

## REPONSE A L'APPEL A PROJETS LORIA (29 juin 2020)

### HLA-3D-Diff

**Titre du projet :** Approches de géométrie différentielle pour la représentation des variations locales de structure 3D des protéines du système humain de compatibilité tissulaire (HLA).

**Equipes LORIA concernées :** CAPSID, PIXEL

**Participants :** Marie-Dominique Devignes (Capsid), Diego Amaya-Ramirez (doctorant, Capsid), Etienne Corman (Pixel), Pr. Jean-Luc Taupin (Hôpital Saint-Louis, Paris)

### **Contexte scientifique**

Dans le monde vivant, les protéines sont les acteurs principaux des fonctions biologiques. Les outils de la bioinformatique structurale permettent de les appréhender comme des objets tri-dimensionnels (3D) complexes dont la forme extérieure dépend du repliement dans l'espace d'une chaîne linéaire d'acides aminés (ou séquence protéique)<sup>1</sup>. De plus, la surface de cet objet 3D porte les caractéristiques physico-chimiques des acides aminés qui se trouvent à la périphérie de la structure 3D de par le repliement de la séquence protéique. L'activité d'une protéine dépend directement de sa forme, qui elle-même dépend de la nature et de l'ordre des acides aminés dans la séquence protéique. Tout remplacement d'un acide aminé par un autre peut donc induire un changement de forme et de propriétés physico-chimiques de la protéine et donc un changement de son activité.

Le système HLA (Human Leukocyte Antigen), qui gouverne la compatibilité tissulaire en transplantation et greffe (avec celui infiniment plus simple des groupes sanguins ABO), constitue le cadre applicatif de ce projet. Il comprend un ensemble de gènes codant pour des protéines appelées antigènes HLA et regroupées en familles (A, B, C, DR, DQ et DP). Chaque individu possède 16 à 20 antigènes HLA différents, parmi un total de plus de 15000 variants génétiques (ou **allèles**) différant pour au moins un acide aminé au sein de l'espèce ( ~80% de ces allèles sont rares donc peu rencontrés). Le système HLA est responsable du rejet par immunisation du Receveur contre les protéines HLA différentes du Donneur et l'incompatibilité HLA est la principale cause de perte des greffons (1 ; Annexe 1).

### **Problème à résoudre et objectif du projet.**

Il est aujourd'hui très difficile de décrire et de comprendre les différences de formes et de propriétés physico-chimiques des protéines, surtout lorsqu'elles sont locales et bruitées. Or ces différences géométriques sont la clé de compréhension des différences dans l'activité des protéines. Le projet vise à explorer de nouvelles façons de représenter les différences entre les modèles 3D des allèles d'une même famille d'antigènes HLA, en vue d'apprendre à reconnaître les paires d'allèles qui présentent des différences déterminantes pour l'incompatibilité tissulaire en transplantation ou greffe. La diversité des allèles HLA et leur très grand nombre (2) rend impossible actuellement de prédire parmi des allèles HLA différents lesquels sont les plus proches du point de vue de leur

---

1 Il existe 20 acides aminés différents qui constituent comme une sorte d'alphabet pour les séquences protéiques.

capacité à déclencher une réponse anticorps dirigée contre le donneur (DSA pour « Donor-Specific Antibody »), autrement dit de comprendre ce qui définit la puissance immunologique d'un allèle pour déclencher la production d'un DSA et estimer la toxicité de ce DSA.

Il s'agit donc dans le projet HLA-3D-Diff de combiner une expertise en modélisation 3D des protéines (équipe CAPSID) avec une expertise en géométrie différentielle (équipe PIXEL), autour d'un problème d'apprentissage automatique dans le domaine de la transplantation (Hôpital Saint-Louis, Paris).

