protein stability, and the difficulty of producing sufficiently accurate models of protein physics. The evolutionary program has recently provided an alternative, defining constraints on protein structure derived from bioinformatic analysis of protein evolutionary history, homology to solved structures, and pairwise evolutionary correlations. This bioinformatic approach has largely benefited from the sustained growth of experimental protein structures deposited in the Protein Data Bank (PDB), the explosion of genomic sequencing, and rapid development of deep learning techniques to interpret these correlations. Until now, contemporary physical and evolutionary history-based approaches have produced predictions far from experimental accuracy where a close homologue has not been solved experimentally, thus limiting their utility for many biological applications.

This scenario changed drastically in 2021. Simultaneous reports from DeepMind (Alphabet, Inc.) and David Baker (University of Washington) laboratories presented artificial intelligence (AI) algorithms that mined known protein structures to predict unknown structures with unprecedented accuracy: AlphaFold[56] and RoseTTAFold. Both computer codes are freely available free for researchers, together with highly accurate structure predictions for the whole human proteome (https://alphafold.ebi.ac.uk)[57]. This is a breakthrough that solved a fundamental scientific problem pursued for more than 50 years after the first protein structures were determined by x-ray crystallography, and the latter postulate that the structure of a protein was thermodynamically stable. Thanks to this major machine-learning-driven technological advance (arguably one of the greatest ever[58]) scientists can now obtain protein structures from these algorithms without having to crystallize their proteins or access cryo-EM tools.

But this is just the tip of the iceberg… Despite being trained for well-structured single protein chains, AlphaFold Multimer has been recently upgraded to predict both homomeric and heteromeric multi-chain protein complexes with astonishing accuracy[59]. Even more extreme, the Baker lab just now demonstrated that a deep neural network trained exclusively on native sequences and structures can generalize to create new proteins with sequences unrelated to those of native proteins that fold into stable structures[60]. The high similarity of the NMR and crystal structures to the so-called 'hallucinated' structure models demonstrate that deep learning process solves the classic de novo protein design problem, despite having no explicit knowledge of the physics of protein folding. In a similar manner, AlphaDesign enables rapid prediction of completely novel protein monomers starting from random sequences[61].

We are witnessing the AI Big Bang, which is revolutionizing each and every aspect of Science by quickly providing near-exact solutions to previously considered 'unsolvable' problems such as the accurate description of electron density in quantum mechanics[62]. In the years (months?) to come, we will see tremendous progress in the design of proteins featuring functional sites[63]: neutralizing antibodies, receptor traps for escape-resistant viral inhibition, metalloproteins and enzymes, and target binding proteins with designed interfaces[64].

The rise of deep learning and IA is both an exciting stimulus and an unescapable challenge for traditional computational chemists. We must adapt and embrace this technological tsunami… or perish to it.

Microbiota's Role in Innate Immune Memory

Samuel T Pasco, *Inflammation and Macrophages Plasticity Laboratory*

The host's gut microbiota consists of trillions of diverse microbial cells that line the gastrointestinal lumen. This microbial ecosystem contributes to the host's metabolism and provides a constant stimulus to the immune system. Healthy and balanced microbiota contributes to physiological homeostasis, and microbial dysbiosis is linked to many pathologies including inflammatory bowel disease and colorectal cancer. Innate immune cells mediate the gut microbiota-immune system interface by sampling the luminal environment, limiting microbial translocation, and maintaining gastrointestinal homeostasis. The microbiota's behavior, metabolism, and microbially-derived products influence host immunity and physiology and are of great interest to our lab.

Innate immune cells, such as monocytes and macrophages, demonstrate significant functional plasticity despite forming the rapidly responding, nonspecific arm of the human immune system. Though lacking the antigen-specific memory of its adaptive counterpart, previous inflammatory insults can imprint an "innate immune memory" (IIM) state on innate cells, characterized by a functional reprogramming that alters secondary responses upon restimulation[65]. While pro- and anti-inflammatory secondary responses distinguish "training" from "tolerance" programs, respectively, establishment of IIM depends on interdependent metabolic shifts and epigenetic modifications that remodel chromatin structure to imprint a transcriptional memory. The most studied IIM-inducing agents include the BCG vaccine against tuberculosis and the fungal cell wall component β-glucan, though many other stimuli can imprint these effects. Recently, our lab has characterized IIM effects induced by the extraintestinal pathogen Borrelia burgdorferi[66] and the gastrointestinal symbiont Lactiplantibacillus plantarum[67].

IIM effects can persist months after induction even though peripheral innate cells have short lifespans in circulation. While long-lasting BCG- and β-glucan-induced changes in hematopoietic stem and progenitor cells (HSPCs) of the bone marrow (BM) have been characterized, new research shows that IIM effects can be hereditary. Systemic fungal infection in male mice induced DNA methylation of immune gene loci in sperm cells, the offspring of which displayed transcriptional and epigenetic changes associated with the BM myeloid progenitors and improved protection against systemic bacterial infection[68]. Conversely, during pregnancy, infection-induced maternal cytokine release causes epigenetic modifications in fetal intestinal epithelial stem cells, and progeny from infected dams demonstrate protection during gut infection and augmented inflammation during colitis[69]. These results demonstrate that IIM effects can alter non-immune cells and imprint across generations.

Microbiota plays an important but incompletely understood role in immune responses, hematopoiesis, and IIM. Stacy et al have described a "metaorganism memory", or the reprogramming of the microbiota induced by a transient infection. In these experiments, host taurine production induced by gastrointestinal infection expands taurine-utilizing taxa, which inhibit pathogen respiration upon reinfection through taurine-induced sulfide production[70]. While this study did not directly examine the immune system's role in this enhanced colonization

resistance, we believe that aspects of this metaorganism memory are essential for the induction of distinct IIM phenotypes. Indeed, correlations between BCG-induced IIM and specific taxa have been reported[71], but functional studies are lacking. Other research has demonstrated that distinct microbially-produced metabolites directly impact the bone environment. Microbially-derived lactate promotes homeostatic hematopoiesis and erythropoiesis in the BM, while lactate administration induces proliferation of HSPCs similar to other IIM agents[72]. Taken together, microbiota-level metabolic shifts and microbially-derived metabolites have the potential to induce long term changes in the innate immune system.

Our lab uses multi-omic approaches to investigate mechanisms of IIM, and one new technique we aim to integrate into our research methods are metaproteomics. When utilized alongside metagenomic techniques, metaproteomic data can confirm species composition while providing information on metabolism and physiology of both microbial species and the host[73]. Our goal is to utilize this and other next generation technologies like single cell sequencing to elucidate the gut-innate immune-bone marrow axis and the interconnected mechanisms of IIM.

Synthetic Approaches for Streamlined Saponin Adjuvants and Self-Adjuvanting Vaccines

Alberto Fernández-Tejada, *Chemical Immunobiology Laboratory*

The clinical success of anticancer and antiviral vaccines often requires the use of an adjuvant, a substance that helps stimulate the body's immune response to the vaccine, making it work better. However, few adjuvants are sufficiently potent and non-toxic for clinical use; moreover, it is not really known how they work. Current vaccine approaches based on weak carbohydrate and glycopeptide antigens are not being particularly effective to induce the human immune system to mount an effective fight against cancer. Despite intensive research and several clinical trials, no such carbohydrate-based antitumor vaccine has yet been approved for public use. In this context, the research topic in the Chemical Immunology group has a double, ultimate goal based on applying chemistry to address the above clear gaps in the adjuvant-vaccine field.

