



Prédire l'amylose AL grâce à l'IA

Dr. Virginie Pascal

Lab. Immunologie CHU Limoges – UMR CNRS INSERM - CRIBL

Journée organisée par :

Avec le parrainage de :

www.congres-amylose.com





Liens d'intérêts

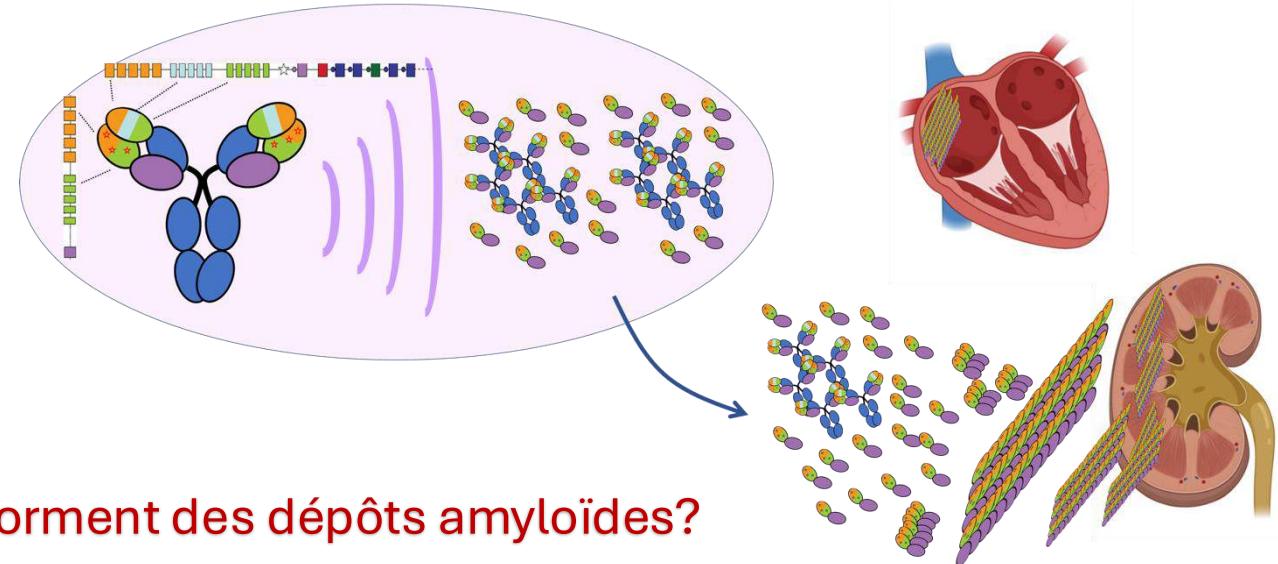
- Aucun



Les chaînes légères au cœur de l'amylose AL

Amylose AL : une maladie de la chaîne légère

- Spécificité immunologique unique : réarrangement + mutations <-> 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?



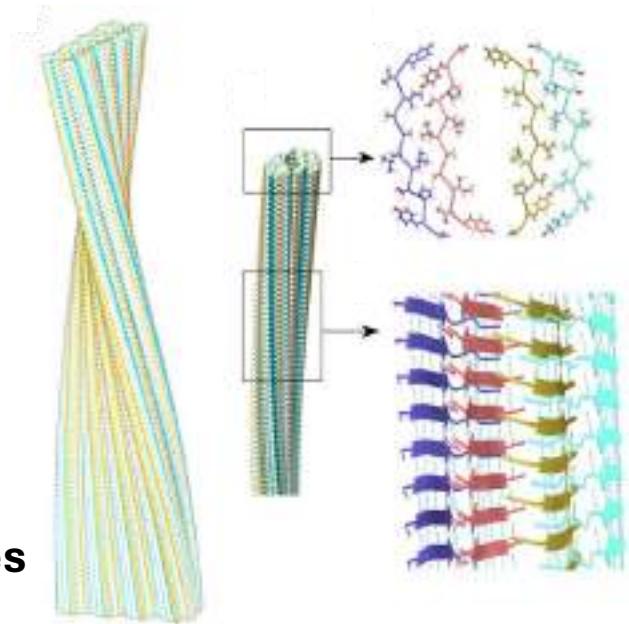
Prédiction de l'amyloïdogénérité : état de l'art