## **Opportunité pour un projet exploratoire LORIA.**

Une thèse Inria-Inserm (Diego Amaya Ramirez) est en cours depuis octobre 2019 dans l'équipe CAPSID (MD Devignes), en collaboration avec l'hôpital Saint-Louis à Paris, dont le Laboratoire d'Histocompatibilité gère 30% de toutes les greffes en France (Pr Jean-Luc Taupin). Le but de la thèse est d'identifier les déterminants structuraux des rejets de greffe pour améliorer la compatibilité entre les receveurs et les donneurs de greffes. L'hôpital Saint-Louis dispose depuis 2008 des données cliniques d'évolution des greffons pour 15000 greffes, de 150 000 résultats de recherche de DSA pour ces Receveurs, ainsi que des données de groupage HLA de ces Receveurs et de leurs Donneurs. Ce laboratoire (parmi les 30 du territoire national) est le plus grand d'Europe (1800 greffes par an) et peut-être du monde, ce qui assure un recrutement unique de situations immunologiques intéressantes pour le projet. Dans l'équipe CAPSID, Diego Amaya Ramirez travaille à construire les modèles 3D pour environ 200 séquences d'allèles HLA étudiés en « single antigen » Luminex<sup>2</sup>. Pour comprendre pourquoi un Receveur s'est révélé compatible ou non avec tel ensemble d'antigènes HLA d'un Donneur, il faut être capable de représenter les différences locales de structure 3D entre les antigènes HLA reconnus par les DSA du Receveur et les antigènes HLA du Donneur. Nous nous sommes donc rapprochés d'Etienne Corman dans l'équipe PIXEL pour étudier avec lui, dans le cadre d'un projet exploratoire Loria, l'intérêt possible de l'utilisation des méthodes de la géométrie différentielle (paramétrisation, cartes fonctionnelles) pour résoudre ce problème.

## **Originalité et nouveauté de l'approche**

### **1. Du point de vue de la recherche en transplantation**

La compatibilité HLA est définie depuis les débuts de la transplantation sur une nomenclature dite sérologique, selon laquelle les antigènes HLA ont été classés en fonction de leur reconnaissance ou absence de reconnaissance par des anticorps de référence de spécificité définie, sélectionnés depuis les années 1950. Cette classification sérologique, toujours en vigueur, est imparfaite et parfois très éloignée de la réalité immunologique.

L'analyse ADN (par séquençage) a permis d'explorer le polymorphisme génétique HLA (les allèles) et d'identifier les 25000 allèles connus à ce jour (dont >15000 ayant un changement d'au moins un acide aminé). Depuis 10 ans des approches d'alignements de séquences reposant sur des modèles structuraux HLA ont permis de d'établir un registre d'éplets<sup>3</sup> candidats ([www.epregistry.com](http://www.epregistry.com)) mais leur validation est complexe et les données internes au Laboratoire d'Histocompatibilité de l'Hôpital Saint-Louis montrent que ce registre est loin d'être complet (données non publiées). Actuellement, l'application du concept d'éplets est

2 <https://www.onelambda.com/en/product/labscreen-single-antigen.html>

3 Un eplet est un groupe d'acides aminés, présent dans un antigène HLA et impliqué dans sa reconnaissance par un anticorps.

principalement limitée à la mesure de la « charge éplétique » qui est le nombre d'éplets absents du Receveur apportés au Receveur par le Donneur. Si ce nombre est positivement corrélé avec le risque de perte de greffon (3, 4, 5), il traduit sans doute plutôt la présence (d'autant plus probable que la liste est longue) d'un ou plusieurs éplets particulièrement délétères. L'identification de ces éplets délétères et la compréhension de leur caractère inducteur de DSA auraient un impact majeur sur la compréhension du rejet.