Saponin adjuvants are plant-based natural products (standing out those derived from the bark of the South American tree Quillaja Saponaria (QS)) that have the ability to increase the immunogenicity of the coadministered antigen, boosting the resulting immune response. QS-21 is one of the most potent adjuvants and is approved as a component of the AS01 adjuvant system (in combination with monophosphoryl lipid A (MPLA) in a liposomal formulation) in shingles (herpes zoster) and malaria vaccines. Moreover, QS-21 has also shown promise as part of the M72/AS01 tuberculosis vaccine Phase III clinical trials. Additional formulations incorporating saponin adjuvants have been explored that have gained a great deal of interest in the context of HIV and SARS-CoV-2 viruses. Notably, the just approved Novavax COVID-19 vaccine includes the saponin-based Matrix-M nanoparticulate adjuvant[74], which not only boosts the immune response against the vaccine but also enables dose-sparing of the protein-based SARS-CoV-2 antigen. Recently, Silva et al. developed a new saponin/MPLA nanoparticle (SMNP) formulation through self-assembly in the presence of phospholipids analogously to immune-stimulatory complexes (ISCOMs)[75]. These SMNPs were superior to other related saponin-based formulations (e.g. AS01), inducing more potent humoral responses in mice and non-human primates vaccinated with a poorly immunogenic HIV antigen. This novel saponin-containing adjuvant was shown to act by enhancing lymphatic flow and antigen delivery to draining lymph nodes, providing new insights into the mechanisms of saponin immunopotentiation. In this framework, this year has brought the publication of a key review article from our group on carbohydrate-based adjuvants and their mechanisms of action[76], where we discuss current knowledge and mechanistic understanding into the proposed mode of action of saponin adjuvants, among others. Moreover, we have also published a structure–activity relationship study on saponin variants with modifications in the linear oligosaccharide domain of QS-21[77] identifying the native rhamnose-xylose motif as necessary for adjuvant activity while also correlating the attenuated antibody responses of these new analogues with their increased conformational flexibility.

Concerning recent developments in synthetic adjuvant–antigen conjugates as potential self-adjuvanting vaccine constructs, a liposomal GM3 tumor carbohydrate antigen vaccine adjuvanted by covalent conjugation of the immunostimulatory glycolipid α-GalCer has been developed[78]. A new self-adjuvanting approach has been recently published consisting of a Mincle agonist (vizantin) as a built-in adjuvant conjugated to the sialyl-Tn glycan cancer antigen[79] generating a robust antibody and cellular immune response that was protective against tumors in mice. From our lab, we have developed the first glycoconjugate construct based on saponin adjuvants, which were covalently linked to the Tn(Thr) glycoamino acid as a prototype cancer-associated antigen[80]. While preliminary immunological evaluation showed modest elicitation of antibodies in mice, structural optimization of this streamlined saponin platform is ongoing for the construction of improved self-adjuvanting anti-cancer vaccines.

Finally, this year we have published a comprehensive review article on the glycosylation of proteins with β-O-N-acetylglucosamine (β-O-GlcNAc) from a chemical viewpoint[81], with emphasis on the structural and mechanistic role of this key post-translational modification implicated in a variety of cellular processes and human diseases.

Looking forward, significant achievements can be anticipated for 2022 from my perspective, including the development of novel, improved saponin adjuvants and chemical probes for molecular level mechanistic investigations. It is expected that critical new insights into the molecular mechanisms of QS-21 and saponin-derived adjuvants can be obtained, albeit it is unlikely that a universal mechanism of action be relevant to the range of saponin adjuvants and new synthetic variants. Finally, further development of the saponin-based self-adjuvanting approach is a major ongoing effort in my laboratory, which will deliver additional synthetic constructs incorporating a range of carbohydrate/peptide-derived antigens for improved vaccine designs.

Ubiquitin-like Modifications in Health and Disease

Rosa Barrio, *Ubiquitin-likes and Development Laboratory*

We are interested on the regulation of developmental processes and diseases by post-translational modifications by the Ubiquitin-like (UbL) family. UbLs, like SUMO, are attached to target proteins altering their function, thus regulating cellular processes like proliferation and transcriptional regulation. Study of UbLs modification is challenging due to the small amounts of a given modified protein and the transient nature of the modification. We developed new proximity proteomics biotin-based strategy, SUMO-ID, to identify interactors of proteins when modified, applicable to any UbL. We are developing strategies to identify UbL modified targets in a subcellular manner.

Importantly, targeted protein degradation (TPD) is getting more and more relevance in biomedicine. We focus on the Spalt-like (SALL) family of transcription factors, necessary for numerous biological processes. Mutations in SALL1 are associated to Townes-Brocks Syndrome (TBS), a rare disease causing kidney defects, deafness and polydactyly. Mutant SALL1 interferes with the function of cilia, which opens new opportunities of intervention.

In 2021 New TBS cases were reported[82]. Special warning was done on the detection of renal anomalies in adults with SALL1 mutations. SALL1 has been involved in stem cell regulation, cancer and SALL1 expression regulation by circular RNA circ_0043265 and miR-1243 was documented[83, 84, 85, 86]. In the case of SALL4, it has been found involved in several cancer types[87,88,89] and self-renewal of stem cells[90]. Importantly, SALL4 reads A/T content of DNA to inhibit expression of genes and stabilizing the pluripotent state[91]. SALL4 remodelates chromatin through the regulation of CECR2, which interacts with the SWI/SNF complex[92]. SUMO has been involved in DNA repair and replication[93,94], regulation of signaling[95], meiosis[96] and pluripotency of stem cells[97]. SUMO1-SIM interaction disruption reduces alpha-synuclein neurotoxicity in model organism[98]. In the field of posttranslational modifications in ciliopathies, new evidences arise on the regulation of ciliogenesis by the ubiquitin system[99,100]. As expected, an explosion of new publications occurred in the field of TPD, some of them concerning the degradation of SALL members[101]. New molecules and strategies to use different E3 ligases have been developed[102,103]. New strategies in proximity proteomics biotin-based strategies have been published to improve efficiency and specificity[104,105,106]. Proximity proteomics has been used in the discovery of interactors of the ubiquitin family machinery, including E3 ligases of ubiquitin or other UbLs or demodifying enzymes[107,108].

The field of TPD is expected to keep growing in the next year, given its capacity to degrade targets previously thought to be undruggable. Furthermore, new strategies for the identification of specific targets of E3 ligases of UbLs is very much needed. Those strategies based on biotin proximity proteomics are very promising given their high specificity and sensitivity.

Transmissible Spongiform Encephalopathies Translated to other Neurodegenerative Diseases Caused by Proteins Showing Prion-like Features

Joaquín Castilla, *Prion Research Laboratory*

Prions are the causal agents of a group of rare and invariably fatal neurodegenerative disorders known as Transmissible Spongiform Encephalopathies (TSE), affecting several mammals including humans. The main event underlying disease development is the misfolding of the cellular prion protein (PrPC) into an alternative pathogenic conformation called PrPSc. This aberrantly folded form or prion presents the capacity to induce its conformation on the cellular counterpart, forming fibrillary aggregates that accumulate in the brain and causes neuronal death in the process. A similar behavior has been described in other proteins associated to highly prevalent neurodegenerative diseases such as Parkinson's disease or Alzheimer's disease. This new point of view is changing the way in which protein-misfolding related neurodegenerative diseases are tackled, with knowledge and methodologies derived from TSE research being increasingly applied to these other diseases.

