



الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
جامعة زيان عاشور-الجلفة
Université Ziane Achour – Djelfa
كلية علوم الطبيعة و الحياة
Faculté des Sciences de la Nature et de la Vie
قسم البيولوجيا
Département en biologie

Mémoire

Projet de Fin d'Etudes en vue d'obtention de Diplôme du Master
Filière : Sciences biologiques
Option : Microbiologie Appliquée

Thème

Étude bioinformatique comparative des protéines et de leurs gènes associés à la maladie d'Alzheimer

Présenté par :

Mlle. BENDRIS Ikram Nour El houda
Mlle. HAOUA Meriem

Soutenu le :

Devant le jury composé de :

Président : BEN CHERIF Karima	Professeur	Univ. Djelfa
Promoteur : SEDDIKI Said	MCA	Univ. Djelfa
Examineurs : SERSEG Talia	MCB	Univ. Djelfa
Examineurs : ABDELALI Salma Kaouthar	Docteur	Univ. Djelfa

Année Universitaire 2025/2026



Dedication

À celui qui a fait entendre à mes oreilles les premières mélodies du Coran, et qui m'a appris que la science est lumière et que le succès est le fruit de la patience et du combat...

À mon cher père, qui a été un soutien indéfectible et une ombre constante, et qui a semé en moi la confiance et la détermination, étant le premier à croire en ma capacité d'atteindre ce jour.

À celle qui m'a porté dans son amour avant même de me porter dans ses bras, et dont les prières m'ont accompagnée à chaque examen, à chaque pas et à chaque moment de faiblesse...

À ma chère mère, qui a toujours rêvé de me voir réussir, et j'espère avoir réalisé aujourd'hui une partie de son rêve et rendu, ne serait-ce qu'un peu, de son immense bienfait irréversible.

À mes frères et sœurs, compagnons de mes jours et partenaires de mon parcours, qui ont partagé avec moi la fatigue avant la joie et ont été un précieux soutien.

À mes honorables enseignants, qui ont éclairé mon chemin par le savoir et la connaissance, et ont laissé en moi une empreinte qui demeurera toute ma vie.

À tous ceux qui m'ont soutenue par un mot bienveillant, un conseil sincère ou une prière discrète dans l'invisible.

Je vous dédie le fruit de plusieurs années de travail, de veilles et de défis, en espérant que ce modeste travail soit une source de fierté pour tous ceux qui m'ont soutenue et ont cru en moi.

Louange à Dieu, par la grâce de qui s'accomplissent les bonnes œuvres .

BENSRIS Ikram nour el houda





Dedication

Je dédie ce travail a mes parents ,à ma famille et à tous
ceux qui m'ont aidé et soutenu durant mes études.avec
toute ma gratitude

HOUA Meriem





Remerciements

Après avoir remercié Allah Tout-Puissant qui m'a donné la santé et la volonté et Patience pour terminer mon travail.

*Tout d'abord, nous exprimons notre gratitude à notre promoteur **Dr. SEDDIKI Said**, professeur à l'Université Ziane Achour, pour nous avoir guidé. Et les instructions et conseils qu'il a donnés pendant notre travail.*

Nous tenons à exprimer notre grande considération et nos profonds remerciements aux membres de jury qui nous ont fait l'honneur de juger notre travail.

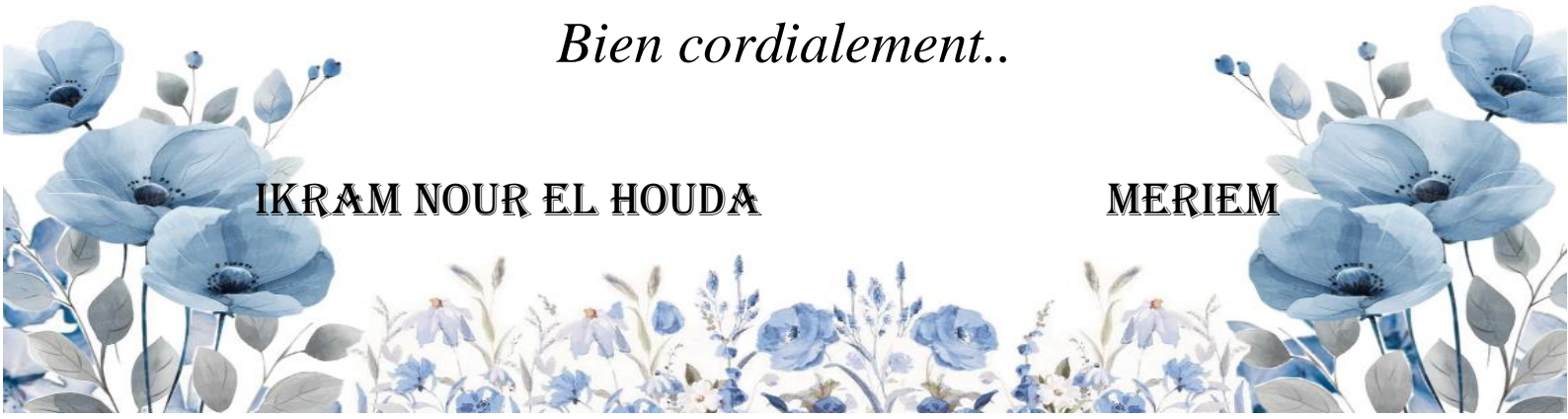
Enfin, nous souhaitons exprimer notre reconnaissance à toutes les personnes qui, de près ou de loin, ont apporté leur soutien à ce projet.

Merci à tous pour votre aide et vos encouragements.

Bien cordialement..

IKRAM NOUR EL HOUDA

MERIEM



Listes des figures

N°	Titre des figures	Page
Figure 1	Bioinformatique comme champ d'application interdisciplinaire ; un modèle d'intégration de la Biologie, de l'Informatique et des Mathématiques .	3
Figure 2	L'impasse de l'analyse manuelle des séquences génomiques ; un aperçu de la complexité pré-bioinformatique .	4
Figure 3	Margaret Dayhoff et l'émergence des premières matrices de substitution ; une réponse à la complexité des séquences biologiques .	5
Figure 4	Principaux outils de la bioinformatique moderne(2026).	8
Figure 5	Rosalind Franklin, la photo 51 et le modèle de Watson-Crick. (<i>Source: Franklin & Gosling, 1953 ; King's College London, 2023</i>).	10
Figure 6	Pionnier du séquençage de première génération : F. Sanger et l'interprétation électrophorétique des fragments d'ADN. (Source : Labster Theory, 2024 ; First-generation sequencing).	10
Figure 7	Chronologie de la génomique moderne : des fondements structuraux à l'ère du Big Data (Conception personnelle, 2026 ; inspirée du NCBI et de l'EBI).	11
Figure 8	Organisation d'un gène eucaryote et processus de maturation de l'ARN (Transcription et Épissage) .Gène : Structure et expression chez les eucaryotes. Récupéré).	13
Figure 9	Caryotype humain et cartographie cytogénétique du locus q21.3 sur le chromosome 21 et localisation du segment critique impliqué dans les corrélations pathologiques "Modifié".	16
Figure 10	Corrélation entre la trisomie 21 et la localisation génomique du gène APP sur le chromosome humain.	17
Figure 11	Contexte génomique et organisation des exons du gène APP humain (Données extraites de NCBI, Assemblage GRCh38.p14) (NCBI, 2024).	18
Figure 12	Organisation structurale et domaines fonctionnels de la protéine précurseur amyloïde (APP). (Réalisée par l'auteur (2026), adaptée d'après Priller et al., (2006) et Selkoe (2001)).	19
Figure 13	Représentation schématique du métabolisme de la protéine APP : Voies amyloïdogène et non-amyloïdogène. (Chow et al., 2010) "Adapté de Chow et al. (2010)"	20
Figure 14	Comparaison anatomique et cellulaire entre un cerveau sain et un cerveau atteint par la maladie d'Alzheimer (Docteur Clic).	24
Figure 15	Plateformes bioinformatiques des principales banques de données et outils de recherche exploités (NCBI, PubMed et BLAST).	31
Figure 16	la plateforme UniProt utilisée pour l'extraction des séquences protéiques	32
Figure 17	logiciel MEGA 12 (Molecular Evolutionary Genetics Analysis) illustrant les différents modules d'analyse (Alignement, Phylogénie, et Distances génétiques).	35
Figure 18	Serveur ProtParam (Expasy)	35
Figure 19	Base de données InterPro	36
Figure 20	Serveur SWISS-MODEL	36
Figure 21	Résultats de l'alignement de la séquence d'ARNm progéniteur du gène APP avec la séquence mutante (mutation Swedish K670N/M671L) à l'aide de l'outil BLASTn.	39