La transplantation est un excellent modèle d'étude, compte tenu des données disponibles (séquences des allèles des patients greffés et de leurs donneurs, profils DSA chez les Receveurs). De plus c'est un modèle d'étude pour lequel des avancées auraient rapidement un impact clinique en adaptant les règles d'attribution des greffons selon une nouvelle définition de la compatibilité immunologique. La thèse Inria-Inserm en cours (Diego Amaya-Ramirez) est adossée à un projet PHRC national (Programme Hospitalier de Recherche Clinique 2016, <https://clinicaltrials.gov/ct2/show/NCT03861962>) d'analyse rétrospective de la compatibilité tissulaire (devenir des greffons en survie et apparition de DSA selon une analyse en antigènes versus allèles versus éplets) chez 15000 patients dépourvus d'anticorps anti-HLA au moment de la greffe, ayant reçu en première greffe de rein/cœur/foie/poumon entre 2008 et 2015. Ce projet HLA-3D-Diff viendrait compléter nos approches PHRC et Inria-Inserm par une nouvelle voie d'exploration complémentaire et novatrice.

## **2. Du point de vue de la représentation des différences de structures des protéines**

Classiquement, la comparaison entre elles de deux structures protéiques voisines consiste à calculer un RMSD (Root Mean Squared Deviation) entre les coordonnées 3D de chaque atome, prises dans un référentiel commun, après superposition des deux structures. Plus le RMSD est grand, plus les structures 3D sont différentes. Cette mesure peut être globale en tenant compte de tous les atomes ou locale dans les zones où l'on attend des variations de structure (6). Le travail de superposition des structures protéiques est très complexe et a donné lieu à de nombreux travaux y compris dans l'équipe Capsid (logiciel Kpax développé par Dave Ritchie ; 7, 8). La localisation des régions qui diffèrent entre deux structures voisines après superposition se fait manuellement et le but de ce projet serait de pouvoir calculer ces différences.

Un autre groupe de méthodes pour comparer entre elles des structures protéiques 3D voisines (notamment celles de la famille HLA) s'appuie sur le degré d'exposition des acides aminés polymorphiques en surface, calculé par des algorithmes estimant la surface exposée au solvant (9, 10), ou estimé par des experts à partir des modèles 3D disponibles. Cependant, ces méthodes ont les défauts de faire une simplification très forte du problème, d'être restreints à un ensemble limité de modèles 3D (notamment pour les protéines HLA classe II), de négliger les propriétés physico-chimiques portées par les acides aminés et de se limiter aux acides aminés polymorphiques en surface. Ce dernier point est particulièrement important car le changement d'un acide aminé à l'intérieur d'une protéine peut modifier le degré d'exposition en surface d'autres acides aminés.

## **3. Du point de vue de la géométrie différentielle.**

La complexité des repliements des protéines rend difficile des comparaisons qui seraient

uniquement basées sur la position spatiale des points des modèles 3D. Au contraire, les méthodes de paramétrisation globale ont l'avantage d'être invariantes aux translations et rotations des objets 3D. L'idée sous-jacente à ces méthodes est d'envoyer une surface 3D vers le plan grâce un découpage et un aplatissement (bien choisi) de celle-ci, de la même façon qu'un planisphère (plan) représente la terre (sphérique). Ces paramétrisations permettent de comparer dans un même domaine plusieurs surfaces. En particulier, pour chaque point d'une surface on peut retrouver l'unique point correspondant sur un autre surface grâce à leur représentation commune dans le domaine paramétrique (11, 12).

Cette technique semble tout à fait adaptée à l'étude des protéines. En effet, si celles-ci présentent des géométries complexes à cause de leurs nombreux repliements, elles ont une structure intrinsèque simple : un arrangement linéaire d'acides aminés. Cela rend le calcul de paramétrisation à la fois robuste et relativement aisé tout créant peu de distorsion qui détruirait une partie de la géométrie. Une première tentative de représentation paramétrique des protéines a été publiée récemment (13).

Une fois les paramétrisations obtenues, toutes les informations liées aux protéines, qu'elles soient géométriques (courbure, convexité, concavité, occlusion) ou physico-chimiques, peuvent être représentées comme l'affectation d'une valeur à chaque point de l'espace de paramètres. Cet espace étant commun à toutes les protéines la comparaison de ces données revient à comparer des images. Non seulement cette méthode préserve l'intégralité des données géométriques, mais elle ouvre la voie à l'utilisation de méthodes d'apprentissage statiques d'imageries pour classifier ou comparer des protéines.