The most relevant breakthrough in the field of TSE happened during 2021 is undoubtedly the resolution at an atomic level of the three dimensional structure of two different prion strains purified from brains of affected animals[109]. However, this kind of work proceeds, with the study of other prion strains as needed to solve the prion code[110,111] To know in detail the structure of this protein is the first step to address the largely unknown molecular mechanisms underlying prion pathogenesis, such as the existence of different strains causing distinct disease subtypes, the role of cofactors on this, or the elements governing interspecies transmission, which would also permit rational design of new diagnostic and therapeutic approaches. With the first detailed prion strain structures in hand, unveiling the molecular basis of prion strain diversity and how it is influenced by environmental and host factors is within reach[112,113]. Several investigations on the role of sequence elements, environmental factors, and cofactors on strain diversification have contributed to this[114,115], as well as studies in vitro and using synthetic prions[116,117]. Similarly, existence of strains has been confirmed in other neurodegenerative disorders and their corresponding atomic structures are also being solved[118,119] as well as their relation with the same cofactors as prions[120,121]. Interestingly, reported interactions between proteins causing other neurodegenerative disease and PrP are also increasing[122,123], opening new possibilities to determine neurotoxicity mechanisms, which could be mediated by PrP. Thus, apart from translating research and diagnostic methods, which is already occurring[124], we could soon expect translation of therapeutic approaches. In terms of therapies in development, new potential anti-prion molecules have been described[125,126] and new therapeutic targets have been discovered[127,128]. However, the most promising advances are related to co-expression of misfolding-resistance variants[129] and lowering PrPC amounts, a strategy that has been already translated to Parkinson's disease[130]. It is also noteworthy the importance of the epidemiological surveillance of TSE, as proven by the various reports[131], including that of Spain[132], published this year. These reports could unravel epidemiological outbreaks of prion infections, responding for example to iatrogenic transmission[133,134], what may be also relevant in the future for other diseases as their transmissibility is being reported[135,136]. Along this line, report of cases with unusual features is also of relevance in TSE since they could help understand disease mechanisms, detect new hereditary forms[137,138] or co-occurrence of the apparently intimately related neurodegenerative disorders[139].

Thus, in upcoming years, more interrelations and interactions between prions and other proteins associated to neurodegenerative disorders are expected. These will contribute to further blurring the line between prions other amyloidogenic proteins and to the implementation of more techniques originally developed for TSE, for other diseases. In addition, we expect that the strain phenomenon will start to be understood as high-resolution structures of different strains are solved. Being related to interspecies transmission of prions, unveiling strain feature determinants could also shed light on the factors governing it, making easier the rational design of new therapies, likely based on prion propagation interference through introduction of misfolding-resistant prion proteins.

Cryo-EM and cellular tomography of viral pathogenesis

Nicola G. A. Abrescia, *Structure and Cell Biology of Viruses Laboratory*

Never as this past two years have manifested the importance of the research in virology (and immunology). Readiness of action when such pandemic events occur rely on years of basic knowledge accumulated in time and constant technology development. The study of viruses infecting either Bacteria, Archaea or Eukaryotes – and whether from the basic,

clinical or translation point of views – is always beneficial to society soon or later. The Abrescia Lab focuses on determining the molecular mechanisms governing viral pathogenesis using structural methods. We study virus assembly, virus cell entry, and virus-antibody recognition mechanisms. Deciphering the principles governing virus assembly allows developing molecular strategies for interfering with its morphogenesis. Expanding our structural knowledge of the virus world (Virosphere) also contributes to establish viral relationships that are undetected at sequence level. Elucidating the virus cell-entry processes provide the possibility of jamming this key-lock mechanism. Antibodies are often used to this purpose and understanding the recognition and neutralization processes greatly increase the possibility of tuning such mechanisms. To this end, we use an integrative approach based on X-ray crystallography and Electron Microscopy (EM) techniques and complemented with functional studies. The generated knowledge is intended not only to elucidate molecular mechanisms but also to translate this knowledge into molecular tools (eg. drugs, diagnostics, vaccines) that would improve human and animal health. Our research fulfils the ONE HEALTH concept and approach to societal challenges enabling the development of new intervention strategies.

The field of Structural Virology has seen pivotal advances on several fronts in tackling human and animal health threats. Studies on Coronaviruses and specifically on SARS-CoV-2 have largely occupied the scientific output in the field accruing even more the large list accumulated during 2020. To this end cryo-EM has been pivotal for the fast development of therapeutics[140],[141],[142],[143],[144]. It is worth mentioning that our group is currently involved in structurally studying virus members of animal and human threatening Bunyavirus, Flavivirus and Coronavirus families. In the case of Flaviviruses we are employing cellular structural techniques at the forefront of cryo-EM. Viruses are also miniaturized machines that perform complex biological tasks and understanding how members of the Virosphere are related beyond the similarity in their primary sequence remains a powerful tool to exploit/discover new virus applications. Further, discovering new viruses or engineering new ones for biotech applications as in the case of adeno-associated virus (AAV) can be considered an outstanding scientific advance.

One of the technical advances that has occurred in the Basque region in 2021 has been the installation of a Titan Krios G4 electron microscope at the Biofisika Institute constituting the Basque Resource for Electron Microscopy (BREM). This has been in part possible thanks to the pioneering efforts made by CIC bioGUNE and its scientists to create a fervent community dedicated to biological electron microscopy[145] as also witnessed by our earliest articles advocating the cryo-democratization of cryo-EM[146],[147].

Advances in our field of methodological expertise expected in 2021 will mainly concern:

- HR cryo-EM: it will continue to develop. Smaller molecules (~50 kDa) and faster data collection with K3 and Falcon IV direct detection cameras, Selectris X imaging filters is rendering EM, the technique of choice to determine not only the 3D structure of a wide size range of macromolecules (and attached glycans) but also the conformational dynamic of such proteins. Pharma companies are heavily involved in drug screening programmes using cryo-EM and cryo-EM will be used in the clinical area.
- We – as Abrescia Lab and CIC bioGUNE - are pivotal in the incorporation of biological cryo-FIB milling procedures as a joint effort with the CIC nanoGUNE institute to develop a pipeline for cellular structural cryo-electron tomography in the Basque Country. Further, the combination of soft X-ray tomography and high-resolution correlative microscopy allows a direct correlation of cellular morphological changes with virus entry pathways and consequent antiviral treatment.
- The above structural techniques contribute to provide a 'Google Map' of distinct (pathological) human cells and thus helping to navigate with predictive capabilities in the landscape of human diseases.

Endosomal Trafficking. The Retromer Complex

Aitor Hierro, *Membrane Trafficking Laboratory*

The lab is devoted to the study of the molecular and structural mechanisms that govern intracellular transport via coated membrane vesicles. In particular how self-assembling cage scaffolds can generate unique geometries on and around membranes to sort cargo proteins into discrete trafficking pathways. The group combines X-ray crystallography, cryo-electron microscopy (cryo-EM) and biochemical reconstitution to characterize the mechanisms by which coat proteins, adaptor proteins and other regulatory molecules determine the itinerary of a cargo protein within the cell.

