

Etude de cas

Ministère de l'Enseignement Supérieur et de la Recherche (MESR)
Mission de l'Information Scientifique et Technique et du Réseau Documentaire
(MISTRD)
Contact: Mr Dominique Aussant

IMGT®, le système d'information international en immunogénétique et immunoinformatique

IMGT®, the international ImMunoGenetics information system®: <http://www.imgt.org>

Fondateur et Directeur: Marie-Paule Lefranc
Professeur classe exceptionnelle Université Montpellier 2
Membre Senior de l'Institut Universitaire de France (IUF)
Marie-Paule.Lefranc@igh.cnrs.fr



Domaines scientifiques:

Immunologie, immunogénétique, immunoinformatique, bioinformatique, ingénierie des anticorps

Octobre 2009

Sommaire

| | |
|--|-----------|
| 1. PRESENTATION DE LA STRUCTURE IMGT® ET DE SON SYSTEME DE DONNEES DE RECHERCHE | 5 |
| 1.1. INTITULE, RESPONSABLES ET RATTACHEMENT DE LA STRUCTURE IMGT® | 5 |
| 1.2. LA STRUCTURE IMGT® DANS LE CONTEXTE INTERNATIONAL, NATIONAL, INTERREGIONAL ET REGIONAL | 6 |
| 1.2.1. International | 6 |
| 1.2.2. National | 6 |
| 1.2.3. Interrégional | 6 |
| 1.2.4. Régional | 6 |
| 1.3. DONNEES DE RECHERCHE D'IMGT® | 7 |
| 1.3.1. Domaines spécifiques d'IMGT® | 7 |
| 1.3.2. Cellules et molécules de la réponse immunitaire adaptative | 7 |
| 1.3.3. Types de données expertisées et gérées par IMGT® | 9 |
| 1.3.4. Nécessité d'une standardisation et d'une ontologie: IMGT-ONTOLOGY | 9 |
| 1.4. COMPOSANTS D'IMGT® | 13 |
| 1.4.1. Les bases de données IMGT® | 15 |
| 1.4.2. Les outils en ligne IMGT® | 18 |
| 1.4.3. Les ressources Web IMGT® | 19 |
| 1.5. UTILISATEURS D'IMGT® | 19 |
| 2. CARACTERISATION DU SYSTEME DE DONNEES DE RECHERCHE IMGT® (OBJECTIFS, HISTORIQUE, CONTEXTE) | 20 |
| 2.1. OBJECTIFS | 20 |
| 2.2. HISTORIQUE | 21 |
| 2.3. CONTEXTE D'IMGT® AU NIVEAU DES FLUX DE DONNEES ET DES COLLABORATIONS | 22 |
| 2.3.1. Contexte d'IMGT® et flux de données entrantes et sortantes | 22 |
| 2.3.2. Relations entre IMGT® et ses différents collaborateurs et utilisateurs | 23 |
| 3. FONCTIONNALITES COUVERTES PAR LE SYSTEME DE DONNEES DE RECHERCHE IMGT® (PRODUCTION, GESTION, ARCHIVAGE, PRESERVATION, MISE A DISPOSITION, DIFFUSION, ANALYSE, EXPLOITATION...) | 25 |
| 3.1. PRODUCTION | 25 |
| 3.1.1. IMGT/LIGM-DB: production de séquences nucléotidiques annotées | 25 |
| 3.1.2. IMGT IMGT/3Dstructure-DB: production de structures 3D annotées | 25 |
| 3.2. GESTION | 28 |
| 3.3. ARCHIVAGE | 28 |
| 3.3.1. Bases de données | 28 |
| 3.3.2. Outils et ressources Web | 29 |
| 3.4. PRESERVATION | 29 |
| 3.4.1. En externe | 29 |
| 3.4.2. En interne | 29 |
| 3.5. MISE A DISPOSITION ET DIFFUSION DES DONNEES D'IMGT® | 29 |
| 3.5.1. Site Web | 29 |
| 3.5.2. Distribution des données d'IMGT/LIGM-DB par FTP | 29 |
| 3.5.3. Accès aux données d'IMGT/LIGM-DB par SRS et ARSA | 29 |
| 3.5.4. Accès aux données par DAS à l'EBI et LinkOut au NCBI | 30 |
| 3.5.5. Comparaison de séquences par BLAST et FASTA | 30 |
| 3.5.6. Références croisées | 30 |
| 3.5.7. Aides spécifiques et conseils aux utilisateurs et API | 30 |
| 3.6. ANALYSE ET EXPLOITATION | 31 |
| 4. INFRASTRUCTURES DU SYSTEME DE DONNEES DE RECHERCHE IMGT® (TECHNIQUES, FINANCIERES, HUMAINES, REGLEMENTAIRES, PARTENARIATS...) | 32 |
| 4.1. INFRASTRUCTURES TECHNIQUES ET MOYENS CPU | 32 |
| 4.1.1. Moyens en équipement | 32 |
| 4.1.2. Capacité totale de stockage | 32 |
| 4.1.3. Nombre total de CPU | 32 |
| 4.2. FINANCEMENTS PUBLICS DE LA STRUCTURE IMGT® | 32 |
| 4.2.1. Contrats Européens | 33 |
| 4.2.2. Financements Nationaux | 34 |
| 4.2.3. Financements Régionaux | 36 |
| 4.3. FINANCEMENTS INDUSTRIELS DE LA STRUCTURE IMGT® | 37 |
| 4.3.1. Contrats types et licences | 37 |

| | |
|---|-----------|
| 4.3.2. Partenaires industriels | 37 |
| 4.4. RESSOURCES HUMAINES DE LA STRUCTURE IMGT® | 38 |
| 4.5. REGLEMENTATION ET VALORISATION DE LA STRUCTURE IMGT® | 39 |
| 4.5.1. Accréditations | 39 |
| 4.5.2. Marques déposées | 39 |
| 4.5.3. Brevets | 40 |
| 4.5.4. Contrats de licence et savoir..... | 40 |
| 4.5.5. Certificats logiciels | 40 |
| 4.5.6. Domaines Internet | 40 |
| 4.5.7. Prix ADER Sciences de la vie..... | 40 |
| 4.6. PARTENARIATS SCIENTIFIQUES | 41 |
| 4.6.1. Nomenclature IMGT® pour les projets génomes et protéomes | 41 |
| 4.6.2. Pronostic dans les leucémies lymphoïdes chroniques..... | 42 |
| 4.6.3. Humanisation d'anticorps par la méthodologie «IMGT Collier de Perles» | 43 |
| 4.6.4. ImmunoGrid | 44 |
| 5. IDENTIFICATION DE L'AUTONOMIE ET/OU DE LA DEPENDANCE EXTERIEURE DES PARTIES DU SYSTEME IMGT ® | 45 |
| 6. PROJETS FUTURS ET ANALYSE DES EVENTUELS BESOINS D'INFRASTRUCTURES EXTERIEURES | 45 |
| 6.1. PROJETS FUTURS DE LA STRUCTURE IMGT® | 45 |
| 6.1.1. Annotation automatique de grandes séquences génomiques des locus IG et TR: IMGT/LIGMotif | 46 |
| 6.1.2. Analyse haut débit des séquences réarrangées des IG et TR: IMGT/HighV-QUEST..... | 47 |
| 6.1.3. La démarche qualité à IMGT®..... | 48 |
| 6.2. ANALYSE DES BESOINS D'INFRASTRUCTURES ET DE POSTES D'INGENIEURS..... | 50 |
| 6.2.1. Reconnaissance d'IMGT® en tant que nœud thématique du projet européen ELIXIR..... | 50 |
| 6.2.2. Nécessité de pérenniser IMGT® par la création de postes d'ingénieurs | 50 |

1. Présentation de la structure IMGT® et de son système de données de recherche

1.1. Intitulé, responsables et rattachement de la structure IMGT®

IMGT®, le système d'information international en immunogénétique et immunoinformatique.

IMGT®, the international ImMunoGenetics information system®, <http://www.imgt.org>

Créé en juin 1989 à Montpellier (France).

Fondateur et Directeur

Marie-Paule Lefranc
Professeur classe exceptionnelle Université Montpellier 2
Membre Senior de l'Institut Universitaire de France (IUF)
Marie-Paule.Lefranc@igh.cnrs.fr



IMGT, Laboratoire d'Immunogénétique Moléculaire, LIGM
Institut de Génétique Humaine, IGH, UPR CNRS 1142,
141, rue de la Cardonille, 34396 Montpellier cedex 5
Tel: 04-99-61-99-65, Fax:04-99-61-99-01

Responsable informatique

Duroux Patrice, Ingénieur de recherche CNRS, Patrice.Duroux@igh.cnrs.fr

Responsable bioinformatique

Giudicelli Véronique, Ingénieur d'études Université Montpellier 2, Veronique.Giudicelli@igh.cnrs.fr

Rattachement administratif

1. Université Montpellier 2, Faculté des Sciences de Montpellier, Place Eugène Bataillon, 34095 Montpellier cedex 5, <http://www.univ-montp2.fr/>

2. CNRS, Délégation Languedoc-Roussillon DR13, 1919, route de Mende, 34293 Montpellier cedex 5, <http://www.cnrs.fr/languedoc-roussillon/index.htm>

Institut des Sciences Biologiques, 3, rue Michel-Ange, 75794 Paris cedex 16 <http://www.cnrs.fr/sdv/>

3. Instituts Thématiques Multi-Organismes :
ITMO Technologies pour la Santé et Génétique,
ITMO Génomique et Bioinformatique.

4. Section C.N.U: 65 (Biologie cellulaire).

5. Section CNRS: 24 (Interactions cellulaires).

Secteurs disciplinaires concernés/mots clés

Immunologie, immunogénétique, immunoinformatique, bioinformatique, ingénierie des anticorps.

1.2. La structure IMGT® dans le contexte international, national, interrégional et régional

IMGT® a un positionnement très concurrentiel en immunoinformatique (Figure 1)

1.2.1. International

IMGT®, the international ImMunoGeneTics information system® (<http://www.imgt.org>), créé en 1989 à Montpellier par le Laboratoire d'ImmunoGénétique Moléculaire (LIGM) des Professeurs Marie-Paule et Gérard Lefranc (Université Montpellier 2 et CNRS), est le premier et, actuellement, le seul système d'information intégré en immunogénétique et immunoinformatique (Figure 1). Il n'existe pas d'équivalent en Europe, aux Etats-Unis, au Japon, et nulle part ailleurs dans le monde. IMGT® est la référence internationale en immunogénétique et immunoinformatique.

IMGT® est:

- marque déposée du CNRS (France, Union Européenne, Canada et Etats-Unis).
- membre institutionnel académique de *l'International Medical Informatics Association* (IMIA) depuis 2006.
- certifié 'HONcode' (*Health on the Net foundation Code of Conduct for medical and health Web sites*), depuis 1999.
- présent dans le Guide Européen de l'Innovation.
- partenaire du projet européen '*The European virtual human immune system*', 'ImmunoGrid' EU 6th PCRTD (IST-2004-028069).
- participant au projet européen '*European Life sciences Infrastructure for Biological Information*' (ELIXIR).

1.2.2. National

IMGT® est:

- plate-forme Nationale RIO: plate-forme de Recherche Bioinformatique dans le cadre de la réunion inter-organismes entre l'INSERM, le CNRS, le CEA et l'INRA, depuis la création de RIO en 2001 et jusqu'en 2007, date de la création du GIS IBiSA.
- plate-forme Nationale IBiSA: plate-forme de Recherche Bioinformatique dans le cadre des Infrastructures en Biologie Santé et Agronomie et de la Coordination des plates-formes de recherches en sciences du vivant, depuis la création du GIS IBiSA en 2007.
- membre du Réseau National des plates-formes Bioinformatique (ReNaBi).
- membre du Groupement de Recherche (GDR) CNRS n° 3003, Bioinformatique Moléculaire (BiM).
- membre du Groupement de Recherche (GDR) CNRS n° 3260, Anticorps et ciblage thérapeutique (ACCITH), 2009-2012.

1.2.3. Interrégional

IMGT® est plate-forme bioinformatique du Cancéropôle Grand Sud-Ouest (GSO).

1.2.4. Régional

IMGT® est:

- Grand Plateau Technique pour la Recherche (GPTR) Région Languedoc-Roussillon, depuis la création des GPTR en 2005.
- Plan Pluri-formation (PPF) Université Montpellier 2 depuis 1999.
- plate-forme Bioinformatique de l'Institut Fédératif de Recherche (IFR3) 'Communications cellulaires normales et pathologiques'.
- plate-forme Bioinformatique du GIS Genopole® Montpellier Languedoc-Roussillon, jusqu'à la création du GIS IBiSA en 2007.

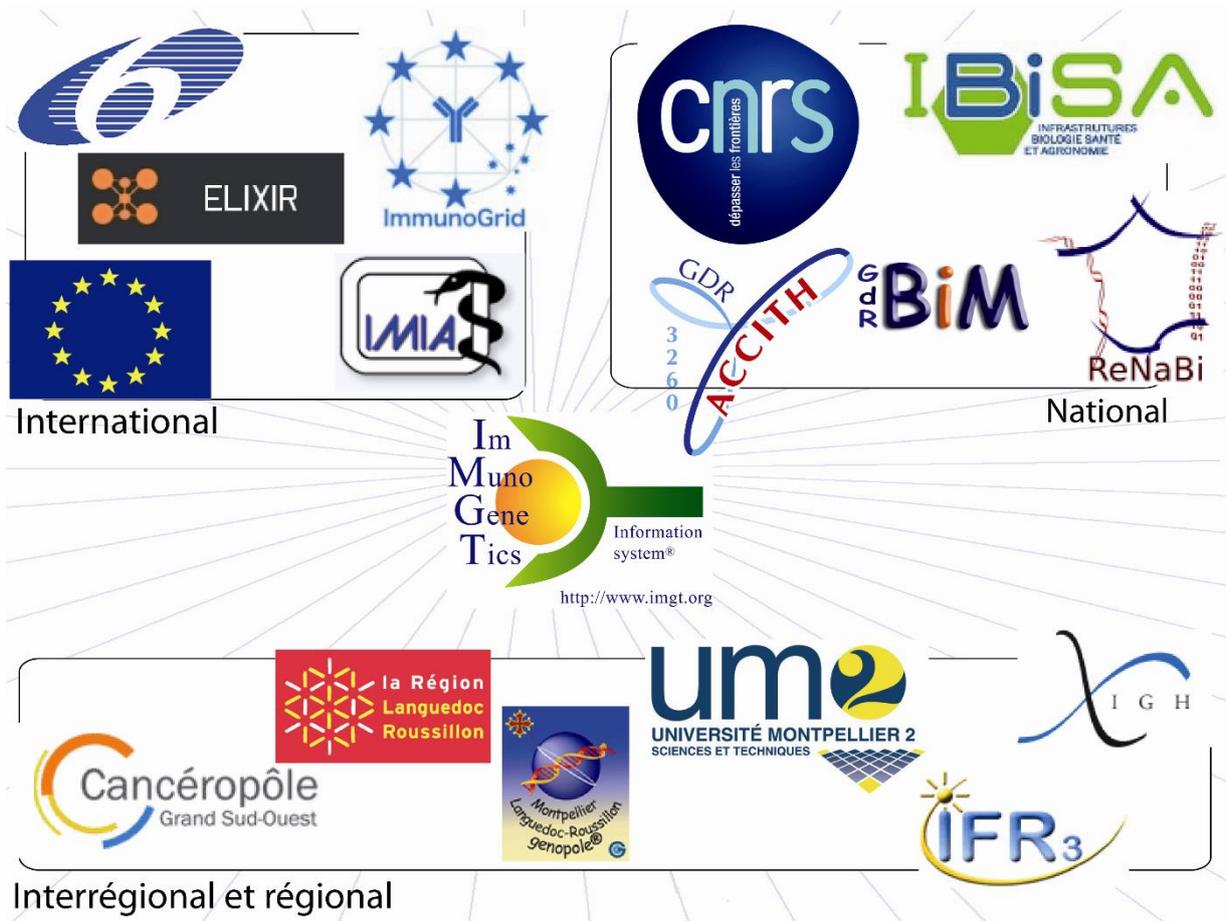


Figure 1. La structure IMGT® dans le contexte international, national, interrégional et régional.

1.3. Données de recherche d'IMGT®

IMGT®, the international ImMunoGeneTics information system® (<http://www.imgt.org>), a pour mission de fournir une source unique de connaissances en immunogénétique et immunoinformatique, au niveau international.

1.3.1. Domaines spécifiques d'IMGT®

IMGT® est spécialisé dans les séquences et les données génétiques, génomiques et structurales des immunoglobulines (IG) ou anticorps, des récepteurs T (TR), du complexe majeur d'histocompatibilité (CMH) (en anglais MHC, pour *major histocompatibility complex*) des vertébrés, des protéines des superfamilles IgSF (*immunoglobulin superfamily*) et MhcSF (*MHC superfamily*) et des protéines apparentées du système immunitaire (RPI, pour *Related Proteins of the Immune system*) des vertébrés et invertébrés.

Mots clés: Immunologie, immunogénétique, immunoinformatique, bioinformatique, ingénierie des anticorps, immunoglobuline (IG), anticorps, récepteur T (TR), réponse immunitaire adaptative, humanisation des anticorps, complexe majeur d'histocompatibilité (CMH), superfamille des Immunoglobulines (IgSF), superfamille des CMH (MhcSF), major histocompatibility complex (MHC), human leucocyte antigen (HLA).

1.3.2. Cellules et molécules de la réponse immunitaire adaptative

La réponse immunitaire adaptative (Figure 2) est apparue il y a 450 millions d'années chez les vertébrés à mâchoires (Gnathostomata).

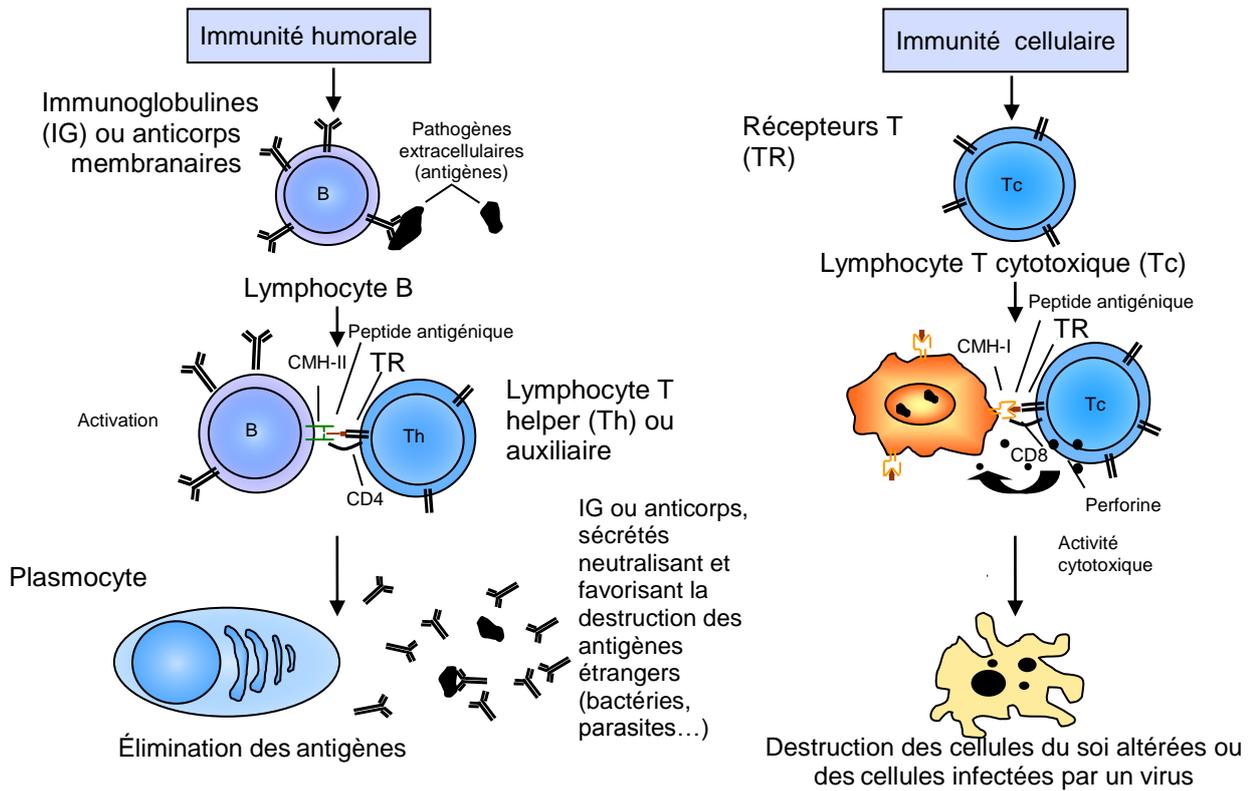


Figure 2. Cellules et molécules de la réponse immunitaire adaptative.

| | Nombre d'individus | Nombre total d'IG et TR |
|---|-----------------------------|-------------------------|
|  | 1 | 4.10^{12} |
|  | 100 | 4.10^{14} |
| France  | 66 millions 66.10^6 | 2.10^{20} |
| Europe  | 733 millions 733.10^6 | 3.10^{21} |
| Monde  | 6,8 milliards $6,8.10^9$ | 27.10^{21} |

Figure 3. Evaluation du répertoire des récepteurs d'antigènes (IG et TR) différents de la réponse immunitaire adaptative.

Les cellules spécifiques de la réponse immunitaire adaptative comprennent les lymphocytes B (et les plasmocytes qui en dérivent) et les lymphocytes T. Les molécules spécifiques de la réponse immunitaire adaptative comprennent les immunoglobulines (IG) ou anticorps, les récepteurs T (TR) et les protéines du CMH. Ces cellules et molécules interviennent dans les deux axes de la réponse immunitaire adaptative: l'immunité humorale et l'immunité cellulaire (Figure 2). Les IG et les TR sont les «récepteurs d'antigènes», présents à la surface des lymphocytes B et T, respectivement, et, dans le cas des IG, sécrétés par les plasmocytes. Ce sont ces molécules qui vont reconnaître de façon spécifique les antigènes des bactéries, des virus, des parasites et des cellules tumorales (cancéreuses) et favoriser leur neutralisation et leur destruction.

Les données immunogénétiques moléculaires correspondant à ces récepteurs ne sont pas gérées dans les bases généralistes, en raison de la complexité de leur synthèse et de leur énorme diversité. Nous estimons que le répertoire potentiel immunitaire de l'humanité à un instant 't' est de $27 \cdot 10^{21}$ anticorps et récepteurs T différents (Figure 3). Il n'est certes pas possible de gérer globalement les informations d'un très grand nombre d'individus mais il est nécessaire et indispensable de créer les bases de données et outils qui puissent permettre l'analyse et l'interprétation du répertoire de tout individu, quelles que soient son origine géographique et sa situation immunitaire. IMGT® a été créé pour répondre à ce besoin et grâce à l'expertise internationalement reconnue du Laboratoire d'ImmunoGénétique Moléculaire des professeurs Marie-Paule et Gérard Lefranc dans le domaine de l'immunologie et de l'immunogénétique.