Figure 22	Représentation graphique de la distribution des meilleurs alignements séquentiels du gène <i>APP</i> via BLAST.	40
Figure 23	Visualisation de l'alignement multiple des séquences (MSA Viewer) du gène <i>APP</i> montrant le site de la mutation.	40
Figure 24	Résultats de l'alignement de séquences nucléotidiques par BLASTN pour le gène <i>APP</i> chez différentes espèces.	42
Figure 25	Graphique comparatif montrant les similitudes et les différences du gène <i>APP</i> chez : <i>Mus musculus</i> ; <i>Rattus norvegicus</i> ; <i>Canis lupus familiaris</i> ; <i>Pan troglodytes</i> et <i>Macaca mulatta</i> , en référence à <i>Homo sapiens</i> .	43
Figure 26	Visualisation de l'alignement multiple des séquences (MSA Viewer) du gène <i>APP</i> montrant les similitudes et les différences du gène <i>APP</i> chez différentes espèces.	43
Figure 27	Arbre phylogénétique des séquences nucléotidiques du gène <i>APP</i> de <i>Homo sapiens</i> et ses orthologues chez d'autres espèces de mammifères. sous le logiciel MEGA 12.	44
Figure 28	Arbre phylogénétique basé sur l'analyse de la séquence codée du gène <i>APP</i> , montrant le degré de parenté génétique.	45
Figure 29	Cartographie topologique détaillée de la protéine <i>APP</i> sauvage (770 aa) en mode d'affichage complet (<i>Full Display Mode</i>) via la plateforme InterPro, illustrant l'organisation globale des macro-domaines et des motifs structuraux conservés.	48
Figure 30	Cartographie topologique détaillée de la protéine <i>APP</i> mutée Swedish (KM670/671NL) via InterPro, montrant une conservation stricte et une identité de 100% de l'architecture des domaines par rapport à la forme sauvage.	49
Figure 31	Prédiction de la structure tridimensionnelle (3D) de la protéine <i>APP</i> humaine normale et mutante complète (version standard composée de 770 acides aminés) via le serveur SWISS-MODEL.	51
Figure 32	Superposition structurale 3D globale entre l' <i>APP</i> sauvage (représentée en vert) et la variante mutée Swedish (représentée en bleu) sous PyMOL, illustrant un repliement spatial identique et une conservation macro-conformationnelle totale.	53
Figure 33	Visualisation détaillée et zoomée sous PyMOL du site de la mutation Swedish (représentée en rouge et en mode bâtonnets) au niveau des positions 670 et 671 , illustrant l'impact micro-conformationnel local au niveau du site de coupure de la β -sécrétase.	54
Figure 34	Résultats de l'alignement de la séquence d'ARNm progéniteur du gène <i>APOE</i> $\epsilon 3$ avec la séquence mutante <i>APOE</i> $\epsilon 2$ et $\epsilon 4$ à l'aide de l'outil BLASTn.	56
Figure 35	Représentation graphique de la distribution des meilleurs alignements séquentiels du gène <i>APOE</i> via BLAST.	57
Figure 36	Résultats de l'alignement de séquences nucléotidiques par BLASTN pour le gène <i>APOE</i> chez différentes espèces.	59

Figure 37	Graphique comparatif montrant les similitudes et les différences du gène <i>ApoE</i> chez : <i>Mus musculus</i> ; <i>Rattus norvegicus</i> ; <i>Canis lupus familiaris</i> ; <i>Pan troglodytes</i> et <i>Macaca mulatta</i> , en référence à <i>Homo sapiens</i> .	60
Figure 38	Visualisation de l'alignement multiple des séquences (MSA Viewer) du gène <i>APP</i> montrant les similitudes et les différences du gène <i>APOE</i> chez différentes espèces.	60
Figure 39	Arbre phylogénétique des séquences nucléotidiques du gène <i>APOE</i> de <i>Homo sapiens</i> et ses orthologues chez d'autres espèces de mammifères. sous le logiciel MEGA 12.	62
Figure 40	Arbre phylogénétique basé sur l'analyse de la séquence codée du gène <i>ApoE</i> , montrant le degré de parenté génétique.	63
Figure 41	Cartographie topologique détaillée de la protéine <i>ApoE</i> $\epsilon 2$ sauvage (317 aa) en mode d'affichage complet (<i>Full Display Mode</i>) via la plateforme InterPro, illustrant l'organisation globale des macro-domaines et des motifs structuraux conservés.	66
Figure 42	Cartographie topologique détaillée de la protéine <i>ApoE</i> $\epsilon 2$ sauvage (317 aa) en mode d'affichage complet (<i>Full Display Mode</i>) via la plateforme InterPro, illustrant l'organisation globale des macro-domaines et des motifs structuraux conservés.	66
Figure 43	Cartographie topologique détaillée de la protéine <i>ApoE</i> $\epsilon 4$ sauvage (317 aa) en mode d'affichage complet (<i>Full Display Mode</i>) via la plateforme InterPro, illustrant l'organisation globale des macro-domaines et des motifs structuraux conservés.	67
Figure 44	Prédiction de la structure tridimensionnelle (3D) de la protéine polymorphes du gène <i>ApoE</i> ($\epsilon 2$, $\epsilon 3$, $\epsilon 4$) complète (version standard composée de 317 acides aminés) via le serveur SWISS-MODEL.	69
Figure 45	Correspondance structurale 3D globale entre la protéine et <i>APOE</i> $\epsilon 2$ (vert), <i>APOE</i> $\epsilon 3$ (gris) et <i>APOE</i> $\epsilon 4$ (rouge), obtenue à l'aide du logiciel PyMOL, démontrant une représentation spatiale identique et une conservation complète de la structure globale	70
Figure 46	Superposition structurale des polymorphes de l'apolipoprotéine E à l'aide du logiciel PyMOL, montrant la superposition complète des structures axiales des hélices (hélices $-\alpha$) d' <i>APOE</i> $\epsilon 2$ (vert), d' <i>APOE</i> $\epsilon 3$ (gris) et d' <i>APOE</i> $\epsilon 4$ (rouge).	71
Figure 47	Une comparaison microscopique moléculaire montrant les sites d'acides aminés spécifiques (positions 112 et 158) responsables de la différenciation structurale entre les trois formes de protéine <i>APOE</i> ; la molécule de cystéine (CYS-112) apparaît dans <i>APOE</i> $\epsilon 2$ et <i>APOE</i> $\epsilon 3$, tandis qu'elle est remplacée par l'arginine (ARG-112) dans la forme <i>APOE</i> $\epsilon 4$.	72

Liste des tableaux

N°	Titre des tableaux	Page
Tableau 01	Gènes clés impliqués dans la maladie d'Alzheimer et l'évaluation de leur impact génétique (Alzheimer's Association, 2026 ; Scheltens et al., 2021)	27
Tableau 2	Gènes APP de comparaison pour chacun des <i>Mus musculus</i> ; <i>Rattus norvegicus</i> ; <i>Canis lupus familiaris</i> ; <i>Pan troglodytes</i> et <i>Macaca mulatta</i> faisant référence à <i>Homo sapiens</i> .	41
Tableau 3	Analyse comparative des caractéristiques physiques et chimiques de la protéine APP normale et de la protéine APP affectée par la mutation Swedish (KM670/671NL) en se servant de l'outil ExPASy ProtParam.	46
Tableau 4	Caractéristiques des allèles du gène APOE chez l'humain	55
Tableau 5	Gènes APOE de comparaison pour chacun des <i>Mus musculus</i> ; <i>Rattus norvegicus</i> ; <i>Canis lupus familiaris</i> ; <i>Pan troglodytes</i> et <i>Macaca mulatta</i> faisant référence à <i>Homo sapiens</i> .	58
Tableau 6	Analyse comparative des caractéristiques physiques et chimiques de la protéine ApoE en se servant de l'outil ExPASy ProtParam. ($\epsilon 2$, $\epsilon 3$, $\epsilon 4$).	64
Tableau 7	Comparaison des caractéristiques des isoformes de l' <i>ApoE</i> ($\epsilon 2$, $\epsilon 3$, $\epsilon 4$)	68

Liste des Abréviations

Abréviaton	Signification
Aβ	Peptide Bêta-Amyloïde
ADN	Acide Désoxyribonucléique
AICD	Domaine Intracellulaire de l'APP
APOE	Apolipoprotéine E
APP	Protéine Précurseur de l'Amyloïde
ARN	Acide Ribonucléique
ARNm	Acide Ribonucléique messenger
BACE1	Bêta-Sécrétase 1
BLAST	<i>Basic Local Alignment Search Tool</i>
EBI	Institut Européen de Bioinformatique
EMBL	Laboratoire Européen de Biologie Moléculaire
EOAD	Forme familiale précoce de la maladie d'Alzheimer (<i>Early-Onset Alzheimer's Disease</i>)
HGP	Projet Génome Humain (<i>Human Genome Project</i>)
kb	Kilobase
KPI	Inhibiteur de sérine protéase de type Kunitz (<i>Kunitz Protease Inhibitor</i>)
MA	Maladie d'Alzheimer
MEGA	<i>Molecular Evolutionary Genetics Analysis</i>
MSA	Alignement multiple de séquences (<i>Multiple Sequence Alignment</i>)
NCBI	<i>National Center for Biotechnology Information</i>
NGS	Séquençage de nouvelle génération (<i>Next-Generation Sequencing</i>)
NHGRI	<i>National Human Genome Research Institute</i>
PAM	Mutation acceptée par point (<i>Point Accepted Mutation</i>)
PDB	<i>Protein Data Bank</i>
PSEN1	Préséniline 1
PSEN2	Préséniline 2
sAPPα	Fragment soluble alpha de l'APP
UCSC	Navigateur génomique UCSC (<i>University of California, Santa Cruz</i>)
UniProt	<i>Universal Protein Resource</i>
pI	Point isoélectrique
Da	Poids moléculaire
II	Indice d'instabilité
GRAVY	Grand Average of Hydropathicity