The number of scientific papers published in the field of retromer during 2021, according to PubMed, is 74. Work published by Leneva[148] presents the structural architecture of fungal and metazoan retromer assembled with sorting nexin SNX3 and a model cargo using cryo-electron tomography. The structures show that the retromer core retains its arched shape, however the peripheral association/organisation with the membrane is different than when bound to SNX-BAR proteins. In another study, Zhang[149] reveals how SNX1 is organized on the membrane and provides a detailed molecular understanding of how SNX1 achieves membrane deformation through BAR and PX domains, and a linker region that connects these two domains. Daniloski Z[150] reports the identification of several protein complexes including retromer in SARS-CoV-2 infection. Interestingly, the authors found that Rab7a, which controls retromer recruitment to endosomes and its function in cargo sorting, is involved in cell surface expression of ACE2 likely by sorting ACE2 in endosomal vesicles. Chen KE[151] reports a novel series of macrocyclic peptides that bind retromer with high affinity and specificity structurally mimicking known interactors. Of note, the authors identify one peptide (RT-L4) that binds retromer at the Vps35-Vps26 interface without simnifically affecting known interactors, thus of potential use as molecular chaperone for therapeutic intervention and/or new tool for the study of retromer.

The process of tubular endosomal budding and trafficking is responsible for the subcellular localization of hundreds of cargo proteins such as signaling receptors, nutrient transporters, ion channels and adhesion molecules. Not surprisingly, numerous mutations affecting cargo recognition and recycling have been associated to a variety of diseases, most of which affect the nervous system. In the future, detailed description of how disease mutations compromise tubule-based endosomal sorting is certainly required. This knowledge is of pivotal importance for future therapeutic intervention as it might provide new targets for the development of scaffolding drugs that either promote or inhibit protein-protein interactions to interfere with specific trafficking pathway(s) without affecting others.

TECHNOLOGIES

Single Cell Proteomics: a Dream Come True?

Felix Elortza, *Proteomics Platform*

We can define proteomics as de discipline which goal is to catalogue and characterize the total collection of protein isoforms from a cell, tissue, organ or organism. The so called 'proteoforms' are encoded by the same gene but due to different molecular mechanism (e.g. alternative splicing) they end having non-identical amino-acid sequences. Moreover, the same sequenced protein shall receive a plethora of post-translational modifications that will define their fate. However, at the single-cell level, that's an incredibly challenging business. For instance, in genomics each type of nucleic acid behaves chemically in a predictable way. But the proteome has a vast assortment of different chemistries, interactions, dynamics, localizations and abundances. And with no equivalent technique over proteins to PCR amplification of DNA, any technique to detect proteins must be sensitive enough to identify and characterize them, regardless the material a cell contains. Thus, single cell proteomics was just a dream few years ago.

Thinking about how to face protein detection at single cell level, antibodies have been and still are very useful tools. Antibody based protein detection methods can be very sensitive and their use is relatively straight forward. Nevertheless, there can be drawbacks since some antibodies show low affinity and may lack specificity. Besides, it is a targeted method where you can only find what you are looking for. Lately sample handling miniaturization and mass spectrometer developments have allowed to scientific community to get unprecedented achievements in single cell proteomic analysis. Among minimal sample handling innovations proteoCHIP is one design (Cellenion). Another approach is the nanoPOTS (nanodroplet processing in one pot for trace samples). This system is like a nanolitre-scale microtiter plate fabricated onto the surface of a microscope slide. Each 'well' is a hydrophobic circle about one millimetre in diameter, with a small hydrophilic 'pedestal' at the centre at which cells are deposited and prepared. The reaction volumes are performed in less than 200 nanolitres, so the entire sample preparation process is optimized for minimal protein loss[152].

On the analytical side, there are two main mass spectrometry based approaches. One was first described by Slavov´s group and has been termed SCoPE-MS (single-cell proteomics by mass spectrometry). A major feature of the

methodology is a mass spectrometry's version of barcoding based on already commercially available isobaric tags. These are molecules with identical masses that fragment into differently sized ions inside a mass spectrometer[153]. The other has been implemented by Mann´s group where they combined sample preparation miniaturization, a novel mass spectrometer now called tims TOF SCP (SPC: single cell proteomics; Bruker) and sophisticated software they managed to perform single cell label free proteomic analysis. So far both approaches describe around 1000 proteins identified and quantified (preliminary data). If the field evolves the way it has done last years, this number surely will be soon expanded.

At Proteomics Platform at CIC bioGUNE we have been working to improve our sample preparation methods to minimize the starting material together with optimizing specific mass spectrometry acquisition methods. Preliminary results show that we can tackle single human oocyte proteomic analyses, which means that we have achieved sub microgram levels of starting material. We are sure that sooner than later these methods for single cell proteomic analyses will settle and democratize. This for sure will help to make some not so time long ago impossible studies feasible, speeding up the understanding of yet hidden molecular mechanisms affecting diseases such as cancer.

Beyond Next Generation Sequencing

Ana M Aransay, *Genome Analysis Platform*

Biomedical research is based to a large extent on genomic and bioinformatic strategies, which are in constant and speedy development. Therefore, every year, the Genome Analysis Platform at CIC bioGUNE keeps a close eye on all protocols under development and their applications. Following these manners, during 2021, we have studied the transcriptomic composition and levels in pools of single-cells, after setting-up 10x Chromium System's sc-mRNAseq protocols from 10x Genomics and running them in several live systems originated from human, mouse, chicken and gecko species. Preliminary analysis of these data has been carried out, from which some very interesting findings are emerging, and now, we are developing advance bioinformatic tools to correctly interpret those results.

On the other hand, spatial-transcriptomics strategies uncover how the complex and rare cell populations interact in specific tissues or organs. However, it has been proved to be a very challenging technique regarding sample preparation, because each specific tissue requires very tightly controlled conditions before transcriptomes are captured from the studied sample. In collaboration with Carlos Larrea's team in the Central Hospital of Asturias (Spain), we keep on testing different approaches to improve and make efficient this procedure, which is clue to properly characterize the gene expression through 2D-tissues. In addition to mentioned set-ups and standard external services, the platform has carried out several fruitful collaborations that have yielded noteworthy scientific reports in genomic and functional advances in organ transplant, immune response related to cardiovascular disorders and other organ failures, as well as in rare and digestive disorders[154],[155].

Moreover, during the last months, we have been working hard to understand new-fangled next generation sequencing (NGS) techniques, emphasizing the ones related to the characterization of nucleotide modifications, both at DNA and RNA level. In our opinion, the company Oxford Nanopore Technologies offers the protocols that best align our goals, as their strategy for sequencing nucleic acids (DNA and RNA) is based on the electric current that flows through the array or nanopores embedded in their devices. The current values are then decoded to basecalling, determining the DNA or RNA sequence in real time, as well as detecting any sort of modification, if any, that each nucleotide has undergone. With this technology, we will start characterizing the methylation and acetylation status of both genomic and mitochondrial DNA, the latter being very important to understand the redox capacity of targeted systems.

Metabolomics based Toxicity

Juan Manuel Falcón and Sebastiaan van Liempd, *Metabolomic Platform*

In the ten-year anniversary of the Metabolomics Platform in CIC bioGUNE, we supported numerous research projects in their metabolite analysis. To do so, we have used high resolution liquid chromatography coupled with high resolution mass spectrometry (hrLCMS) to target the metabolites of interest. Our core assays are focused on metabolic members of the methionine cycle, polyamide synthesis, transsulfuration pathway and the tricarboxylic acid (TCA, Krebs) cycle. Although we have developed several assays on request and in close collaboration with the client.

Recently we have added stable-labeled flux analysis methods to our portfolio. The essence of these methods is the tracking of stable labelled atoms from a precursor metabolite throughout a metabolic pathway. In our platform we use 13C5-methionine and 13CD3-methionine to probe the methionine cycle while we determine fluxes trough the TCA cycle with 13C6-glucose and 13C5-glutamine. However, if a labeled precursor is available, we can track most pathways with our hrLCMS setup. The great advantage of these stable labelled experiments is the absence of background signals of endogenous metabolites. And when the system is measured over an adequate time course, the dynamics of the pathway will be clearly revealed. An example of how these methods can be used is described in a recent paper in which we show how pharmacological intervention normalizes glucose homeostasis in a non-alcoholic steatohepatitis mouse model[156].