1.3.3. Types de données expertisées et gérées par IMGT®

Les données de recherche expertisées et gérées par IMGT® comprennent essentiellement les données moléculaires des IG, TR et CMH de la réponse immunitaire adaptative et se sont étendues progressivement aux protéines RPI des superfamilles IgSF et MhcSF.

Ces données sont principalement de trois types:

- données génétiques (séquences nucléotidiques et protéiques),
- données génomiques (organisation des gènes et des locus),
- données structurales (structures bidimensionnelles (2D) et tridimensionnelles (3D)).

Ces données sont indispensables en recherche fondamentale, médicale et vétérinaire pour l'analyse et la compréhension des mécanismes de la synthèse des immunoglobulines (IG) ou anticorps et des récepteurs T (TR), l'analyse du répertoire immunitaire en situations normales et pathologiques et l'étude des actions biologiques à la suite de la reconnaissance spécifique des antigènes des agents pathogènes et des tumeurs par les récepteurs d'antigènes (Figure 4).

L'analyse et la compréhension des remarquables propriétés de la réponse immunitaire adaptative (spécificité, mémoire immunitaire) permettent des développements considérables en médecine (applications cliniques en diagnostic et thérapeutique) et en biotechnologie (développement de vaccins, ingénierie moléculaire et humanisation des anticorps) (Figure 4), ce qui explique l'intérêt des laboratoires pharmaceutiques ('Biotech', 'Big Pharma') pour les données d'IMGT®.

1.3.4. Nécessité d'une standardisation et d'une ontologie: IMGT-ONTOLOGY

La biosynthèse au niveau moléculaire et la génétique des IG et des TR et le polymorphisme du CMH sont particulièrement complexes et l'une des premières tâches d'IMGT® a été définir une spécification formelle et standardisée des termes à utiliser dans le domaine de l'immunogénétique. Ceci a conduit à la création d'IMGT-ONTOLOGY, la première ontologie en immunogénétique et immunoinformatique.

IMGT-ONTOLOGY comprend un vocabulaire contrôlé et les règles d'annotation qui sont indispensables pour garantir l'exactitude, l'uniformité et la cohérence d'IMGT®. IMGT-ONTOLOGY permet aux scientifiques et aux cliniciens d'utiliser des termes identiques, tout en contrôlant le sens de chacun d'eux. En fournissant un référentiel sémantique qui peut être inclus dans un grand nombre d'ontologies de biologie moléculaire, IMGT-ONTOLOGY est d'une grande aide pour accroître l'interopérabilité entre les bases de données spécialisées et généralistes.

Les axiomes et concepts d'IMGT-ONTOLOGY

IMGT-ONTOLOGY comprend sept axiomes : 'IDENTIFICATION', 'CLASSIFICATION', 'DESCRIPTION', 'NUMEROTATION', 'LOCALIZATION', 'ORIENTATION', et 'OBTENTION' (Figure 5). Ils constituent l'IMGT-ONTOLOGY formelle (Formal IMGT-ONTOLOGY) ou IMGT-Kaleidoscope. Ces sept axiomes ont permis de définir les concepts correspondants d'IMGT-ONTOLOGY et les règles de la charte scientifique d'IMGT® (IMGT Scientific chart).

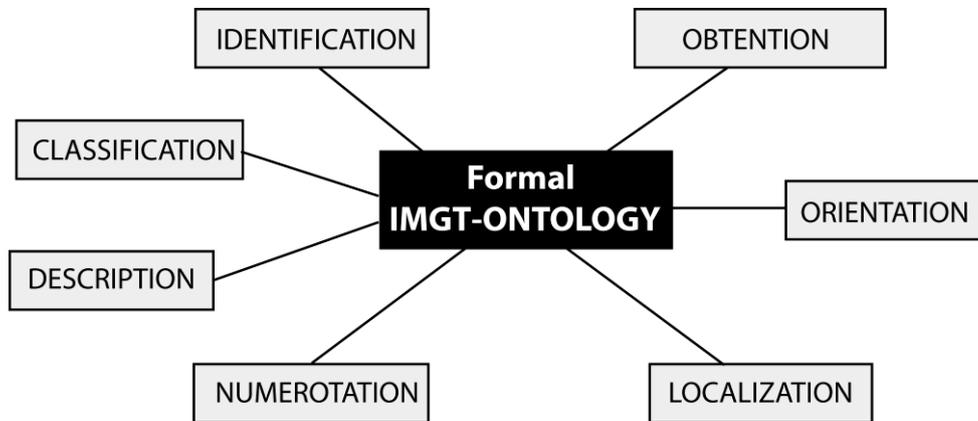


Figure 5. Axiomes de l'IMGT-ONTOLOGY formelle (Formal IMGT-ONTOLOGY) ou IMGT-Kaleidoscope.

Utilisation des concepts d'IMGT-ONTOLOGY

Les concepts générés par les axiomes d'IMGT-ONTOLOGY sont utilisés pour l'expertise et la gestion des données IMGT®. Ces concepts permettent ainsi l'annotation standardisée des séquences (Figure 6), la gestion générique 'des séquences aux structures 3D' (Figure 7) et la gestion générique des domaines quel que soit leur type (V, C et G) (Figure 8).

| Key | Location/Qualifiers |
|--------------------|--|
| L-V-D-J-C-SEQUENCE | <1..356> /partial /db_xref="taxon:9606" /cell_type="mononuclear cells" /isolation_source="chronic lymphocy patient" /isolate="P103H" /IMGT_note="automatically annotated with IMGT tools" /tissue_type="peripheral blood" /organism="Homo sapiens" 1..356 |
| V-D-J-REGION | /partial /protein_id="AAZ08839.1" /translation="QVQLQQWAGLLKPSSETLSLTCVAVYGESFSGYYWVIRQPP GKGLEWIGEINHSGSTNYNPSLKSRTVISVDTSKNQFSLKLTSTVTAADTAVYYCA RGYPDTPVVRYYYYGMDVWGQ" |
| V-REGION | 1..293 /allele="IGHV4-34*01, putative" /gene="IGHV4-34" /CDR_length="{0..7,20}" /putative_limit="3' side" /translation="QVQLQQWAGLLKPSSETLSLTCVAVYGESFSGYYWVIRQPP GKGLEWIGEINHSGSTNYNPSLKSRTVISVDTSKNQFSLKLTSTVTAADTAVYYCA R" |
| FR1-IMGT | 1..75 /AA_IMGT="1 to 26, AA 10 is missing" /translation="QVQLQQWAGLLKPSSETLSLTCVAVY" |
| 1st-CYS | 64..66 |
| CDR1-IMGT | 76..99 /AA_IMGT="27 to 34" /translation="GESFSGYY" |
| FR2-IMGT | 100..150 /AA_IMGT="39 to 55" /translation="WVIRQPPGKLEWIGE" |
| CONSERVED-TRP | 106..108 |
| CDR2-IMGT | 151..171 |

Figure 6. Annotation standardisée des séquences utilisant les concepts d'IMGT-ONTOLOGY générés à partir des axiomes. Extrait d'un fichier à plat d'IMGT/LIGM-DB annoté par l'outil interne IMGT/Automat.

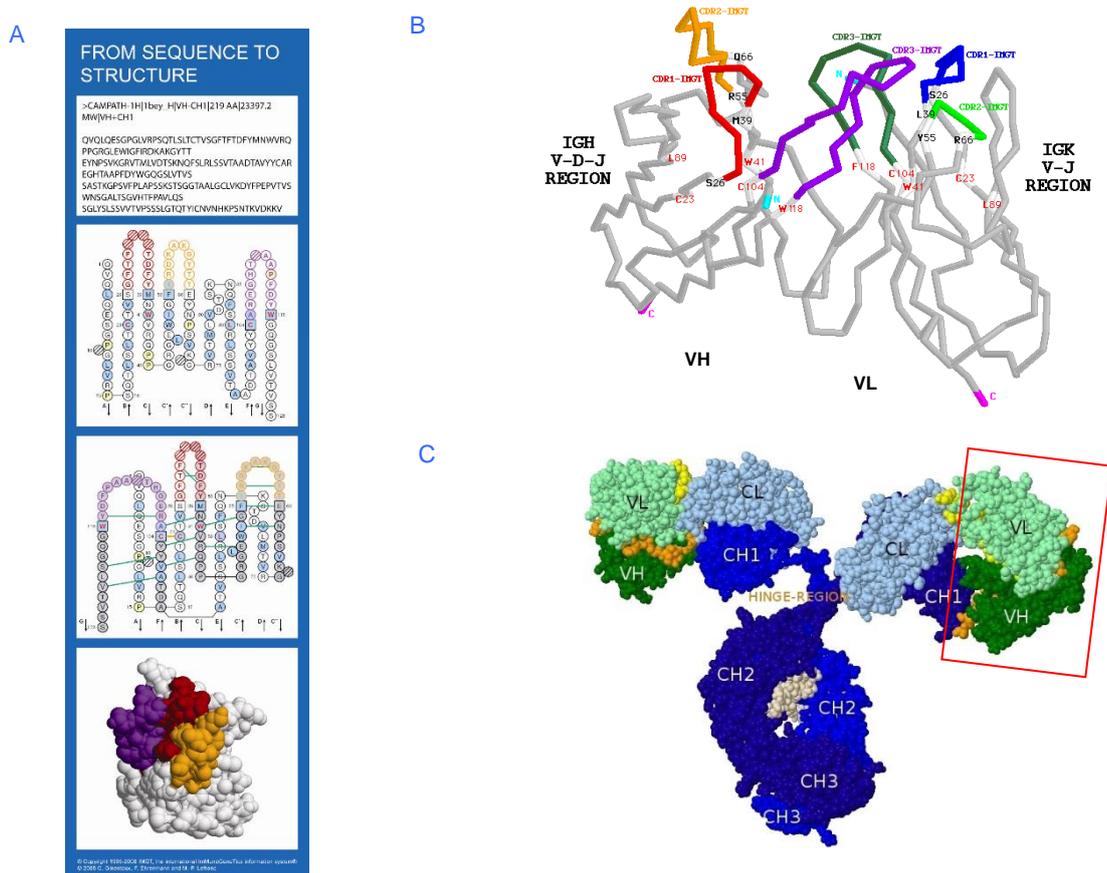


Figure 7. Gestion générique des séquences aux structures 3D. Les mêmes concepts sont utilisés qu'il s'agisse de la description (A) d'un domaine en tant que séquence, représentation (2D) (IMGT Collier de Perles) ou structure 3D, (B) de deux domaines constituant le site de reconnaissance d'un récepteur d'antigène ou (C) de la structure 3D complète d'un anticorps.

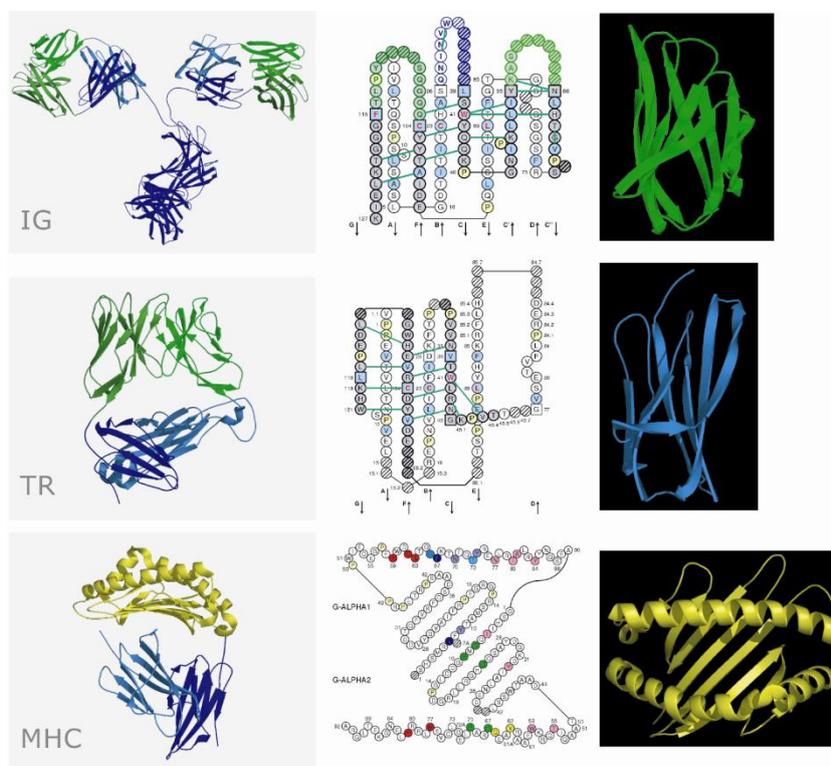


Figure 8. Gestion générique des domaines. Représentation d'une molécule d'IG, TR et MHC, avec les 'IMGT Colliers de Perles' correspondant à un domaine de type variable (V), un domaine de type constant (C) et deux domaines de type 'groove' (G) (sillon en français). Les mêmes concepts sont aujourd'hui utilisés pour les domaines V-like, C-like et G-like des protéines appartenant aux superfamilles IgSF et MhcSF.

1.4. Composants d'IMGT®

Construit sur les concepts d'IMGT-ONTOLOGY, IMGT® est rapidement devenu le système d'information international de référence en immunogénétique et immunoinformatique. IMGT® comprend, en octobre 2009 (Figure 9):

- six bases de données : de séquences (données génétiques), de gènes (données génomiques) et de structures 2D et 3D (données structurales) (représentées par des cylindres),
 - une quinzaine d'outils interactifs en ligne (représentées par des rectangles),
 - des ressources Web (IMGT Marie-Paule page, composée de plus de 10.000 pages Web).
- La page d'accueil d'IMGT®, <http://www.imgt.org> (Montpellier, France) (Annexe 1), permet l'accès à l'ensemble des données immunogénétiques.

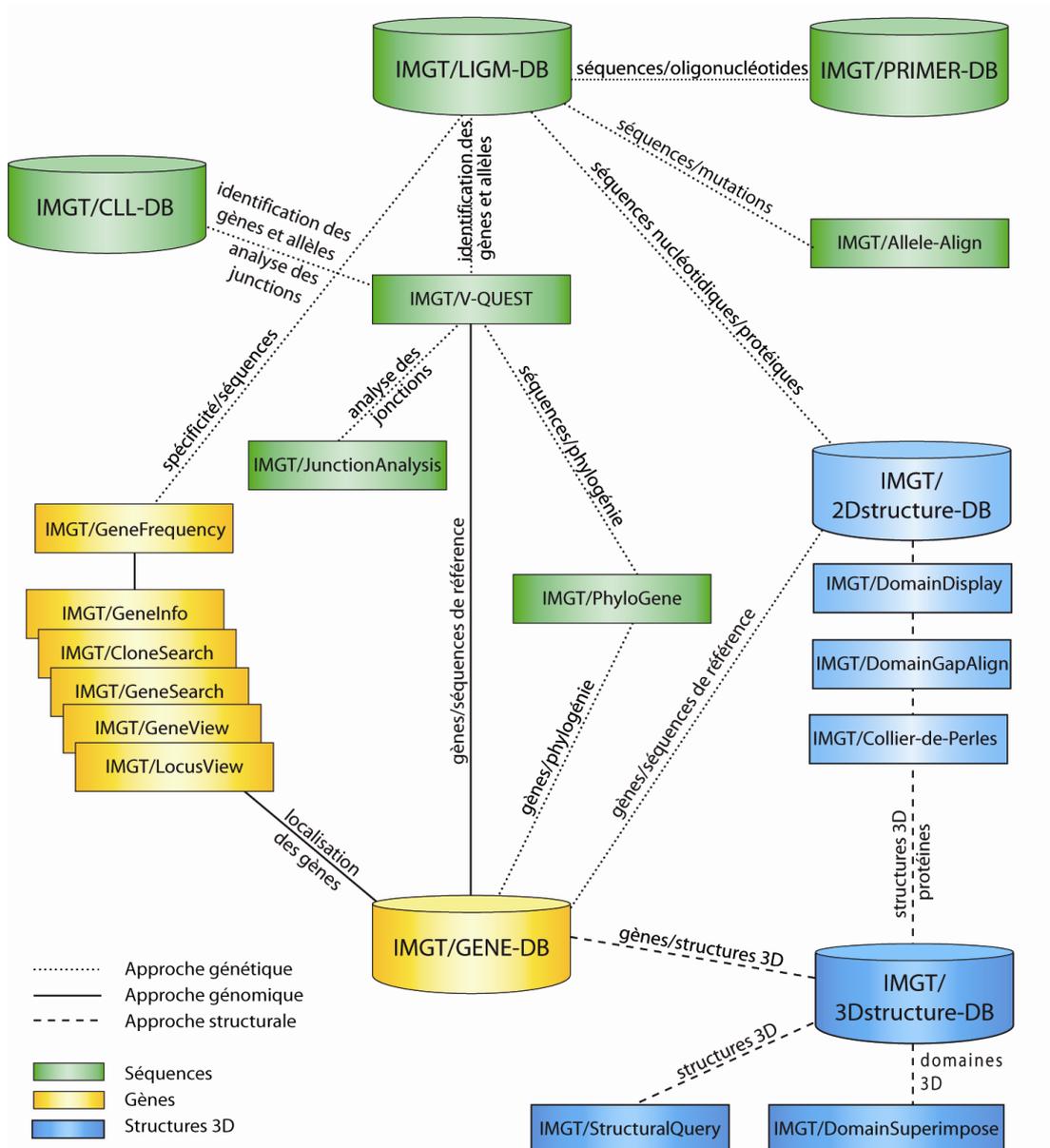


Figure 9. IMGT®, the international ImMunoGeneTics information system® (<http://www.imgt.org>).

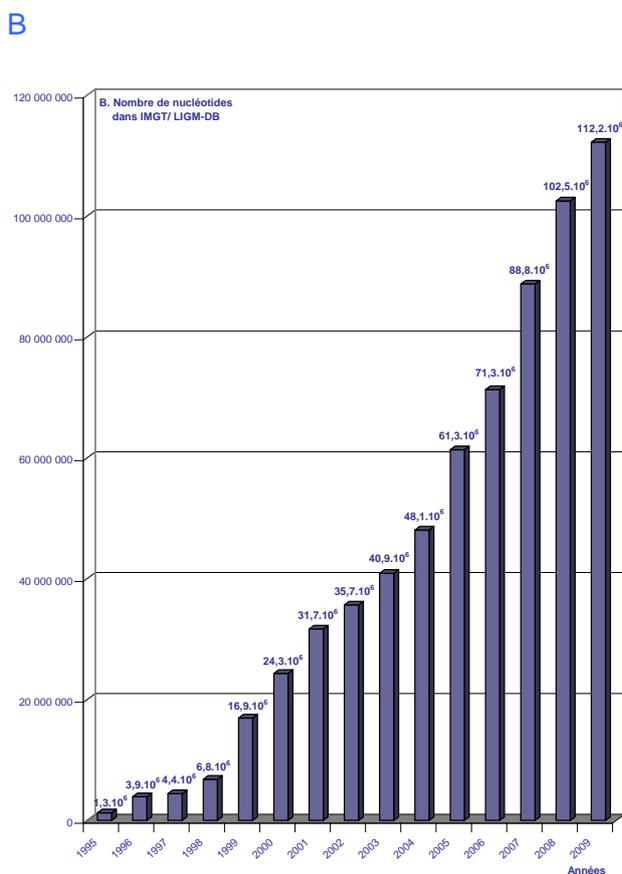
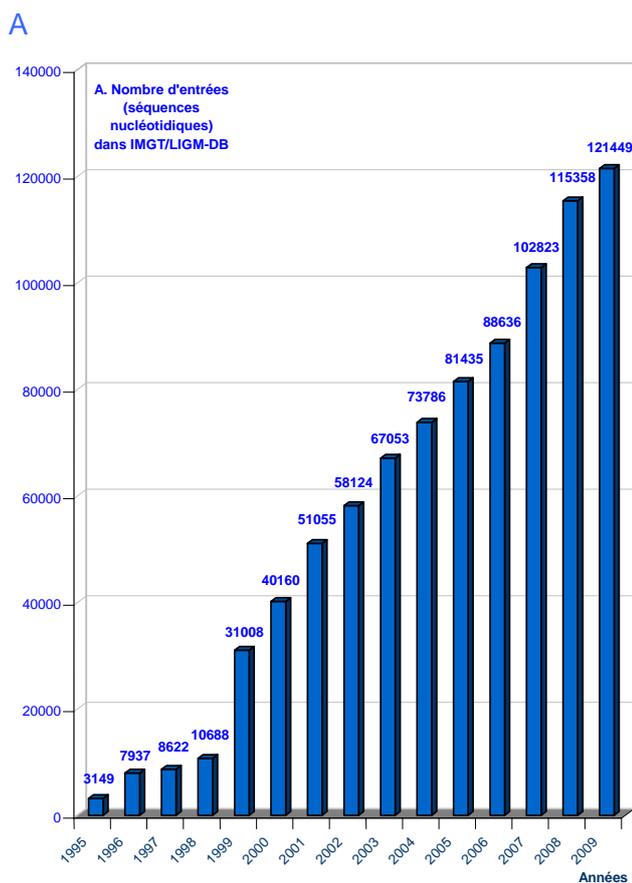


Figure 10. Evolution d'IMGT/LIGM-DB de 1995 à 2009. (A) Nombre d'entrées (séquences nucléotidiques d'IG et de TR) par an. (B) Nombre de nucléotides par an. L'augmentation entre 1998 et 1999 correspond à l'entrée de séquences non gérées auparavant par LIGM.

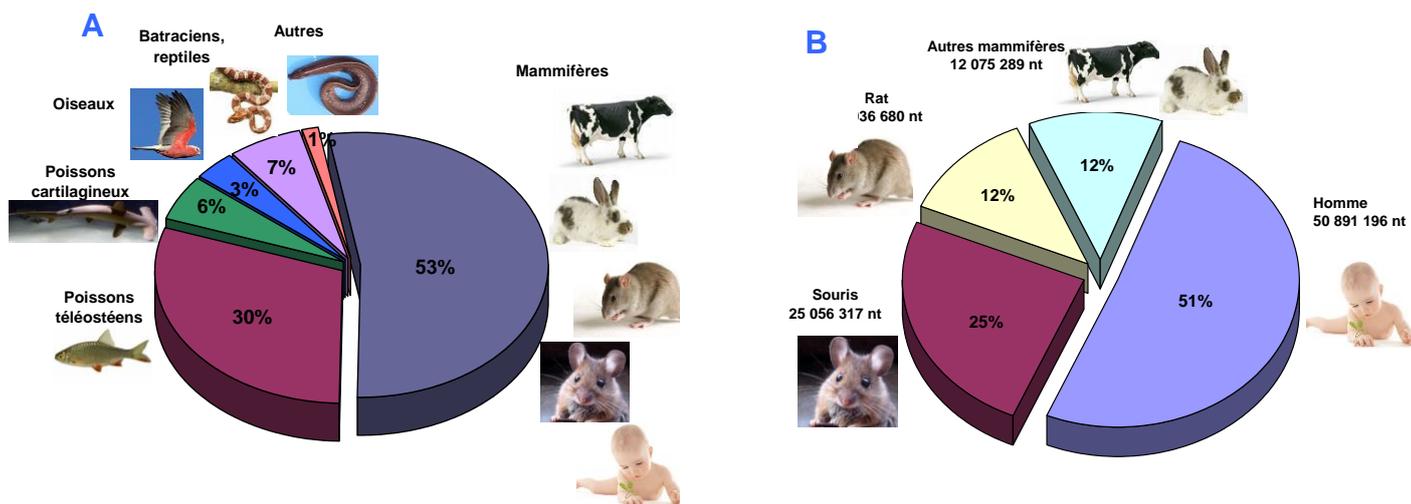


Figure 11. Répartition des données dans IMGT/LIGM-DB. (A) Nombre d'entrées (séquences nucléotidiques d'IG et TR) selon les grandes divisions de vertébrés. (B) Nombre de nucléotides (nt) entre l'homme, la souris, le rat et les autres mammifères.