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 β)* -> *Alzheimer..*
- **AGGRESCAN**, régions hydrophobes agrégantes : *Tau* -> *Alzheimer..*
- **FoldAmyloid**, structures amyloïdes en feuillet β : *Prions* -> *Creutzfeldt-Jakob, A β , Tau...*



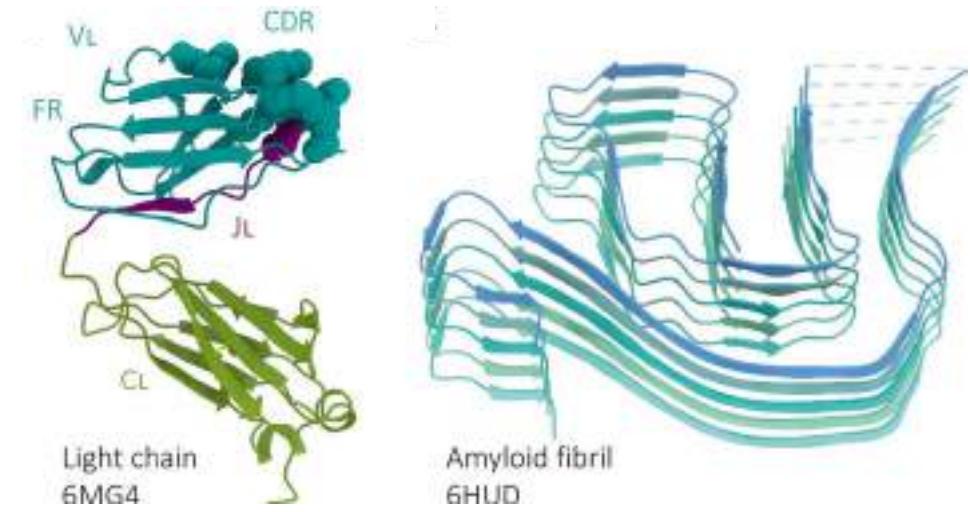
-> Peut-on utiliser ces mêmes outils pour les chaînes légères ?