Besides our commissioned work, the Metabolomics Platform has its own lines of research. These projects are mostly based on so-called untargeted metabolomics. The idea is that by measuring a sufficiently big part of a metabolome, observed metabolic differences upon a challenge can reveal underlying mechanisms. For example, we used this to determine the mechanism of how hepatic extracellular vesicles change the blood metabolome and so influence endothelial tissue[157]. In another untargeted study we showed via changes in the metabolome how a defect in a protein involved in vesicular transport likely affects neurotransmission in the hippocampus[158]. And a recent publication from our team describes how an untargeted metabolomics study revealed why an inbred mouse strain was highly vulnerable to influenza A viral infection[159].

Looking at the future, we are aiming to set-up assays focused on drug metabolism. Especially cytochrome P450 inhibition, reactive intermediate trapping and liver toxicity have our interest. These assays are part of the pre-clinical drug discovery pipeline and thus of great importance to the pharmaceutical industry. With these innovations we are in line with the future direction of CIC bioGUNE.

Landmarks in structural biology

Isaac Santos and **Adriana Rojas**, *Electron Microscopy Platform*

The cryo-EM revolution is today an undeniable fact in structural biology. The number of structures solved using this technique has kept growing. Remarkably in 2021 the quality in the deposited structures has had a jump, being the structures

between 1-4 Angstroms the majority for the first time[160]. Technical advances in cryo-EM are continuing at a tremendous rate; the combination of stable and brighter electron sources, improved electron optics, more stable cryo stages, ultra-clean vacuum and fast highly sensitive direct electron counting cameras, make present-day cryo-data collection particularly efficiently. Moreover, the strengthening of the informatics resources (clusters, GPUs, etc.) and the improvement of the algorithms used for data processing are helping tremendously to push the processing of the data. Another aspect that is becoming more evident is the ability to classify large heterogeneous datasets, sorting out many different conformation classes that make possible to observe biological process in action. Single-particle analysis (SPA) made already possible to classify this structural heterogeneity based on computational averaging of thousands of images of identical particles. However, a good biochemical purification of the sample is still required. Moreover, it is also possible to reconstruct subcellular structures reaching sub-nanometre resolution using the cryo-electron tomography (Cryo-ET) technique that functions by collecting a series of projection images through an object from different angles[161]. Many biological structures have been recently accomplished using subtomogram averaging. A protocol consists of aligning and averaging subvolumes within a tomogram containing an object of interest in an iterative mode[162]. Compared to the SPA, one of the handicaps of this technique is the low number of tomograms collected per session, which has changed during this year with the implementation of a protocol able to save time during tilt series collection; the fast-incremental single-exposure (FISE) method. In this process, the camera acquires in a single long exposure, blanking the beam when the stage is tilting and unblanking when the stage is stationary, eliminating the camera's downtimes every time it starts. Specimen shifts throughout tilting can also be calibrated and compensated by applying tilt angle dependent image shifts and defocus[163]. On the other hand, the new Cryo-ET collection approach named 'beam image-shift electron cryo-tomography (BISECT) substantially accelerates data collection speed by using beam-image shift to multiply the number of areas imaged at each stage position. This is possible by applying a geometrical high precision targeting and performing per-tilt astigmatic CTF estimation[164]. Combining these protocols allows obtaining a high-resolution tomogram in less than 5 min. But the strength of Cryo-TEM is not only limited to the technique per se, as it presents powerful combinations with other structural imaging tools that make possible the observation of the biological specimens in their natural state[165]. One of these is Liquid cell TEM (LC-EM), which allows the observation of processes in a liquid state that cannot be imaged with conventional TEM or other techniques[166]. The combination of Cryo-ET with a technique called cryo-focused-ion-beam (cryo-FIB) milling can cut samples such as cells into extremely thin slices known as lamellae, which at the same time can be melded with fluorescent images by labeling specific proteins[167]. Nevertheless, a tomogram is less than 0.1 % of a cell, which means that many organelles can be only viewed in tiny segments. To fill this gap, some researchers are developing machine learning to elucidate how various cellular components interact[168]. Other researchers have recently solved this problem by combining cryo-ET with another technique called soft-X-ray tomography, which can capture images of whole cells, deciphering biological questions such as the mechanics of SARS-CoV-2 infection[169].

Yet, the most immediate milestone achieved by Cryo-EM is probably its use for the pharmaceutical industry. New computational implements combined with the high-resolution Cryo-EM data provide critical data for understanding ligand-protein interactions. Thus, cryo-EM has turned into an exceptional Structure-based drug discovery (SBDD) tool[170]. The speed of response of Cryo-EM in medicine design has been evidenced in the development of vaccines against SARS-CoV-2. In less than two years, hundreds of structures have been characterized regarding the biological functions of the glycosylation in SARS-CoV-2 proteins as well as the human receptor ACE2 in different variants of the virus[171].

Another promising application in the drug discovery is the use of microcrystal electron diffraction (MicroED) for structure determination of natural products and proteins in complex with small molecules. That has been increased due to the improvement of the speed of data collection and structure determination[172]. As an example, for some small organic molecules, it has been possible to determine the structure within less than one hour[173]. One of the weakness of the MicroED has been the sample preparation. Still, a method in which excess liquid is removed through the EM grid with the assistance of pressure, called pressure-assisted backside blotting (Pre-assis) has been used to increase blotting efficiency for viscous samples[174]. Even in presence of 44% polyethylene glycol (PEG) 400, grids showing a good distribution of hydrated microcrystals in a thin layer of vitreous ice could be obtained. On the other hand, Martynowycz and coworkers[175] used the protein crystalline lamellae and microED to assess the effects of ice thickness the quality of data. They demonstrated that high-quality structures could be obtained from samples up to 2× mean free path (MFP) regardless of the acceleration voltage. Corresponding to physical sample thicknesses of ~430, 540, and 640 nm for the accelerating voltages of 120, 200, and 300 kV, respectively. These results provide a benchmark of the ideal specimen thickness that could be extrapolated to other cryo-EM methods.

Three-dimensional Cryo-Electron Microscopy

Mikel Valle, *Cryo-EM of biologicals macromolecules Laboratory*

This September a new high-resolution electron microscope has arrived at the Biophysics Unit. It is a Titan G4 equipped with a gatan K3 direct detection camera. This is the best performing equipment in the cryoEM field and will be available for users after the initial installation, adjustments and testing (probably in early 2022). It will improve our access to this type of microscopes and our scientific performance.

A recent development on image processing tools for cryoEM has been presented on new Relion 4.0 software. The launch of this version has suffered a long delay due to the sars-cov2 pandemic. It includes the implementation of new algorithms that significantly improve the classification of images. It also delivers a completely new approach of the sub-tomogram averaging technique. This Relion 4.0 has been released in a beta version that will require some collaborative work to remove bugs and improve it. We have a strong line of research on the structure of flexible filamentous plant viruses and their VLPs. Recently a new work of VLPs from alternanthera mosaic virus (AltMV) has been published[176]. These VLPs are produced in plants, but the viral CP recruit ssRNAs that are available on the cell. It is a new type of VLPs described that keep the native helical symmetry of the virions.