1.4.1. Les bases de données IMGT®

IMGT® comprend six bases de données:

- **IMGT/LIGM-DB.** Créée à Montpellier en 1989, et sur le Web depuis juillet 1995, IMGT/LIGM-DB est la base de données IMGT® des séquences nucléotidiques des IG et TR de l'homme et des autres espèces de vertébrés, avec traduction des séquences entièrement annotées. En octobre 2009, IMGT/LIGM-DB contient 137.963 séquences de 236 espèces (Figures 10 et 11).

- **IMGT/PRIMER-DB.** Sur le Web depuis février 2002, IMGT/PRIMER-DB est la base de données IMGT® des oligonucléotides (primers ou amorces). Ces oligonucléotides servent d'amorce pour l'amplification génique par PCR des gènes IG et TR, lors de la construction de banques combinatoires (*single chain Fragment variable* (scFv), *Fragment antigen binding* (Fab), *phage display*) et dans les biotechnologies par microarray. En octobre 2009, IMGT/PRIMER-DB contient 1864 entrées.

- **IMGT/GENE-DB.** Sur le Web depuis janvier 2003, IMGT/GENE-DB est la base de données IMGT® de gènes. IMGT/GENE-DB contient des informations sur tous les gènes et allèles des IG et TR de l'homme et de la souris. Les gènes et allèles d'autres espèces (rat, lapin, etc.) sont ajoutés progressivement en fonction de l'avancée des annotations des locus des IG et TR par LIGM (Figure 12).

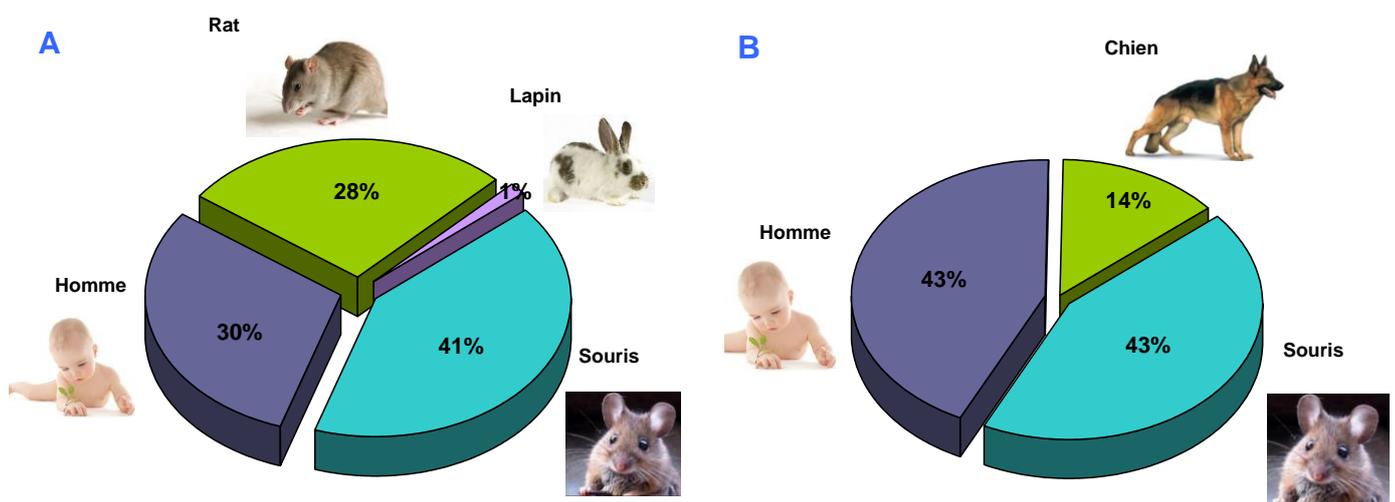


Figure 12. Répartition du nombre d'entrées (gènes) dans IMGT/GENE-DB selon les espèces et le type de récepteur. (A) Gènes d'immunoglobulines (IG). (B) Gènes de récepteurs T (TR).

Des liens directs réciproques existent entre IMGT/GENE-DB et Entrez Gene au *National Center for Biotechnology Information* NCBI (ceci est le seul exemple de liens directs effectués par Entrez Gene au NCBI sur une base de données externe). IMGT/GENE-DB est la référence internationale pour la nomenclature des gènes d'IG et TR et travaille en étroite collaboration avec le *Human Genome Organisation* (HUGO) *Nomenclature Committee* (HGNC). En octobre 2009, IMGT/GENE-DB contient 5.363 entrées (2.168 gènes et 3.195 allèles).

- **IMGT/3Dstructure-DB.** Sur le Web depuis novembre 2001, IMGT/3Dstructure-DB est la base de données IMGT® de structures 3D. Ces structures comprennent celles des IG, TR, MHC, RPI et de leurs complexes: peptide/MHC (pMHC), TR/pMHC, IG/antigène (IG/Ag) (Figure 13). En octobre 2009, IMGT/3Dstructure-DB contient 1.702 fichiers de coordonnées atomiques.

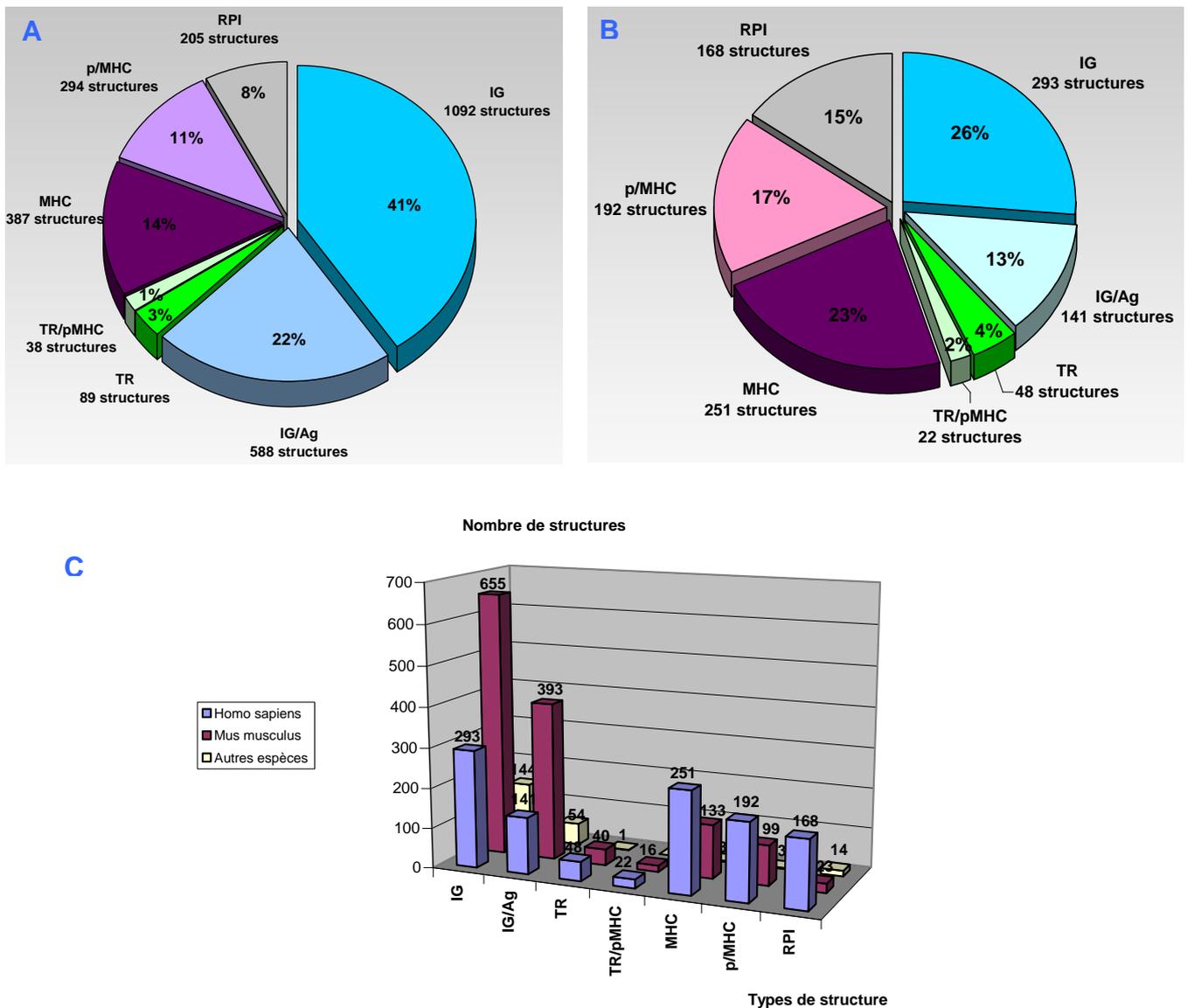


Figure 13. Répartition du nombre d'entrées (structures 3D) dans IMGT/3Dstructure-DB selon le type de structures et leur origine. (A) Toutes espèces représentées dans la base. (B) Pour l'homme (*Homo sapiens*). (C) Par comparaison entre l'homme, la souris (*Mus musculus*) et les autres espèces.

La plupart des données structurales sont de l'ordre de dizaines de milliers (Tableau 1), cependant certaines données telles que les *IMGT Residue @Position cards* atteignent déjà un million et demi, et les nombres augmentent très rapidement en fonction du nombre d'entrées dans IMGT/3Dstructure-DB, lequel a triplé de 2004 à 2009.

Tableau 1. Exemples de type de données gérées dans IMGT/3Dstructure-DB.

| Type de données | Nombre |
|---|-----------|
| IMGT Collier de Perles: | |
| sur un plan | 9 000 |
| sur deux plans, avec liaisons hydrogène | 22 000 |
| Analyse des contacts: | |
| Sites de contact entre peptide et MHC | 15 000 |
| Paires de contacts entre domaines | 45 000 |
| IMGT Residue@Position cards | 1 500 000 |
| Résultats d'IMGT/DomainGapAlign | 9 000 |

- [IMGT/2Dstructure-DB](#). Créée en novembre 2008, IGMT/2Dstructure-DB est la base de données IGMT® de structures 2D. Cette base est actuellement gérée, au niveau informatique, en tant que section d'IMGT/3Dstructure-DB. Elle contient les séquences protéiques des anticorps monoclonaux et des protéines de fusion pour les applications immunes (FPIA pour *Fusion Protein for Immune Applications*). Ces séquences sont soumises par les sociétés pharmaceutiques au programme de l'Organisation Mondiale de la Santé (OMS) pour l'obtention d'un nom générique ou dénomination commune internationale (DCI) (en anglais, *World Health Organization/International Nonproprietary Name (WHO/INN) programme*).

- [IMGT/CLL-DB](#). Créée en novembre 2007, IGMT/CLL-DB contient les données qui résultent des analyses d'IMGT/V-QUEST des séquences réarrangées d'IG de patients atteints de leucémie lymphoïde chronique (LLC, en anglais *Chronic Lymphocytic Leukemia* ou CLL), ainsi que les données cliniques de ces patients d'intérêt pour l'interprétation des séquences. La base IGMT/CLL-DB contient des données non encore publiées, son accès est donc sécurisé et limité aux membres du groupe IGMT/CLL-DB (IMGT/CLL-DB Group). L'entrée des séquences se fait grâce à l'outil IGMT/V-QUEST, décrit dans la partie 1.4.2.

Les caractéristiques des bases de données IGMT/LIGM-DB, IGMT/PRIMER-DB, IGMT/GENE-DB et IGMT/3Dstructure-DB, en réponse à l'enquête ELIXIR (octobre 2008) sont présentées dans l'annexe 2.

1.4.2. Les outils en ligne IGMT®

IMGT® comprend quinze outils interactifs en ligne pour l'analyse:

- [des séquences:](#)

IMGT/V-QUEST (V-QUERy and STandardization), l'outil le plus utilisé d'IMGT®, tant sur le plan national qu'international, fournit une analyse minutieuse et novatrice des séquences réarrangées V-J et V-D-J des IG et TR qui résultent des réarrangements des gènes variables (V), de jonction (J) et de diversité (D) lors de la synthèse des IG et des TR. De plus IGMT/V-QUEST intègre IGMT/JunctionAnalysis pour l'analyse de leurs jonctions.

IMGT/Allele-Align permet la comparaison de deux allèles.

IMGT/PhyloGene calcule et représente les arbres phylogénétiques des séquences nucléotidiques d'IG et de TR.

- [des gènes:](#)

IMGT/GeneView, IGMT/GeneSearch, IGMT/LocusView et IGMT/CloneSearch permettent de localiser un gène donné dans un locus d'IG ou de TR, de visualiser un locus et de rechercher les clones correspondants. IGMT/GenInfo, développé par le Laboratoire TIMC-ICH de Grenoble, permet d'obtenir des informations graphiques sur les gènes TR de l'homme et de la souris.

IMGT/GeneFrequency permet de représenter, sous forme d'histogrammes, la fréquence des gènes utilisés dans les séquences réarrangées d'IMGT/LIGM-DB, selon des critères d'espèces, de locus, de spécificités.

- [des domaines protéiques:](#)

IMGT/DomainDisplay affiche les domaines protéiques de type V (variable), de type C (constant) et de type G (groove, sillon en français) selon la numérotation unique IGMT.

IMGT/DomainGapAlign permet d'aligner une séquence protéique de domaine avec la séquence de référence la plus proche et de créer les «gaps» selon la numérotation unique IGMT.

IMGT/Collier-de-Perles fournit la représentation standardisée en deux dimensions ou «IMGT Colliers de Perles» des domaines de type V, C et G des séquences protéiques.

- [des données structurales:](#)

IMGT/StructuralQuery permet de faire des requêtes dans IGMT/3Dstructure-DB, basées sur une description structurale.

IMGT/DomainSuperimpose superpose deux structures 3D de domaines de type V, C ou G, contenus dans IGMT/3Dstructure-DB.

1.4.3. Les ressources Web IMGT®

Les ressources Web IMGT® comprennent plus de 10.000 pages HTML, organisées en plusieurs sections:

- 'IMGT Scientific chart' fournit le vocabulaire contrôlé et les règles d'annotation, basés sur les concepts d'IMGT-ONTOLOGY.
- 'IMGT Répertoire' fournit une interface conviviale aux données soigneusement expertisées qui concernent le génome, le protéome, le polymorphisme et la structure des IG et TR de l'homme et des autres vertébrés.
- 'IMGT Index' et 'IMGT Education' fournissent des ressources pédagogiques en immunologie et immunogénétique utiles pour les chercheurs, les enseignants et les étudiants.
- 'IMGT Bloc-notes' fournit de très nombreux liens hypertexte sur les serveurs Web spécialisés en immunologie, génétique, biologie moléculaire et bioinformatique.
- 'IMGT Immunoinformatics page' fournit les liens sur les bases de données, outils en ligne et ressources Web en immunoinformatique.
- 'IMGT Medical page' facilite l'accès aux données IMGT® d'intérêt médical.
- 'IMGT Veterinary page' fournit aux vétérinaires les informations d'IMGT® sur les génomes, séquences et structures 3D des IG et TR des espèces animales (domestiques, sauvages, ou utilisées en tant que modèle animal).
- 'IMGT Biotechnology page' fournit un ensemble d'informations sur l'ingénierie des anticorps (*phage display*), les anticorps de camélidés, l'humanisation des anticorps monoclonaux thérapeutiques.

1.5. Utilisateurs d'IMGT®

IMGT® est reconnu internationalement pour la richesse et la qualité de ses données scientifiques et son interface conviviale. Le nombre d'équipes utilisatrices d'IMGT® est supérieur à 80.000. Ces équipes académiques et industrielles, appartiennent à des domaines d'activité de recherche professionnelle variés et sont réparties dans le monde entier, à parts égales entre les Etats-Unis et Canada, l'Europe et le reste du monde (Figure 14). Ceci représente plus de 150.000 requêtes par mois, et un nombre de sessions de travail sans cesse croissant:

240.000 (2007)

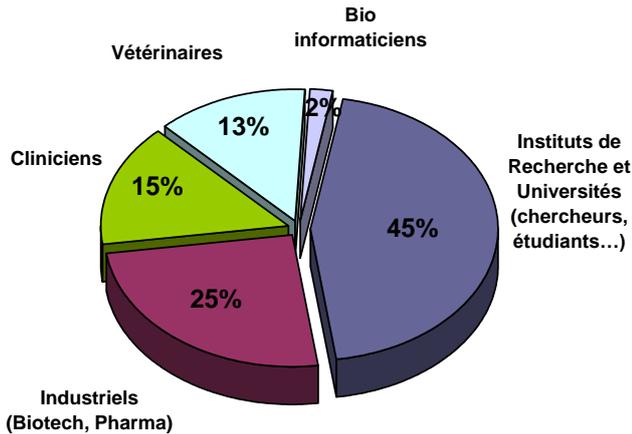
260.000 (2008)

280.000 (2009, prévu sur les chiffres au 15/10/2009).

IMGT® est utilisé par des chercheurs d'équipes académiques et industrielles dans de multiples domaines de recherche:

- (1) recherche fondamentale,
- (2) recherche médicale (analyse des répertoires des anticorps et des sites de reconnaissance des récepteurs T dans les réponses immunitaires normales contre les agents infectieux et les cellules malignes, et lors des réponses anormales au cours de maladies autoimmunes et les syndromes prolifératifs (leucémies, lymphomes, myélomes)),
- (3) recherche vétérinaire (étude des répertoires des IG et TR dans les espèces domestiques et sauvages),
- (4) recherche génomique (étude de la diversité et de l'évolution des gènes de la réponse immunitaire adaptative),
- (5) recherche en biologie structurale (évolution des domaines des protéines des superfamilles IgSF et MhcSF des vertébrés et invertébrés),
- (6) biotechnologies relatives aux projets de l'*Human Proteome Organisation* (HUPO) et à l'ingénierie des anticorps (scFv, Fab, banques combinatoires, *phage display*, anticorps chimériques, humanisés et humains),
- (7) recherche pour le diagnostic, le pronostic et le suivi thérapeutique des leucémies, lymphomes et myélomes (identification du ou des clone(s) malin(s) et évaluation de la maladie résiduelle),
- (8) approches thérapeutiques (greffes, immunothérapie, vaccinologie).

A



B

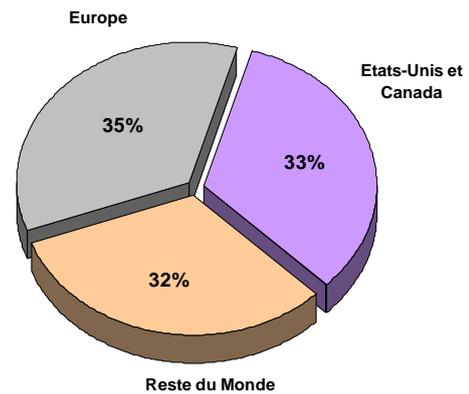


Figure 14. Répartition des utilisateurs d'IMG T®. (A) En fonction de leur activité de recherche professionnelle. (B) En fonction de leur localisation géographique.

2. Caractérisation du système de données de recherche IMG T® (objectifs, historique, contexte)

2.1. Objectifs

Un organisme humain synthétise plus de 2.10^{12} anticorps et plus de 2.10^{12} récepteurs T différents pour faire face aux agressions extérieures des virus, bactéries, parasites et pour lutter contre ses propres cellules malignes. C'est par des mécanismes complexes de réarrangements au niveau de l'ADN et, dans le cas des IG, de mutations somatiques qu'un nombre aussi important de récepteurs d'antigènes peut être obtenu. En raison de la complexité de leur synthèse et de leur incroyable diversité, les molécules du système immunitaire, et en particulier les anticorps, doivent être traitées d'une façon spécifique. Le système d'information IMG T® a été créé par Marie-Paule Lefranc, à Montpellier, en 1989, à la suite du Congrès «Human Gene Mapping 10» à New Haven (Etats-Unis), avec pour objectif de gérer cette complexité des gènes, séquences et structures des IG et TR. Ces «premiers» objectifs sont décrits dans 'Pôle Biologie-Informatique' de juillet 1990.

Extrait de 'Pôle Biologie-Informatique'. Une structure de travail dans le sud de la France, entre le Centre National Universitaire Sud de Calcul, les laboratoires d'informatique, et la communauté scientifique 'Biologie-Santé'. Juillet 1990. C3NI, projet joint CNUSC-IBM

Titre du projet: « Constitution d'une banque de données et analyse des régions variables des immunoglobulines et récepteurs T ».

Objectifs:

1. Constitution d'une banque de données 'SEQUENCES' et son utilisation pour:
 - a. l'analyse des régions hypervariables et des mutations somatiques
 - b. l'étude des régions de jonction et de reconnaissance des antigènes
 - c. l'assignement des gènes variables à des sous-groupes bien définis
 - d. Une proposition de nomenclature des gènes variables pour Human Gene Mapping 11 (1991)
 - e. la comparaison des séquences entre espèces
 - f. des études de modélisation
2. Constitution d'une banque de données 'SONDES' et 'RFLP' des gènes variables des immunoglobulines et des récepteurs T.

Ces objectifs initiaux ont été atteints avec la création de IMG T/LIGM-DB et d'IMG T Répertoire. Depuis, IMG T/LIGM-DB est en développement constant (avec des données de 236 espèces, en octobre 2009).

A ces objectifs se sont ajoutées l'analyse et la gestion:

- des données de séquences, gènes et polymorphismes génétiques, liées au séquençage de nouveaux génomes,

- des données de structures 2D et 3D, issues de la recherche et des développements biotechnologiques et pharmaceutiques (analyse des spécificités des récepteurs d'antigènes, ingénierie et humanisation des anticorps, etc.).
- des données de répertoires des IG et TR exprimés dans des conditions physiologiques normales et pathologiques: infections virales, tumeurs, leucémies, lymphomes, maladies autoimmunes etc.
- et plus récemment des données de séquences réarrangées des IG et des TR, liées aux nouvelles techniques de séquençage haut débit (Next-Gen sequencing).