## ***Déroulement prévisionnel et moyens demandés***

Le projet se déroulera sur une année universitaire (2020-2021).

Il impliquera des réunions de discussion régulières entre les partenaires (présentations mutuelles de l'état de l'art de chaque point de vue, avancement des travaux)

Un stage M2 (Master Maths-Info, orienté géométrie différentielle) est prévu en 2020-2021 pour explorer concrètement les méthodes de paramétrisation sur un jeu de données de variants protéiques HLA préparé par Diego.

Quelques mois d'ingénieur seraient souhaitables aussi pour structurer en base de données les caractéristiques des milliers de sérums disponibles à l'hôpital Saint-Louis, ainsi que leurs différences selon les divers modes de représentation étudiés, afin de permettre l'extraction de jeux de données pour l'apprentissage automatique.

### **Récapitulatif des moyens demandés**

Stage M2 (Equipe PIXEL) : 3,2 K€

CDD ingénieur (Equipe CAPSID) : 4 à 6 mois

## ***Pertinence par rapport à l'appel à projets Loria***

Ce projet répond à deux objectifs de l'appel à projet :

(1) Faire émerger des thèmes de recherche nouveaux  
et (4) Soutenir les nouveaux arrivants (Etienne Corman).

## Références

1. Lefaucheur C, Loupy A. [Antibody-Mediated Rejection of Solid-Organ Allografts](#). *N Engl J Med*. 2018 Dec 27;379(26):2580-2582.
2. Tambur, A.R., and Claas, F.H. 2015. HLA epitopes as viewed by antibodies: what is it all about? *Am J Transplant* 15:1148-1154.
3. Wiebe, C., Pochinco, D., Blydt-Hansen, T.D., Ho, J., Birk, P.E., Karpinski, M., Goldberg, A., Storsley, L.J., Gibson, I.W., Rush, D.N., et al. 2013. Class II HLA epitope matching-A strategy to minimize de novo donor-specific antibody development and improve outcomes. *Am J Transplant* 13:3114-3122.
4. McCaughan JA, Battle RK, Singh SKS, Tikkanen JM, Moayedi Y, Ross HJ, Singer LG, Keshavjee S, Tinckam KJ. Identification of risk epitope mismatches associated with de novo donor-specific HLA antibody development in cardiothoracic transplantation. *Am J Transplant*. 2018 Dec;18(12):2924-2933.
5. Snanoudj R, Kamar N, Cassuto E, Caillard S, Metzger M, Merville P, Thierry A, Jollet I, Grimbert P, Anglicheau D, Hazzan M, Choukroun G, Hurault De Ligny B, Janbon B, Vuiblet V, Devys A, Le Meur Y, Delahousse M, Morelon E, Bailly E, Girerd S, Amokrane K, Legendre C, Hertig A, Rondeau E, Taupin JL. Epitope load identifies kidney transplant recipients at risk of allosensitization following minimization of immunosuppression. *Kidney Int*. 2019 in press.
6. Shanthirabalan S, Chomilier J and Carpentier M. Structural effects of point mutations in proteins. *Proteins - Structure, Function and Bioinformatics*, Wiley, 2018, 86 :853-867. 10.1002/prot.25499. Hal-01909365
7. Ritchie DW, Ghoorah AW, Mavridis L, Venkatraman V. Fast protein structure alignment using Gaussian overlap scoring of backbone peptide fragment similarity. *Bioinformatics*. 2012;28(24):3274-3281. doi:10.1093/bioinformatics/bts618
8. Ritchie DW. Calculating and scoring high quality multiple flexible protein structure alignments. *Bioinformatics*. 2016;32(17):2650-2658. doi:10.1093/bioinformatics/btw300
9. Klausen MS, Jespersen MC, Nielsen H, et al. NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning. *Proteins*. 2019;87(6):520-527.
10. Mirabello C, Pollastri G. Porter, PaleAle 4.0: high-accuracy prediction of protein secondary structure and relative solvent accessibility. *Bioinformatics*. 2013;29(16):2056-2058.
11. Aigerman N, Poranne R and Lipman Y. Seamless Surface Mappings. ACM SIGGRAPH 2015.
12. Aigerman N and Lipman Y. Orbifold Tutte Embeddings. ACM SIGGRAPH Asia 2015.
13. Lam KC and Lui LM. Optimized quasiconformal parameterization with user-defined area distortions. *Commun. Math. Sci*. 2017, 15: 2027-2054.