Nuclear Magnetic Resonance Methodological Advances

Tammo Diercks, *NMR Platform*

NMR relaxation measurements track molecular dynamics and exchange processes on wide timescales divided by the rotational correlation time tau_R for overall molecular tumbling:

Fast local motions can be resolved if their correlation time tauC is much shorter than tauR. Nanoparticle assisted NMR spin relaxation[177] extends the traceable tauC limit by extending the

protein's tauR via transient contacts with large nanoparticles. This was demonstrated for 15N[178] and methyl 13C[179] relaxation data. The bias from remote dipole-dipole interactions on 13C relaxation[180] and on deriving tauR[181] was analysed. Optimised experiments and sampling schemes to measure all differential relaxation terms in isolated 13CH3 groups were presented[182] and compared with 13CHD2 relaxation[183]. As molecular dynamics (MD) trajectories can cover local motions with tauC ≤ 10 ns, comparing MD and NMR derived order parameters provides unique atomistic insight into the thermodynamics, kinetics, and nature of molecular motions and interactions. New SPINACH software predicts NMR relaxation from MD trajectories[184]. Fuzzy interactions between H4 histone tails and DNA were revealed by MD and 15N relaxation data[185] NMR derived sidechain order parameters were shown to be less reliable for methyl groups while solvent exposed sidechain amide groups form only transient H-bonds[186].

Slow local motions with exchange (correlation) times ex larger than tauR derive from conformational or chemical exchange processes and are monitored by R2 or R1ρ relaxation dispersion experiments. New applications include the use of dynamic nuclear polarization (DNP) to monitor ligand binding via 13C R2 dispersion at natural abundance[187] or self-assembly of amyloids via temperature dependent 15N R2 dispersion[188]. A new experiment measures aromatic 1H R1ρ dispersion by site selective 13C-1H labelling in deuterated aromatics[189]. Advances in data analysis include algebraic expressions for R2 dispersion in n-site chemical exchange[190] and artificial intelligence (AI) based selection of the best fitting model for up to 5-site chemical exchange[191]. The new RING NMR dynamics software analyses both fast and slowmotion sensing relaxation data[192]. Aromatic ring flip frequencies were derived from 1H signal line shape analysis combined with MD[193].

Molecular interaction studies by NMR deliver rich, atomically resolved information and can be monitored via diverse NMR observables of distinct sensitivity. NMR derived diffusion and fast HN/H2O exchange rates revealed autonomous pre-association of ribosomal assembly factors (Schedlbauer et al. 2021, doi: 10.1126/sciadv.abf7547) and ligand-induced long-range effects in galectin-1[194]. The oligomeric assembly of a mitochondrial protease chaperon was probed by methyl TROSY chemical shift perturbation (CSP)[195], structural and dynamic effects of Na+ vs. K+ binding of a cation channel by 15N relaxation and CSP[196], quantified millisecond exchange in Huntingtin-SH3 interaction by relaxation dispersion and CSP[197], and the stereochemistry and energetics of aromatic (CH/π)-π interactions by NOESY/ROESY data[198]. Residue specific pKa values were measured for tailored protein mutants to probe interactions between charged residues[199].

New NMR methodology includes further developments of looped projected spectroscopy (L-PROSY) to enhance the signals of H in fast exchange with H2O, e.g. an extension to 2D NOESY/TOCSY[200] and 3D HSQC-NOESY (HETMAT)[201,202] experiments. These revealed critical intermolecular OHglycan-Hprotein[203] and NHRNA-NHRNA NOE con¬tacts for sequential assignment in SARS-CoV-2 derived RNA[204]. Time saving concatenations of NMR experiments were presented, e.g. a NOAH combination of HSQC and HSQC-TOCSY[205] used also for NMR metabolomics[206], the NORD concept[207], and sensitivity enhanced seHCACO with simultaneous IPAP acquisition[208]. New, short broadband excitation pulses can mitigate effects of relaxation and homonuclear coupling[209]. Finally, we presented FOSY NMR to zoom onto selected protein signals with concatenated sensitivity optimised, spin state selective hetero-nuclear transfer schemes[210]. Sparse NMR data sampling (NUS) for massive time saving requires optimised sampling schemes and algorithms for reliable, artefact free data reconstruction. The impact of clustered sparsity on Poisson-gap sampling[211] and optimally interleaved sampling of 2D relaxation data into a pseudo-3D spectrum[212] were now analysed.

Homonuclear decoupling increases spectral resolution and intensity, but remains challenging especially in the direct 1H dimension. Here, PSYCHE and perfect BASH homodecoupling methods were compared by a sensitivity enhanced EASY-ROESY experiment[213]. For virtual homodecoupling in an indirect 13C dimension by data processing, new AI algorithms (FID-Net)[214] and Compressed Sensing reconstruction with deconvolution were developed.

Heteronuclear decoupling is easy to implement due to clearly distinct frequency ranges. Still, interference in simultaneous multi-band decoupling may cause accidental perturbations that can be avoided by de-synchronising and interleaving[215]. Software for NMR spectral analysis increasingly implements AI, e.g. to enhance signal deconvolution and peak picking in heavily overlapped 2D spectra (DEEP picker)[216]. The new SPARKY plugin iPick[217] detect and validate peaks while CHESPA/CHESCA-SPARKY[218] maps protein allostery. MAUS[219] assigns methyl signals from NOESY spectra and known protein structure. Conformationally averaged 3JHH' coupling constants can now be better used as restraints for MD to refine protein structures[220]. Chemical shifts bear wide information that continues to be explored. Thus, CSI-LSTM uses AI for refined secondary structure prediction[221] while CheSPI infers small populations of residual secondary structure in IDPs[222]. Sidechain 13C shifts were used to infer both conformational order in β-sheets of amyloids vs barrel-like membrane proteins[223] and global protein dynamics (from methionine methyl 13C shifts)[224]. Diffusion ordered (DOSY) NMR on large objects like amyloid fibrils is biased by tensorial coupling between translational and rotational diffusion, which may now be deconvoluted[225]. A relaxation optimised DOSY experiment with concomitant long-range H,X INEPT transfer[226] and ultrafast DOSY for compound separation[227] were presented. Slice selective excitation (SSE) underlies both DOSY and ultrafast single-scan (US) NMR methods. The latter was now combined with DNP for further acceleration[228]. SSE can also enhance frequency selective excitation, as shown by the GEMSTONE-NOESY[229] and GEMSTONE-TOCSY[230] experiments. Saturation Transfer Difference (STD) NMR is a powerful technique to map the binding epitope on weakly binding ligands. We used residue specific 13C labelling to break the degeneracy in repeating LacNAc oligomers and applied 2D STD-13C-HSQC to evince the epitope recognised by various galectins[231]. STD was also used to analyse the ligand discrimination by cellodextrin phosphorylase[232] and for chemosensing of water contaminants by hybrid nanoreceptors.

