



Prédire l'amylose AL grâce à l'IA

Dr. Virginie Pascal

Lab. Immunologie CHU Limoges – UMR CNRS INSERM - CRIBL





Liens d'intérêts

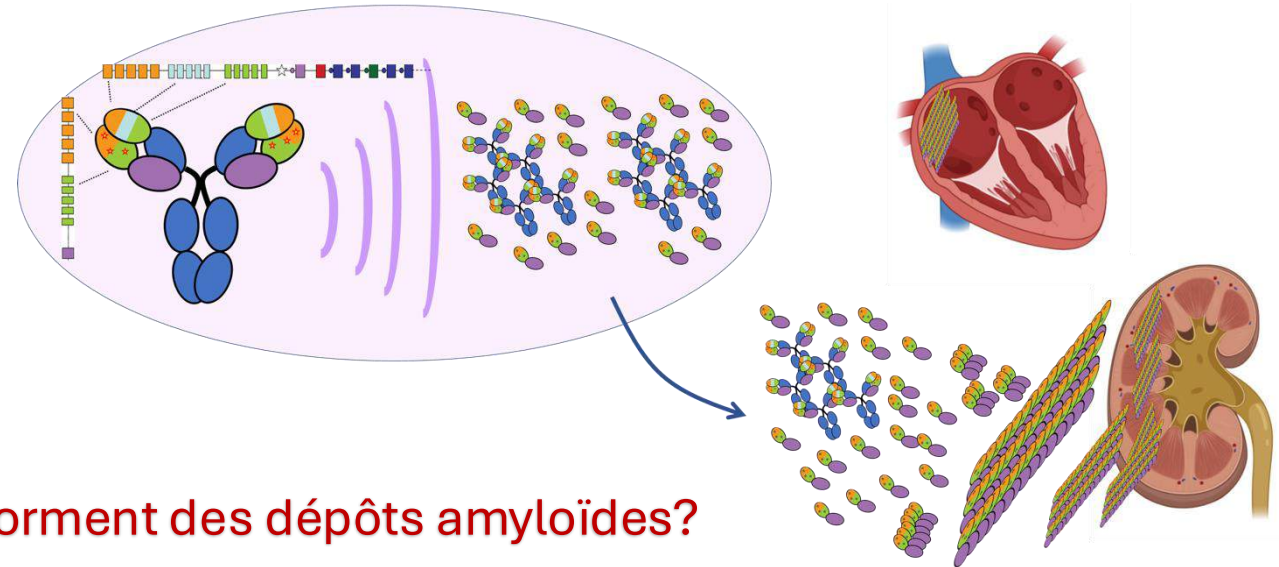
- Aucun



Les chaînes légères au cœur de l'amylose AL

Amylose AL : une maladie de la chaîne légère

- Spécificité immunologique unique : réarrangement + mutations \leftrightarrow Pathogène
- Production excessive par un clone B / plasmocytaire
- Chaîne légère d'Immunoglobuline
 - ❖ à la fois **fonctionnelle**... et **pathogène**
 - ❖ responsable directement des dépôts



-> Pourquoi certaines chaînes légères forment des dépôts amyloïdes?



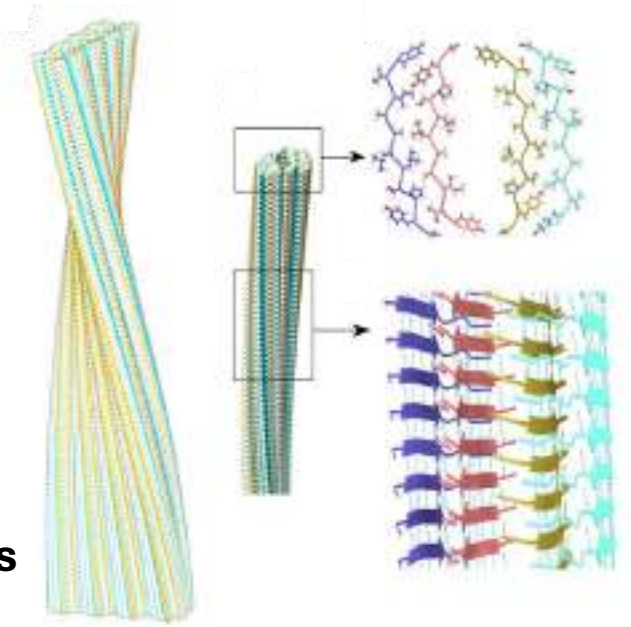
Prédiction de l'amyloïdogénicité : état de l'art