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

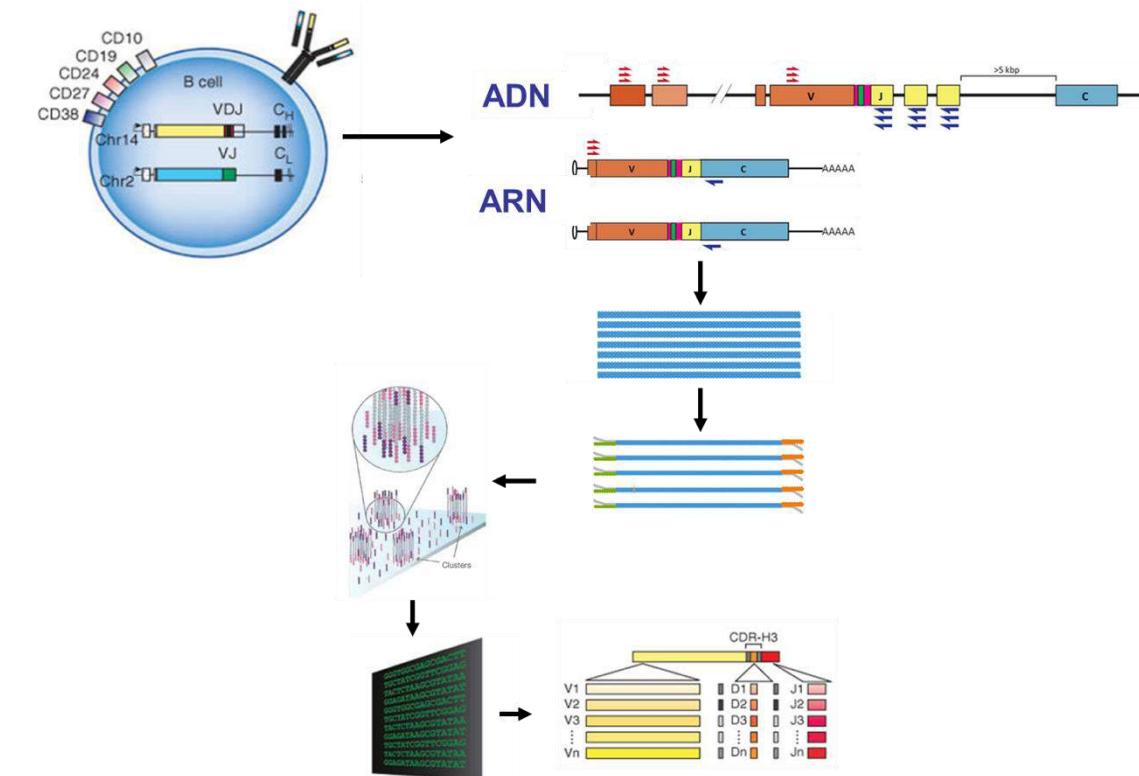




À l'ère du séquençage haut débit des immunoglobulines (NGS)

Le NGS a profondément modifié notre approche:

- Accès à la **séquence nucléotidique complète** des chaînes légères monocloniales
- 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 prédition par IA



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

- **Créeée 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

AMYLOID LIGHT CHAIN DATABASE

[Home](#) [Search](#) [Statistics](#) [About](#) [Contact](#) [Latest News](#)

WELCOME

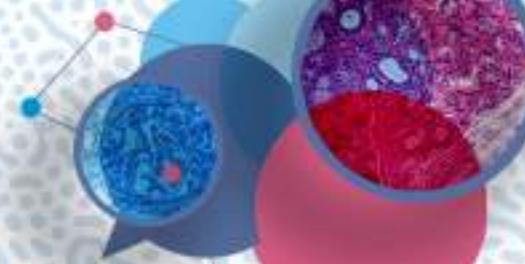
AL-Base is a curated database of antibody light chain sequences associated with diseases known as plasma cell dyscrasias (PCD), primarily immunoglobulin light chain amyloidosis, also known as AL amyloidosis. AL-Base is intended to facilitate studies into the mechanisms of diseases involving monoclonal immunoglobulin proteins. The hypothesis underlying this goal is that the primary sequence of the light chain is likely to be a major determinant of its propensity to form amyloid fibrils and cause disease. De-identified light chain nucleotide and protein sequences from patient-derived bone marrow samples are organized according to the disease from which they originate and the light chain genes involved. Tools are provided for basic search and visualization functions, and both nucleotide and protein sequences can be exported for further analyses. Please use the [search page](#) to explore the database.

AL AMYLOIDOSIS

Amyloidosis is a condition where normally soluble proteins misfold and aggregate in the body as amyloid fibrils, leading to progressive tissue toxicity and organ failure. AL amyloidosis is one of the most [common forms of systemic amyloidosis](#), involving deposition of antibody light chains in multiple organs. AL amyloidosis is most often caused by an aberrant population of clonal plasma cells in the bone marrow, which secrete the light chain that aggregates. Untreated disease is rapidly fatal, as patients can develop heart and kidney failure as well as liver, intestinal, soft tissue or neurologic symptoms.

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 séquence d'AA**

❖ VLAmY-Pred (Sci Rep 2021)

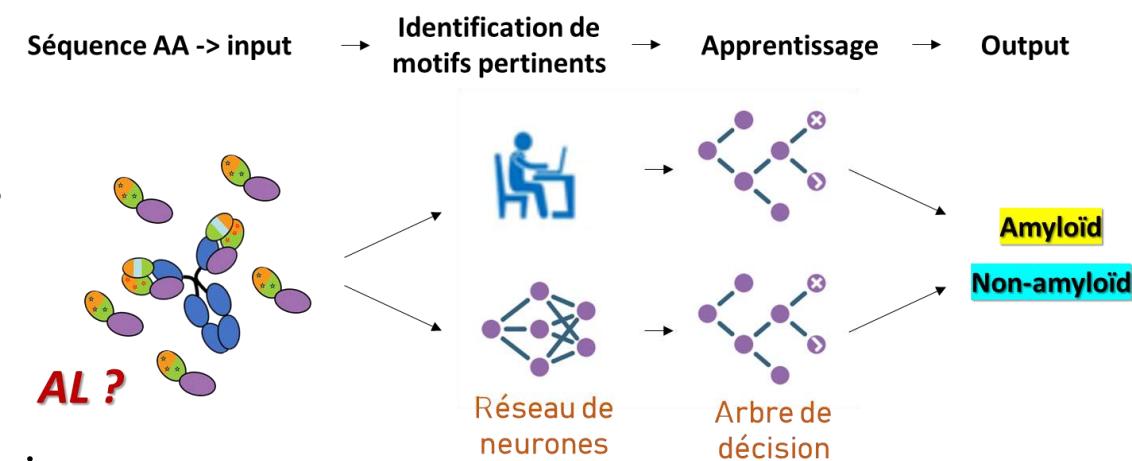
- Machine learning -> descripteurs de séquence VL
- Perf. modérées (~ 79% précision)