SOMMAIRE

Dédicaces	I
Remerciements	II
Liste des Figures	IV
Liste des Tableaux.....	V
Liste des Abréviations	VI
INTRODUCTION GÉNÉRALE	1

PARTIE I THÉORIQUE

CHAPITRE I : Bioinformatique et la maladie d'Alzheimer

I. Bioinformatique générale:	3
I.1. Définition de la bioinformatique:	3
I.2. Origine et évolution historique de la bioinformatique	3
I.2.1. Avant l'apparition de la bioinformatique:	3
I.2.2. L'Héritage de Margaret Oakley Dayhoff :	4
I.2.3. Émergence des bases de données biologiques	6
I.3. Objectifs de la bioinformatique:.....	6
I.3.1 Analyse des gènes et des protéines.....	7
I.3.2. Identification des mutations associées aux maladies.....	7
I.3.3. Étude des relations évolutives (Phylogénie)	7
I.3.4. Développement de nouveaux médicaments (Drug discovery).....	7
I.3.5. Compréhension des maladies complexes.....	7
I.4. Principaux outils de la bioinformatique:	7
I.4.1. Outils d'analyse des séquences	8
I.4.2. Outils d'analyse phylogénétique	8
I.4.3. Bases de données:	8
II. Bioinformatique génomique:.....	9
II.1. Définition bioinformatique génomique:	9
II.2. Évolution Historique de la bioinformatique génomique:	9
II.3. Le Génome : organisation et structure:	12
II.3.1 Définition du génome:	12
II.3.2 Structure du gène:.....	12
3.3 Éléments régulateurs:	13
II.4. L'analyse bioinformatique du génome.....	14
II.4.1 Outils d'analyse de séquences	14
II.4.2 Annotation et applications	14
II.5. Génomique Comparative:.....	15
II.5.1. Définition:	15
III. Le gene <i>APP</i> :.....	16
III.1. Présentation du Gène <i>APP</i> (Amyloid Precursor Protein)	16
III.2 Structure Génomique et Exons du gène <i>APP</i>	17
III.3. Organisation Structurale de la Protéine <i>APP</i>	18

III.4. Métabolisme et Rôles Physiologiques de la Protéine <i>APP</i>	19
III.4.1 La Voie Non-Amyloïdogène (Physiologique et Neuroprotectrice)	19
III.4.2 La Voie Amyloïdogène (Pathologique)	20
III.4.3 Signalisation Intracellulaire et Adhésion.....	21
III.5. Lien entre le métabolisme de <i>l'APP</i> et la Maladie d'Alzheimer	21
III.5.1 La Cascade Amyloïde	21
III.5.2 Neurotoxicité et Mort Cellulaire	22
III.6.1. Les Mutations du Gène <i>APP</i>	22
III.6.2 Mode de Transmission et Caractéristiques Cliniques	23
III.6.3 Facteurs de Risque Génétiques : Le rôle de <i>l'APOE</i>	23
IV. La Maladie d'Alzheimer : Contexte Clinique et Moléculaire	23
IV.1 Définition et Épidémiologie	23
IV.2 Physiopathologie :	24
IV.3 Les formes de la maladie	25

PARTIE II PRATIQUE

Chapitre II : Matériel et Méthodes

II.1. Objectif de l'étude et Choix de l'Organisme Modèle:	26
A. Objectif de l'étude :	26
B. Choix de l'organisme modèle:	26
II.2. Choix des gènes étudiés:	26
II.3. Matériel:	29
II.3.1. Bases de données biologiques :	29
1. Définition du Centre national d'information sur la biotechnologie (NCBI)	29
2. Base de données UniProtKB (Ressource universelle des protéines).....	31
II.3.2. Notepad++ (Éditeur de texte)	31
II.3.3. BLAST (Basic Local Alignment Search Tool).....	32
II.3.4. Outil MAFFT (Multiple Alignment using Fast Fourier Transform).....	32
II.3.5. MEGA12 (Analyse génétique de l'évolution moléculaire, version 12)	33
II.1.6. Serveur ProtParam (Expasy).....	35
II.1.7. Base de données InterPro	35
II.1.8. Serveur SWISS-MODEL	36
II.1.9. PyMOL (Molecular Graphics System)	36
III.4. Méthodes	37
II.4.1. BLAST	37
II.4.2. MEGA 12.....	37
II.4.3. Serveur ProtParam (Expasy).....	38
II.4.4. Plateforme InterPro	38
II.4.5. Serveur Swiss Model.....	38
II.4.6. Programme PyMOL	38

Chapitre III : Résultats et discussion

I. Résultats :	39
----------------------	----

I.1- Analyse comparative du gène <i>APP</i> (Protéine précurseur de l'amyloïde).....	39
I.1.1- Comparaison des séquences nucléotidiques des gènes <i>APP</i> :	41
I.1.2- Étude comparative représentative	43
I.1.3- Alignement multiple des séquences (AMS)	44
I.1.4- Étude de classification moléculaire des gènes <i>APP</i> codant.....	45
I.2- Analyse et Caractérisation de la Protéine <i>APP</i>	46
I.2.1.- Étude des propriétés physico-chimiques (ProtParam)	46
I.2.2.- Analyse des domaines protéiques (InterPro)	47
I.2.3.- Structure tridimensionnelle (SWISS-MODEL + PyMOL)	50
1. Modélisation conforme via le serveur SWISS-MODEL	50
2. Appariement spatial global et calcul de l'indice RMSD à l'aide du logiciel	
PyMOL	52
3. Analyse de la localisation structurale et de l'environnement atomique du site de	
mutation	54
I.3- Analyse comparative du gène <i>APOE</i> (Apolipoprotéine E).....	58
I.3.1- Comparaison des séquences nucléotidiques des gènes <i>APOE</i> :	55
I.3.2- Étude comparative représentative	60
I.3.3- Alignement multiple des séquences (AMS)	61
I.3.4- Étude de classification moléculaire des gènes <i>APOE</i> codant.....	63
I.4- Analyse et Caractérisation de la Protéine <i>APOE</i>	64
I.4.1.- Étude des propriétés physico-chimiques (ProtParam)	64
I.4.2.- Analyse des domaines protéiques (InterPro)	66
I.4.3.- Structure tridimensionnelle (SWISS-MODEL + PyMOL)	69
1. Modélisation conforme via le serveur SWISS-MODEL	69
2. Appariement spatial global et calcul de l'indice RMSD à l'aide du logiciel	
PyMOL	70
3. Analyse de la localisation structurale et de l'environnement atomique du site de	
mutation	72
II. Discussion:	74
II.1. Discussion de l'analyse comparative du gène <i>APP</i>	74
II.2. Discussion de l'analyse comparative du gène <i>APOE</i> (Apolipoprotéine E)	74
II.3. Discussion comparative entre les espèces des gènes <i>APP</i> et <i>ApoE</i>	75
II.4. Discussion du gène <i>APP</i>	75
II.5. Discussion du gène <i>ApoE</i>	76
II.6. Discussion générale de l'analyse phylogénétique et de la comparaison	
moléculaire des gènes (<i>APP</i> et <i>APOE</i>).....	77
II.7. Discussion structurale et évolutive de la protéine et du gène <i>APP</i> mutant et	
normal.....	78
II.8. Discussion Évolutive et Structurale de la Protéine et du Gène <i>APOE</i>	79
 CONCLUSION GÉNÉRALE.....	81
RÉFÉRENCES.....	83
ANNEXES	
Résumé / Abstract	

Introduction

Introduction:

Cette étude utilise une méthode analytique moderne, *in silico*, pour identifier les gènes et les variants génétiques associés à la maladie d'Alzheimer. Elle vise à explorer les modifications moléculaires et les marqueurs génétiques liés à cette maladie grâce à des approches informatiques avancées (Goodwin *et al.*, 2016). La maladie d'Alzheimer (MA) demeure l'une des maladies neurodégénératives les plus répandues et les plus mortelles au monde, représentant la majorité des cas de démence et touchant des millions de personnes chaque année. En raison du vieillissement démographique accéléré, cette maladie représente un fardeau sanitaire, social et économique considérable à l'échelle mondiale. Selon les statistiques de l'Organisation mondiale de la Santé, plus de 55 millions de personnes vivent actuellement avec la démence, et les projections indiquent que ce nombre atteindra environ 139 millions d'ici 2050 (OMS, 2023).

Sur le plan biologique, la maladie d'Alzheimer se caractérise par un déclin progressif et irréversible des fonctions cognitives telles que la mémoire, le langage et la capacité de raisonner de manière autonome (Scheltens *et al.*, 2021 ; Jack *et al.*, 2024). Cette neurodégénérescence est définie histologiquement par deux phénomènes principaux : l'accumulation de peptides bêta-amyloïdes extracellulaires formant des « plaques séniles », et l'agrégation intracellulaire de protéines tau hyperphosphorylées sous forme d'« enchevêtrements neurofibrillaires » (Hardy & Higgins, 1992). Pour expliquer ces déficits, la recherche actuelle se concentre sur les facteurs génétiques, et deux gènes clés aux mécanismes d'action différents se distinguent : le gène *APP* (protéine précurseur de l'amyloïde) et le gène *APOE* (apolipoprotéine E).