Des outils développés pour d'autres protéines amyloïdes

- Basés sur la **séquence protéique**
- Identification de **régions à risque d'agrégation**
- Lien séquence → feuillets β → fibrilles amyloïdes

Ex d'outils utilisés / Maladies neurodégénératives et amyloses ATTR héréditaires

- **TANGO**, propension à former des feuillets β : *α -synucléine* → *Parkinson*, *transthyrétine*
- **WALTZ**, motifs amyloïdogènes courts : *Peptide β -amyloïde ($A\beta$)* → *Alzheimer*..
- **AGGRESCAN**, régions hydrophobes aggrégantes : *Tau* → *Alzheimer*..
- **FoldAmyloid**, structures amyloïdes en feuillet β : *Prions* → *Creutzfeldt-Jakob*, *$A\beta$* , *Tau*...



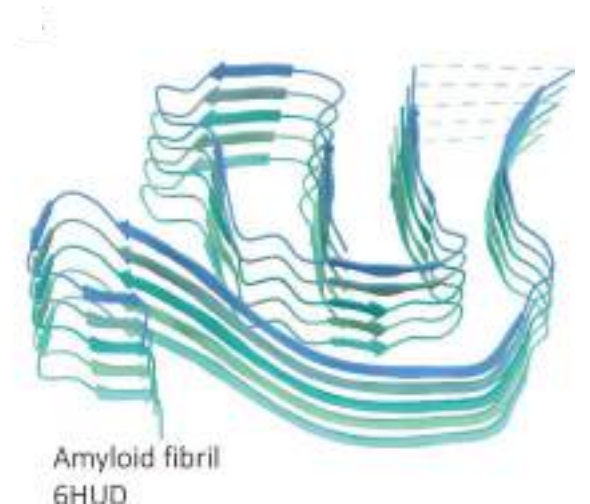
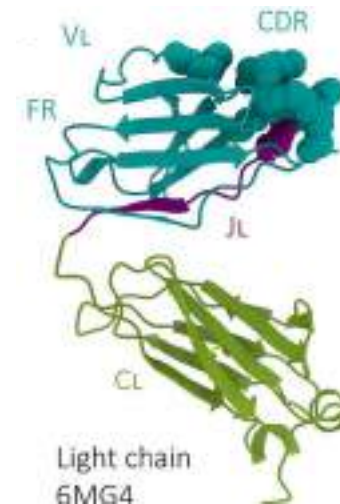
-> Peut-on utiliser ces mêmes outils pour les chaînes légères ?



Pourquoi les chaînes légères sont un défi à part

Les chaînes légères : un cas particulier

- Outils historiques développés pour des protéines peu variables
- Chaînes légères : extrême diversité des séquences (κ / λ , segments V et J, mutations)
- Architecture particulière des Ig (domaines CDR)
- Agrégation influencée par des facteurs non strictement séquentiels