L'analyse des données, la standardisation de leur annotation et l'évolution constante du système IMGT® ont été possibles grâce à la création d'IMGT-ONTOLOGY, la première ontologie dans le domaine. Les standards basés sur l'IMGT Scientific chart' et les concepts d'IMGT-ONTOLOGY ont été approuvés par l'Organisation Mondiale de la Santé (rapports WHO-IUIS, 2007, 2008).

IMGT® est aujourd'hui le système de référence international en immunogénétique et immunoinformatique, utilisé en recherche fondamentale, médicale et biotechnologique.

2.2. Historique

IMGT®, the international ImMunoGeneTics information system®, <http://www.imgt.org>, a été créé en 1989, à Montpellier par Marie-Paule Lefranc (Université Montpellier 2 et CNRS).

1989-1994 :

- En 1989, un premier prototype du système d'information en immunogénétique (LIGM-DB) est développé pour les séquences d'immunoglobulines (IG) et récepteurs T (TR) par le Laboratoire d'ImmunoGénétique Moléculaire (LIGM, CNRS et Université Montpellier 2, Montpellier), en collaboration avec le Centre National Universitaire Sud de Calcul (CNUSC, Montpellier, Sophie Creuzet et Denys Chaume) et le Laboratoire d'Informatique, de Robotique et de Micro-électronique de Montpellier (LIRMM, CNRS et Université Montpellier 2, Montpellier, Isabelle Mougenot et Patrice Déhais).

- Dès 1990, une collaboration est établie avec EMBL (Graham Cameron).

- En 1992, la base de données devient véritablement européenne avec la collaboration de l'European Molecular Biology Laboratory (EMBL, Rainer Fuchs, l'Université de Cologne (Institute for Genetics, Werner Müller), l'Imperial Cancer Research Foundation (ICRF, Julia Bodmer) et l'obtention d'un premier financement EU de l'Union Européenne BIOMED1 (BIOCT930038).

- En 1993, une démonstration à Kobe (Japon) au congrès 'Human Genome Mapping Workshop 93' (HGM 93), 14-17 novembre 1993 est possible grâce à la connexion de 36 nœuds par France Télécom entre la France et le Japon (seule démonstration en ligne du Congrès).

- En 1994, une démonstration à Barcelone (Espagne) au congrès '12th European Immunology Meeting, EFIS 94', 14-17 juin 1994 est réalisée grâce au transport du serveur par camion à Barcelone.

1995-à ce jour:

- En 1995, a lieu la première démonstration en ligne sur Internet d'IMGT® à San Francisco (Etats-Unis) au congrès '9th International Congress of Immunology', 22-29 juillet 1995. Il s'agissait de la première connexion sur Internet dans un congrès international d'Immunologie. Bien que ce congrès regroupait 9.000 participants dont 2.500 exposants, IMGT® était l'unique connexion sur Internet du congrès.

- En 1996, une démonstration en ligne sur Internet a lieu à Science en Fête à Montpellier 11-13 octobre 1996. Il s'agissait de la première connexion sur Internet réalisée sur la Place de la Comédie à Montpellier. Cette démonstration a été renouvelée à Science en Fête en 1997 (10-12 octobre), et en 1999 (21-23 octobre), Fête de la Science en 2000 (16-22 octobre).

- En 1998, une démonstration en ligne a eu lieu à Hawaï (Etats-Unis) au congrès 'Pacific Symposium on Biocomputing', 4-9 janvier 1998.

- En 1999, pour ses 10 ans, IMGT® est en couverture de Nucleic Acids research (Database Issue).
- En 2001, mise en ligne de la base de données [IMGT/3Dstructure-DB](#) (novembre 2001).
- En 2002, mise en ligne de la base de données [IMGT/Primer-DB](#) (février 2002).
- En 2003, mise en ligne de la base de données [IMGT/GENE-DB](#) (janvier 2003).
- En 2007, création de la base de données IMGT/CLL-DB (novembre 2007).
- En 2008, mise en ligne de IMGT/2Dstructure-DB en tant que section de la base de données [IMGT/3Dstructure-DB](#) (novembre 2008).

2.3. Contexte d'IMGT® au niveau des flux de données et des collaborations

2.3.1. Contexte d'IMGT® et flux de données entrantes et sortantes

IMGT® réceptionne des données de l'extérieur et les met à disposition après expertise et annotation, sur le site IMGT® (Figure 15). Il s'agit de:

- données nucléotidiques:

- Réception par LIGM des fichiers à plat (*flat files*) des séquences nucléotidiques en provenance d'EMBL-Bank de l'European Bioinformatics Institute (EBI) <http://www.ebi.ac.uk/>.
- Mise à disposition par LIGM des fichiers à plat annotés d'IMGT/LIGM-DB par *File Transfer Protocol* (FTP) sur le serveur FTP du CINES.
- Les fichiers à plat annotés d'IMGT/LIGM-DB sur le serveur du CINES permettent les mises à jour des serveurs *Sequence Retrieval System* (SRS) en Europe (en particulier celui d'EBI), aux Etats-Unis et au Japon et le serveur *All-round Retrieval of Sequence and Annotation* (ARSA) au *DNA Data Bank of Japan* (DDBJ) (Japon).

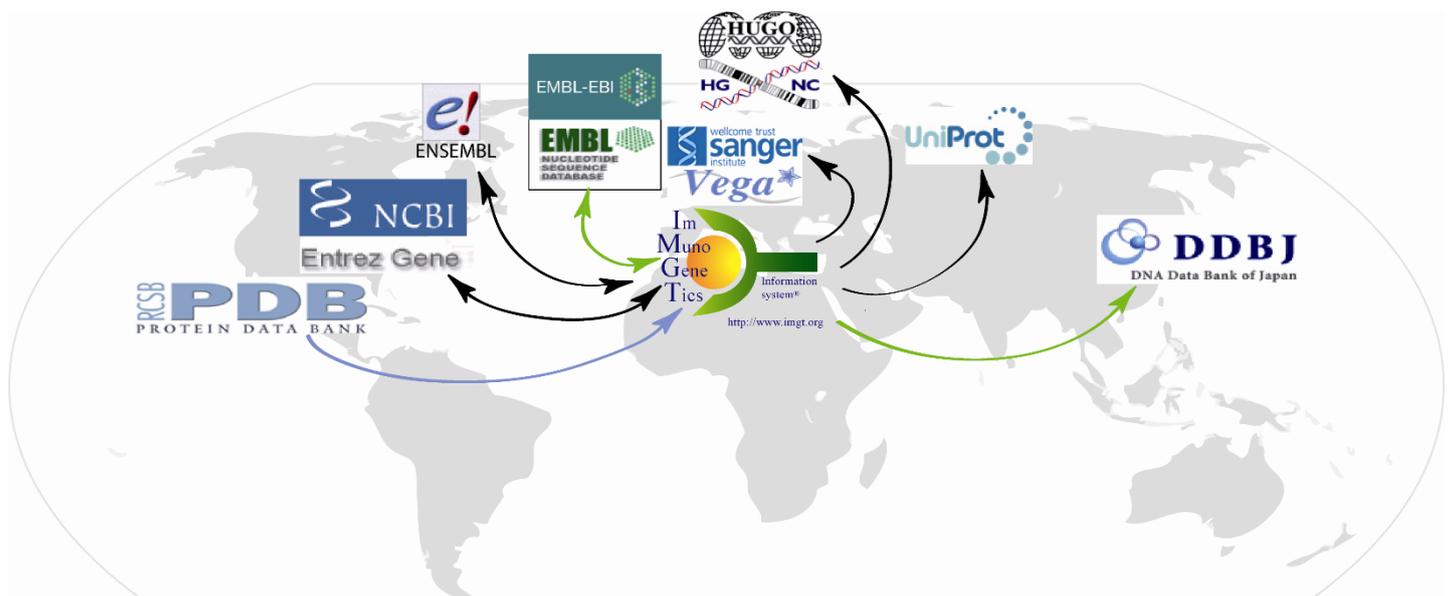


Figure 15. Contexte d'IMGT® au niveau des flux des données entrantes et sortantes. Les flux de données sont représentés par des flèches colorées: données nucléotidiques en vert, données génomiques en noir et données structurales en bleu.

- données protéiques:

- Envoi par LIGM des séquences protéiques de référence des gènes constants d'IG et TR humains à UniProt (serveur ExPASy au Swiss Institute of Bioinformatics) par courriel.

- données génomiques:

- Récupération par LIGM des séquences génomiques sur les serveurs Ensembl à l'EBI et Entrez Gene au NCBI.

- Transfert par LIGM de la nomenclature des gènes des IG et TR au HGNC, à *Vertebrate Genome Annotation* (Vega) au Wellcome Trust Sanger Institute, à Ensembl à l'EBI et à Entrez Gene au NCBI
- Etablissement des références croisées.
- Les séquences de référence et positions des gènes IMGT sur les chromosomes sont également disponibles sur les pistes «IMGT Genes» mises en place par LIGM sur le serveur *Distributed Annotation System* (DAS) d'Ensembl à l'EBI.

- *données structurales:*

- Réception par LIGM des fichiers à plat des structures 3D en provenance de *Research Collaboratory for Structural Bioinformatics* (RCSB) *Protein Data Bank* (PDB) <http://www.rcsb.org/> par FTP.
- Mise à disposition par LIGM des structures annotées dans IMGT/3Dstructure-DB.

2.3.2. Relations entre IMGT® et ses différents collaborateurs et utilisateurs

Les relations entre IMGT® et ses différents collaborateurs et utilisateurs sont illustrées dans la figure 16.

Sur cette figure, sont précisés :

- *le type de données entrantes et sortantes*

(s): séquence.

(d): autres données. Il s'agit des données autres que les séquences.

- *le type de transfert*

(C): courriel,

(F): FTP,

(W): Web

(S): Scp (pour *SSH CoPy*), un protocole de transfert de fichiers entre machines UNIX via le protocole sécurisé *Secure Shell* (SSH).

- *le format des données*

(1): fasta,

(2): flat file (fichier à plat),

(3): excel,

(4): texte,

(5): format pdb,

(6): autres sources servlet.

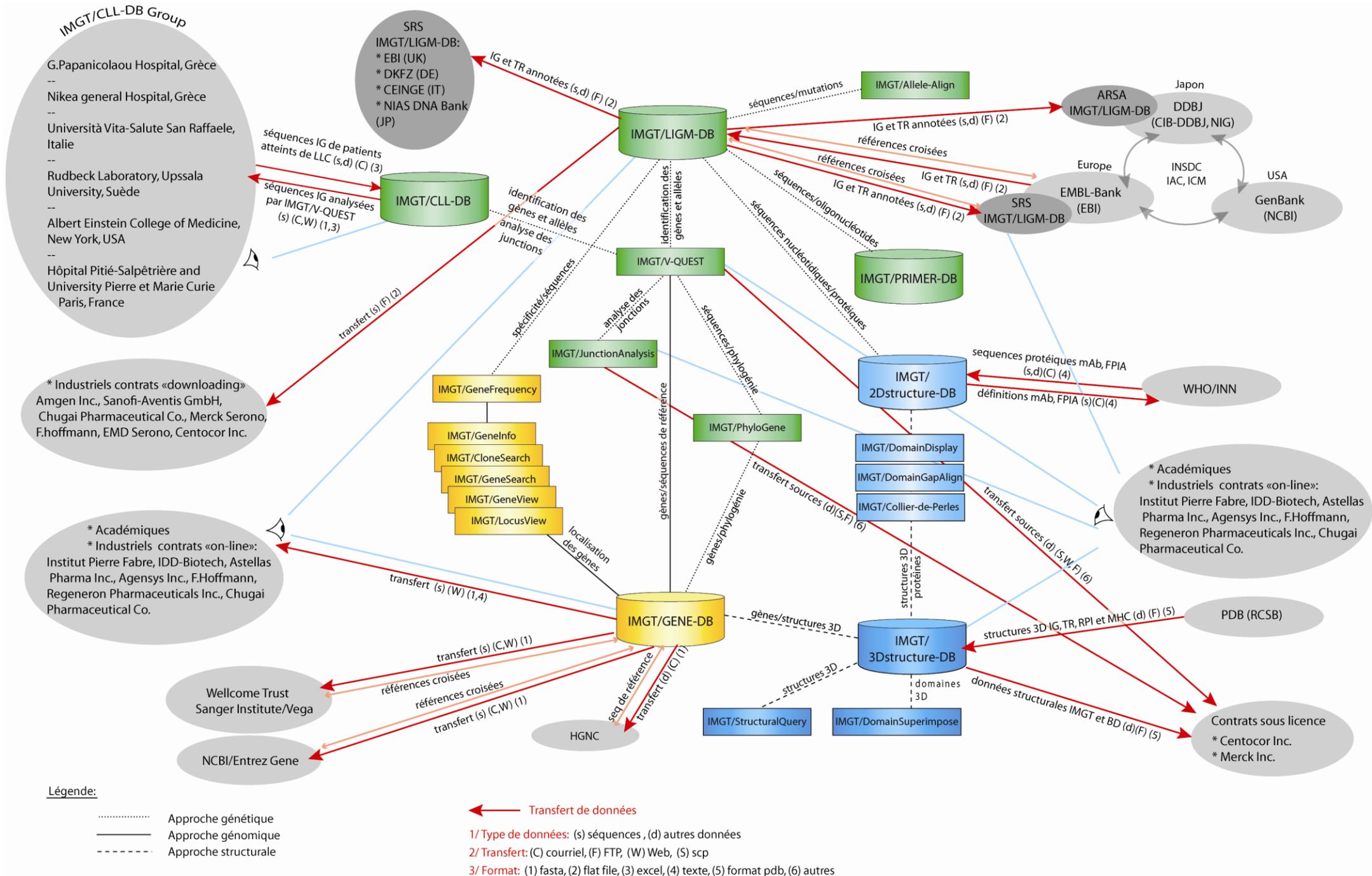


Figure 16. Contexte d'IMGT® au niveau des flux de données et relations entre IMGT® et ses différents collaborateurs et utilisateurs.

3. Fonctionnalités couvertes par le système de données de recherche IMGT® (production, gestion, archivage, préservation, mise à disposition, diffusion, analyse, exploitation...)

3.1. Production

La production IMGT® est représentée par les données scientifiques expérimentales (séquences, gènes et structures 3D) en immunogénétique et en immunoinformatique, après expertise par IMGT®.

Les données sont annotées selon les standards IMGT® en terme de classification (nomenclature), de description (labels) et de numérotation ('IMGT unique numbering') et sont ensuite mises à disposition sur le site <http://www.imgt.org>, dans les bases de données IMGT®, les outils créés et produits par IMGT®, et les ressources Web IMGT® (10.000 pages Web, données pédagogiques, pages fixes, expertise scientifique...)

3.1.1. IMGT/LIGM-DB: production de séquences nucléotidiques annotées

Les données récupérées par IMGT® à partir d'EMBL-Bank (EBI) représentent la totalité des séquences IG et TR soumises aux bases de données généralistes. En effet, le contenu des 3 bases de données généralistes EMBL-Bank (EBI), GenBank (NCBI) et DNA Data Bank of Japan (DDBJ) est identique grâce aux échanges quotidiens des entrées, organisé dans le cadre de l'*International Nucleotide Sequence Database Collaboration* (INSDC) *International Advisory Committee* (IAC) et *International Collaborative Meeting* (ICM).

Après un filtre basé sur des mots clés, les données brutes (séquences) des IG et TR sont importées d'EMBL-Bank (EBI). Ces données sont expertisées et annotées par LIGM et entrées dans IMGT/LIGM-DB (Figure 17).

3.1.2. IMGT IMGT/3Dstructure-DB: production de structures 3D annotées

Les données brutes (séquences, fichiers de coordonnées, description des atomes, structures 3D) des IG, TR, MHC et RPI et de leurs complexes (en particulier IG/Ag, pMHC et TR/pMHC) sont importées de la base RCSB PDB. Ces données sont alors expertisées, annotées par LIGM et entrées dans IMGT/3Dstructure-DB (Figure 18).

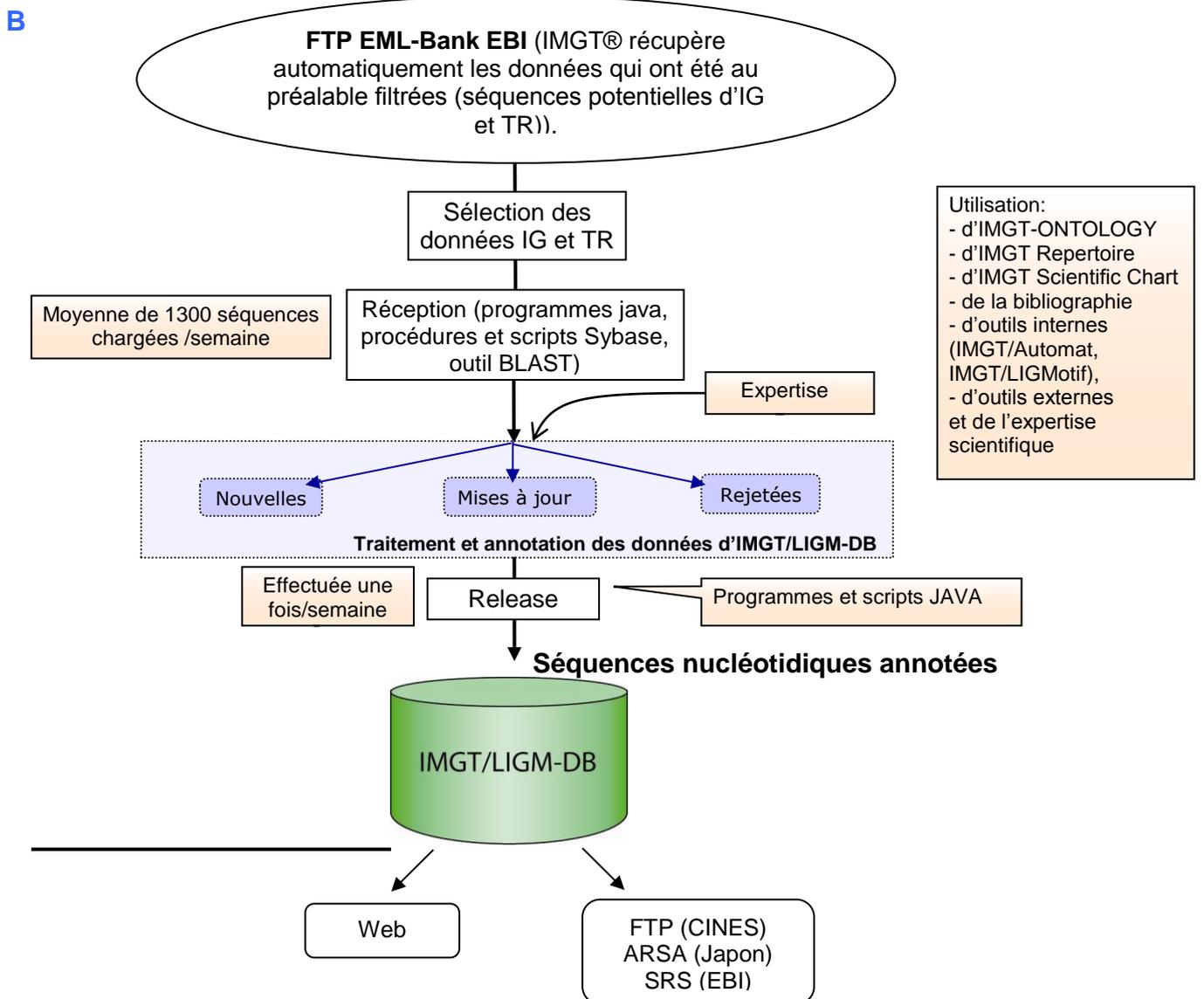
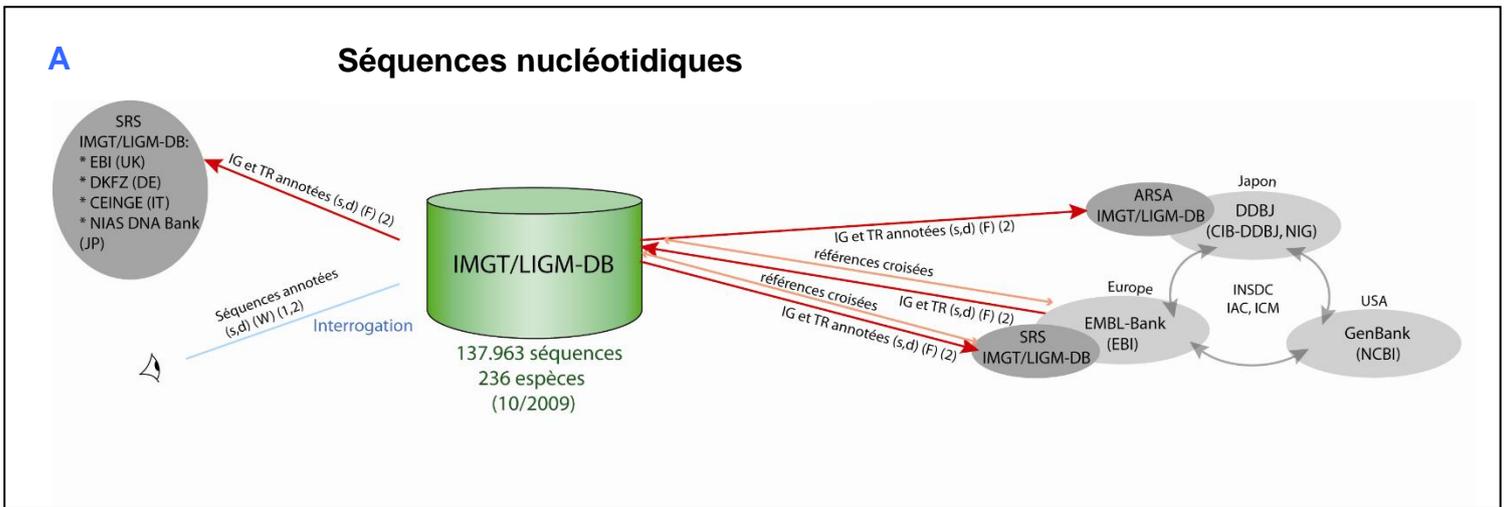


Figure 17. Production de données d'IMGT/LIGM-DB. (A) Flux d'entrées entrantes et sortantes, (B) Production des séquences nucléotidiques annotées.

Type de données: (s) séquences, (d) autres données dans un fichier à plat (flat file) comprenant les mots clés, les définitions, les références bibliographiques, les éléments descriptifs désignés 'Features', lesquels correspondent aux 'Labels' d'IMGT®. Type de transfert: (F) FTP, (W) Web. Type de format: (1) fasta, (2) flat file.