Les mutations ponctuelles du gène *APP*, situé sur le chromosome 21, sont directement responsables des formes familiales précoces et rares de la maladie, à transmission autosomique dominante. Ces mutations entraînent un clivage anormal de la protéine et stimulent la production du peptide toxique $A\beta_{1-42}$ (Lanoiselée *et al.*, 2017). En revanche, le gène *APOE*, situé sur le chromosome 19, est le principal déterminant génétique des formes sporadiques courantes ; son allèle variant *APOE* $\epsilon 4$ contribue à une élimination cellulaire altérée des dépôts amyloïdes et augmente le risque de forme sévère de la maladie (Jack *et al.*, 2018 ; Mormino *et al.*, 2018).

Grâce à la vaste gamme de technologies de séquençage génétique et de bioinformatique, l'analyse *in silico* est devenue un outil indispensable pour étudier avec précision les profils génétiques des maladies neurodégénératives, sans nécessiter de procédures de laboratoire complexes (Goodwin *et al.*, 2016).

Les plateformes de génomique en ligne telles que NCBI, UniProt et GenBank, associées à des outils comme BLAST, MAFFT et MEGA12, ainsi qu'à des outils de prédiction fonctionnelle comme SWISS-MODEL et InterPro, constituent de puissantes ressources pour l'extraction de séquences génétiques, la détection d'altérations nucléotidiques et même protéiques, la prédiction

Introduction

des effets structuraux des mutations, l'extraction des propriétés physico-chimiques des protéines à l'aide de ProtParam (Expasy) et l'évaluation des relations évolutives entre les allèles sauvages et mutants (**Karczewski & Snyder, 2018**).

Malgré ces progrès, la complexité de la variation génétique chez les patients soulève une question cruciale :

Dans quelle mesure les outils bioinformatiques peuvent-ils identifier et caractériser avec précision les mutations et variants pathogènes des gènes *APP* et *APOE* ?

En utilisant ces gènes comme modèle d'étude, quel est le degré de similarité et de relation évolutive de ces gènes humains par rapport à ceux d'autres organismes ?

Partant de ce constat, la présente étude vise à mettre en œuvre un flux de travail informatique complet pour extraire les séquences des gènes *APP* et *APOE* des bases de données, réaliser un alignement multiple de séquences afin de localiser les variants, les interrompre à l'aide de BLAST, puis prédire les conséquences structurales et fonctionnelles de ces mutations à l'aide d'InterPro, de SWISS-MODEL et de PyMOL. Grâce à cette approche intégrée, cette étude cherche à approfondir notre compréhension des mécanismes moléculaires sous-jacents à la maladie d'Alzheimer et à contribuer à l'élaboration de futures stratégies de médecine de précision et de thérapies ciblées (**Cummings et al., 2024**).

PARTIE I
THÉORIQUE:

Chapitre I :

*La bioinformatique et la
maladie d'Alzheimer*

I. Bioinformatique générale:

I.1. Définition de la bioinformatique:

La bioinformatique est un domaine multidisciplinaire qui fusionne biologie, informatique et mathématiques pour analyser et interpréter les données biologiques complexes, en particulier les séquences d'ADN, d'ARN et de protéines, grâce à des outils computationnels avancés (Lesk, 2019).

En termes simples, il s'agit l'utilisation de l'informatique pour résoudre des problèmes biologiques ardues, tels que :

- Identifier les gènes responsables de maladies (ex. le gène *APP* dans la maladie d'Alzheimer)
- Comparer des séquences (sequence alignment)
- Analyser la structure et la fonction des protéines

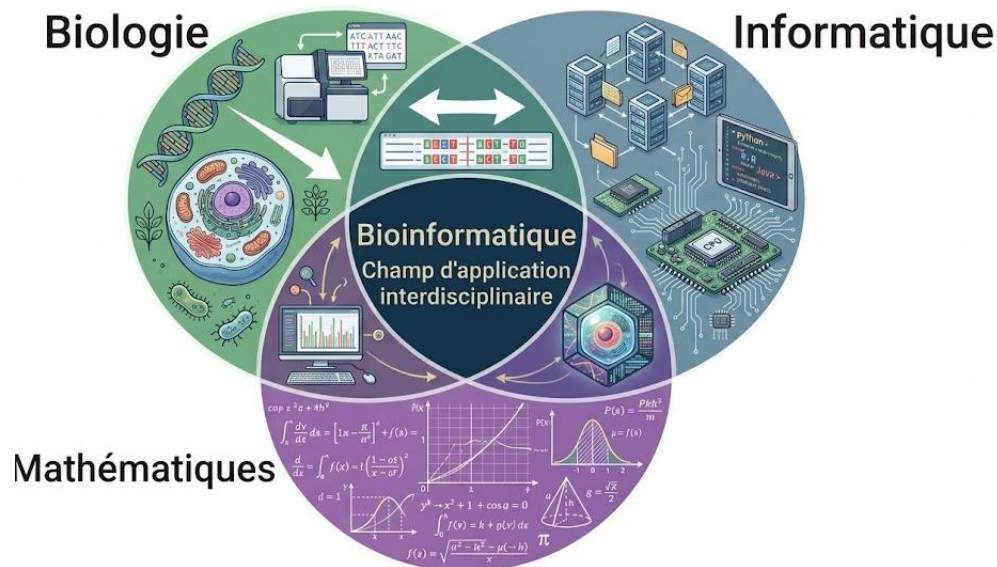


Figure 1: Bioinformatique comme champ d'application interdisciplinaire ; un modèle d'intégration de la Biologie, de l'Informatique et des Mathématiques.

I.2. Origine et évolution historique de la bioinformatique

I.2.1. Avant l'apparition de la bioinformatique:

Au début des sciences de la biologie moléculaire, les chercheurs dépendaient entièrement des méthodes expérimentales traditionnelles, où :

- L'analyse et la comparaison de deux séquences simples d'ADN ou de protéines pouvaient prendre des années, à cause de l'immense volume de données à traiter manuellement.
- Gérer de grandes quantités d'informations biologiques relevait de l'impossible, faute d'outils adaptés.
- Identifier les gènes liés aux maladies, tels que le gène *APP* responsable de la maladie d'Alzheimer – était un travail de longue haleine, limité à quelques cas seulement.

Ces obstacles rendaient toute comparaison globale entre gènes ou protéines hors de portée, transformant la recherche en un processus extrêmement lent et laborieux (David W. Mount., 2004).



Figure 2 : L'impasse de l'analyse manuelle des séquences génomiques ; un aperçu de la complexité pré-bioinformatique

I.2.2.L'Héritage de Margaret Oakley Dayhoff : Pionnière de la Bioinformatique Moderne

L'émergence de la bioinformatique en tant que discipline scientifique rigoureuse est indissociable des travaux visionnaires de Margaret Oakley Dayhoff (1925-1983). À une époque où les ressources computationnelles étaient limitées, elle a su transformer des données biologiques brutes en informations structurées et comparables.

1. La naissance des premières bases de données biologiques En 1965, Dayhoff a publié l'**Atlas of Protein Sequence and Structure**, qui constitue la première tentative mondiale de compilation systématique des séquences protéiques connues (Dayhoff, 1965). Ce recueil est aujourd'hui considéré comme l'ancêtre direct des banques de données modernes telles que GenBank ou UniProt (NCBI, 2024).

2. Optimisation du traitement des données : Le code à une lettre Afin de faciliter le stockage informatique et la comparaison visuelle des séquences, elle a introduit le système de codification des acides aminés par une seule lettre (**Dayhoff et al., 1978**). Ce passage d'une nomenclature à trois lettres à une lettre unique a permis d'optimiser les algorithmes de recherche de similitudes, une norme qui prévaut encore dans l'analyse génomique actuelle (**Mount, 2001**).

3. Modélisation de l'évolution : Les matrices PAM L'une de ses contributions les plus fondamentales reste le développement des matrices de substitution PAM (*Point Accepted Mutation*). En analysant statistiquement les fréquences de mutations acceptées au cours de l'évolution, elle a fourni un cadre mathématique pour évaluer l'homologie entre les séquences (**Dayhoff et al., 1978**). Ces matrices constituent le moteur algorithmique des outils d'alignement tels que BLAST, essentiels pour l'étude de la conservation de gènes complexes tels que le gène APP (**Mount, 2001**).

4. Reconstruction de l'histoire évolutive Dayhoff a été parmi les premières à utiliser des approches algorithmiques pour inférer des arbres phylogénétiques à partir de séquences moléculaires, prouvant ainsi que l'information génétique contient les archives de l'histoire du vivant (**Dayhoff, 1966**).



Figure 3: Margaret Dayhoff et l'émergence des premières matrices de substitution ; une réponse à la complexité des séquences biologiques .

I.2.3.Émergence des bases de données biologiques

Avec l'évolution rapide de la biologie moléculaire, la nécessité de disposer de moyens efficaces pour organiser et stocker des volumes massifs de données biologiques est devenue impérative. Cela a conduit à l'apparition de bases de données spécialisées qui constituent aujourd'hui les piliers fondamentaux de la bioinformatique.

Parmi les bases de données les plus importantes, on peut citer :

- **GenBank** : une base de données mondiale dédiée au stockage des séquences d'acide désoxyribonucléique (ADN), gérée par le National Center for Biotechnology Information (**Benson et al., 2013**).
- **Protein Data Bank (PDB)** : la base de référence internationale spécialisée dans la structure tridimensionnelle des protéines et des macromolécules biologiques (**RCSB, 2000**).