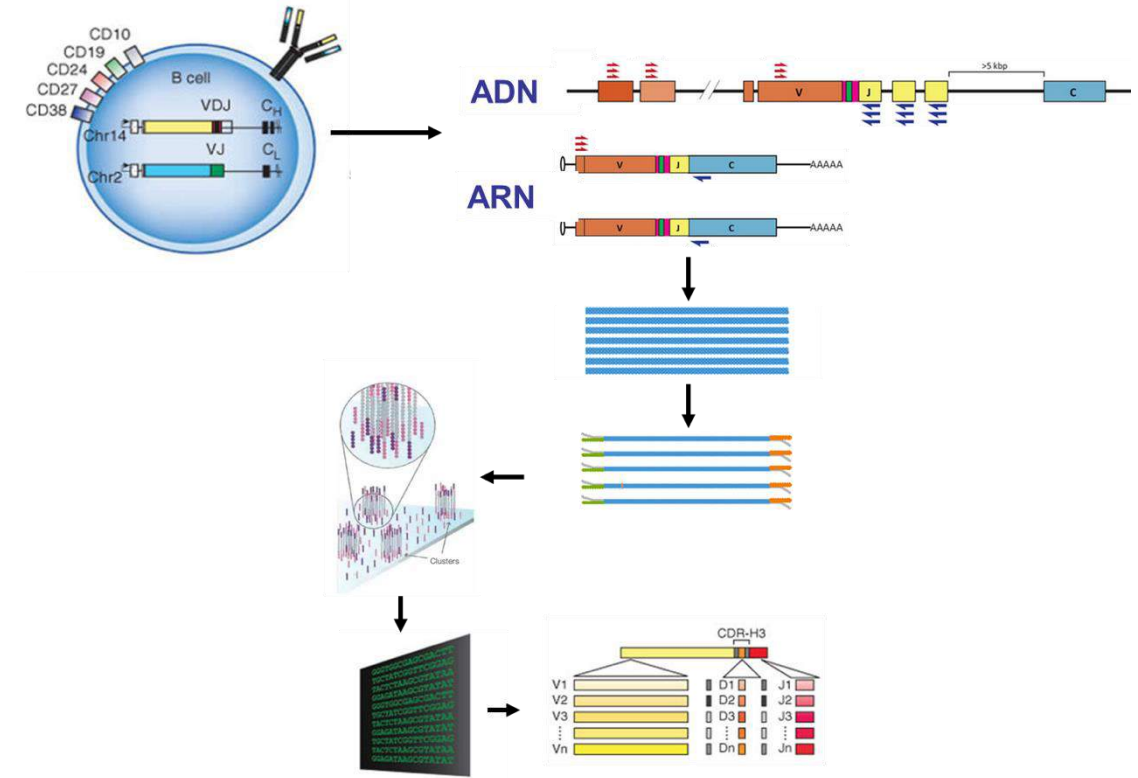


À l'ère du séquençage haut débit des immunoglobulines (NGS)

Le NGS a profondément modifié notre approche:

- Accès à la **séquence nucléotidique complète** des chaînes légères monoclonales
- Reconstruction de la **séquence en acides aminés** de la protéine pathogène
- Analyse fine :
 - ❖ Réarrangements V-J
 - ❖ Mutations somatiques
 - ❖ Caractéristiques physico-chimiques
- Constitution de **banques de séquences de chaînes légères AL**

→ un prérequis pour les approches de prédiction par IA





AL-Base : une ressource de référence pour l'amylose AL

- **Créée en 2009**, régulièrement enrichie (dernière mise à jour : **2025**)
- Base de données regroupant plus de 2 200 séquences issues de patients AL, MM, MGUS, LCDD, POEMS, MW, etc



https://wwwapp.bumc.bu.edu/BEDAC_ALBase

-> Base de référence indispensable pour le développement de modèles de prédiction spécifiques aux immunoglobulines



Modèles d'IA dédiés à l'amylose AL : où en est-on ?

Plusieurs modèles d'IA dédiés → estimation du risque amyloïdogène d'une chaîne légère

- Entraînés principalement à partir de l'**AL-Base**
- **3 ex** qui exploitent différentes **caractéristiques des séquences** d'AA



❖ VLAmY-Pred (Sci Rep 2021)

<https://web.iitm.ac.in/bioinfo2/vlami-pred/>

- Machine learning → descripteurs de séquence VL
- Perf. modérées (~ 79% précision)



❖ LICTOR (Nat Commun 2021)