Heteronuclear NMR may offer a molecular specificity, spectral dispersion and simplicity superior to 1H. We have exploited these benefits by quantitative 31P NMR to obtain a comprehensive overview of the phosphorome from mouse liver extracts[233]. Also, 31P NMR is often used in hydrolysis studies on NTPs[234]. Powerful 19F NMR continues to find widespread applications and was recently used in fragment screening to prove the drugability of RNA[235]. Fluorine tags were introduced, e.g., on metal-binding pharmacophores to study their binding by carbonic anhydrase[236], on lanthanide binding tags for proteins to assess their conformational landscape[237], and as conjugated fluorocarbon chains on a GPCR binding peptide to analyse its micelle formation[238]. Site specific fluorine labelling of Trp[239] and Phe in a cold shock protein allowed to quantify its affinity to DNA or other proteins by 19F NMR[240]. Fluorine labelled glycans[241] and LacNAc derivatives[242] were synthesised to probe their lectin binding. 1JF,C scalar coupling constants were shown to increase with the strength of nearby intermolecular halogen bonds[243] while 19F chemical shifts correlate with the fluorine's capacity to act as H-bond acceptor[244]. Finally, 77Se NMR enabled the specific observation of selenoglycan recognition[245] and selenocysteine selenic acid formation in the catalytic cycle of glutathione peroxidase[246]. NMR sample preparation comprises several critical steps. New isotopic labelling methods for proteins base on the use of 2H/13C labelled acetolactate for linearised Leu/Val sidechains[247], synthesis of only locally deuterated 13CH3 Leu[248],

or on reverse (I,L,F,W,Y,K) labelling for sparse 13C labelling[249]. An improved protocol for 13CH3 labelling in deuterated proteins[250] and the synthesis of 13C,2H,19F labelled Trp[251] were reported. Specific protein tags were introduced by genetic encoding of trimethylsilylmethoxycarbonyl-Lys[252], by attaching tButyl to Cys, by cell-free synthesis of selenoproteins as reactive tag acceptors[253], or as a new chiral lanthanide tag for Cys[254]. The efficient production of functional GPCR protein in E.coli[255] and suitability of 4-F-DMPC nanodiscs for high-resolution NMR[256] may facilitate NMR studies on membrane proteins.

Pharmaceutical NMR advances include a comprehensive assessment of protein stability in pharmaceutical formulations by 1H NMR[257], a 13C-HSQC based method to identify isoAsp and C-terminal Asp formation from protein degradation[258], site-directed methyl labelling to test GPCR functionality by 13CH3 TROSY[259], and fragment-based screening by 15N-HSQC to identify an extension of the canonical binding site on RH domain[260].

Front Page Image:
https://www.freepik.es/vectores/animal">Vector de Animal creado por brgfx - www.freepik.es

Editing:
Donatello Castellana

Middle Page Image:
https://www.freepik.es/vectores/tecnologia'>Vector de Tecnología creado por starline - www.freepik.es

REFERENCES

1 PMID: 33419761
2 PMID: 34312556
3 DOI: 10.1200/JCO.2021.39.15_suppl.2534
4 doi:10.1038/s43018-021-00291-9
5 doi:10.1038/s41556-020-00613-6
6 doi:10.1038/s43018-021-00272-y
7 doi:10.15252/embr.202051981
8 doi:10.1038/s43018-021-00221-9
9 doi:10.1038/s41588-021-00887-y
10 doi:10.1038/s41467-021-25739-5
11 doi:10.1038/s41586-021-04075-0
12 doi:10.3390/cancers13051118
13 doi:10.1158/0008-5472.can-20-2811
14 doi:10.1016/j.ccell.2021.05.015
15 doi:10.1016/j.canlet.2021.03.013
16 doi:10.1016/j.cell.2020.11.018
17 PMID: 34079126
18 PMID: 34079128
19 https://doi.org/10.1038/s41586-021-03525-z
20 https://doi.org/10.3390/cancers11070908
21 PMID: 34546040
22 PMID: 32976798
23 PMID: 32737464
24 PMID: 33340459
25 PMID: 33296678
26 Doi: 10.1038/s41467-021-21801-4
27 DOI: 10.1038/s41587-020-0742-6
28 DOI: 10.1126/science.abg5159
29 Doi: 10.1111/febs.15832
30 PMID: 34485379
31 PMID: 32866920
32 PMID: 34630283
33 PMID: 34626833
34 PMID: 34099675
35 PMID: 32553762
36 PMID: 34190319
37 PMID 34475573
38 PMID 31999256
39 PMID 33119106
40 PMID 34083811
41 PMID 34363016
42 doi: 10.1038/s41592-020-01028-9
43 doi: 10.1016/j.jbc.2021.100448
44 doi: 10.1002/cpz1.104
45 doi: 10.1016/j.cell.2021.04.023
46 doi: 10.1038/s41586-021-03925-1
47 doi: 10.1021/acscentsci.0c01669
48 doi: 10.1073/pnas.2026102118
49 doi: 10.1016/j.jbc.2021.100966
50 doi: 10.1093/glycob/cwab063;
51 doi: 10.1093/glycob/cwab032
52 doi: 10.3389/fmicb.2021.731026
53 doi: 10.1016/j.jaut.2020.102575
54 doi: 10.3389/fchem.2021.664097
55 doi: 10.1002/anie.202106056
56 doi: 10.1038/s41586-021-03819-2
57 doi: 10.1038/s41586-021-03828-1
58 doi: 10.1126/science.abn5795
59 doi : 10.1101/2021.10.04.463034
60 doi: 10.1038/s41586-021-04184-w
61 doi: 10.1101/2021.10.11.463937
62 doi: 10.1126/science.abj6511
63 doi: 10.1101/2021.11.10.468128
64 doi: 10.1101/2021.09.04.459002
65 https://doi.org/10.1038/s41590-020-00845-6
66 https://doi.org/10.1371/journal.pbio.3001062
67 https://doi.org/10.1080/19490976.2021.1939598
68 https://doi.org/10.1038/s41590-021-01052-7
69 https://doi.org/10.1126/science.abf3002
70 https://doi.org/10.1016/j.cell.2020.12.011
71 https://doi.org/10.1186/s13059-021-02482-0
72 https://doi.org/10.1038/s12276-021-00667-y
73 https://doi.org/10.1371/journal.ppat.1009245
74 https://ec.europa.eu/commission/presscorner/detail/en/ip_21_6966
75 doi: 10.1126/sciimmunol.abf1152
76 doi: 10.1038/s41570-020-00244-3
77 doi: 10.1002/chem.202004705
78 doi: 10.1021/acs.jmedchem.0c01186
79 doi: 10.1039/d1sc05736g
80 doi: 10.1039/d1cc04459a
81 doi: 10.1039/d0cs01275k
82 PMID: 33478437
83 PMID: 34795231,
84 PMID: 34621739
85 PMID 34411579
86 PMID 34257535
87 PMID: 34706681
88 PMID 34782707
89 PMID 34732693
90 PMID: 33961790
91 PMID: 33406384
92 PMID: 33144328
93 PMID: 34879279
94 PMID 34585421
95 PMID: 33480129
96 PMID: 33502312
97 PMID: 34673284
98 PMID: 33444530
99 PMID: 34368969
100 PMID 34299191
101 PMID: 33470442
102 PMID: 34279939
103 PMID 33970635
104 PMID: 33909971
105 DOI: 10.1101/2021.06.16.448656
106 DOI: 10.1101/2021.07.01.450794
107 PMID: 34702444
108 PMID 33378226
109 doi:10.1016/j.molcel.2021.08.011
110 doi:10.1016/j.biosystems.2021.104542
111 doi:10.1101/2021.08.10.455830
112 doi:10.1016/j.conb.2021.07.010
113 doi:10.1007/s00401-021-02310-6
114 doi: 10.1371/journal.ppat.1009232
115 doi:10.3390/ijms22105075
116 doi:10.1073/pnas.2100968118
117 doi:10.1016/j.jbc.2021.101344
118 doi:10.1101/276436
119 doi:10.1007/s00401-020-02163-5
120 doi:10.3389/fmolb.2021.671458
121 doi:10.1007/978-1-0716-1398-6_60
122 doi: 10.1016/j.jbc.2021.100499
123 doi:10.1002/mds.28774
124 doi:10.1007/s00401-021-02312-4
125 doi:10.3390/ijms222111742
126 doi:10.1016/j.jbc.2021.101073
127 doi:10.1371/journal.ppat.1010013
128 doi:10.3390/ijms22010465
129 doi:10.1128/mBio.03508-20
130 doi:10.3389/fnagi.2021.698979
131 doi:10.1159/000520736
132 doi:10.1080/19336896.2021.1933873
133 doi:10.1038/s41582-021-00488-7
134 doi:10.1007/s00401-020-02231-w
135 doi:10.1007/s00401-021-02326-y
136 doi:10.1186/s40478-021-01266-8
137 doi:10.1212/WNL.0000000000012737
138 doi:10.1186/s12883-021-02274-w
139 doi:10.1111/ene.14624
140 PMID: 34799235
141 PMID: 34077195
142 PMID: 34562174
143 PMID: 34058526