❖ LICTOR (Nat Commun 2021)

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

❖ AB-Amy (v1 & v2) (2023–2025) /

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

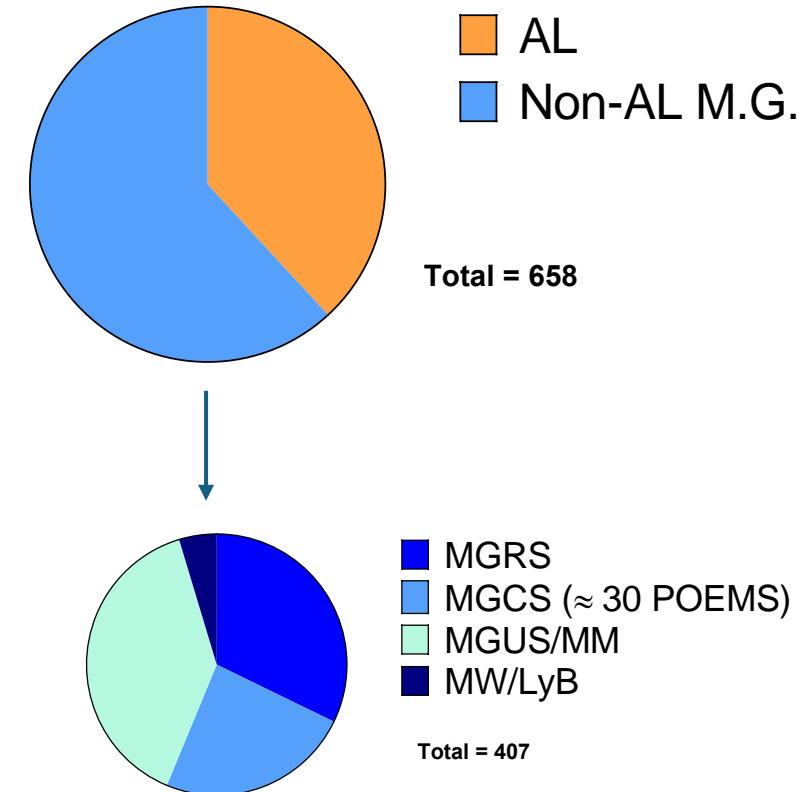




Une approche ancrée dans la pratique clinique : l'apport du NGS au CNR Amylose AL et autres MGCS/MGRS

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”

