IMGT®, the international ImMunoGeneTics information system®, Laboratoire d'ImmunoGénétique Moléculaire LIGM

Marie-Paule Lefranc and Gérard Lefranc

Our researches concern molecular immunogenetics, bioinformatics and rare human genetic diseases. We are studying the genetics, structures, functions and repertoires of the immunoglobulins (IG) of the B lymphocytes and plasmocytes, and of the T cell receptors (TR) on the T lymphocytes, which are essential components of the adaptive (specific) immunity in human and other vertebrates. In 1989, we created **IMGT®, the international ImMunoGeneTics information system®,** http://imgt.cines.fr (University Montpellier 2 and CNRS). IMGT®, a CNRS registered trademark (France, E.U., Canada, U.S.A.), is the international reference in immunogenetics and immunoinformatics. Other researches, with the Unit of Medical Genetics, St-Joseph University, Beirut (Lebanon), concern **rare recessive genetic diseases** in consanguineous Lebanese families.

For the period 2005-2007, the research activity of the IMGT team resulted in:
- 2 PhD theses: Quentin Kaas, December 2005, Post-doctoral researcher at the University of Queensland, Brisbane, Australia; Elodie Duprat, December 2005, Lecturer 'Maître de Conférences' at the University Paris 7)
- 33 publications in international journals - 10 chapters of books - 7 proceedings.
- 52 communications in congresses (11 regional, 12 national and 29 international).
- 54 invited lectures (3 regional, 10 national and 41 international):
Europe (25) (of which 7 as panel moderators and round table), United States (6), Tunisia (3), Switzerland (2), Brazil (1), Corea (1), Taiwan (2), India (1).
- Prix Rammal 2006 (Euroscience) to Gérard Lefranc for his contribution on rare genetic diseases in Lebanese and Tunisian consanguineous families (see below).

This high-quality integrated knowledge resource is specialized in the IG, TR and major histocompatibility complex (MHC) of vertebrate species, in the immunoglobulin superfamily (IgSF) and MHC superfamily (MhcSF) proteins, and in related proteins of the immune system (RPI). IMGT® provides a common access to expertly annotated nucleotide and protein sequences, structural data, and genetic information. IMGT® includes six databases (IMGT/LIGM-DB, a comprehensive database of more than 114,239 IG and TR sequences from human and from 208 other vertebrate species in November 2007, IMGT/GENE-DB, IMGT/3Dstructure-DB, etc.), Web resources which consist of 10,000 HTML pages, and fifteen interactive tools (IMGT/V-QUEST, IMGT/JunctionAnalysis, etc.). Since July 1995, IMGT® is available on the Web at http://imgt.cines.fr. The IMGT® Web server at Montpellier is accessed by more than 80,000 sites a year. IMGT® has an exceptional response with more than 150,000 requests per month, the users being divided equally between Europe, the United States and the rest of the world.

At the ***international*** level, IMGT® is the first and the only integrated information system in immunogenetics and immunoinformatics. There is no equivalent in Europe, in Japan, in the United States and nowhere else in the world (China, India). IMGT® site is the only foreign site referenced by the National Center for Biotechnology Information (NCBI) in the United States. IMGT® is partner of the ImmunoGrid (FP6-2004-IST-4) STREPS "The European Virtual Human Immune System Project" in the 6th EC Framework Programme, participant via ReNabi to the European Life sciences Infrastructure for Biological Information (ELIXIR) project, Academic institutional member of the International Medical Informatics Association (IMIA) since 2006.
At the ***national*** level, IMGT® is Bioinformatics Platform RIO since RIO creation (CNRS, INSERM, CEA, INRA) in 2001, member of the National Network of Bioinformatics Platforms (ReNaBi), member of the GIS IBiSA since the creation of the GIS in 2007, member of the Molecular Bioinformatics GDR (BiM), partner in the ANR ACI Informatics, Mathematics, Physics in Biology (ACI IMPBio 2004-2007), GIS AGENAE (2004-2007), ANR BIOSYS (2006-2010) and ANR FLAVORES (2008-2010).
At the ***regional*** level***,*** IMGT® is Plan Pluri-formation University Montpellier 2 since 1999, 'Grand Plateau Technique pour la Recherche Région' (GPTR) Languedoc-Roussillon, member of the GIS Genopole Montpellier Languedoc-Roussillon, Bioinformatics platform of the IFR3 Normal and pathological cell Communications, Bioinformatics Platform of the Cancéropôle Grand Sud-Ouest (GSO).