<http://lictor.irb.usi.ch/lictor/>

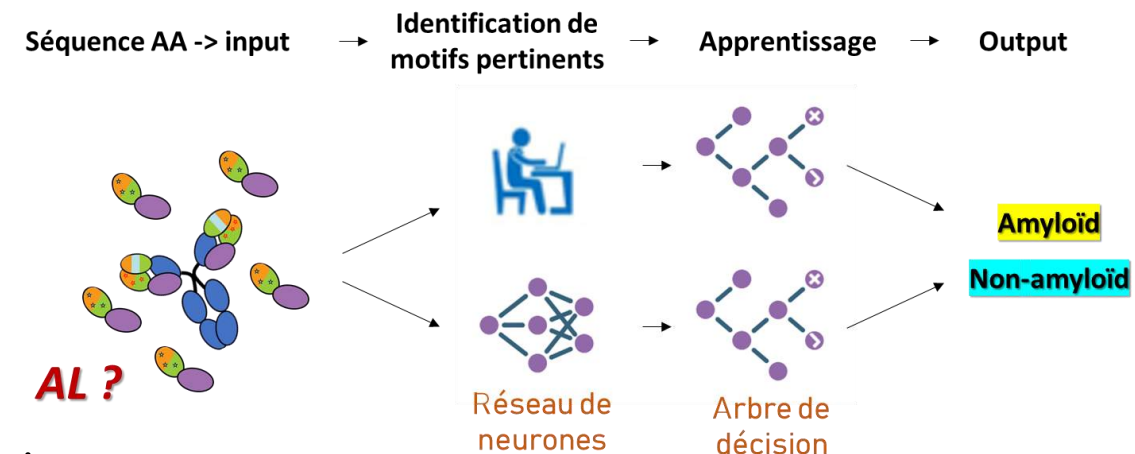
- Random forest → mutations somatiques
- Perf. intermédiaire (~ 83% précision)



❖ AB-Amy (v1 & v2) (2023–2025) /

<http://i.uestc.edu.cn/AB-Amy2>

- Apprentissage motifs amyloïdogènes
- Meilleures performances ~93 % de précision

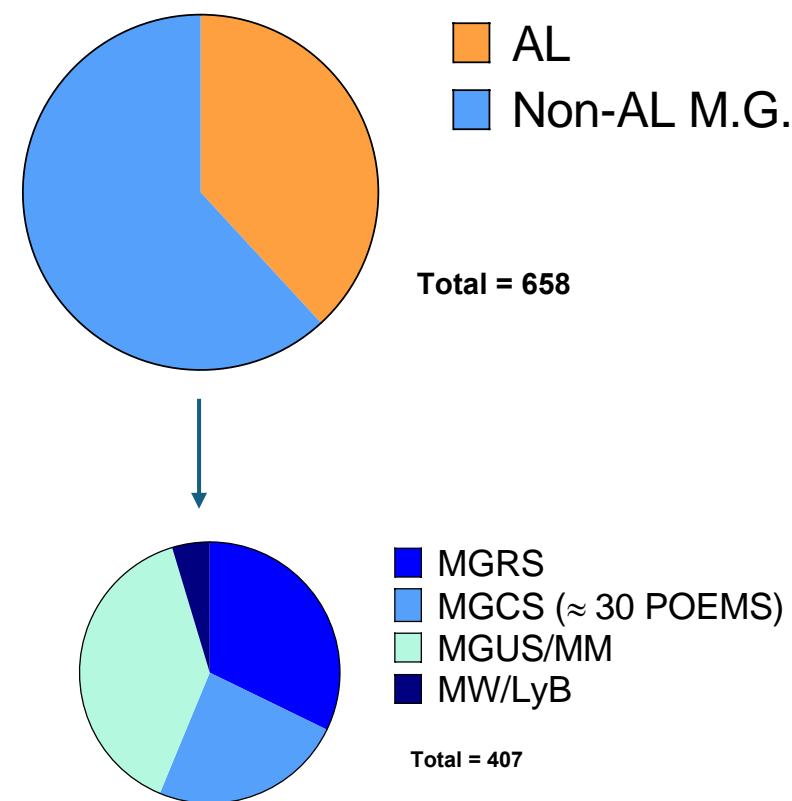




Une approche ancrée dans la pratique clinique : l'apport du NGS au CNR Amylose AL et autres MGRS/MGCS

Analyse du répertoire des immunoglobulines par NGS (RACE RepSeq)