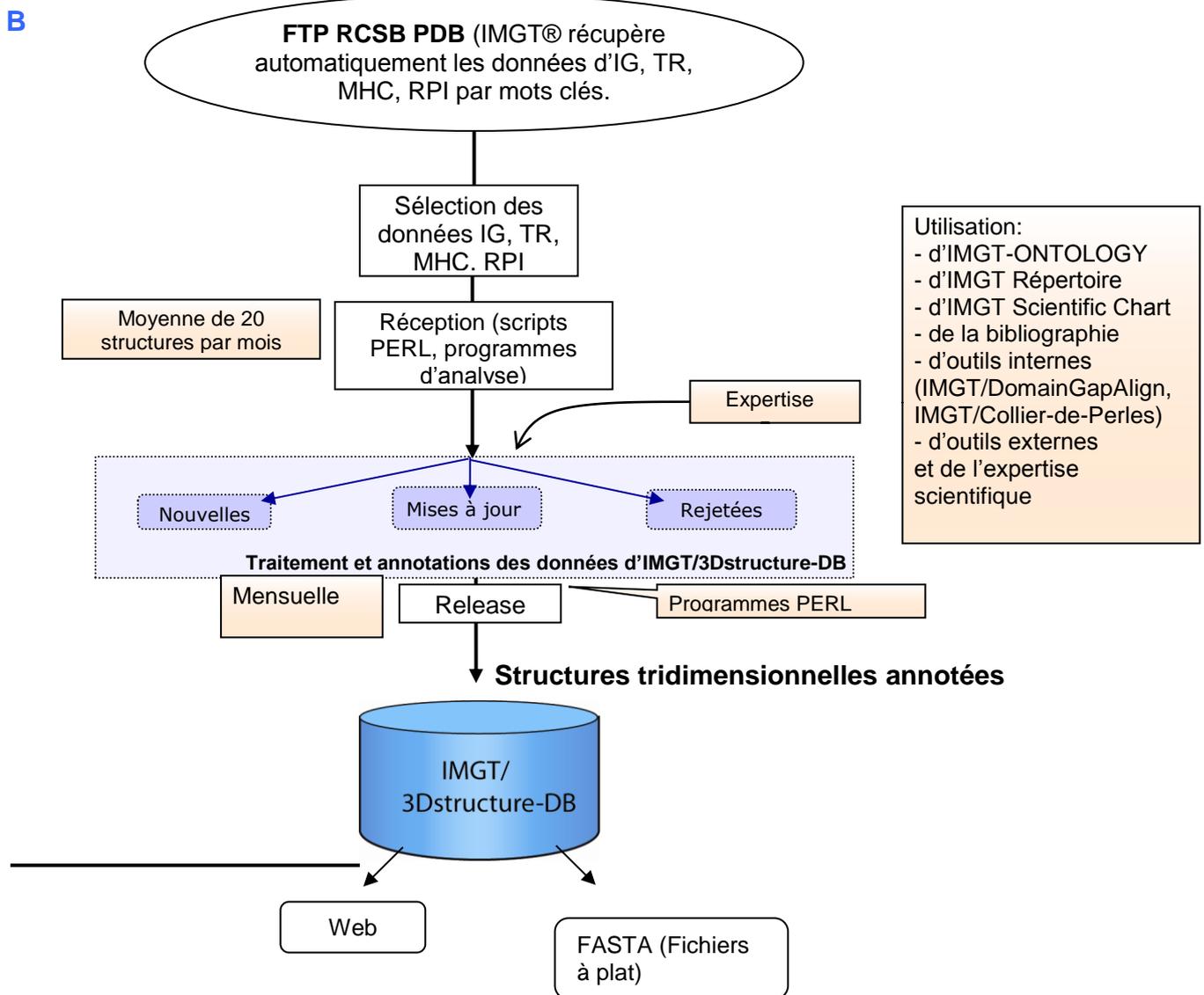
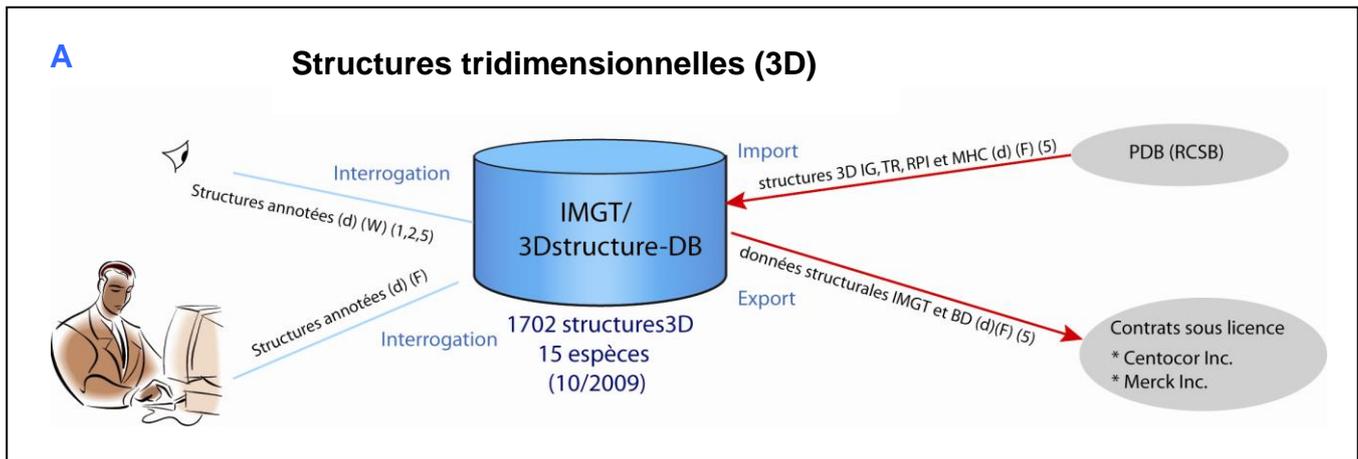


Figure 18. Production de données d'IMGT/3Dstructure-DB. (A) Flux d'entrées entrantes et sortantes. (B) Production de structures tridimensionnelles annotées.

Type de données: (d) données protéiques et structurales des récepteurs et de leurs chaînes et domaines constitutifs comprenant les différents contacts, les données de visualisation 3D, la comparaison avec la nomenclature IMGT, les références bibliographiques et les liens externes. Type de transfert: (F) FTP, (W) Web. Type de format: (1) fasta, (2) flat file.

3.2. Gestion

La gestion des bases de données et outils IMGT® (analyse, conception et développement bioinformatique) ainsi que celle des ressources Web est effectuée par LIGM sur les serveurs IMGT® (Figure 19).

Les outils et les bases de données sont développés au sein de la structure dans le cadre du processus de recherche et développement puis «déposés» au CINES pour la mise à disposition.

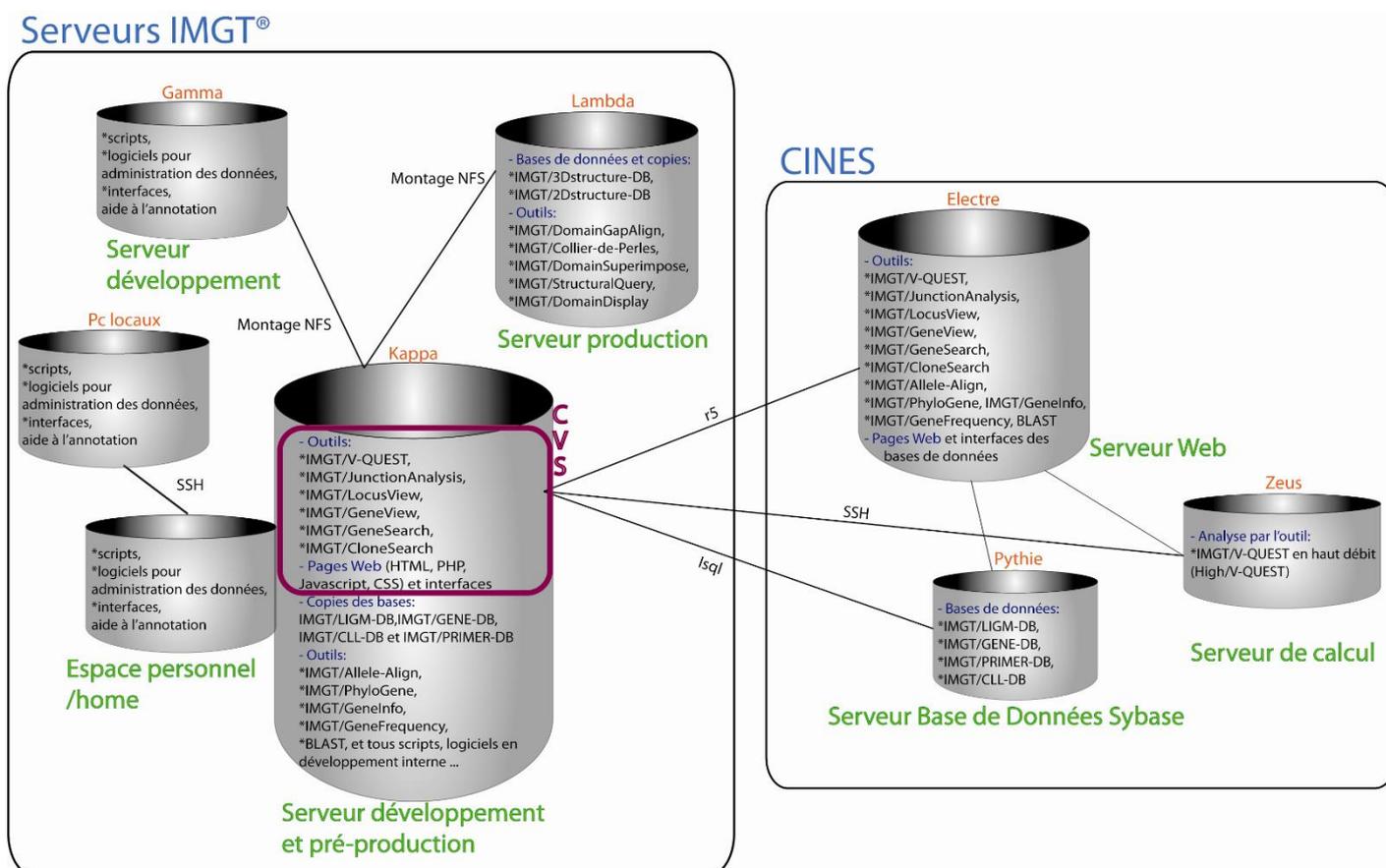


Figure 19. Relations entre les différents serveurs utilisés par IMGT® pour la gestion des données.

Montage NFS: Le service NFS (network File System) permet le partage d'un système de fichiers sur un réseau Linux. CVS: Concurrent Versions System, un système de gestion de versions, libre. r5: Synchronisation/mise à jour des pages Web par la technique r5. SSH: Secure Shell, programme informatique et protocole de communication sécurisé. IsqL: Interactive Structured query langage (SQL), outil permettant la communication en langage SQL.

3.3. Archivage

3.3.1. Bases de données

Chaque semaine, en interne, une copie (dump) des bases de données IMGT/LIGM-DB, IMGT/GENE-DB, IMGT/PRIMER-DB et IMGT/CLL-DB est effectuée et archivée sur 2 serveurs différents, l'un appartenant à la structure IMGT® (serveur Kappa) et l'autre appartenant à l'IGH. Chaque mois, une copie des bases de données IMGT/3Dstructure-DB et IMGT/2Dstructure-DB est effectuée et archivée sur le serveur Lambda de la structure IMGT®.

Le CINES effectue régulièrement de son côté une copie des bases de données gérées sur leur serveur Pythie par le système de gestion des bases de données Sybase. Les bases de données IMGT® concernées sont IMGT/LIGM-DB, IMGT/GENE-DB, IMGT/PRIMER-DB et IMGT/CLL-DB.

3.3.2. Outils et ressources Web

Les différentes versions des outils publics en ligne et la totalité des ressources Web ('pages fixes') sont archivées depuis 2002. La gestion de versions par CVS se fait sur le serveur Kappa pour:

- les outils: IMGT/V-QUEST, IMGT/JunctionAnalysis, IMGT/LocusView, IMGT/GeneView, IMGT/GeneSearch, IMGT/CloneSearch,
 - toutes les pages Web : HTML, PHP, Javascript, CSS,
 - les applications Web servant directement aux interfaces d'IMGT/LIGM-DB et d'IMGT/GENE-DB.
- Les outils qui ne sont pas distribués et les outils internes, non accessibles sur le site Web IMGT®, ne font pas l'objet de versions CVS.

3.4. Préservation

3.4.1. En externe

La préservation des données de recherche IMGT® à l'extérieur est concrétisée par la récupération des données IMGT® par d'autres organismes (EBI, Institut Pasteur, NCBI...) sur leur serveur FTP (Fichiers à plat (flat files), séquences IMGT®).

3.4.2. En interne

La préservation est faite par le CINES, avec possibilité de générer des réplicats et copies sur bandes.

3.5. Mise à disposition et diffusion des données d'IMGT®

L'accès à IMGT® est libre, gratuit pour les académiques, sous licences et contrats avec le CNRS pour les industriels (les types de contrats disponibles pour les sociétés industrielles sont décrits dans la partie Valorisation).

3.5.1. Site Web

La consultation d'IMGT® sur le site Web, à la page d'accueil d'IMGT®: <http://www.imgt.org> (contact: Marie-Paule.Lefranc@igh.cnrs.fr), peut prendre trois formes différentes: soit une requête via une interface utilisateur pour questionner une des bases de données IMGT®, soit le traitement d'une séquence nucléotidique ou protéique ou d'une structure 3D par un des quinze outils spécifiques IMGT®, ou encore une recherche d'informations dans les ressources Web IMGT®. Le site IMGT® est également accessible via EBI.

3.5.2. Distribution des données d'IMGT/LIGM-DB par FTP

Les données de la base de données IMGT/LIGM-DB peuvent être transférées par FTP (gratuit pour les académiques, sous contrat avec le CNRS pour les industriels).

Les fichiers à plat (flat files), mis à jour et distribués de manière hebdomadaire par LIGM, sont disponibles sur les serveurs FTP anonymes du CINES (<ftp://ftp.cines.fr/IMGT/>). Les organismes externes tels qu'EBI et DDBJ peuvent alors les importer sur leur propre FTP (<ftp://ftp.ebi.ac.uk/pub/databases/imgt/>).

3.5.3. Accès aux données d'IMGT/LIGM-DB par SRS et ARSA

IMGT/LIGM-DB est disponible par *Sequence Retrieval System* (SRS) sur différents serveurs, par exemple, ceux de :

- European Bioinformatics Institute (EBI), <http://srs.ebi.ac.uk/srs6bin/cgi-bin/wgetz?-page+top>
- Columbia University (CU), New York (Etats-Unis), <http://walnut.bioc.columbia.edu/srs7/>
- Indiana University, IUBio (Etats-Unis), <http://iubio.bio.indiana.edu/srs/>

- Deutsche Krebsforschungszentrum (DKFZ), Heidelberg (Allemagne), <http://genius.embnet.dkfz-heidelberg.de/menu/cgi-bin/srs7.1.3.1/wgetz>
- CEINGE Biotechnologie Avanzate (Italie), <http://bioinfo.ceinge.unina.it/framepage.php?link=srs>
- National Institute of Agrobiological Sciences (NIAS) DNA Bank, Tsukuba (Japon), <http://srs.dna.affrc.go.jp/srs8/srs>

IMGT/LIGM-DB est disponible par ARSA (*All-round Retrieval of Sequence and Annotation*) au DNA Data Bank of Japan (DDBJ) National Institute of Genetics (NIG) (Japon) <http://arsa.ddbj.nig.ac.jp/>

Ces sites utilisent, pour les mises à jour, les fichiers à plat d'IMGT/LIGM-DB disponibles sur le serveur FTP anonyme du CINES et de l'EBI.

3.5.4. Accès aux données par DAS à l'EBI et LinkOut au NCBI

- IMGT® a mis en place les données annotées sur le serveur DAS 'IMGT Genes' du Genome Browser Ensembl à l'EBI. Ce serveur fournit les positions des gènes IG et TR humains sur les chromosomes. Des liens pour les différents locus sont disponibles à partir d'IMGT Répertoire <http://www.imgt.org/textes/IMGTindex/Ensembl.html>.

- L'*United States National Library of Medicine* (US NLM) au *National Center for Biotechnology Information* (NCBI) a créé des liens ([LinkOut](#)) à partir de la page des résumés PubMed sur les numéros d'accès IMGT/LIGM-DB. Ces liens sont mis à jour automatiquement au NCBI chaque semaine, lors de la release hebdomadaire des fichiers à plat d'IMGT/LIGM-DB, et un message de confirmation est envoyé au LIGM.

3.5.5. Comparaison de séquences par BLAST et FASTA

Les séquences peuvent être comparées aux séquences issues de la base de données IMGT/LIGM-DB, par les outils d'alignements BLAST et FASTA, sur les serveurs de l'EBI (<http://www.ebi.ac.uk/Tools/blast2/nucleotide.html>) et de l'Institut Pasteur (<http://mobylye.pasteur.fr/cgi-bin/MobylyePortal/portal.py?form=fasta>, <http://mobylye.pasteur.fr/cgi-bin/MobylyePortal/portal.py?form=blast2>).

Des comparaisons de séquences par BLAST sont possibles également sur le site d'IMGT®, avec les séquences d'IMGT/LIGM-DB et avec les séquences de référence d'IMGT/GENE-DB (<http://www.imgt.org/blast/blast.html>)

3.5.6. Références croisées

Des références croisées facilitant l'interopérabilité entre IMGT® et les bases généralistes ont été mis en place:

- entre HGNC et IMGT/GENE-DB (noms des gènes IG et TR humains)
- entre Entrez Gene (NCBI) et IMGT/GENE-DB (noms des gènes IG et TR humains et de souris)
- entre Vega du Wellcome Trust Sanger Institute et IMGT/GENE-DB (noms des gènes IG et TR humains)
- entre UniProt (Expasy) et IMGT/GENE-DB (séquences protéiques des gènes constants des IG et TR)
- entre Sequence Ontology (SO) et IMGT-ONTOLOGY (labels IMGT® standardisés).

3.5.7. Aides spécifiques et conseils aux utilisateurs et API

Des aides spécifiques et des conseils sont fournis aux utilisateurs en réponse à leur demande. Afin de faciliter l'intégration des données d'IMGT® dans des applications développées par d'autres laboratoires, nous avons construit une API (*Application Programming Interface*) pour l'accès à la base et à ses logiciels.

Tout nouvel outil IMGT® mis à la disposition des utilisateurs (en accès libre pour les académiques, et sous contrat pour les industriels) fait l'objet, de la part d'IMGT®, d'une demande de protection à l'Agence pour la Protection des Programmes (APP, <http://app.legalis.net/>) (organisation européenne des auteurs de logiciels et concepteurs en technologies de l'information) (voir paragraphe 4.5.5). Grâce à cette démarche, IMGT® conserve l'exclusivité intellectuelle de ses développements et données.

IMGT® et son logo sont des marques déposées du CNRS (voir paragraphe 4.5.2). L'utilisation des bases de données et outils IMGT® est couverte par un copyright ©. Les logiciels et les données IMGT® sont fournis comme un service à la communauté scientifique pour être utilisés à des fins de recherche et à des fins éducatives. Les particuliers peuvent imprimer ou sauvegarder des portions d'IMGT® pour leur usage personnel. Toute autre utilisation du matériel IMGT® nécessite au préalable une autorisation écrite du directeur d'IMGT® et des institutions juridiques (CNRS et/ou Université Montpellier 2). IMGT® est déchargé de toute responsabilité en cas de préjudice et/ou dommage aux personnes ou aux biens qui résulterait d'un produit défectueux, d'une négligence ou de toute utilisation de méthodes, produits, instructions ou idées contenues dans IMGT® <http://www.imgt.org/textes/Warranty.html>.

Les industriels sous licences (CENTOCOR Inc. Johnson & Johnson, MERCK & Co Inc.) ont la possibilité d'exploiter les résultats issus de l'utilisation des bases de données et outils IMGT® qu'ils ont en interne, selon les termes des licences et contrats signés avec le CNRS.

Toutes les données immunologiques et immunoinformatiques issues de l'expertise IMGT® sont publiquement accessibles, exceptées celles fournies par les cliniciens et qui concernent les patients atteints de CLL. Les données organisées et gérées dans la base de données IMGT/CLL-DB sont soumises aux règles («Bylaws») mises en place par l'IMGT/CLL-DB Group.

En réponse à certaines attentes des utilisateurs et lorsque l'information est pertinente pour l'ensemble de la communauté scientifique, des réponses aux questions des utilisateurs sont rendues publiques dans la section *Frequently asked questions* (FAQ) <http://www.imgt.org/textes/FAQ/>.

4. Infrastructures du système de données de recherche IMGT® (techniques, financières, humaines, réglementaires, partenariats...)

4.1. Infrastructures techniques et moyens CPU

4.1.1. Moyens en équipement

Les moyens en équipement de la structure d'IMGT® sont décrits dans le tableau 2 (page suivante).

4.1.2. Capacité totale de stockage

Total: 6 Téra-octets (To) dont:

- 3 serveurs IGH : 300 Giga-octets (Go) + 2 x 130 Go + 140 Go + 4 x 300 Go (= ~ 2 To)
- 2 serveurs CINES: 50 Go (sans la partie Sybase)
- 20 postes = 20 x 250 Go = 4 To

4.1.3. Nombre total de CPU

Total : 32 CPU (central processing unit) dont:

- 3 serveurs IGH = 2+2+4 = 8 CPU
- 2 serveurs CINES = 2+2 = 4 CPU
- 20 postes = 20 CPU

Tableau 2. Nature des équipements de la structure IMGT®.

| Nature équipement | Gestion (Localisation) | Date d'acquisition | Contrat de maintenance | % utilisation |
|--|------------------------|--------------------|------------------------|---------------|
| Serveur DELL/PE6850 (Gamma) 4 processeurs | IMGT® (IGH) | 2007 | Sous garantie | 100% |
| Serveur DELL/PE6850 (Lambda) 2 processeurs | IMGT® (IGH) | 2006 | Sous garantie | 100% |
| Serveur DELL/PE7250 (Kappa) 2 processeurs connexions, stabilisateurs | IMGT® (IGH) | 2004 | Non | 100% |
| Baie à sauvegarde DELL | IMGT® (IGH) | 2007 | Non | 100% |
| Baie 4 disques DELL/PV220S pour serveurs et armoire DELL/4210 | IMGT® (IGH) | 2004 | Non | 100% |
| Serveur production 4 processeurs (Electre) et serveur Base données 2 processeurs (Pythie) | CINES (CINES) | 2006 | Oui | 100% |
| Serveur de Calcul IBM ANAKIN(16 proc) et HERA (32 proc) accessibles par le frontal d'accès Zeus à disposition depuis janvier 2009 (DARI 2009) | CINES (CINES) | 2009 | Oui | 100% |

4.2. Financements publics de la structure IMGT®

4.2.1. Contrats Européens

- FP6 (6th PCRDT) Programme: IST (2006-2009)

Pays: Italie

Numéro du contrat: IST-2004-028069

Titre du projet: ImmunoGrid – The European Virtual Human Immune System Project.