Ces plateformes ont profondément transformé la recherche scientifique en offrant plusieurs avantages :

1. Le stockage organisé de millions de séquences biologiques, facilement accessibles et interrogeables (**NCBI, 2018**).
2. La facilitation des analyses comparatives de séquences (*Sequence comparison*) grâce à des outils informatiques avancés tels que **BLAST** et **MEGA**.
3. La réduction du temps d'analyse, passant de plusieurs années de travail manuel à quelques secondes seulement, ce qui a considérablement accéléré l'identification des gènes pathogènes et l'étude des relations évolutives entre les espèces.

I.3. Objectifs de la bioinformatique:

La bioinformatique constitue un domaine clé dans les sciences modernes, visant à exploiter les outils informatiques pour analyser et interpréter les données biologiques complexes. Ses principaux objectifs sont les suivants :

I.3.1. Analyse des gènes et des protéines

La bioinformatique permet d'étudier les séquences d'ADN et de protéines afin d'identifier leur structure, leur fonction et leurs interactions, contribuant ainsi à une meilleure compréhension des mécanismes cellulaires (David W. Mount., 2004).

I.3.2. Identification des mutations associées aux maladies

Grâce aux outils d'analyse génétique, il est possible de détecter les mutations responsables de nombreuses pathologies, ce qui favorise le diagnostic précoce et l'amélioration des stratégies thérapeutiques (Lesk, 2019).

I.3.3. Étude des relations évolutives (Phylogénie)

L'analyse comparative des séquences biologiques permet de reconstituer les relations évolutives entre les organismes et de mieux comprendre l'évolution des gènes au cours du temps (David W. Mount., 2004).

I.3.4. Développement de nouveaux médicaments (Drug discovery)

La bioinformatique joue un rôle essentiel dans la conception de médicaments en permettant d'identifier des cibles thérapeutiques et de modéliser les interactions moléculaires, réduisant ainsi le temps et les coûts de développement (Lesk, 2019).

I.3.5. Compréhension des maladies complexes

C'est le cas de la maladie d'Alzheimer, pour laquelle les approches bioinformatiques permettent d'analyser les gènes impliqués, notamment le gène *APP* (*Amyloid Precursor Protein*), afin de mieux comprendre les mécanismes moléculaires de la neurodégénérescence (NCBI, 2018).

I.4. Principaux outils de la bioinformatique:

L'analyse bioinformatique repose sur l'utilisation de logiciels et de plateformes spécialisés permettant de traiter les données génomiques et protéiques. Les outils les plus couramment utilisés sont :

I.4.1. Outils d'analyse des séquences

- **BLAST** (*Basic Local Alignment Search Tool*) : Il permet de comparer une séquence donnée (ADN ou protéine) avec des millions de séquences stockées dans les bases de données afin d'identifier les similarités, prédire la fonction des gènes et détecter les mutations (Altschul et al., 1990).
- **Clustal Omega** : Cet outil est utilisé pour l'alignement multiple de séquences (*Multiple Sequence Alignment - MSA*). Il permet d'identifier les régions conservées, d'étudier les relations fonctionnelles et de préparer les données pour les analyses phylogénétiques (Sievers et al., 2011).

I.4.2. Outils d'analyse phylogénétique

- **MEGA** (*Molecular Evolutionary Genetics Analysis*) : C'est un logiciel polyvalent qui permet de construire des arbres phylogénétiques, d'analyser la divergence génétique et d'étudier les relations évolutives entre les espèces ou les gènes (Kumar et al., 2018).

I.4.3. Bases de données:

- **NCBI** : Fournit un accès centralisé à plusieurs bases de données, notamment **GenBank**, ainsi qu'à des outils d'analyse essentiels comme BLAST (NCBI, 2018).
- **UniProt** : Une base de données spécialisée dans les séquences et les fonctions des protéines, largement utilisée pour l'annotation et l'analyse fonctionnelle précise (UniProt, 2023).



Figure 4 : Principaux outils de la bioinformatique moderne (2026).

II. Bioinformatique génomique:

La bioinformatique génomique est une branche avancée de la bioinformatique qui se concentre sur l'étude des génomes complets des organismes vivants à l'aide des ordinateurs et des outils logiciels (Pevsner, 2015). Ce domaine vise à **analyser et comprendre les séquences d'ADN et d'ARN**, identifier les gènes fonctionnels, étudier les mutations génétiques et comprendre l'évolution des organismes au niveau génomique (Lesk, 2019).

La bioinformatique génomique se distingue par sa capacité à :

- Détecter les maladies génétiques et analyser les schémas de mutations (Koonin & Galperin, 2003).
- Comparer les gènes entre différentes espèces afin de comprendre les relations évolutives.
- Contribuer au développement de la médecine personnalisée, à l'amélioration des cultures agricoles et à la recherche scientifique, en offrant une compréhension complète des fonctions biologiques à l'échelle du génome (Mount, 2004).

II.1. Définition bioinformatique génomique:

La **bioinformatique génomique** est une sous-discipline de la bioinformatique dédiée à l'analyse, l'annotation et la comparaison des séquences issues du génome, en utilisant des outils computationnels et des bases de données à grande échelle (Pevsner, 2015). Elle permet de décrypter l'organisation des gènes et les variations génétiques au sein des populations (NHGRI, 2024).

II.2. Évolution Historique de la bioinformatique génomique:

La bioinformatique génomique s'est développée à travers une série d'avancées scientifiques majeures, marquant le passage de la biologie descriptive à la biologie computationnelle. Elle trouve son origine en **1953** avec la découverte de la structure en double hélice de l'**ADN**, posant ainsi les bases de la génomique moderne (Watson & Crick, 1953). Il est désormais fondamental de souligner que cette découverte capitale n'aurait pas été possible sans les travaux cruciaux de Rosalind Franklin. Bien que Watson et Crick aient publié le modèle théorique, les données

expérimentales décisives — notamment la célèbre **Photo 51 "Figure 05 "** obtenue par diffraction des **rayons X** — **Figure 05**) ont été générées par Franklin. Son rôle prépondérant a été longtemps minimisé avant d'être pleinement reconnu par la communauté scientifique internationale comme une contribution indispensable à la compréhension du support moléculaire de l'hérédité (**Franklin & Gosling, 1953 ; Maddox, 2002 ; Nature Education, 2014**).



Figure 05 : Rosalind Franklin, la photo 51 et le modèle de Watson-Crick.
(Source: **Franklin & Gosling, 1953 ; King's College London, 2023**).

En **1977**, une nouvelle étape est franchie avec l'introduction de la première méthode de séquençage chimique par Frederick Sanger **Figure 06**, permettant enfin l'analyse directe et précise des séquences nucléotidiques (**Sanger et al., 1977**).



Figure 06 : Pionnier du séquençage de première génération : F. Sanger et l'interprétation électrophorétique des fragments d'ADN. (Source : **Labster Theory, 2024 ; First-generation sequencing**).

La structuration massive de ces données biologiques débute réellement dans les années **1980**, notamment avec la création de GenBank en 1982 par Walter Goad, suivie de la fondation du National Center for Biotechnology Information (NCBI) en 1988 (**NCBI, 2024 ; Strasser, 2008**).

Ces institutions ont permis d'instaurer un cadre mondial pour le stockage et le partage universel des séquences génomiques.

En 1990, le Human Genome Project (HGP) marque un tournant historique en visant le séquençage complet du code génétique humain. Ce projet titanesque aboutit en 2001 à la publication de la première version de référence du génome humain (NHGRI, 2024). Enfin, depuis 2010, l'émergence des technologies de Next-Generation Sequencing (NGS) a révolutionné le domaine. En permettant un séquençage massif, rapide et à haut débit, le NGS a propulsé la génomique dans l'ère du Big Data, nécessitant aujourd'hui des outils computationnels et des algorithmes de plus en plus sophistiqués pour l'analyse et l'interprétation des données biologiques (Pevsner, 2015 ; EBI, 2024).

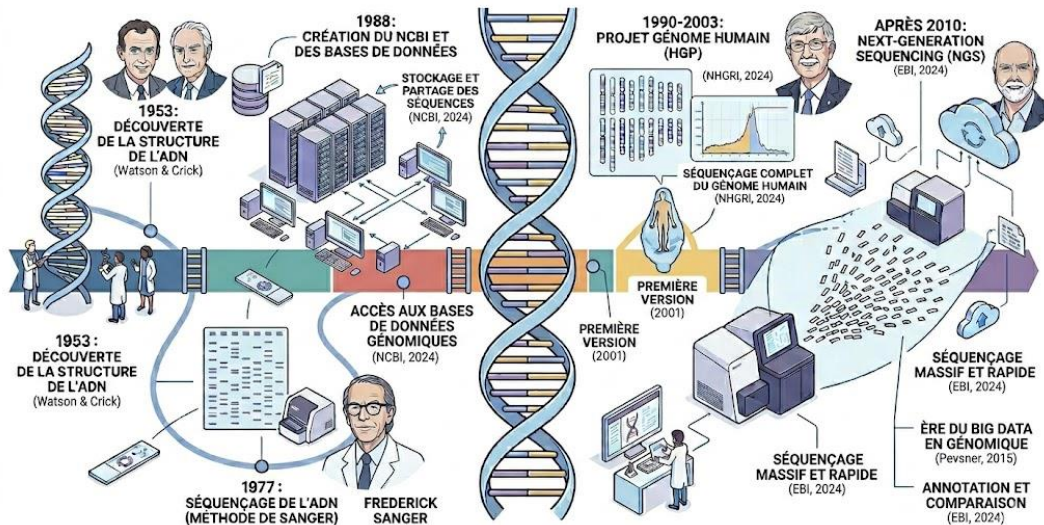


Figure 07 : Chronologie de la génomique moderne : des fondements structuraux à l'ère du Big Data (Conception personnelle, 2026 ; inspirée du NCBI et de l'EBI).