- Activité de routine intégrée à la prise en charge clinique
- Indications ciblées :
 - ❖ composantes monoclonales de **faible abondance**
 - ❖ situations diagnostiques complexes
 - ❖ suivi moléculaire (MRD)
- **658 séquences de chaînes légères** de patients **AL** et **Non-AL**
- Corrélation systématique :
 - ❖ données cliniques
 - ❖ données biologiques
 - ❖ caractéristiques des chaînes légères



→ Une base de données orientée “Ig pathogènes”

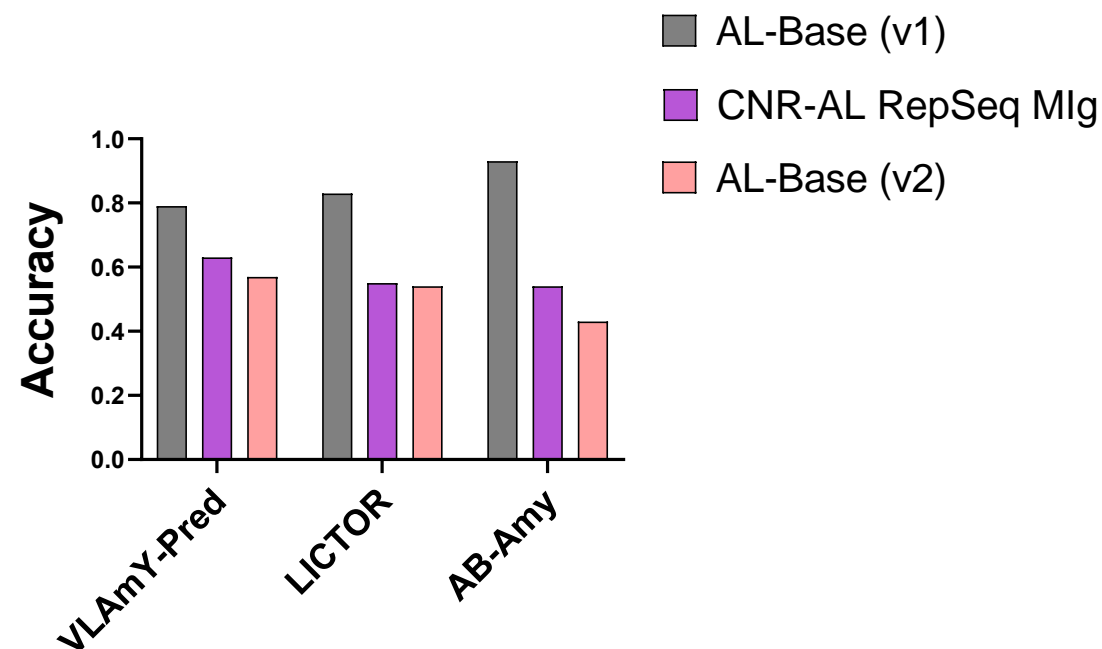


Évaluation des modèles de prédiction existants sur notre base clinique

Outils évalués: **VLAmy-Pred** · **LICTOR** · **AB-Amy**

Constats principaux:

- Performances élevées rapportées sur les **jeux de données d'apprentissage**
- **Diminution significative** des performances sur les **jeux de test indépendants** :
 - ❖ Séquences des patients du **CNR Amylose AL**
 - ❖ Après mise à jour d'**AL-Base**



-> Modèles actuels insuffisants pour une utilisation en routine clinique



Vers une approche combinée pour la prédiction de l'AL

Input = Base CNR

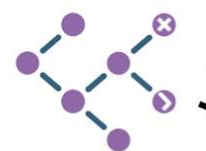
- Séquences validées par cliniciens et biologistes
- Forte représentation d'Ig pathogènes
- Données clinico-bio associées



paramètres physico-chimiques + clinico-biologiques



AL ?