Acronyme du projet: ImmunoGrid

Type de contrat: Specific Targeted Research Project

Durée: 36 mois (01/02/06-31/01/09)

Coût total du projet: 2.622.274 € - Contribution de l'UE: 1.951.042 €

Nombre de partenaires: 8, CINECA (Consortium Interuniversitario per la Gestion del centro di Calcolo Elettronico dell'Italia Nord Orientale, Italie), [CNRS-IMGT \(France\)](#), Université de Bologne (Italie), Université de Queensland (Australie), Université de Catania (Italie), Danmarks Tekniske Universitet (Danemark), Birkbeck College Université de Londres (Royaume-Uni), CNR (Consiglio Nazionale dell Ricerche, Italie).

Coordinateur du projet: Elda Rossi (CINECA)

CNRS-IMGT: Coûts éligibles 420.446 € - Contribution de l'UE: 221.223 €

- FP5 (5th PCRDT) Programme: LIFE QUALITY (2000-2003)

Pays: France

Numéro du contrat: GLG2-CT-2000-01287

Titre du projet: Integrated system to manage the enormous complexity of the genome and proteome of the specific immune response: IMGT, the international ImMunoGeneTics database. (Quality of Life and Management of Living Resources Call identifier 1999/C64/14)

Acronyme du projet: IMGT

Type de contrat: Cost-sharing contracts

Durée: 39 mois (01/09/00-30/11/03)

Coût total du projet: 2.819.481 € - Contribution de l'UE: 1.250.000 €

Nombre de partenaires: 5, [CNRS-LIGM](#), EMBL-EBI, CR UK, EUROGENTEC SA, BPRC,

Coordinateur du projet: Marie-Paule Lefranc (CNRS)

CNRS-LIGM Coûts éligibles : 1.809.481 € – Contribution de l'UE: 540.000 €

- FP4 (4th PCRDT) Programme: BIOTECH2 (1996-1999)

Pays: France

Numéro du contrat: BIO4CT960037 (BIO4960037)

Titre du projet: Common access to immunogenetics data: the integrated immunogenetics database.

Acronyme du projet: IMGT

Durée: 36 mois (01/11/96-31/10/99)

Nombre de partenaires: 6, CNRS-IMGT, EMBL.EBI, ICRF, Univ Cologne, BPRC, EUROGENTEC S.A.

Coordinateur du projet: Marie-Paule Lefranc (CNRS)

- Programme: BIOMED1 (III.4 Data-handling and databases 1993) (1993-1996)

Pays: Allemagne

Numéro du contrat: BIOCT930038 (GENE-CT93-0038) DG 12 SSMA

Titre du projet: Development of a European Integrated Database of Immunological Interest.

Acronyme du projet: IMGT, ImMunoGeneTics database

Durée: 30 mois (01/11/93-30/04/96)

Contribution de l'UE: 435.000 ECU

Nombre de partenaires: 4, EMBL, Université Montpellier 2.LIGM, ICRF, Univ Cologne

Coordinateur du projet: Rainer Fuchs (EMBL)

Coordinateur scientifique: Marie-Paule Lefranc, Université Montpellier 2-LIGM: Contribution EU: 200.000 ECU

4.2.2. Financements Nationaux

- Financement IBISA (Infrastructures Biologie Santé et Agronomie) 2009-2011 (en cours)

Pays: France

Décision d'aide: 2008-37

Titre du projet: IMGT®, the international ImMunoGeneTics information system®.

Durée: 24 mois (12/01/09-11/01/2011)

Financement IMGT: 200.000 € (4 CDD)

Nombre de partenaires: 1

- Agence Nationale de Recherche (ANR) BIOSYS 2006-2011 (en cours)

Pays: France

Numéro du contrat: ANR-06-BIOSYS-005-01

Acronyme du projet: IMGT

Titre du projet: IMGT-ONTOLOGY and IMGT-Choreography: foundations of a systemic approach for immunogenetics and immunoinformatics.

Durée: 48 mois (11/12/06-10/01/11)

Financement IMGT: 240.000 €

Nombre de partenaires: 1

- Agence Nationale de Recherche (ANR) FLAVORES 2008-2010 (en cours)

Pays: France

Numéro du contrat: ANR-07-GANI-002

Acronyme du projet: FLAVORES

Durée: 36 mois (02/01/08-31/12/10)

Coût total projet: 1.427.253 € - Coût financement: 428.236 €

Financement IMGT: 15.000 €

Nombre de partenaires: 5, INRA LGP, INRA VIM, Université de Bourgogne CRI, CNRS-IMGT, SYSAAF

- Plan Pluri-Formation Université Montpellier 2 (PPF) depuis 1999 (en cours)

Pays : France

* 2007-2010 (en cours)

Titre du projet: Maintien en France d'IMGT®.

Durée: 48 mois

Financement IMGT: 95.940 € d'équipement et 100.000 € de vacances

Nombre de partenaires : 1

* 2003-2006

Titre du projet: Création d'une structure afin de maintenir IMGT en France.

Durée: 48 mois

Financement IMGT: 68.000 € d'équipements et 55.250 € de vacances

Nombre de partenaires: 1

* 1999-2002

Titre du projet: IMGT

Durée: 48 mois

Financement IMGT: 425.000 F

Nombre de partenaires: 1

- ACI Informatique, Mathématiques, Physique en Biologie (ACI IMPBio 2004-2007)

Pays: France

Numéro de contrat: IMP82-2004

Titre projet: Maintien en France de IMGT, référence internationale en immunogénétique et immunoinformatique, <http://www.imgt.org/textes/IMGTinformatio//IMPBIO2004/>

Durée: 36 mois (20/09/04-10/09/07)

Financement IMGT: 135.000 € TTC

Nombre de partenaires: 24 non financés (soutien au projet IMGT): 8 partenaires français et 16 partenaires étrangers.

- Groupements d'Intérêt Scientifique (GIS) AGENAE (2004-2007)

Pays: France

Numéro de contrat: 753558/00 (FNS) 8279

Durée: 36 mois (21/07/04-20/07/07)

Financement IMGT: 12.000 € TTC

- Institut Universitaire de France (IUF) (2002-2007)

Pays: France

Durée: 60 mois

Financement IMGT: 76.225 € TTC

- Genopole RNG (2006-2007)

Pays: France

Durée: 24 mois

Financement IMGT: 123.635 € TTC

- Action incitative inter-EPST Bioinformatique (CNRS) (2002)

* 2002

Pays : France

Durée: 12 mois

Financement IMGT: 10.000 € TTC

Nombre de partenaires: 2, CBS Montpellier, CNRS-IMGT

* 2002

Pays : France

Durée: 12 mois

Financement IMGT: 5.488 € TTC

Nombre de partenaires: 2, LIRMM Montpellier, CNRS-IMGT

- Genopole Montpellier Languedoc-Roussillon (1999-2001)

Pays : France

Numéro de contrat: Notification 00N60/0909

Durée: 36 mois
Financement IMGT: 753.000 F
Nombre de partenaires: 1

- ARC 9494 Association pour la recherche sur le Cancer (1999-2000)

Pays : France
Durée: 24 mois
Financement IMGT: 300.000 F TTC
Nombre de partenaires: 1

4.2.3. Financements Régionaux

- Grands Plateaux Techniques pour la recherche (GPTR), Région Languedoc-Roussillon

Pays: France
* 2008-2012 (en cours)
GPTR, Grands Equipements pour l'Evolution Technologique et l'Ouverture Scientifique (GEPETOS), Région Languedoc-Roussillon
Titre projet: Mise en place d'une démarche qualité fondée sur IMGT-ONTOLOGY et conciliant créativité, innovation et valorisation au sein d'IMGT®.
Durée: 48 mois (31/10/08 - 31/12/2012)
Financement IMGT: 100.000 € TTC
Nombre de partenaires: 1

* 2007-2009

Titre du projet: IMGT-Choréographie: création d'un système d'interactions dynamiques entre les bases de données, les outils et les ressources Web d'IMGT.
Durée: 36 mois (01/01/07-31/12/09)
Financement IMGT: 90.000 € TTC, 75.258 € HT
Nombre de partenaires: 1

- BIOSTIC-LR (2004-2005)

Pays: France
Titre projet: IMGT-Choréographie: création d'un système d'interactions dynamiques entre les bases de données, les outils et les ressources Web d'IMGT.
Durée: 24 mois
Financement IMGT: 40.000 HT
Nombre de partenaires: 1

- Région Languedoc-Roussillon (1999-2000)

Pays: France
* 2000
Numéro de contrat: Etude R086
Durée: 12 mois
Financement IMGT: 207.296.84 F HT

* 1999

Numéro de contrat: Dossier 99015193
Durée: 12 mois
Financement IMGT: 106.000 F HT

- BQR Université Montpellier 2, UM2 (1999)

Pays: France
Durée: 12 mois
Financement IMGT: 100.000 F

4.3. Financements industriels de la structure IMGT®

4.3.1. Contrats types et licences

Des contrats types, rédigés avec le Service du Partenariat et de la Valorisation du CNRS, sont signés par des industriels lorsqu'ils demandent à utiliser IMGT®.

Trois types de contrats sont possibles:

- 1) accès en ligne (on-line) et consultation de l'ensemble du portail Internet,
- 2) téléchargement FTP de données (fichiers à plat des séquences d'IMGT/LIGM-DB et des séquences de référence d'IMGT/GENE-DB),
- 3) installation en interne des systèmes d'information IMGT/V-QUEST, IMGT/JunctionAnalysis, IMGT/3Dstructure-DB et IMGT/DomainGapAlign (contrats associés aux licences logicielles).

En fonction de l'origine des utilisateurs, ces trois types de contrats sont utilisables de la façon suivante :

- Académiques: l'accès en ligne au site IMGT® libre et gratuit.
 - Petites et moyennes entreprises (PME) en biotechnologies ('Biotech'): deux types de contrats sont possibles selon les besoins, l'un pour l'accès en ligne d'IMGT®, l'autre pour le téléchargement FTP d'IMGT/LIGM-DB (base de données).
 - Grands Laboratoires Pharmaceutiques ('Big Pharma'): installation en interne d'IMGT/V-QUEST, IMGT/JunctionAnalysis, IMGT/3Dstructure-DB et IMGT/DomainGapAlign, sous licences (gérées par France Innovation Scientifique et Transfert (FIST SA) et par la Direction Politique Industrielle (DPI) du CNRS national) et contrats d'assistance et de collaboration (gérés par le Service du Partenariat et de la Valorisation de la Délégation Languedoc-Roussillon du CNRS, DR13).
- Des prestations de service sont également possibles, en particulier dans le cadre d'humanisation d'anticorps.

Ces contrats permettent le financement des CDD IMGT® (biocurateurs, bioinformaticiens, annotateurs). Il est cependant indispensable que le personnel d'IMGT® soit stabilisé de manière urgente sur des postes statutaires afin de maintenir l'expertise d'IMGT® en France.

4.3.2. Partenaires industriels

Une vingtaine de contrats et/ou licences ont été signés entre le CNRS et 13 partenaires industriels depuis mars 2006:

- CENTOCOR Research and Development, Inc. Johnson & Johnson, Malvern, Pennsylvanie PA, Etats-Unis (depuis mars 2006).
- AMGEN Inc., Seattle, WA, Etats-Unis (depuis mars 2006).
- SANOFI-AVENTIS Deutschland GmbH, Frankfurt am Main, Allemagne (depuis janvier 2007).
- INSTITUT DE RECHERCHE PIERRE FABRE, Boulogne-Billancourt, France (depuis avril 2007).
- MERCK & Co., Inc. Whitehouse Station, NJ, Etats-Unis (septembre 2007-septembre 2009)
- IDD-BIOTECH, Dardilly, France (depuis novembre 2007).
- CHUGAI Pharmaceutical Co., Ltd, Tokyo, Japon (depuis septembre 2008).
- ASTELLAS PHARMA Inc., Tokyo, Japon (depuis janvier 2009).
- AGENSYS Inc., Santa Monica, Californie CA, Etats-Unis (depuis février 2009).
- MERCK SERONO S.A., Genève, Suisse (depuis mars 2009).
- F. HOFFMANN-LA ROCHE AG, Bâle, Suisse (depuis avril 2009).
- EMD SERONO RESEARCH CENTER, Inc., Turnpike, Billerica, Massachusetts, Etats-Unis (depuis septembre 2009).
- REGENERON PHARMACEUTICALS Inc., Tarrytown, New York, Etats-Unis (depuis septembre 2009).

Des contrats de prestations de service ont été signés avec 3 autres partenaires industriels.

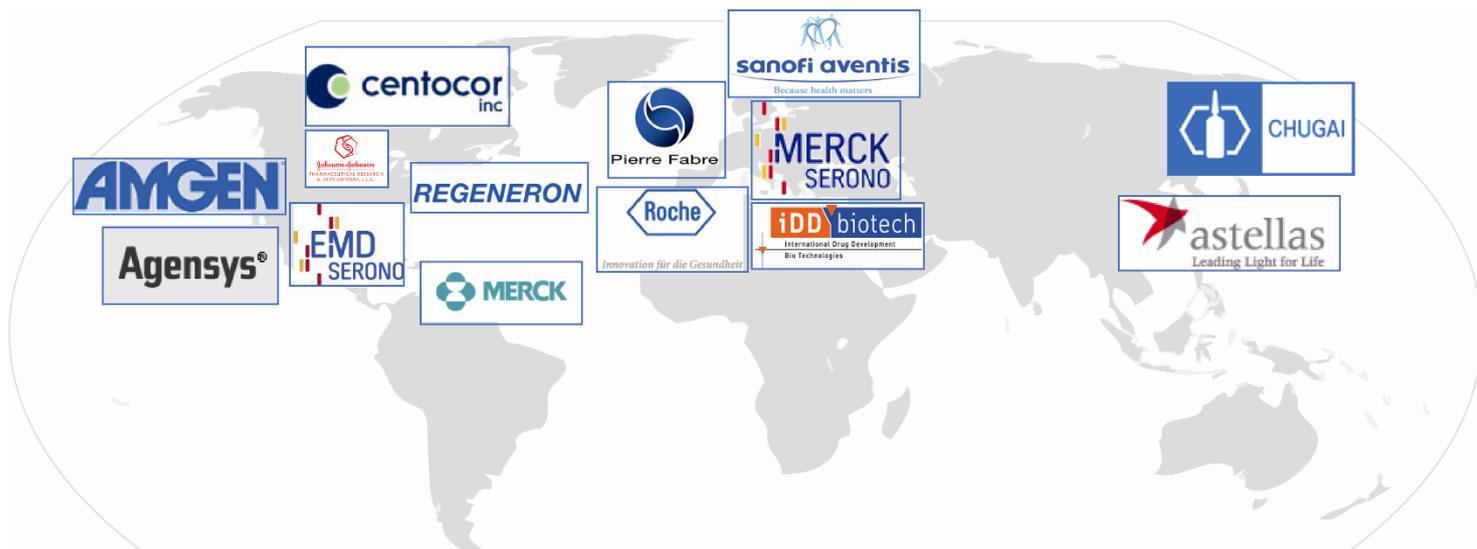


Figure 21. Industries pharmaceutiques en contrat et/ou licence avec le CNRS-IMGT® depuis mars 2006.

4.4. Ressources humaines de la structure IMGT®

L'équipe IMGT® comprend 18 personnes: 6 statutaires et 12 CDD. Les grades et les fonctions au sein de la structure IMGT® sont précisés dans le Tableau 3.

Tableau 3. Personnel de la structure IMGT®.

| Nom-Prénom | Grade et organisme employeur | Fonction au sein d'IMGT® |
|--------------------------|--|---|
| LEFRANC Marie-Paule | Professeur des Universités classe exceptionnelle. Membre Senior de l'IUF, Université Montpellier 2 | Fondateur et Directeur |
| LEFRANC Gérard | Professeur Emérite des Universités classe exceptionnelle, Université Montpellier 2 | Conseiller Scientifique |
| DUROUX Patrice | Ingénieur de Recherche (RIO), CNRS | Responsable Informatique, Maintenance et développement des bases de données et outils |
| GIUDICELLI Véronique | Ingénieur d'Etude, Université Montpellier 2 | Responsable Bioinformatique, Recherche et développement IMGT-ONTOLOGY |
| JABADO-MICHALOUD Joumana | Ingénieur d'Etude (RIO), CNRS | Gestion, réception et annotation des données d'IMGT/LIGM-DB IMGT Répertoire |
| GINESTOUX Chantal | Technicien classe supérieure TCS, CNRS | Responsable du développement et de la maintenance du site Web Ressources Web |
| ALAMYAR Eltaf | CDD (IE IBISA), CNRS | Développement et déploiement d'IMGT/HighV-QUEST (IMGT/V-QUEST haut-débit) |
| BELLAHCENE Fatena | CDD (IE), CNRS | Annotation des données d'IMGT/GENE-DB et outils d'analyse de génomes |
| FOLCH Géraldine | CDD (IE), CNRS | Annotation des données d'IMGT/LIGM-DB et outils d'analyse de séquences |

| | | |
|--------------------|---------------------------|--|
| LACAN Amandine | CDD (IE IBISA), CNRS | Développement d'interface d'aide à l'annotation de grandes séquences génomiques |
| LE ROY Christophe | CDD (IE IBISA), CNRS | Intégration des données dans IMGT/2Dstructure-DB et annotation |
| OUARAY Zohra | CDD (IE), UM2 | Développement d'outils de visualisation 3D dans IMGT/3Dstructure-DB |
| POIRON Claire | CDD (IE), CNRS | Annotation de données de grandes séquences IG et TR génomiques et outils d'analyse |
| REGNIER Laëtitia | CDD (IE), CNRS | Contrôle de qualité des bases de données et outils, Responsable Management de la Qualité |
| SCHUMENG Marie | CDD (IE IBISA), CNRS | Ingénieur Qualité, mise en place de la démarche Qualité norme ISO9001 |
| BROCHET Xavier | CDD, Post doctorant, CNRS | Développement de la base IMGT/CLL-DB et des outils d'analyse associés |
| LANE Jérôme | CDD (IE), Doctorant, CNRS | Développement de logiciels pour l'analyse de génomes (IMGT/LIGMotif) |
| EHRENMANN François | CDD (IE), Doctorant, CNRS | Maintenance, développement de logiciels. Développement d'IMGT/3Dstructure-DB |

4.5. Réglementation et valorisation de la structure IMGT®

4.5.1. Accréditations

- IMGT®, the international ImMunoGenetics information system®, a été approuvé comme Membre Institutionnel Universitaire de l'International Medical Informatics Association (IMIA), le 10 novembre 2006.

- IMGT® est en conformité avec la charte du 'HONcode'

- IMGT® est labellisé "Plate-forme Nationale de Recherche Bioinformatique" Plate-forme RIO recherche inter-organismes (INSERM, CNRS, CEA et INRA), depuis la création de RIO en décembre 2001 et Plate-forme Bioinformatique IBISA depuis la création du GIS IBISA en 2007.

4.5.2. Marques déposées

1/ Marques déposées du CNRS "IMGT"

- Marque française n°00/3004406 du 31 janvier 2000 (BOP1 n°00/27 NL du 07/07/2000).
- Marque communautaire n°001607274 du 6 avril 2000.
- Marque Canada n°1/058.361, enregistrée le 3 août 2004.
- Marques Etats-Unis, US n°76/071.478 et US n°76/975731, enregistrées le 4 novembre 2003.

2/ Marques déposées du CNRS "ImMunoGeneTics" et vignette en couleurs (logo)

- Marque française n°00/3005152 du 3 février 2000 (BOP1 n°00/27 NL du 07/07/2000).
- Marque communautaire n°001608074 du 6 avril 2000.
- Marque Canada n°1/058.360, enregistrée le 3 août 2004.
- Marques Etats-Unis, US n°76/071.477, enregistrée le 24 juin 2003 et US n°76/975485 enregistrée le 5 août 2003.

4.5.3. Brevets

- **WO952672** “Recombinant baculovirus and use thereof in the production of monoclonal antibodies”. Brevet déposé le 31 janvier 1994 par Proteine Performance, publié le 3 août 1995. Baculovirus recombinant et son utilisation pour la production d’anticorps monoclonaux. Ce brevet revendique des vecteurs recombinants, contenant des cassettes, pour l’expression d’anticorps monoclonaux en cellules d’insectes.

- **US6852507** “Recombinant baculovirus and use thereof in the production of monoclonal antibodies”. Etats-Unis d’Amérique – Demande de brevet US19970975982 du 21 novembre 1997, par INRA and CNRS, publié le 8 février 2005.

- **WO9608564** “Humanised antibody to integrin chain beta-1” Brevet déposé le 12 septembre 1994 par INRA et le CNRS, publié le 21 mars 1996. Ce brevet protège le procédé de production de l’anticorps monoclonal humanisé K20 et les propriétés biologiques de ce prototype.

4.5.4. Contrats de licence et savoir

“Contrat de licence et de savoir” (n° L01044) signé le 27 juin 2001 entre le CNRS, l’Université Montpellier 2 et le Babraham Bioscience Technologies (Cambridge, UK) pour l’exploitation d’anticorps humains par des souris transgéniques. Ces souris ont été obtenues par transfection de constructions dérivées de clones YACs que nous avons isolés et caractérisés dans le Laboratoire.

4.5.5. Certificats logiciels

Cinq logiciels ont été déposés à l’Agence pour la Protection des Programmes (APP):

* IDDN.FR.001.300031.000.S.P.2006.000.31235 pour IMGT/JunctionAnalysis, juillet 2007.

* IDDN.FR.001.300030.000.S.P.2006.000.31235 pour IMGT/3Dstructure-DB et IMGT/StructuralQuery, juillet 2007.

* IDDN.FR.001.300033.000.S.P.2006.000.31235 pour IMGT/V-QUEST, juillet 2007.

* Demande en cours pour IMGT/DomainGapAlign, 2009.

* Demande en cours pour IMG/LIGMotif, 2009.

4.5.6. Domaines Internet

Les domaines Internet IMGT® (Tableau 4) sont gérés par INDOM (<http://www.indom.com/indom/>).

Tableau 4. Domaines Internet IMGT®

| Nom | Expiration |
|-----------|-------------------|
| imgt.fr | 9 avril 2019 |
| imgt.org | 3 février 2013 |
| imgt.info | 28 septembre 2015 |
| imgt.eu | 28 mars 2017 |

4.5.7. Prix ADER Sciences de la vie

- Prix ADER Sciences de la Vie “Innovation-Recherches-Entreprise” 1994 décerné à Marie-Alix Poul et Marie-Paule Lefranc, et à Proteine Performance S.A., pour l’expression de l’anticorps monoclonal humanisé K20 en cellules d’insectes.

- Nomination au prix ADER Sciences de la Vie “Innovation-Recherches-Entreprise” 2003, Marie-Paule Lefranc (IMGT®) et Eurogentec S.A., pour la création de IMGT/PRIMER-DB.