II.3. Le Génome : organisation et structure:

II.3.1. Définition du génome:

Le génome représente l'intégralité de l'information génétique d'un organisme, codée sous forme d'ADN. Il englobe non seulement les séquences codantes (gènes) **Figure 08**, mais également les régions non codantes, incluant les éléments régulateurs et structurels (**Brown, 2017**). Chez *Homo sapiens*, le génome nucléaire est organisé en 23 paires de chromosomes et contient environ 20 000 à 25 000 gènes codants pour des protéines. Toutefois, ces séquences codantes ne représentent qu'environ 1,5 % du génome total, le reste étant composé d'ADN non codant essentiel à la régulation génique et à la plasticité évolutive (**International Human Genome Sequencing Consortium, 2004**).

II.3.2. Structure du gène:

Un gène eucaryote est une unité transcriptionnelle complexe, définie par une alternance de segments fonctionnels **Figure 08** :

- **Exons** : Segments d'ADN conservés dans l'ARN messager mature après l'épissage, destinés à être traduits en protéines ou à former des ARN fonctionnels.
- **Introns** : Séquences non codantes intercalaires, éliminées lors du processus d'épissage (splicing).

Contrairement aux conceptions initiales, les introns ne sont pas de l'« ADN poubelle » ; ils jouent des rôles cruciaux dans la régulation de l'expression via l'épissage alternatif "**Figure 08**" (permettant à un seul gène de coder plusieurs protéines) et dans la stabilité structurelle du génome (**Pevsner, 2015**). Chez l'humain, la taille moyenne des introns dépasse largement celle des exons, une complexité structurale qui distingue les eucaryotes supérieurs des procaryotes (**Alberts et al., 2017**).

3.3. Éléments régulateurs:

Le contrôle spatio-temporel de l'expression génique repose sur des séquences cis-régulatrices spécifiques :

- **Promoteur** : Région située immédiatement en amont (5') du site d'initiation de la transcription, servant de plateforme de recrutement pour l'ARN polymérase et les facteurs de transcription généraux.
- **Enhancer (Enshanteur)** : Séquence régulatrice à distance capable d'augmenter significativement le taux de transcription d'un gène cible, indépendamment de sa distance ou de son orientation.
- **Silencer (Silencieux)** : Séquence capable de réprimer ou d'inhiber l'expression génique en bloquant l'accès de la machinerie de transcription.

Ces éléments assurent la coordination complexe nécessaire au développement cellulaire et aux réponses physiologiques, faisant le pont entre la structure statique du génome et sa fonction biologique dynamique (Lodish et al., 2021).

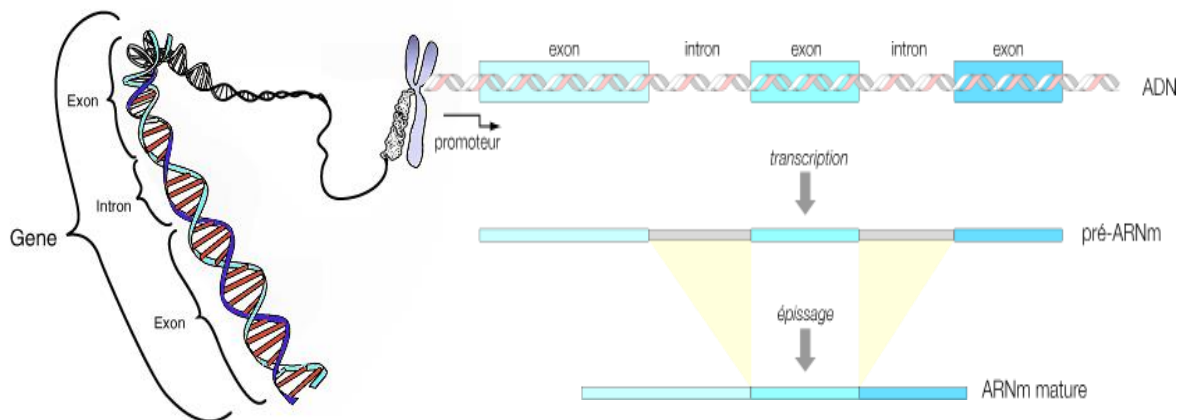


Figure 08 : Organisation d'un gène eucaryote et processus de maturation de l'ARN (Transcription et Épissage). **Gène : Structure et expression chez les eucaryotes. Récupéré).**

II.4. L'analyse bioinformatique du génome

L'analyse bioinformatique du génome consiste à exploiter des outils computationnels pour traiter, comparer et interpréter les séquences d'ADN à grande échelle. Elle permet de transformer des données brutes issues du séquençage en informations biologiques structurées, essentielles pour comprendre la structure, la fonction et l'évolution des gènes (Pevsner, 2015).

II.4.1 .Outils d'analyse de séquences

Parmi les algorithmes les plus sollicités, nous citons : "**Figure 4** "

- **BLAST (Basic Local Alignment Search Tool)** : permet de comparer rapidement une séquence d'ADN ou de protéine avec des bases de données mondiales (GenBank) afin d'identifier des homologies de séquences et d'inférer des fonctions biologiques potentielles (Altschul et al., 1990).
- **Clustal Omega** : utilisé pour réaliser des alignements multiples de séquences (MSA). Cet outil est indispensable pour l'étude des relations phylogénétiques et l'identification des régions hautement conservées au cours de l'évolution (Sievers & Higgins, 2018).
- **Navigateurs Génomiques (Genome Browsers)** : à l'instar de UCSC Genome Browser ou Ensembl, ils permettent de visualiser l'organisation spatiale du génome, incluant la cartographie des gènes et des éléments régulateurs (promoteurs, enhanceurs) (Kent et al., 2002).

II.4.2. Annotation et applications

Ces outils sont au cœur de l'annotation génomique, processus consistant à identifier les gènes au sein d'une séquence brute et à leur attribuer des fonctions biologiques précises. Ainsi, la bioinformatique génomique occupe une place centrale dans la recherche biomédicale contemporaine, facilitant l'émergence de la médecine personnalisée et le développement de nouvelles solutions en biotechnologie (NCBI, 2024).

II.5. Génomique Comparative:

II.5.1. Définition:

La génomique comparative est l'étude comparative de la structure et de la fonction des génomes de différentes espèces, dont les buts principaux sont de mieux comprendre comment les différentes espèces ont évolué, quels sont les effets de la sélection sur l'organisation et l'évolution des génomes, ainsi que de déterminer les fonctions des gènes et des régions non codantes du génome (Raphaël, 2010). La génomique comparative est une branche de la génomique qui vise à caractériser les similitudes et les différences des caractéristiques génomiques et à tracer leur gain et leur perte le long de différentes lignées évolutives, comprendre les forces évolutives telles que la mutation et la sélection qui régissent les changements de ces caractéristiques génomiques, et découvrir comment l'évolution génomique peut nous aider à combattre les maladies, restaurer la santé environnementale, gagner de l'argent, etc. (Xia, 2013). Les analyses comparatives en génomique peuvent se concentrer sur la similarité et les différences entre l'annotation ou entre la séquence de deux génomes ou plus (Herrero et *al.*, 2015).

À titre d'exemple, le génome du Chimpanzé (*Pan troglodytes*) présente une identité de séquence d'environ 98 à 99 % avec celui d'*Homo sapiens*, ce qui en fait un modèle phylogénétique clé pour l'étude de l'évolution humaine. De même, la Souris (*Mus musculus*) est un organisme modèle prédominant en recherche biomédicale en raison de la forte conservation de nombreux gènes orthologues entre le rongeur et l'homme (Brown, 2017).

Cependant, malgré ces homologies génétiques, il existe des variations significatives dans l'organisation structurale des gènes, notamment au niveau de la taille des introns et du nombre d'exons. Ces variations peuvent influencer les profils d'expression génique (via l'épissage alternatif) et expliquer ainsi les disparités phénotypiques observées entre les espèces. Ainsi, la génomique comparative démontre que la similarité des séquences n'implique pas systématiquement une identité fonctionnelle, la régulation et la structure des gènes jouant un rôle déterminant.

L'analyse du génome, combinée aux approches de génomique comparative, constitue un levier essentiel pour décrypter les bases moléculaires des pathologies humaines. L'étude de la structure

des gènes (exons/introns) et de leur conservation interspécifique permet d'identifier des mutations délétères, d'élucider les mécanismes moléculaires sous-jacents aux maladies et de développer de nouvelles stratégies diagnostiques et thérapeutiques ciblées (Lodish et *al.*, 2021).