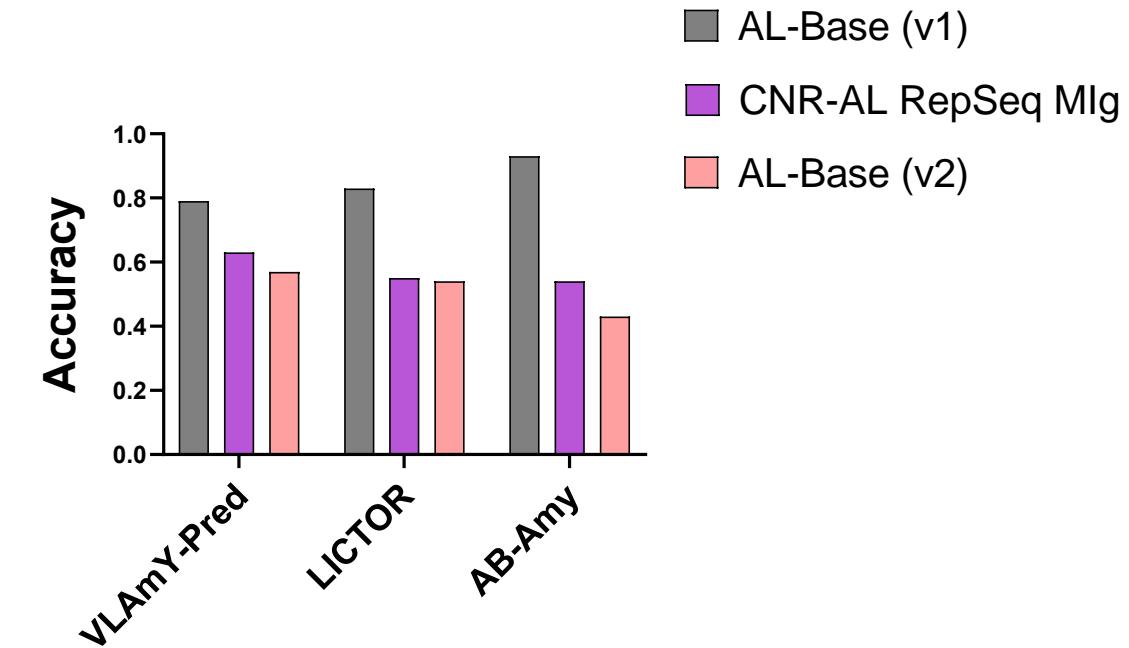


Évaluation des modèles de prédition existants sur notre base clinique

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 prédition de l'AL

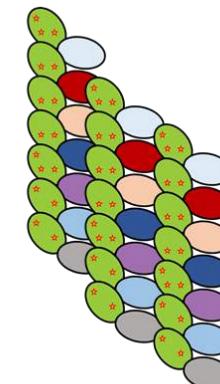
Input = Base CNR

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



Input = AL-Base

- Large cohorte de séquences



paramètres physico-
chimiques + clinico-
biologiques



AL ?

→ Apprentissage →

Arbre de
décision

Réseau de
neurones

Deep Learning sur
séquence protéique

Test sur AL-Base
Précision ≈ 0,75

Décision combinée :

- Confrontation prédictions + scores de certitude
- **Précision globale ≈ 0,88**

Test sur base CNR
Précision ≈ 0,77



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



Outil de prédition 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 prédition, 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



Filière de santé Maladies Rares Immuno-Hémato-génétiques

Recherche pour la Prévention de la Santé

e



Sébastien Bender



Christophe Sirac



Vincent Javaugue



Séléna Teillaud



Steve Jatiere



Frank Bridoux



Arnaud Jaccard



Murielle Roussel



Paco Derouault



Jean-Baptiste Drouet



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



Biology of Plasma cells, Immunopathology and Cancer BioPIC

Laurent DELPY
Research Director CNRS

Christophe SIRAC
University Professor

Administration

Partnership & Valorization
Project structuring

Ig deposition diseases

Exploration of Ig toxicity for plasma cells

Regulation of RNA surveillance and splicing in B and plasma cells

Fatouma TOURE
University Professor Hospital Practitioner

Jean-Claude ALDIGIER
Professor Emeritus

Frank BRIDOUX
University Professor Hospital Practitioner

Arnaud JACCARD
University Professor Hospital Practitioner

Vincent JAVAUGUE
Associate Professor Hospital Practitioner

Virginie PASCAL
University Hospital Assistant

Vincent GUIGONIS
University Professor Hospital Practitioner

Matthieu FILLOUX
Hospital Practitioner

Sébastien BENDER
Hospital Engineer

Christelle OBLET
CNRS Assistant Engineer

Florence GLAUDET
Clinical Research Assistant

G. Roussine CODO
AVRUL Engineer

Murielle ROUSSEL
Hospital practitioner & PhD Student

Marine CAHEN
PhD Student

Catherine HORIOT
PhD Student

Gemma MARTINEZ RIVAS
PhD Student



Remerciements

Platinum sponsor



Premium sponsors



Silver sponsors