IMGT® is used by scientists from both academic and pharmaceutical companies, from very diverse research domains: (1) fundamental research, (2) medical research (repertoire analysis of the antibody sites and of the T cell receptors recognition sites in normal and pathological situations such as autoimmune diseases, infectious diseases, AIDS, leukaemias, lymphomas, myelomas), (3) veterinary research (IG and TR repertoires in domestic and wild life species, animal models for the analysis of the adaptive immune responses, eg mouse, chimpanzee), (4) genomics (study of the genome diversity and evolution of the adaptive immune responses), (5) structural biology (evolution of the domains of the IgSF and MhcSF proteins), (6) biotechnology related to antibody engineering (construction and analysis of single chain Fragment variable (scFv), phage displays, combinatorial libraries, chimeric, humanized and human antibodies), (7) diagnostics (clonalities, detection and follow-up of minimal residual diseases), and (8) therapeutical approaches (grafts, immunotherapy, vaccinology).

IMGT® provides a common access to standardized data that include nucleotide sequences, protein sequences, locus maps, genetic polymorphisms and three-dimensional structures (3D). IMGT® is based on IMGT-ONTOLOGY, the reference ontology in immunogenetics and immunoinformatics. Data are defined according to the IMGT Scientific chart rules, generated from the IMGT-ONTOLOGY axioms and concepts. The success of that standardization has introduced the expression 'IMGT Colliers de Perles' in the United States and in England. IMGT® development includes the creation of dynamic interactions between components of the information system (databases, tools and Web resources), according to the IMGT genetic, genomic and structural approaches. In order to reach that goal, we chose to use the Web Services. The approach of IMGT-Choreography is followed with interest by other laboratories IMGT-Choreography strengthens the position of IMGT as a system of international reference in immunogenetics and immunoinformatics.

2. Genetic diseases in consanguineous families: Public health problem and invaluable contribution to the fundamental research

In Lebanon and in Tunisia, as well as in other countries from North Africa, Near- and Middle-East, consanguinity is common and marriages among relatives occur widely: 25% of all marriages, particularly in rural areas, are between cousins (often between first cousins). In such families, if we consider a rare, or very rare, autosomal recessive allele in the general population, and a common ancestor who is carrier, the mutant gene will have been transmitted from the ancestor to its descendants. Therefore, the probability of receiving this rare allele, identical by descent, from the ancestor -i.e. the probability of being homozygous for that allele, is greatest in the offspring of a marriage between cousins. Thus, rare or very rare autosomal recessive diseases are more frequent in the offspring of such unions and the rarer is the autosomal recessive allele frequency, the higher will be the proportion of patients found to be consanguineous. Therefore, compared with panmictic populations, those with high levels of consanguinity experience much more rare genetic diseases, that remains **a problem of public health**. The occurrence in the consanguineous populations of genetic defects which are almost unknown in panmictic populations, and the generally large size of these families which increases the number of patients, are **invaluable starting points from which to identify unknown genes, their products and their functions**, responsible for unknown steps in signaling or in regulatory pathways. Previously unsuspected links to cell physiology are thus unmasked and can be analyzed. These consanguineous populations also allow to better study the genomic polymorphisms as markers of positive selection or, in contrast, of susceptibility against infectious diseases. At last, Lebanon and Tunisia are ideally located in order to describe the genome evolution and the gene admixture (Y chromosome, mtDNA…) on the pathways of human expansion out of Africa and, then, of numerous migrations and invasions.