## **ANNEXE 1 : Détails sur les mécanismes du rejet de greffe par incompatibilité HLA**

L'activité des protéines HLA est triple :

- 1) Un rôle physiologique de présentation de peptides étrangers au receveur (peptides microbiens ou propres à une tumeur) aux lymphocytes T pour lancer une réponse immunitaire salvatrice. Ceci est dépendant d'une « poche à peptides » dont la forme spatiale et les propriétés physicochimiques conditionnent le répertoire de peptides qui peuvent s'y fixer et donc être présentés. Les 2/3 des polymorphismes alléliques sont localisés dans cette « poche ». Le répertoire peptidique normale du Receveur permet d'éduquer les lymphocytes T vers une tolérance au Soi (absence d'auto-immunité destructrice).
- 2) Un effet collatéral délétère en situation de transplantation non identique en HLA (>95% des transplantations) lorsque au moins une des « poches » du Donneur est différente du receveur. En effet, le répertoire peptidique sera modifié et le greffon sera considéré comme étranger même en l'absence d'infection ou de tumeur, car les lymphocytes T du Receveur ne toléreront pas ces peptides inconnus. Ceci entraînera un rejet dit cellulaire, bien pris en charge par l'arsenal thérapeutique des immunosuppresseurs mais qui fait le lit d'une autre forme du rejet (paragraphe suivant)..
- 3) Une double peine en transplantation liée aux polymorphismes localisés en dehors de la « poche », mais accessibles à la surface de la protéine, ou internes et entraînant des modifications de la forme de la surface de la protéine. Ces changements créeront des cibles de reconnaissance pour des Anticorps anti-HLA spécifiques du donneur (DSA pour Donor-Specific Antibody) produits par des lymphocytes B du Receveur incapables de tolérer ces régions appelées épitopes absentes du Receveur. Les DSA causent le rejet humoral, principale cause immunologique des pertes de greffon, car très difficiles à traiter. La réponse B nécessite une réponse T. En fait un épitope est une surface d'environ 15 nM de rayon, au sein de laquelle se situe l'éplet qui est le (ou les) acide(s) aminé(s) propre(s) au Donneur contribuant la plus grande part de l'affinité de l'interaction avec le DSA. La molécule HLA du Donneur est l'antigène, constitué d'un nombre variable d'épitopes et d'éplets selon la « distance » génétique entre les allèles du Receveur et du Donneur. De plus, la plupart des éplets sont partagés par plusieurs ou un grand nombre d'allèles HLA différents (2).

La solution immunologique pour la transplantation est donc la greffe HLA-identique, qui n'est que très rarement possible. La solution de remplacement est donc un suivi immunologique régulier des patients, à la recherche de DSA en étudiant la réactivité du sérum des Receveurs contre environ 200 des allèles HLA les plus fréquents, avec une méthode très sensible appelée « single antigen » Luminex. Ceci permet d'établir la liste des allèles HLA contre lesquels le Receveur est immunisé avant la greffe (pour sélectionner les donneurs « compatibles ») et après la greffe (pour détecter au plus tôt l'apparition de DSA et renforcer le traitement immunosuppresseur).