III. Le gene *APP* :

III.1. Présentation du Gène *APP* (Amyloid Precursor Protein)

Le gène *APP* code pour la protéine précurseur de l'amyloïde, une protéine transmembranaire de type I "Figure 12" exprimée de façon ubiquitaire, mais particulièrement abondante dans les synapses neuronales. Elle joue un rôle physiologique essentiel dans la plasticité synaptique, l'adhésion cellulaire et la croissance axonale (Priller et *al.*, 2006).

Chez l'humain, le gène *APP* est localisé sur le bras long du chromosome "Figure 09" plus précisément à la position cytogénétique 21q21.3 (NCBI, 2024) "Figure 09".

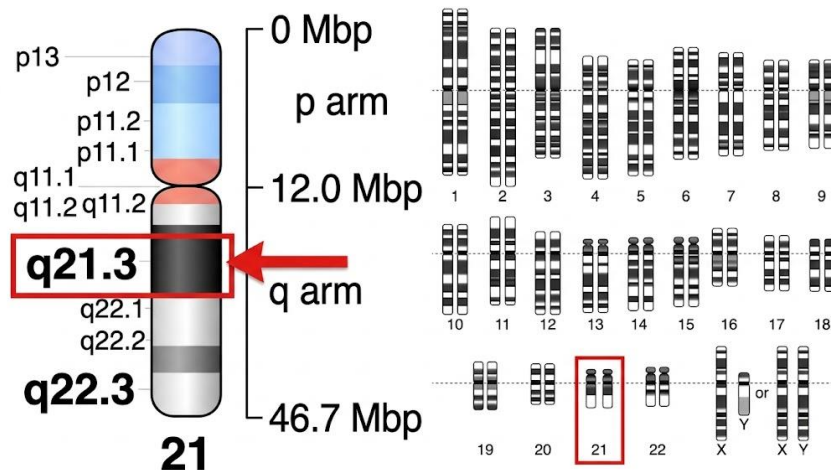


Figure 09 : Caryotype humain et cartographie cytogénétique du locus q21.3 sur le chromosome 21 et localisation du segment critique impliqué dans les corrélations pathologiques "Modifié".

Cette localisation est d'une importance stratégique pour comprendre certaines corrélations pathologiques. En effet, il est cliniquement établi que les individus atteints de trisomie 21 (syndrome de Down) possèdent une troisième copie de ce gène, ce qui entraîne une surexpression constitutionnelle de la protéine *APP* et conduit quasi systématiquement au développement précoce de lésions de type Alzheimer (Wiseman et *al.*, 2015).

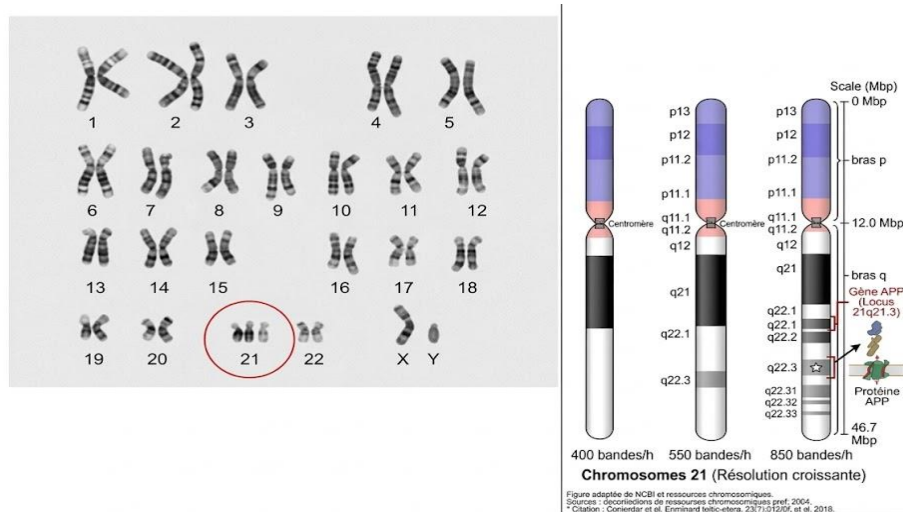


Figure 10 : Corrélation entre la trisomie 21 et la localisation génomique du gène *APP* sur le chromosome humain

III.2 Structure Génomique et Exons du gène *APP*

La structure du gène *APP* est particulièrement complexe et fragmentée, s'étendant sur environ 240 kb d'ADN génomique. Cette complexité permet une régulation fine de son expression à travers différents tissus (NCBI, 2024).

- **Nombre d'Exons :** Bien que les modèles classiques décrivent souvent ce gène comme étant composé de 18 exons (Yoshikai et al., 1990), les bases de données génomiques récentes, telles que NCBI (RefSeq GRCh38.p14) **Figure 11**, identifient désormais jusqu'à 20 exons selon les variantes de transcription répertoriées (NCBI, 2024). Cette variabilité reflète la complexité des régions non codantes et les sites d'initiation de la transcription.
- **Épissage Alternatif (Alternative Splicing) :** Grâce à ce mécanisme post-transcriptionnel, le gène *APP* génère plusieurs isoformes d'ARNm dont la longueur varie de 695 à 770 acides aminés (Kang et al., 1987). Les trois isoformes majeures exprimées chez l'homme sont :
 1. ***APP695* :** C'est l'isoforme prédominante dans le système nerveux central. Elle se distingue par l'absence de l'exon 7 (Sandbrink et al., 1994).
 2. ***APP751* :** Cette variante contient l'exon 7, qui code pour un domaine inhibiteur de sérine protéase de type Kunitz (KPI) (Tanzi et al., 1988).

3. **APP770** : Il s'agit de l'isoforme la plus longue, contenant à la fois les exons 7 et 8. Elle est principalement exprimée dans les tissus périphériques non neuronaux (Kitaguchi *et al.*, 1988).

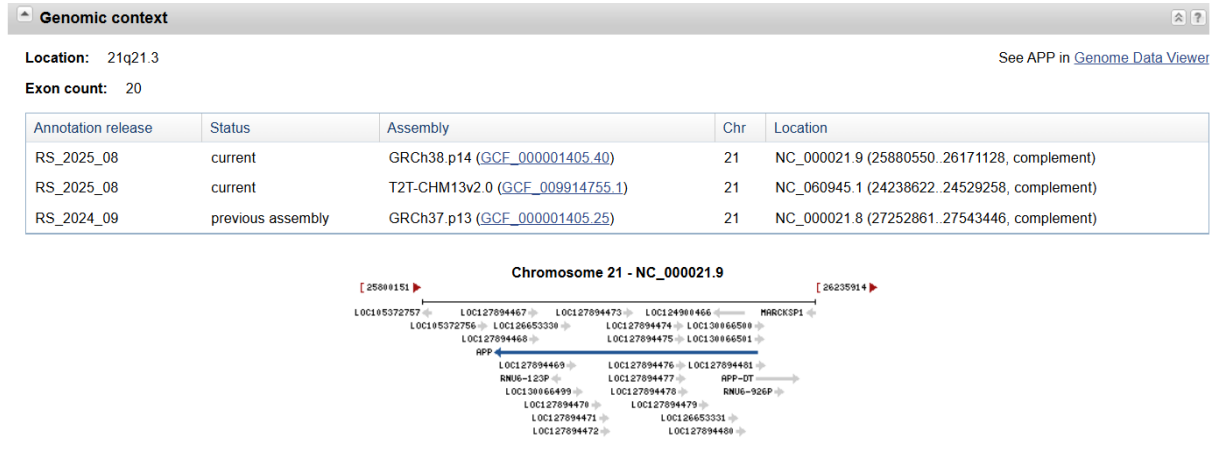


Figure 11 : Contexte génomique et organisation des exons du gène *APP* humain (Données extraites de NCBI, Assemblage GRCh38.p14) (NCBI, 2024).

III.3. Organisation Structurale de la Protéine *APP*

Une fois traduit, le gène *APP* donne naissance à une protéine précurseur transmembranaire de type I. Sa structure mature est hautement organisée et se divise en trois domaines fonctionnels distincts

Figure 12 :

Domaine Extracellulaire (N-terminal) : Il représente la partie la plus volumineuse de la protéine. Ce domaine est impliqué dans des fonctions physiologiques vitales telles que l'adhésion cellulaire, la croissance neuritique et la neuroprotection (Priller *et al.*, 2006). Il contient également des sites de liaison pour des métaux (Cuivre et Zinc) et l'héparine, régulant ainsi l'homéostasie neuronale (Multhaup *et al.*, 1996).

Domaine Transmembranaire : Ce domaine est constitué d'une hélice alpha unique qui traverse la bicouche lipidique. Son importance pathologique est majeure car c'est ici que se loge la séquence du peptide β -amyloïde ($A\beta$), située à cheval entre la membrane plasmique et le milieu extracellulaire (Kang *et al.*, 1987). C'est la coupure séquentielle de ce domaine par les sécrétases (β et γ) qui libère le peptide neurotoxique (Selkoe, 2001).

Domaine Intracellulaire (C-terminal) : Souvent appelé *AICD* (*APP* Intracellular Domain) ou queue cytosolique, ce domaine court est essentiel pour la signalisation intracellulaire. Il interagit avec diverses protéines adaptatrices pour réguler le transport axonal, la transcription génique et l'apoptose (Cao & Südhof, 2001).