Limb-Girdle Muscular Dystrophy type 2A (LGMD2A). LGMD2A is caused by mutations in the calcium-activated cysteine protease calpain 3 leading to progressive atrophy and weakness of the proximal limb, scapular pelvic girdle and trunk muscles. We previously showed that calpain 3 deficiency is associated with a profound perturbation of the NF-κB/IκBα survival pathway and that IκBα, the inhibitor of the NF-κB-Rel transcription factor, might be a substrate of calpain 3: muscular cells of LGMD2A patients from two Lebanese families presented apoptotic features that were associated with subsarcolemmal localization of NF-κB and selective nuclear accumulation of IκBα in TUNEL-positive myonuclei. Thus, the apoptotic cell death event in LGMD2A would be due to a failure in calpain 3-dependent IκBα proteolysis resulting in sequestration of NF-κB in the cytoplasm and subsequent downregulation or non expression of survival genes. Recently we have shown that the expression of the anti-apoptotic factor c-FLIP is dependent of the NF-κB pathway in normal muscle cells and that it is down-regulated in LGMD2A (Benayoun et al., FASEB J., in press). One therapeutic possibility is to envision blocking or slowing down apoptotic event(s) with a directed expression of anti-apoptotic proteins.

Immunodeficiency, Centromeric region instability and Facial anomalies (ICF) syndrome. ICF syndrome (OMIM 242860) is characterized by a variable immunodeficiency, mild facial anomalies, and centromeric decondensation chromosomal instability (breaks, deletions, isochromosomes and multiradial

configurations in mitogen-stimulated lymphocytes) involving chromosomes 1, 9, and 16. Of the few molecular findings reported, the most consistent is hypomethylation of the satellite 2 and satellite 3 regions of these chromosomes, whereas methylation of these sequences is normally almost complete in leukocyte DNA. Other regions of heterochromatin, such as the centromeric, satellite repeats and the inactive X, may also be hypomethylated. The ICF syndrome appears to offer an important model for understanding these methylation patterns. We identified three consanguineous Lebanese ICF patients: two brothers, aged 7 and 5, and the third, aged 14, in another family. The later is homozygous for a mutation in the *DNMT3B* gene (ICF type 1) located on chromosome 20q11-13. The identification of the mutated gene responsible for the ICF brothers (ICF type 2) is now in progress. The methylation patterns are now studied with the Albertina de SARIO's group, at the IGH, and with a team of Seattle. At the immunological level, the B cells defects associated with hypogammaglobulinemia in ICF type 1 are characterized by only naïve and no memory B cells in peripheral blood due to B cell maturation blockage.

Hyper-immunoglobulin E (IgE) syndromes (HIES). HIES (also called Job's syndrome, OMIM 147060 and 243700) are very rare primary immunodeficiencies, characterized by high serum levels of IgE (>2000 IU/ml), eczema, recurrent staphylococcal skin abscesses, mucocutaneous candidiasis, and pneumonia with pneumatocele formation. Most cases are sporadic, but both autosomal dominant forms of HIES (AD-HIES) and autosomal recessive forms (AR-HIES) have been described. The clinical features of sporadic and AD-HIES encompass the immune system, connective tissue, skeleton, and dental development with variations in the severity of the symptoms as facial abnormalities and characteristic facies, hyper-extensibility of joints, scoliosis, osteoporosis, recurrent pathological fractures and retained primary teeth. The AR-HIES is characterized by severe recurrent viral infections, extreme eosinophilia, defect in neutrophil chemotaxis and neurological complications that are often fatal in childhood, but without skeletal or dental abnormalities. Recently, a homozygous mutation in the Tyrosine kinase 2 (*Tyk2*) gene, coding for a non-receptor tyrosine kinase belonging to the Janus kinase (Jak) family, was found in a patient clinically diagnosed with AR-HIES. Tyk2 is essential for type I IFNα signaling and also for IL-6, -10, -12 and -23 pathways as exemplified by the defects in the patient's cells. More recently, dominant-negative mutations in the DNA-binding domain of STAT3 have been identified as responsible for the sporadic and AD form of HIES. Given the STAT3's role in the signaling pathways of the IL-6-, IFN- and IL-2-families cytokines, as well as of IL-5, IL-23, CSF3/G, EGF, CSF1 and leptin, its diminished activity has an impact on the development and functions of multiple organ systems leading to compound clinical manifestations. We have investigated sporadic and AD-HIES forms of HIES. Until now we have found three mutations in the DNA binding domain of STAT3 and another one in the SH2 domain (to be published). Further, we will study the *Tyk2* gene in AR-HIES patients.