4.6. Partenariats scientifiques

Les projets réalisés par des laboratoires utilisant les bases de données et les outils IMGT® sont de l'ordre de plusieurs centaines pour les quatre dernières années (2005-2009), comme l'attestent les très nombreuses publications dans lesquelles IMGT® est cité. La structure IMGT® est elle-même auteur, pour la période 2005-2009, de 38 publications dans des journaux à comité de lecture et de 15 articles dans des ouvrages.

4.6.1. Nomenclature IMGT® pour les projets génomes et protéomes

Laboratoire porteur du projet: IMGT®.

Autres laboratoires: National Center for Biotechnology Information, NCBI, Bethesda (Etats-Unis); European Bioinformatics Institute, EBI (Angleterre); Wellcome Trust Sanger Institute (Angleterre); University of Victoria, British Columbia (Canada); Simon Fraser University, Burnaby (Canada); Institut National de la Recherche Agronomique INRA, Unité de Virologie et Immunologie Moléculaires Jouy-en-Josas (France); Paige Laboratory, University of Massachusetts Amherst, MA (Etats-Unis).

Compléments de financement: Laboratoires;

GIS AGENAE 753558/00 (FNS) 8279 (12.000 €, 21/07/04-20/07/07);

Agence Nationale de la Recherche ANR-07-GANI-002 FLAVORES (15.000 €, 02/01/08-31/12/10).

Depuis le congrès «Human Gene Mapping 10 Workshop (HGM10)», New Haven (USA) (11-17 juin 1989), le comité de nomenclature IMGT (IMGT-NC) est responsable des noms de gènes IG et TR et travaille en étroite collaboration avec le comité de nomenclature de *Human Genome Organisation* (HUGO) *Nomenclature Committee* (HGNC). Ainsi les gènes humains (664 gènes dont 421 IG et 243 TR) ont été approuvés par HGNC en 1999, et ceci avant même que la séquence du Génome Humain soit connue. Ces gènes ont été ajoutés dans IMGT/GENE-DB, Genome Database (GDB, Canada), GeneCards, LocusLink au National Center for Biotechnology Information (NCBI, Bethesda, Etats-Unis) et Entrez Gene (NCBI) quand cette base a remplacé LocusLink. Entrez Gene fait des liens directs sur les entrées d'IMGT/GENE-DB (c'est le seul exemple de liens directs d'Entrez Gene sur une base externe).

Les rapports de l'IMGT-NC sont régulièrement approuvés par le Comité de Nomenclature de l'Union Internationale des Sociétés d'Immunologie qui relève de l'Organisation Mondiale de la Santé (OMS) (*World Health Organization-International Union of the Immunological Societies* (WHO-IUIS)). Les résultats ont été publiés dans 2 livres qui sont devenus les références dans le domaine «The immunoglobulin FactsBook» et «The T cell receptor FactsBook» par M.-P. Lefranc and G. Lefranc (Academic Press, 2001). IMGT® travaille en collaboration avec Entrez Gene (NCBI), '*Vertebrate Genome Annotation*' (Vega) au Wellcome Trust Sanger Institute (Hinxton, Angleterre) et Ensembl à l'EBI, pour les locus des IG et TR des nouvelles espèces séquencées qui font l'objet d'une expertise similaire. En particulier, IMGT® transmet le nom des gènes annotés dans IMGT® à ces centres et vérifie leur attribution correcte selon leur localisation dans les locus.

L'expertise d'IMGT® est très sollicitée pour la mise en place de standards basés sur les concepts IMGT®. Il est intéressant de noter que ces principes se généralisent aux molécules des invertébrés et que le standard d'une numérotation unique, représentée par des Colliers de Perles, a été adopté par plusieurs bases de données de peptides (Knottin, Cybase, PenBase).

Publications à titre d'exemples:

Conrad M.L., Mawer M.A., Lefranc M.-P., McKinnell L., Whitehead J., Davis S.K., Pettman, R. and Koop B.F. The genomic sequence of the bovine T cell receptor gamma TRG loci and localization of the TRGC5 cassette. *Vet. Immunol. Immunopathol.* 2007, 115, 346-356.

Bernard D., Hansen J.D., du Pasquier L., Lefranc M.-P., Benmansour A. and Boudinot P. Costimulatory receptors in jawed vertebrates: conserved CD28, odd CTLA4 and multiple BTLAs. *Dev. Comp. Immunol.* 2007, 31, 255-271.

Gueguen Y., Garnier J., Robert L., Lefranc M.-P., Mougenot I., de Lorgeril J., Janech M., Gross P., Warr G., Cuthbertson B., Barracco M.A., Bulet P., Aumelas, A., Yang Y., Bo D., Xiang J., Tassanakajon A., Piquemal

D. and Bachère E. PenBase, the shrimp antimicrobial peptide penaeidin database: sequence-based classification and recommended nomenclature. *Dev. Comp. Immunol.* 2006, 30, 283-288.

Herzig C., Blumerman S., Lefranc M.-P. and Baldwin C. Bovine T cell receptor gamma variable and constant genes: combinatorial usage by circulating $\gamma\delta$ T cells. *Immunogenetics*, 2006, 58, 138-151.

4.6.2. Pronostic dans les leucémies lymphoïdes chroniques

Laboratoire porteur du projet: G. Papanicolaou Hospital Exokhi, Thessaloniki (Grèce).

Autres laboratoires: Laboratoires du réseau européen ERIC, European Research Initiative on CLL: 130 laboratoires de 26 pays; Nikea General Hospital Keratsini (Grèce); Università Vita-Salute San Raffaele, Milan (Italie); University Medical Center Rotterdam (Pays-Bas); Hôpital Pitié-Salpêtrière Paris (France); Uppsala University, Uppsala (Suède); Faculté de Médecine de Sousse (Tunisie); Feinstein Institute for Medical Research, New York (Etats-Unis).

Compléments de financement: Laboratoires.

La base de données de séquences nucléotidiques d'IMGT®, IMGT/LIGM-DB, et les outils d'analyse de séquences, IMGT/V-QUEST et IMGT/JunctionAnalysis, ont rapidement été adoptés par les laboratoires de recherche, pharmaceutiques et hospitaliers pour l'étude des répertoires. A titre d'exemple le réseau European Research Initiative on CLL (chronic lymphocytic leukemia) (ERIC) qui regroupe dans 26 pays 130 laboratoires travaillant sur les leucémies lymphoïdes chroniques (LLC ou CLL en anglais) a recommandé l'utilisation de IMGT/V-QUEST comme référence, pour la détermination des mutations des gènes IGHV (gènes variables V du locus des chaînes lourdes H des IG) des patients atteints de LLC. Les résultats standardisés d'IMGT/V-QUEST sont cruciaux car le taux de mutations est un facteur pronostic important des LLC: le pronostic est favorable si les gènes sont mutés et défavorable si les gènes ne sont pas mutés.

Un livre «Immunoglobulin Gene Analysis in Chronic Lymphocytic Leukemia» a été publié en 2009, dont les 8 chapitres ont été rédigés par des membres du réseau ERIC, selon les critères scientifiques d'IMGT®.

Une base de données, IMGT/CLL-DB, a été créée à l'usage des équipes de recherche clinique afin de gérer et d'analyser les séquences des IG des patients atteints de LLC (Thèse de Xavier Brochet, décembre 2008). Les annotations des séquences sont entrées dans IMGT/CLL-DB par l'outil d'analyse IMGT/V-QUEST, ce qui fournit des données standardisées de très grande qualité. L'implémentation d'IMGT/CLL-DB a été conçue pour servir de modèle générique utilisable pour la gestion et l'analyse de séquences de patients dans d'autres pathologies.

Trois ateliers internationaux ont été organisés pour l'analyse des séquences dans les LLC selon les critères scientifiques d'IMGT®, le premier à Uppsala (Suède) en 2007, le second à Paris en 2008 et le troisième en 2009 à Thessaloniki (Grèce). Une soixantaine de chercheurs et cliniciens participent à ces ateliers.

Références citant IMGT/V-QUEST: Ces références sont très nombreuses. Nous citons ci-dessous, à titre d'exemples:

Article princeps du réseau ERIC:

Ghia P., Stamatopoulos K., Belessi C., Moreno C., Stilgenbauer S., Stevenson F., Davi F., Rosenquist R., on behalf of the European research Initiative on CLL (ERIC). ERIC recommendations on IGHV gene mutational status analysis in chronic lymphocytic leukemia. *Leukemia*, 2007, 21, 1-3.

Livre publié en 2009:

'Immunoglobulin Gene Analysis in Chronic Lymphocytic Leukemia' (Ghia P., Rosenquist R. and Davi F. eds), Wolters Kluwer Health Italy Ltd, Milan, Italy, 2009, 8 chapters, 143 pages. ISBN 978-88-7556-377-6.

3 ateliers internationaux sur les séquences des LLC:

Educational Workshop on Immunoglobulin Gene Analysis in Chronic Lymphocytic Leukemia, Uppsala, Sweden (14-15 juin 2007).

2nd Educational Workshop on Immunoglobulin Gene Analysis in Chronic Lymphocytic Leukemia, Paris, France (9-10 octobre 2008).

4.6.3. Humanisation d'anticorps par la méthodologie «IMGT Collier de Perles»

Laboratoire porteur du projet: IMGT®.

Autres laboratoires: Laboratoire Français du fractionnement et des biotechnologies LFB, Les Ulis; Institut de Recherche Pierre Fabre, Centre d'Immunologie Pierre Fabre (CIPF), Saint-Julien en Genevois; Immuno-Pharmaco-Génétique des anticorps thérapeutiques, Université François Rabelais, CHRU de Tours, Tours; Centre de Recherche du Service de Santé des Armées (CRSSA), La Tronche Grenoble (France); Merck & Co., Inc., WhiteHouse Station, New Jersey (Etats-Unis) [licence et contrat]; Centocor Inc. Johnson and Johnson, Malvern, Philadelphia (Etats-Unis) [licence et contrat].

Compléments de financement: Contrats industriels;

Centre de Recherche du Service de Santé des Armées (CRSSA);

Région Languedoc-Roussillon Grands Plateaux Techniques pour la Recherche (GPTR), Grands Equipements pour l'Evolution Technologique et l'Ouverture (GEPETOS) (75.258 €, 01/01/07-31/12/09).

Dans le cadre de la standardisation IMGT-ONTOLOGY, IMGT® a proposé le concept d'une numérotation unique des domaines variables des IG et TR, quel que soit le type de récepteur, quel que soit le type de chaîne et quelle que soit l'espèce.

La numérotation unique IMGT a permis la représentation des domaines V en deux dimensions ou «IMGT Collier de Perles» qui permet de délimiter de manière standardisée les boucles hypervariables (ou *complementarity determining regions*, CDR-IMGT) conférant la spécificité aux IG (ou anticorps) et aux TR et les régions charpente (ou *framework regions*, FR-IMGT) du domaine. La base de structures 3D d'IMGT®, IMGT/3Dstructure-DB, et l'outil d'analyse de structures, IMGT/DomainGapAlign, intègrent la numérotation unique IMGT® et la représentation graphique «IMGT Collier de Perles» qui permet de passer facilement d'une séquence protéique à une structure 3D. Dans le cas d'un domaine variable d'IG ou de TR, les trois boucles hypervariables qui assurent la spécificité de l'anticorps ou du récepteur T vis-à-vis de l'antigène sont aisément visualisées.

Aujourd'hui les IMGT Colliers de Perles ont été adoptés par toute la communauté scientifique, et l'expression est en français dans les publications scientifiques. Ces résultats sont particulièrement importants dans le domaine de l'ingénierie et de l'humanisation des anticorps thérapeutiques. Lorsqu'un anticorps murin est injecté à un patient, une réponse immunitaire se produit contre cet anticorps d'une autre espèce, ce qui réduit son efficacité. L'humanisation consiste à greffer les boucles hypervariables murines, qui confèrent la spécificité, sur un anticorps humain. L'outil IMGT/Collier-de-Perles permet de délimiter précisément les parties à greffer. Grâce à ses «IMGT Colliers de Perles» standardisés, IMGT® propose aux laboratoires une méthodologie où l'anticorps humanisé conserve sa spécificité avec le minimum de risque d'immunogénicité. Ce qui explique l'intérêt des industries pharmaceutiques pour les bases et outils d'IMGT®.

Cette approche a permis de revoir la situation de cinq anticorps commerciaux utilisés en oncologie. Plusieurs laboratoires français et étrangers ont signé des conventions avec IMGT® pour l'humanisation d'anticorps selon le protocole IMGT®. C'est également par la méthode «IMGT Collier de Perles», que le CRSSA a apprécié l'utilisation possible, en thérapeutique humaine, d'anticorps de macaque dirigés contre le bacille du charbon *Bacillus anthracis*.

Publications à titre d'exemples:

Magdelaine-Beuzelin C., Kaas Q., Wehbi V., Ohresser M., Jefferis R., Lefranc M.-P., Watier H. Structure-function relationships of the variable domains of monoclonal antibodies approved for cancer treatment. *Critical reviews in Oncology and Hematology*, 2007, 64, 210-225.

Pelat T., Hust M., Laffly E., Condemine F., Bottex C., Vidal D., Lefranc M.-P., Dubel S. and Thullier P. A high affinity, human-like antibody fragment (scFv) neutralising the lethal factor (LF) of *Bacillus anthracis* by inhibiting PA-LF complex formation. *Antimicrob Agents Chemother.* 2007, 51, 2758-2764.

Laffly E., Danjou L., Condemine F., Vidal D., Drouet E., Lefranc M.-P., Bottex C. and Thullier P. Selection of a macaque Fab with human-like framework regions, high affinity, and that neutralizes the protective antigen (PA) of Bacillus anthracis. *Antimicrob. Agents Chemother.* 2005, 49, 3414-3420.

4.6.4. ImmunoGrid

Laboratoire porteur du projet: Consortium Interuniversitario per la Gestion del centro di Calcolo Elettronico dell'Italia Nord Orientale, CINECA, Bologne (Italie).

Autres laboratoires: IMGT® CNRS UPR 1142 (France); Université de Bologne (Italie); Université de Queensland (Australie); Université de Catania (Italie); Danmarks Tekniske Universitet (Danemark); Birkbeck College, Université de Londres (Angleterre); Consiglio Nazionale dell Ricerche, CNR, Rome (Italie).

Compléments de financement: contrat Européen IST-2004-028069 (221.223 €, 01/02/06-31/01/09); ACI Informatique, Mathématiques, Physique en Biologie ACI IMPBio IMP82-2004 (135.000 €, 20/09/04-10/09/07); Agence Nationale de la Recherche ANR-06-BIOSYS-005-01 (240.000 €, 11/12/06-10/01/11).

IMGT® est partenaire du projet européen ImmunoGrid. La finalité du projet, développé dans le domaine de l'immunoinformatique, correspond à l'implémentation du système immunitaire humain virtuel (*Virtual Human Immune System* ou VHIS) en utilisant la technologie des grilles. Ce système virtuel doit permettre de simuler les processus de la réponse immunitaire naturelle. Les outils informatiques développés pourront s'intégrer dans les projets d'immunologie clinique, de conception de vaccins et d'immunothérapie pour le traitement du cancer et des infections chroniques.

L'implication d'IMGT® dans ce projet porte sur la standardisation des connaissances et la définition des concepts fondamentaux en immunogénétique et immunoinformatique qui sont utilisés pour la modélisation des composants et des processus du VHIS. Comme cela a été décrit précédemment, IMGT® a élaboré IMGT-ONTOLOGY, la première ontologie dans le domaine de l'immunogénétique et de l'immunoinformatique, basée sur 7 axiomes "IDENTIFICATION", "CLASSIFICATION", "DESCRIPTION", "NUMEROTATION", "LOCALIZATION", "ORIENTATION" et "OBTENTION". Ces axiomes postulent que les composants et les processus du système immunitaire ainsi que leurs relations doivent être identifiés, décrits, classés, numérotés, localisés et orientés. Ces axiomes constituent l'ontologie formelle d'IMGT-ONTOLOGY ou IMGT-Kaleidoscope. Lors d'une première étape, IMGT® a établi les règles et les principes de standardisation nécessaires à la modélisation des composants moléculaires du VHIS. IMGT® a ensuite défini et précisé les concepts d'identification, de description et de classification des récepteurs d'antigènes (IG et TR) et du MHC. IMGT® a également défini le vocabulaire et les principes de numérotation des domaines de type V et C des IG et TR et des domaines de type G du MHC qui tiennent compte des données structurales et de la position des acides aminés conservés. La structuration de ces connaissances sous forme de concepts et de relations est à la base de l'analyse pour une description standardisée des interactions entre ligands et récepteurs. Cette étape est primordiale pour l'étude des spécificités (reconnaissance de l'antigène, caractérisation des épitopes des cellules B et T, spécificité des IG et TR, utilisation de peptides en vaccinologie et immunothérapie, anticorps humanisés et humains) et pour l'étude des interactions de protéines IgSF et MhcSF.

Par ailleurs, la standardisation des connaissances en immunogénétique réalisée par IMGT® dans le projet ImmunoGrid est implémentée dans toutes les bases de données, ressources Web et outils d'IMGT®. Cette standardisation permet de fédérer les bases de données, ressources et outils d'IMGT® en vue de répondre à des questions biologiques complexes, selon trois approches biologiques: génomique, génétique et structurale. Cet ensemble constitue IMGT-Choreography (ACI IMPBio IMP82-2004). L'implémentation d'IMGT-ONTOLOGY en langage OWL en vue de sa distribution est en cours d'édition (éditeur d'ontologie Protégé <http://protege.stanford.edu/>, Stanford Center for Biomedical Informatics Research, Californie, Etats-Unis) (ANR-06-BIOSYS-005-01).

Publications à titre d'exemples

Duroux P., Kaas Q., Brochet X., Lane J., Ginestoux C., Lefranc M.-P. and Giudicelli V. IMGT-Kaleidoscope, the Formal IMGT-ONTOLOGY paradigm. *Biochimie*, 2008, 90, 570-583.

5. Identification de l'autonomie et/ou de la dépendance extérieure des parties du système IMGT®

Les principales sources de données extérieures sont EMBL-Bank à l'EBI (en ce qui concerne les séquences nucléotidiques d'IMGT/LIGM-DB) et RCSB PDB (en ce qui concerne les données structurales d'IMGT/3Dstructure-DB) (Tableau 5). Les bases de données et les outils IMGT® sont hébergés par le CINES (Centre Informatique National de l'Enseignement Supérieur) à Montpellier, pour leur mise à disposition.

Tableau 5. Autonomie et/ou dépendance extérieure des parties du système IMGT®

| Données | Bases de données et outils | Bases de données | Outils en ligne | Autonomie complète sur l'origine des données | Origine interne des données | Dépendances externes |
|--------------|----------------------------|------------------|-----------------|--|-------------------------------------|------------------------|
| génétiques | IMGT/LIGM-DB | * | | - | - | EMBL-Bank (EBI) |
| | IMGT/Primer-DB | * | | - | IMGT/LIGM-DB IMGT/GENE-DB | |
| | IMGT/CLL-DB | * | | - | IMGT/V-QUEST | IMGT/CLL-DB Group |
| | IMGT/V-QUEST | | * | oui | IMGT/GENE-DB | |
| | IMGT/JunctionAnalysis | | * | oui | IMGT/ GENE-DB | |
| | IMGT/PhyloGene | | * | oui | IMGT/GENE-DB | |
| | IMGT/Allele-Align | | * | oui | - | |
| génomiques | IMGT/GENE-DB | * | | oui | IMGT/LIGM-DB | |
| | IMGT/LocusView | | * | oui | IMGT/GENE-DB | |
| | IMGT/GeneView | | * | oui | IMGT/GENE-DB | |
| | IMGT/GeneSearch | | * | oui | IMGT/GENE-DB | |
| | IMGT/CloneSearch | | * | oui | IMGT/GENE-DB | |
| | IMGT/GenelInfo | | * | - | IMGT/GENE-DB | TIMC-ICH (Grenoble) |
| | IMGT/GeneFrequency | | * | oui | IMGT/LIGM-DB IMGT/GENE-DB | |
| structurales | IMGT/3Dstructure-DB | * | | - | Annotations IMGT® | RCSB (PDB) |
| | IMGT/StructuralQuery | | * | oui | IMGT/3Dstructure-DB | |
| | IMGT/DomainSuperimpose | | * | oui | IMGT/3Dstructure-DB | |
| | IMGT/2Dstructure-DB | * | | - | IMGT/3Dstructure-DB IMGT/GENE-DB | Programme WHO/INN |
| | IMGT/Collier-de-Perles | | * | oui | - | |
| | IMGT/DomainGapAlign | | * | oui | IMGT/2Dstructure-DB | |
| | IMGT/DomainDisplay | | * | oui | IMGT/2Dstructure-DB | |

6. Projets futurs et analyse des éventuels besoins d'infrastructures extérieures

6.1. Projets futurs de la structure IMGT®

Tout en poursuivant les projets décrits dans la section 4.6, de nouvelles méthodologies sont développées en priorité pour maintenir la compétitivité internationale d'IMGT®:

- l'annotation automatique de grandes séquences génomiques des IG et des TR par le développement d'un logiciel IMGT/LIGMotif qui prend en compte les concepts d'IMGT-ONTOLOGY.
- l'analyse haut débit des séquences réarrangées des IG et des TR par IMGT/V-QUEST version haut débit ou IMGT/HighV-QUEST.

En parallèle, nous mettons en place une démarche qualité au sein d'IMGT en vue de la certification ISO 9001:2008.

6.1.1. Annotation automatique de grandes séquences génomiques des locus IG et des TR: IMGT/LIGMotif

L'annotation par IMGT® de séquences génomiques des locus codant les IG et les TR répond à une demande accrue des laboratoires de recherche et des industriels travaillant dans des domaines aussi variés que la recherche fondamentale, la recherche médicale et vétérinaire. Ces laboratoires souhaitent des données génomiques et géniques précises et de qualité afin, d'une part, d'analyser le répertoire des IG et TR de sujets sains ou malades (maladies autoimmunes, maladies infectieuses...) ou de modèles animaux, et, d'autre part, de comparer la diversité génomique et étudier l'évolution de la réponse immunitaire adaptative. Ces données d'immunogénétique sont également un outil précieux pour les laboratoires impliqués dans les projets HUPO, l'ingénierie des anticorps, le diagnostic des leucémies et des lymphomes et l'établissement d'approches immunothérapeutiques.

L'annotation par IMGT® des locus IG et TR des génomes nouvellement séquencés ou en cours de séquençage est donc très attendue au niveau international. Une nouvelle méthodologie est mise en place au sein de l'équipe IMGT® afin de faire face à la grande quantité de données (génomes complets de plus en plus nombreux et fréquents), tout en garantissant la qualité et la précision des données.