Test sur AL-Base
Précision $\approx 0,75$

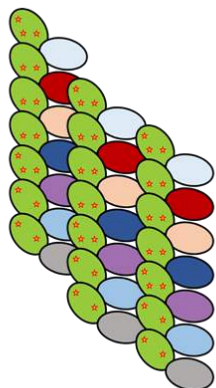
Arbre de décision

← Apprentissage →

Réseau de neurones

Input = AL-Base

- Large cohorte de séquences



Deep Learning sur
séquence protéique



Test sur base CNR
Précision $\approx 0,77$

Décision combinée :

- Confrontation prédictions + scores de certitude
- **Précision globale $\approx 0,88$**



Une stratégie d'aide à la prédiction de l'amylose AL



Outil de prédiction combinant :

- Données clinico-biologiques et moléculaires
- Approches de machine learning et de deep learning
- Déploiement envisagé en **2026**
 - **Accès libre en ligne**

Prochaines étapes

- Évaluation sur jeux de données indépendant -> Extension de la base CNR (+ ~250 nouvelles séquences en cours)

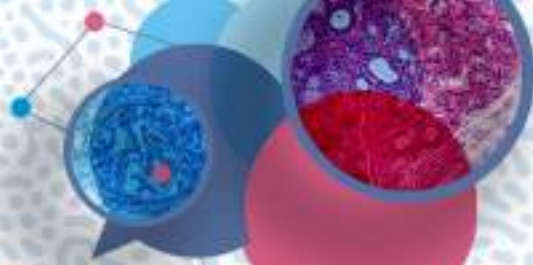
Objectif

- Développer un **outil d'aide à la décision, complémentaire de l'expertise clinique**, non substitutif

Au-delà de la prédiction, exploiter les modèles pour :

- Mieux comprendre la **physiopathologie** des chaînes légères amyloïdogènes
- Générer de nouvelles **hypothèses biologiques**

-> La séquence seule n'explique pas tout, mais peut ouvrir des pistes sur d'autres déterminants...



Merci



Alexia Rinsant, Sihem Kaaki, Cécile Ory, Estelle Desport, JM Goujon, Pr Guy Touchard



Sébastien Bender



Christophe Sirac



Vincent Javaugue



Séléna Teillaud



Steve Jatiere



Frank Bridoux



Arnaud Jaccard



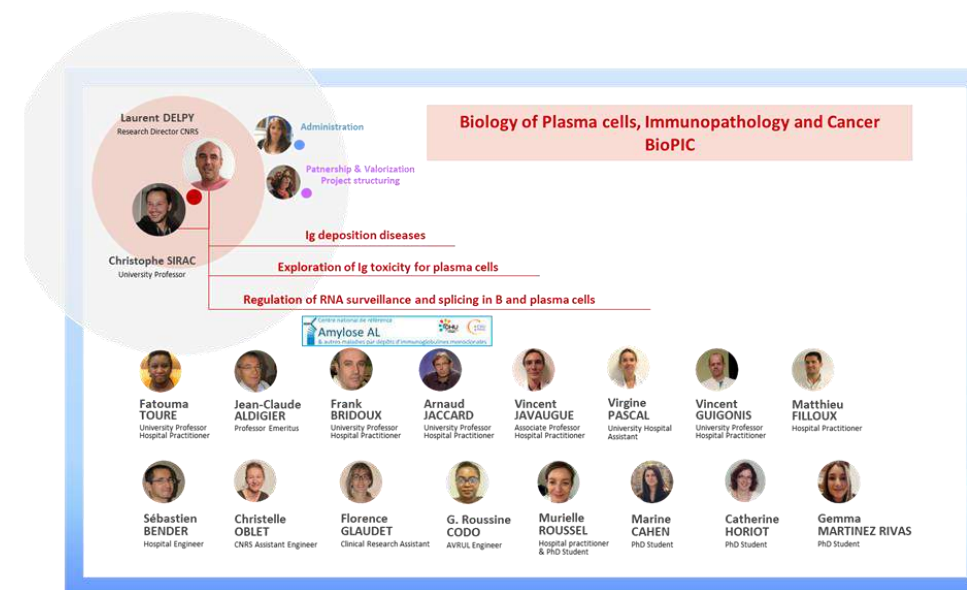
Murielle Roussel



Paco Derouault



Jean-Baptiste Drouet





Remerciements

Platinum sponsor



Premium sponsors



Silver sponsors