Autosomal Recessive Osteopetrosis (ARO). ARO (also called malignant or infantile osteopetrosis, OMIM 259700) are severe hereditary bone diseases whose cellular basis is in the osteoclast, but whose molecular defect is heterogeneous, with neural involvement for the most severe form. Mutations of either *TCIRG1* (T-cell immune regulator1, vacuolar proton pump VPP alpha subunit 3, OMIM 604592), or *ClCN7* (Chloride channel 7, OMIM 602727), or *OSTM1* (Osteopetrosis associated transmembrane protein 1, OMIM 607649; grey lethal in the mouse) genes have

been until now recognized as responsible for the ARO. Approximately 55% of patients with clinical diagnosis of ARO show mutations in the *TCIRG1* gene. VPP plays a fundamental role in acidifying the osteoclast–bone interface, which is a prerequisite for bone mineral resorption. This acidification is also hampered by a defect in the *ClCN7* gene which is mutated in 10% of ARO patients. The third form of ARO, due to mutations of the *OSTM1* gene, has been described so far in only two ARO patients, due to a very short life expectancy. The function of the OSTM1 protein is much less clear. TCIRG1-dependent ARO patients have a severe homogeneous phenotype; their nervous system involvement (hydrocephalus and cranial nerve defects) is secondary to the compression because of skull deformities. On the other hand, there is increasing evidence that patients with recessive *ClCN7* mutations (*ClCN7*-dependent ARO), besides the bone manifestations, show a primary severe neurological defect (retinopathy and progressive cortical atrophy in addition to the secondary neural defects). We performed the clinical and molecular analysis of seven ARO patients (among them, three new cases due to mutations of the *OSTM1* gene), belonging to six families originating from Near-East (Lebanon and Syria) (Pangrazio et al., 2006; Souraty et al., 2007). This study allowed us to suggest that *OSTM1*-dependent ARO, besides the bone manifestations, shows a primary severe neurological defect (retinopathy and cortical atrophy in addition to secondary neural defects) due to lysosomal storage disease (Alroy et al., 2007). Also we are investigating a mild osteopetrosis in two consanguineous adult sisters.

Hepatic veno-occlusive disease with immunodeficiency syndrome (VODI), VODI (OMIM 235550) is a primary immunodeficiency associated with hepatic vascular occlusion and fibrosis. VODI is associated with an 85% mortality if unrecognized and untreated with intravenous immunoglobulins. We identified a 15 years old girl from a Lebanese family with a lymphopenia, a humoral deficiency (she receives IVIG every 3 or 4 weeks), a very weak proliferation of her T cells. In this consanguineous family, and in another Lebanese family showing the same syndrome, brothers and sisters as well as several cousins died of this disease at a very young age. Histologic study revealed that spleen and lymph nodes were devoid of germinative centers. The thymus was also relatively poor in lymphocytes. The B cell immunodeficiency was characterized by severe hypogammaglobulinemia and absent memory B cells, consistent with a block in B cell differentiation. The T cell immunodeficiency was due to reduced numbers of memory T cells. The *VODI* gene is mapped to chromosome 2q36.3-37.1 and it encodes Sp110, a 110-kDa Speckled immunoregulatory protein, expressed in T and B lymphocytes, lymph nodes, spleen and liver. Mutation screening revealed that the girl is homozygous for a single-base deletion responsible for a stop codon at position 228. Subsequently, the same mutation was found at the heterozygous state in another family in the healthy parents and sisters of deceased children. Due to the stop codon, the mutated Sp110 protein is truncated and lacks the C-terminal region and the LXXLL (aa 525 to 529) nuclear hormone receptor binding motif. The wild type protein is associated with the PML nuclear bodies that have a role in chromatin structure, are present in areas of active DNA replication, transcription and repair and are also involved in apoptosis, cell cycle control and immunity. The mutated Sp110 offers an exceptional opportunity to better understand its fundamental functions in the cell biology.