IMGT/LIGMotif

L'annotation d'un locus IG ou TR ne peut pas faire appel aux logiciels et programmes classiques étant donné la structure spécifique de ces gènes et la difficulté d'interprétation de leur organisation en familles multigéniques avec de nombreux pseudogènes. Afin de faciliter le travail d'annotation, nous avons développé un prototype IMGT/LIGMotif qui permet l'identification des gènes et l'annotation de grandes séquences génomiques. IMGT/LIGMotif peut être utilisé via l'interface Web IMGT ou en ligne de commande. De plus, il offre une visualisation graphique des résultats via le logiciel Artémis (<http://www.sanger.ac.uk/Software/Artemis/>)

L'objectif est, dans un premier temps, de valider ce programme sur des séquences de 700 kb. Après validation, nous mettrons en place une analyse à haut débit sur des lots de séquences génomiques pouvant atteindre 1 Mégabase (1Mb), ce qui nécessitera le développement d'un service Web efficace.

L'outil devra être optimisé pour:

- l'utilisation des ressources informatiques (mémoire).
- le temps d'exécution
- la gestion du chargement, l'analyse des séquences et la restitution des résultats.
- la traçabilité des différentes requêtes soumises.

Genres et espèces concernés dans le projet

Les genres et espèces concernés par l'annotation des locus IG et TR correspondent d'abord à ceux dont le génome a été nouvellement séquencé et pour lesquels il y a une forte demande. Il s'agit du chien (*Canis lupus familiaris*) en tant que modèle animal, du rat (*Rattus norvegicus*) très utilisé pour l'obtention d'anticorps monoclonaux, du lapin (*Oryctolagus cuniculus*) largement utilisé comme modèle d'étude et pour la production d'anticorps. Ensuite nous analyserons les locus IG et TR des autres genres et espèces dont les génomes auront été séquencés.

Dans un premier temps nous choisirons des locus de petite taille (locus TRG du chien 700 kb, locus IGL du rat 450 kb) que nous annoterons de façon manuelle et en parallèle avec la version test de IMGT/LIGMotif. Cela nous permettra d'évaluer l'efficacité de IMGT/LIGMotif (estimation du gain de temps, qualité des annotations obtenues, facilité d'utilisation...).

Dans un deuxième temps, nous testerons les performances d'IMGT/LIGMotif en terme d'identification de gènes et d'annotations sur un grand locus de plusieurs Mégabases. Nous choisirons le locus IG Heavy (IGH) du rat qui couvre 5 Mb sur le chromosome 6 (6q31,32) et dans lequel nous avons localisé et identifié 232 séquences dont 119 d'intérêt. L'annotation manuelle montre que ces séquences contiennent 342 gènes IGHV, 20 gènes IGHD, 4 gènes IGHJ et 8 gènes IGHC.

Si la qualité des résultats est satisfaisante selon les critères IMGT®, nous procéderons directement à l'annotation automatique du locus IG Kappa (IGK) du rat dont la taille est très élevée (environ 7 Mb) et nous poursuivrons avec les locus IGH, IGK et IGL du lapin.

IMGT/LIGMotif facilitera le processus d'annotation d'IMGT/LIGM-DB. Les étapes ultérieures: entrée des données dans IMGT/GENE-DB, création d'IMGT Répertoire (Alignments of alleles, Tables of alleles, Gene tables, Protein displays, IMGT Colliers de Perles) font l'objet d'une recherche d'automatisation (scripts PHP, liens vers les bases de données, fichiers CSV...) tout en respectant les concepts et labels d'IMGT-ONTOLOGY. Cela devrait représenter un gain de temps considérable et fournir une qualité d'annotation équivalente à celle d'annotations manuelles.

6.1.2. Analyse haut débit des séquences réarrangées des IG et TR: IMGT/HighV-QUEST

IMGT/V-QUEST ('V-QUERy and STandardization') analyse les séquences nucléotidiques réarrangées des IG et TR en tenant compte de la structure particulière des domaines V des récepteurs d'antigènes (IG et TR). Il permet d'identifier précisément les gènes et les allèles impliqués dans les séquences nucléotidiques IG et TR, de caractériser les mécanismes complexes qui interviennent lors du réarrangement de ces gènes (N-diversité, mutations somatiques) et propose une analyse détaillée de la jonction (en intégrant l'outil IMGT/JunctionAnalysis). IMGT/V-QUEST est l'outil le plus utilisé d'IMGT®. Il offre un choix très vaste d'options et de résultats à l'utilisateur. Il est disponible pour l'analyse des séquences humaines et murines et de quelques espèces de primates, chondrichthyens et téléostéens. Les capacités actuelles d'IMGT/V-QUEST en ligne sont de 50 séquences par lot.

Compte tenu de l'accroissement total de ces ressources en amont (augmentation des données du séquençage: *Next-Gen sequencing technology*), les laboratoires de recherche sont en mesure de séquencer l'ensemble du répertoire IG et TR exprimé chez des individus sains ou dans le cas de pathologies. IMGT® reçoit de nombreuses demandes de laboratoires publics et d'industriels pour l'analyse simultanée de centaines de milliers de séquences.

Afin de répondre à ce besoin, LIGM a entrepris le développement d'IMGT/V-QUEST haut débit (IMGT/HighV-QUEST). La production de données de haut débit et de haute qualité permettra à IMGT® de maintenir sa compétitivité vis-à-vis du monde de l'université et de l'industrie.

IMGT/V-QUEST haut débit (IMGT/HighV-QUEST)

La mise en place d'IMGT/V-QUEST haut débit (IMGT/HighV-QUEST) va de pair avec une réorganisation intrinsèque de l'outil pour optimiser ses performances et l'intégration de nouvelles fonctionnalités inhérentes à l'analyse d'un grand nombre de séquences.

Cette étape nécessite le développement de méthodologies spécifiques, de contraintes de sécurité et de confidentialité des données pour:

- la soumission de fichiers volumineux contenant les séquences (services Web ou programmes clients avec procédures d'authentification, de contrôle de validité des données).
- l'analyse des données en mode différé sur un ou plusieurs serveurs de calcul. Elle nécessite l'installation et la configuration du serveur de calcul, la création de procédures de soumission, d'organisation et de surveillance des jobs et des opérations. IMGT® a obtenu l'attribution de ressources du Grand Equipement National de Calcul Intensif (GENCI) pour du calcul intensif au Centre Informatique National de l'Enseignement Supérieur (CINES), soit 70.000 heures de calcul en 2009). Une nouvelle demande d'attribution de ressources informatiques (DARI) au GENCI a été faite en novembre 2009 pour l'année 2010.
- la restitution des résultats: procédures de stockages, mise à disposition pour les utilisateurs, contrôle des accès, procédures de nettoyage de l'espace de stockage.

Nous attendons d'IMGT/HighV-QUEST qu'il facilite en aval l'exploitation statistique des résultats, et permette l'étude de la réponse immunitaire adaptative à grande échelle en corrélation avec l'expression de gènes IG et TR.

Prototype, matériel et logiciels

Prototype: Le prototype comprend:

1. La mise en place d'un service Web pour la soumission des séquences par des utilisateurs ou des programmes clients:
 - il devra authentifier l'utilisateur (par son adresse e-mail) afin de lui proposer un suivi des analyses en cours et l'avertir lorsque les résultats seront disponibles. Des fonctions de contrôle de l'authentification seront nécessaires.
 - la soumission des séquences s'effectuera par le chargement d'un fichier regroupant l'ensemble des séquences en format FASTA. Leur format devra être systématiquement contrôlé.
 - il enregistrera les paramètres d'entrée nécessaires à l'analyse de IMGT/V-QUEST et les paramètres de présentation des résultats souhaités par l'utilisateur.
2. L'installation et la configuration du serveur de calcul (création des procédures de soumission de séquences en 'batch', organisation des jobs et surveillance du système des opérations).
3. Le développement des procédures d'enregistrement des caractéristiques des requêtes et des analyses pour le suivi et les statistiques internes d'utilisation (nombre de séquences soumises dans la requête, taille moyenne des séquences...).
4. Le développement des procédures pour la restitution des résultats. Les résultats seront stockés temporairement sur un serveur de fichier. L'utilisateur sera averti par e-mail de leur disponibilité et de la date à laquelle ils seront effacés. La mise en place de contraintes de sécurité pour l'accès aux résultats sera indispensable pour respecter la confidentialité des données: l'accès devra être limité aux propriétaires des données.

Matériel et logiciels: Les moyens matériels et logiciels nécessaires prévus comprennent:

- Serveur de développement, environnement Eclipse pour le développement Java.
- Serveur Web, Apache/Tomcat d'application, Web service, technologie RMI.
- Serveur de calcul (parallélisé), technologie MPI, la possibilité d'utiliser les grilles de calcul sera envisagée.
- Serveur de fichiers pour le stockage temporaire des résultats.
- Serveur de messagerie.
- Serveur SSH/FTP pour la récupération des résultats.
- Outils de surveillance système, outils statistiques et gestionnaires de bases de données pour le suivi interne et l'historique des analyses.

6.1.3. La démarche qualité à IMGT®

En tant que système d'information intégré dans le domaine de l'immunogénétique et de l'immunoinformatique, IMGT® répond à la demande des utilisateurs (clients), répartis à parts égales entre les Etats-Unis et Canada, l'Europe et le reste du monde. Les ressources d'IMGT® (outils, bases de données, pages Web) sont utilisées par des chercheurs d'équipes académiques

et industrielles: en recherche fondamentale, en recherche médicale, en recherche vétérinaire, en génomique, pour les diagnostics, en biotechnologie relative aux projets HUPO et à l'ingénierie des anticorps et dans les approches thérapeutiques. IMGT® s'inscrit ainsi dans différentes «missions» liées les unes aux autres qui permettent de fournir un ensemble de ressources et de services variés aux utilisateurs. La politique d'IMGT® est de mettre en place une Démarche Qualité, en respectant les normes en vigueur dans les domaines scientifiques et ceux du système de management (normes ISO 9001:2008) et en les appliquant à nos procédures de travail. Ce travail passe par la définition et la description de nos utilisateurs et de nos procédures (processus), tout en respectant la qualité des données et les exigences des utilisateurs en matière de qualité. La rédaction d'un manuel qualité et la certification sont parmi les objectifs de ce projet.

IMGT® a su intégrer différentes attestations de conformité en regard des diverses réglementations :

- Au niveau scientifique, IMGT® suit les règles de la nomenclature internationale et est le site de référence de *Human Genome Organisation (HUGO) Nomenclature Committee (HGNC)* pour les IG et TR.
- Au niveau du site Web, IMGT® a le contrôle de qualité HONcode (*Health on the Net foundation (HON) Code of Conduct for medical and health Web sites*), depuis le 26 juillet 1999.
- Au niveau des ontologies, tout le personnel d'IMGT® est formé aux critères de la Charte Scientifique d'IMGT-ONTOLOGY. Une spécification formelle des termes utilisés en immunogénétique, ainsi que de leurs relations, a été développée par le Laboratoire d'ImmunoGénétique Moléculaire (LIGM) afin d'assurer la précision, la consistance et la cohérence d'IMGT®.
- Au niveau de l'Organisation Mondiale de la Santé (OMS), Marie-Paule Lefranc est expert du programme de Dénomination Commune Internationale (DCI) des Substances pharmaceutiques (*World Health Organization/International Nonproprietary Name (WHO/INN) programme*).
- IMGT® est responsable du comité de nomenclature des IG et TR au sein du comité de nomenclature de l'OMS-Union Internationale des Sociétés d'Immunologie (*WHO-International Union of Immunological Societies, IUIS*). Le rapport 2007 a été approuvé au congrès International en Immunologie à Rio de Janeiro, Brésil (août 2007) et publié dans *Immunogenetics (2007)* et *Developmental Comparative Immunology (2008)*.

Toutes les personnes de l'équipe IMGT®, travaillent en collaboration selon 5 grands axes (ou groupes de travail) scientifiques et de recherche :

- Expertise scientifique,
- Annotations,
- Développement d'outils (bioinformatique, informatique) et de bases de données,
- Maintenance informatique,
- Développement Web et communication.

A partir de ces 5 axes, nous avons, en premier lieu, dégagé 3 grands processus généraux qui nous permettent de définir notre démarche qualité à IMGT® :

- Processus de management (stratégie: ressources humaines, système qualité: système d'information, suivi des indicateurs).
- Processus de réalisation ou opérationnels, très nombreux et très variés parmi lesquels le processus d'annotation des fichiers à plat (flat file) de séquences, de gènes et de structures 3D, le développement de logiciels (IMGT/V-QUEST...), la gestion de bases de données.
- Processus support, formations, expertise scientifique...

IMGT® reconnu en tant que standard international dans son domaine scientifique souhaite, par sa démarche qualité, satisfaire à la fois aux exigences de ses utilisateurs et partenaires et à la réglementation en vigueur, tout en poursuivant sa démarche de recherche scientifique et de créativité.

6.2. Analyse des besoins d'infrastructures et de postes d'ingénieurs

6.2.1. Reconnaissance d'IMGT® en tant que nœud thématique du projet européen ELIXIR

La mission d'ELIXIR (*European Life Sciences for Biological Information Infrastructure*) est de construire et d'exploiter une infrastructure durable pour l'information biologique en Europe afin de soutenir la recherche en sciences du vivant et ses applications à la médecine et à l'environnement, aux bio-industries et à la Société. Un ensemble de workprogrammes (WP) a été mis en place pour étudier la faisabilité du projet et IMGT® participe à deux d'entre eux, WP2 (*ELIXIR Strategy for Data Resources Committee*) et WP7 (*Data Integration & Interoperability*). Un ensemble de 'nœuds nationaux' et de 'nœuds thématiques' est prévu dans cette infrastructure européenne et il est évident qu'IMGT® y a sa place en tant que 'nœud thématique'.

La collaboration d'IMGT® au niveau européen avec EMBL (EBI) existe depuis 1990, c'est-à-dire depuis près de 20 ans, et remonte quasiment à la création d'IMGT®. IMGT® a été coordinateur de trois projets européens de 1993 à 2003 (BIOMED2, BIOTECH, 5th PCRDT) dans lesquels EMBL-EBI a été partenaire. La structure IMGT® est reconnue comme le fleuron de l'Europe en immunogénétique et immunoinformatique. Cette reconnaissance est effective et IMGT® figure en tant que nœud thématique proposé sur le document *Stakeholders Meeting* d'ELIXIR (Copenhague, 20 mai 2009). Cette reconnaissance d'IMGT® est un atout pour le positionnement international de la recherche française en immunoinformatique.

Afin de garantir cette visibilité sur le long terme, il est crucial cependant que le Ministère fasse de son côté la démarche officielle de proposer IMGT® en tant que nœud thématique d'ELIXIR (la démarche ne pouvant venir que des gouvernements nationaux).

L'intégration d'IMGT® en tant que «nœud thématique» d'ELIXIR est indispensable afin de:

- pérenniser la structure IMGT®,
- préserver les partenariats avec les bases généralistes EMBL-Bank et Ensembl (EBI), Vega (Wellcome Trust Sanger Institute) et UniProt (Swiss Institute of Bioinformatics (SIB)),
- maintenir IMGT®, en tant que référence internationale en immunogénétique et immunoinformatique.

IMGT®, système international en immunoinformatique et immunogénétique, apporte une expertise unique pour les données génétiques, génomiques et structurales des molécules de la réponse immunitaire adaptative, et ce depuis 20 ans.

6.2.2. Nécessité de pérenniser IMGT® par la création de postes d'ingénieurs

La création d'au moins quatre postes statutaires d'ingénieurs d'études en analyse de données (en 2010-2011) et d'un poste IR en bioinformatique est indispensable pour préserver, au sein d'IMGT®, l'expertise acquise au cours des années par les biocurateurs, annotateurs et bioinformaticiens qui assurent actuellement ces fonctions. Ces recrutements sont cruciaux pour assurer le maintien d'IMGT® en France et pérenniser la structure. Cette structure est vivante, unique dans le monde, internationalement reconnue, capable de drainer du financement des industriels et des contrats de recherche pour assurer en partie le financement de son personnel mais **cette dynamique ne peut être maintenue que si l'expertise actuelle est préservée par des titularisations.**

Abréviations

2D: bidimensionnel(le).

3D: tridimensionnel(le).

ACCITH: Anticorps et ciblage thérapeutique, GDR CNRS n°3260.

ACI: Action Concertée Incitative

Ag: antigène.

ANR: Agence Nationale de Recherche (<http://www.agence-nationale-recherche.fr/>).

API: Application Programming Interface.

APP: Agence pour la Protection des Programmes (<http://app.legalis.net/>).

ARSA: All-round Retrieval of Sequence and Annotation.

BiM: Bioinformatique Moléculaire, GDR CNRS n°3003 (<http://www.gdr-bim.u-psud.fr/>).

BLAST: Basic Local Alignment Search Tool.

C: constant.

CDR: complementarity determining region.

CEA: Commissariat à l'énergie atomique (<http://www.cea.fr/>).

CINECA: Consortium Interuniversitario per la Gestione del centro di Calcolo Elettronico dell'Italia Nord Orientale

CINES: Centre Informatique National de l'Enseignement Supérieur (<http://www.cines.fr/>).

CLL: Chronic lymphocytic leukemia (LLC en français).

CMH: Complexe majeur d'histocompatibilité (MHC en anglais).

CNRS: Centre National de la Recherche Scientifique (<http://www.cnrs.fr/>).

CNUSC: Centre National Universitaire Sud de Calcul, devenu CINES.

CPU: Central Processing Unit.

CRSSA: Centre de Recherche du Service de Santé des Armées.

CSS: Cascading Style Sheets

CU: Columbia University.

CVS: Concurrent Versions System.

D: Diversité (diversity en anglais).

DARI: Demande d'Attribution de Ressources Informatiques (<https://www.edari.fr/>).

DAS: Distributed Annotation System.

DCI: Dénomination Commune Internationale (INN en anglais).

DDBJ: DNA Data Bank of Japan (<http://arsa.ddbj.nig.ac.jp/>).

DKFZ: Deutsche Krebsforschungszentrum.

DPI: Direction Politique Industrielle.

DR: Délégation régionale.

EBI: European Bioinformatics Institute (<http://www.ebi.ac.uk/>).

EFIS: European Federation of Immunological Societies.

EMBL: European Molecular Biology Laboratory.

ELIXIR: European Life sciences Infrastructure for Biological Information (<http://www.elixir-europe.org/page.php?page=home>).

ERIC: European Research Initiative on CLL (<http://www.ericll.org/>).

EU: European Union (UE en français).

Fab: Fragment antigen binding.

FAQ: Frequently asked questions.

FIST SA: France Innovation Scientifique et Transfert (<http://www.fist.fr/>).

FPIA: Fusion Protein for Immune Applications.

FP: Framework Program.

FR: Framework region.

FTP: File Transfer Protocol.

G: groove (sillon en français).

GDR: Groupement de Recherche.

GEPETOS: Grands Equipements pour l'Evolution Technologique et l'Ouverture Scientifique.

GENCI: Grand Equipement National de Calcul Scientifique.

GIS: Groupement d'Intérêt Scientifique.

Go: Giga-octet.

GPTR: Grand Plateau Technique pour la Recherche.

GSO: Grand Sud-Ouest.

HGM: Human Gene Mapping.

HGNC: The HUGO Nomenclature Committee (<http://www.genenames.org/>).

HLA: Human leukocyte antigen.

HONcode: Health On the Net foundation Code of Conduct for medical and health Web sites (<http://www.hon.ch/HONcode/French/>).

HTML: Hypertext Markup Language

HUGO: Human Genome Organisation (<http://www.hugo-international.org/>).

HUPO: Human Proteome Organisation (<http://www.hupo.org/>).

IAC: International Advisory Committee.
ICRF: Imperial Research Cancer Foundation, devenu Cancer Research.
IBiSA: Infrastructures en Biologie Santé et Agronomie (<http://www.ibisa.net/>). Coordination des plates-formes de recherches en sciences du vivant.
ICH: Laboratoire d'Immunochimie
ICM: International Collaborative Meeting.
IFR: Institut Fédératif de Recherche.
IG: Immunoglobuline (ou anticorps).
IGH: Institut de Génétique Humaine.
IgSF: Immunoglobulin superfamily (superfamille des immunoglobulines en français).
IMGT: IMGT®, the international ImMunoGeneTics information system® (<http://www.imgt.org>)
IMIA: International Medical Informatics Association (<http://www.imia.org/>).
INN: International Nonproprietary Name (DCI en français) <http://www.who.int/>.
INRA: Institut National de la Recherche Agronomique <http://www.inra.fr/>.
INSDC: The International Nucleotide Sequence Database Collaboration.
INSERM: Institut National de la Santé et de la Recherche Médicale.
Isql: Interactive structured query langage.
IST: Information Society Technologies.
IUF: Institut Universitaire de France.
IUIS: International Union of Immunological Societies.
J: Jonction (joining en anglais).
LIGM: Laboratoire d'ImmunoGénétique Moléculaire (<http://www.imgt.org/textes/IMGTinformation/LIGM.html>).
LIRMM : Laboratoire d'Informatique, de Robotique et de Micro-électronique de Montpellier
LLC: Leucémie lymphoïde chronique (CLL en anglais).
mAb: Monoclonal antibody.
MHC: Major histocompatibility complex (CMH en français).
Mb: Mégabase.
MESR: Ministère de l'Enseignement Supérieur et de la Recherche.
MhcSF: MHC superfamily (superfamille des CMH en français).
MISTRD: Mission de l'Information Scientifique et Technique et du Réseau Documentaire.
NCBI: National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>).
Next-Gen: Next-generation sequencing technology.
NFS: Network File System.
NIAS: National Institute of Agrobiological Sciences.
NIG: National Institute of Genetics.
nt : Nucleotide.
OMS: Organisation Mondiale de la Santé (WHO en anglais).
OWL: Web Ontology Language
PDB: Protein Data Bank.
PCR: Polymerase chain reaction.
PCRDT: Programme cadre pour la recherche et le développement technologique (FP en anglais).
PHP: Hypertext Preprocessor
PPF: Plan Pluri-Formation.
PME: petites et moyennes entreprises.
pMHC: peptide/MHC.
r5: technologie de transfert sécurisé de fichiers.
RCSB: Research Collaboratory for Structural Bioinformatics (<http://www.rcsb.org/>).
ReNaBi: Réseau National des plates-formes Bioinformatique (<http://www.renabi.fr/>).
RFLP: Restriction fragment length polymorphism.
RIO: Réunion Inter-Organismes.
RPI: Related Proteins of the Immune system
scFv: single chain Fragment variable.
Scp: SSH CoPy.
SIB: Swiss Institute of Bioinformatics.
SO: Sequence Ontology (<http://www.sequenceontology.org/>).
SQL: Structured Query Language.

SRS: Sequence Retrieval System.
SSH: Secure Shell (programme informatique et protocole de communication sécurisé).
TIMC: Laboratoire technique de l'Imagerie, de la Modélisation et de la Cognition
To: Téra-octet.
TR: Récepteur T (T cell receptor en anglais).
UPR: Unité Propre de Recherche.
UE: Union Européenne (EU en anglais).
US NLM: United States National Library of Medicine (<http://www.nlm.nih.gov/>).
V: variable.
Vega: Vertebrate Genome Annotation.
VHIS: Virtual Human Immune System.
WHO: World Health Organization (OMS en français).
WP: workprogramme.