144 PMID: 32695405
145 PMID: 28580910
146 PMID: 27467721
147 https://www.deia.eus/opinion/tribuna-abierta/2017/01/19/potente-mirada-euskadi-futuro/551456.html
148 doi: 10.1126/sciadv.abf8598
149 doi: 10.1073/pnas.2022614118
150 doi: 10.1016/j.cell.2020.10.030
151 doi: 10.1126/sciadv.abg4007
152 DOI: 10.1021/acs.analchem.0c01551
153 DOI: 10.1186/s13059-018-1547-5
154 doi: 10.1038/s41467-021-26807-6
155 doi: 10.3389/fcell.2021.715868
156 PMID: 32982112
157 PMID: 28211494
158 PMID: 28701731
159 PMID: 34637789
160 PMID: 33744287
161 PMID: 34466785
162 PMID: 33087931
163 PMID: 31425790
164 PMID: 33785757
165 PMID: 34703004
166 PMID: 32218950
167 PMID: 34466785
168 PMID: 31949053
169 PMID: 34330917
170 PMID: 34281791
171 PMID: 34782609
172 PMID: 34895659
173 PMID: 30325568
174 PMID: 34413316
175 PMID: 34873060
176 doi:10.3390/v13050885
177 doi: 10.1021/acs.jctc.1c00086
178 doi: 10.1002/anie.202007205
179 doi: 10.1021/jacs.1c04687
180 doi: 10.1007/s10858-021-00368-8
181 doi: 10.1007/s10858-021-00379-5
182 doi: 10.1016/j.jmr.2021.106937
183 doi: 10.1021/acs.jpcb.1c00989
184 doi: 10.1016/j.jmr.2020.106891
185 doi: 10.1002/anie.202012046
186 doi: 10.1002/cbic.202000674
187 doi: 10.1002/anie.202109430
188 doi: 10.1021/jacs.0c13289
189 doi: 10.1007/s10858-021-00382-w
190 doi: 10.1016/j.jmr.2020.106846
191 doi: 10.1016/j.jmr.2021.107003
192 doi: 10.1007/s10858-020-00350-w
193 doi: 10.1002/cbic.202000517
194 doi: 10.1016/j.biochi.2021.05.008
195 doi: 10.1073/pnas.2025022118
196 doi: 10.1038/s41467-021-26538-8
197 doi: 10.1021/jacs.1c04786
198 doi: 10.1002/chem.202100890
199 doi: 10.1021/jacs.0c10789
200 doi: 10.1021/jacs.1c04512
201 doi: 10.1021/jacs.1c01914
202 doi: 10.1016/j.jmr.2021.107083
203 doi: 10.1021/jacs.1c04512
204 doi: 10.1002/anie.202015948
205 doi: 10.1016/j.jmr.2021.107027
206 doi: 10.1021/acs.analchem.0c05205
207 doi: 10.1002/anie.202102487
208 doi: 10.1016/j.jmr.2020.106878
209 doi: 10.1016/j.jmr.2021.106928
210 doi: 10.1002/anie.202102758
211 doi: 10.1007/s10858-021-00385-7
212 doi: 10.1007/s10858-021-00369-7
213 doi: 10.1016/j.jmr.2020.106900
214 doi: 10.1007/s10858-021-00366-w
215 doi: 10.1007/s10858-021-00360-2
216 doi: 10.1038/s41467-021-25496-5
217 doi: 10.1016/j.jmr.2021.106995
218 doi: 10.1093/bioinformatics/btaa781
219 doi: 10.1038/s41467-021-20984-0
220 doi: 10.1007/s10858-020-00355-5
221 doi: 10.1007/s10858-021-00383-9
222 doi: 10.1007/s10858-021-00374-w
223 doi: 10.1007/s10858-021-00364-y
224 doi: 10.1002/cbic.202000701
225 doi: 10.1002/anie.202102408
226 doi: 10.1016/j.jmr.2020.106892
227 doi: 10.1039/d0cc07757g
228 doi: 10.1039/d1cc03079e
229 doi: 10.1002/anie.202011642
230 doi: 10.1039/d0cc08033k
231 doi: 10.1002/anie.202106056
232 doi: 10.1002/chem.202102039
233 doi: 10.1002/hep.31676
234 doi: 10.1002/cbic.202000698
235 doi: 10.1002/cbic.202000476
236 doi: 10.1039/d1cc01231b
237 doi: 10.1039/d1cc00791b
238 doi: 10.1002/cbic.202000601
239 doi: 10.1039/d1ob00611h
240 doi: 10.1002/cbic.202100304
241 doi: 10.1002/anie.202102690
242 doi: 10.1002/chem.202101752
243 doi: 10.1021/jacs.1c04477
244 doi: 10.1002/chem.202100301
245 doi: 10.1002/chem.202003143
246 doi: 10.1021/jacs.1c02383
247 doi: 10.1007/s10858-021-00370-0
248 doi: 10.1002/anie.202016070
249 doi: 10.1007/s10858-021-00372-y
250 doi: 10.1007/s10858-021-00357-x
251 doi: 10.1039/d1ob00611h
252 doi: 10.1021/jacs.0c11971
253 doi: 10.1002/cbic.202000785
254 doi: 10.1002/chem.202101143
255 doi: 10.1007/s10858-020-00354-6
256 doi: 10.1039/d1cc01018b
257 doi: 10.1021/acsptsci.0c00188
258 doi: 10.1007/s10858-020-00356-4
259 doi: 10.1002/cbic.202000541
260 doi: 10.1002/cbic.202000740

CIC bioGUNE
MEMBER OF BASQUE RESEARCH & TECHNOLOGY ALLIANCE

www.cicbiogune.es

Funding Agencies:

EUSKO JAURLARITZA GOBIERNO VASCO

Bizkaia foru aldundia diputación foral

ikerbasque Basque Foundation for Science

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN

Instituto de Salud Carlos III

HFSP

NIH National Institutes of Health

imi innovative medicines initiative

erc European Research Council

European Commission

HORIZON 2020

FUNDACIÓN aecc INVESTIGACIÓN CONTRA EL CÁNCER

bioef

Interreg POCTEFA CTP

Fundación BBVA

Fundación AstraZeneca

Obra Social Fundación "la Caixa"

fero Fundación de Investigación Oncológica

ACMBI Asociación de Cáncer de Mama de Bizkaia

INTERNATIONAL BRAIN IBRO RESEARCH ORGANIZATION

Fundación Jesús Gangoiti

Asociación Española CJD Enfermedades Priónicas

Mizutani Foundation for Glycoscience

CIMIT® Consortia for Improving Medicine with Innovation & Technology

Fundación Tatiana Pérez de Guzmán el Bueno

JOHN TEMPLETON FOUNDATION

BRTA BASQUE RESEARCH & TECHNOLOGY ALLIANCE