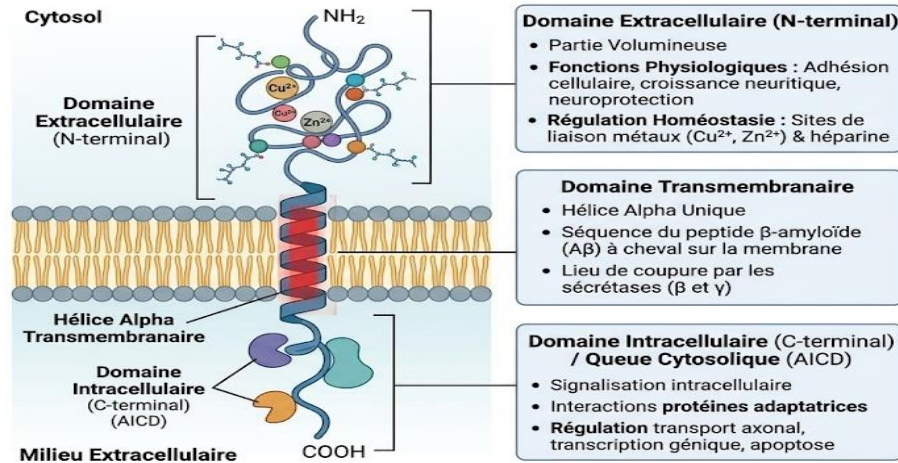


Figure 12 : Organisation structurale et domaines fonctionnels de la protéine précurseur amyloïde (*APP*). (Réalisée par l'auteur (2026), adaptée d'après Priller et al., (2006) et Selkoe (2001)).

III.4. Métabolisme et Rôles Physiologiques de la Protéine *APP*

Le devenir de la protéine *APP* (Amyloid Precursor Protein) au sein du système nerveux n'est pas un simple processus de dégradation, mais un mécanisme biologique finement régulé. Ce processus, appelé clivage protéolytique, permet la libération de fragments actifs (peptides) ayant des fonctions essentielles. On distingue deux voies métaboliques aux conséquences opposées :

III.4.1 La Voie Non-Amyloïdogène (Physiologique et Neuroprotectrice)

Cette voie est prédominante dans des conditions homéostatiques normales. Elle est initiée par l'action de l' α -sécrétase **Figure 13** (Esch et al., 1990).

- **Mécanisme** : L'enzyme clive l'*APP* au sein même de la séquence du peptide amyloïde, ce qui empêche intrinsèquement la formation du peptide toxique A β **Figure 13**.
- **Objectifs Biologiques** : Ce clivage libère un fragment soluble appelé *sAPP α* . Ce dernier agit comme un facteur de croissance (*Growth factor*) qui favorise la survie neuronale et

stimule la formation des synapses (synaptogenèse). Cette voie est donc vitale pour le maintien de la plasticité cérébrale (Priller *et al.*, 2006).

III.4.2 La Voie Amyloïdogène (Pathologique)

À l'inverse, cette voie conduit à la pathogénèse de la maladie d'Alzheimer. Elle nécessite l'intervention coordonnée de deux enzymes : la β -sécrétase (BACE1) et la γ -sécrétase **Figure 13** (Vassar *et al.*, 1999).

- **Mécanisme** : Ces enzymes découpent l'APP de manière à libérer le peptide β -amyloïde ($A\beta$). **Figure 13**.
- **Conséquences** : Les peptides produits (notamment l'isoforme $A\beta_{42}$) possèdent une forte propension à s'agréger pour former des plaques amyloïdes. Ces dépôts extracellulaires perturbent la communication synaptique et déclenchent une cascade neurodégénérative conduisant à la mort cellulaire (Hardy & Selkoe, 2002).

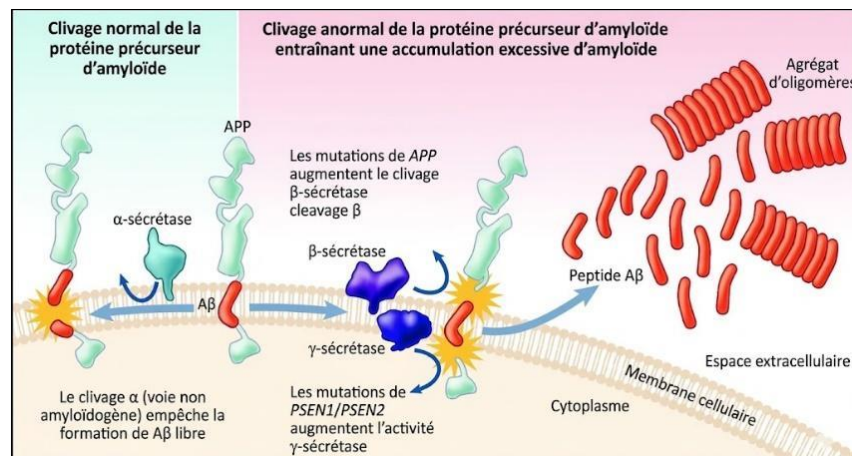


Figure 13 : Représentation schématique du métabolisme de la protéine APP : Voies amyloïdogène et non-amyloïdogène. (Chow *et al.*, 2010) "Adapté de Chow *et al.* (2010)".

III.4.3 Signalisation Intracellulaire et Adhésion

Outre la production de $sAPP\alpha$, le clivage par la γ -sécrétase libère un fragment cytosolique appelé *AICD* (*APP Intracellular Domain*).

- **Rôle de messenger** : Le fragment *AICD* est capable de se déplacer vers le noyau pour réguler l'expression de certains gènes, prouvant que *l'APP* intervient dans la communication nucléaire (Cao & Südhof, 2001).
- **Rôle structural** : Avant son clivage, l'APP complète agit comme une molécule d'adhésion cellulaire, facilitant les interactions entre les neurones et contribuant à l'organisation structurelle du tissu nerveux (Young-Pearse et al., 2008).

En somme, le clivage de la protéine APP est un mécanisme régulé permettant de produire des fragments fonctionnels impliqués dans la neuroprotection, la signalisation et l'organisation neuronale. Le gène *APP* n'est pas pathogène en soi ; c'est le déséquilibre métabolique favorisant la voie amyloïdogène qui détermine l'initiation de la maladie d'Alzheimer.

III.4. Lien entre le métabolisme de l'APP et la Maladie d'Alzheimer

Le rôle de la protéine *APP* dans l'étiologie de la maladie d'Alzheimer repose sur un déséquilibre du processus de clivage, favorisant la voie amyloïdogène au détriment de la voie physiologique. Ce basculement métabolique déclenche une cascade d'événements neurotoxiques :

III.4.1 La Cascade Amyloïde

Le mécanisme pathologique suit une séquence linéaire de dégradation neuronale (Hardy & Higgins, 1992) :

1. **Production excessive** : Le clivage séquentiel par les β et γ -sécrétases libère le peptide $A\beta$ en quantités anormales (Vassar et al., 1999).
2. **Agrégation et Fibrillation** : Les monomères d' $A\beta$ (particulièrement l'isoforme $A\beta_{42}$) s'assemblent en oligomères solubles, puis en fibrilles insolubles.

3. **Formation des Plaques** : Ces fibrilles s'accumulent dans le milieu extracellulaire pour former les plaques amyloïdes (ou plaques séniles), caractéristiques histopathologiques majeures de la maladie (**Hardy & Selkoe, 2002**).

III.4.2 Neurotoxicité et Mort Cellulaire

La présence de ces dépôts amyloïdes n'est pas inerte ; elle active des processus délétères secondaires :

- **Stress Oxydatif et Inflammation** : Les plaques stimulent une réponse immunitaire locale (microglieuse) et augmentent la production de radicaux libres (**Selkoe, 2001**).
- **Dysfonction Synaptique** : Avant même la mort cellulaire, les oligomères d'A β altèrent la communication entre les neurones, entraînant les premiers troubles de la mémoire (**Walsh et al., 2002**).
- **Neurodégénérescence** : L'accumulation de toxicité conduit inévitablement à la perte synaptique et à la mort neuronale progressive.

Il est fondamental de souligner que le gène *APP* n'est pas pathologique en soi ; ses fonctions physiologiques sont essentielles au développement et à la plasticité cérébrale (**Müller et al., 2017**). La pathologie résulte exclusivement d'une anomalie de son traitement protéolytique, orientant le métabolisme vers la production de fragments neurotoxiques.

III.5.1 Les Mutations du Gène *APP*

Certaines mutations spécifiques modifient le site de reconnaissance des sécrétases, favorisant ainsi la voie amyloïdogène au détriment de la voie physiologique :

- **La Mutation Swedish (KM670/671NL)** : Située à l'extrémité N-terminale du domaine A β , elle augmente l'affinité de l'*APP* pour la β -sécrétase, entraînant une production massive de peptides amyloïdes (**Citron et al., 1992**).
- **La Mutation London (V717I)** : Localisée près du site de coupure de la γ -sécrétase, elle modifie le ratio des peptides produits en faveur de l'isoforme A β 42, reconnue pour sa forte toxicité et sa propension à l'agrégation (**Goate et al., 1991**).

- **La Mutation Arctic (E693G) :** Cette mutation intra-amyloïde favorise la formation rapide de protofibrilles, accélérant ainsi la neurodégénérescence (Nilsberth *et al.*, 2001).

III.5.2 Mode de Transmission et Caractéristiques Cliniques

Les mutations du gène *APP* suivent un mode de transmission autosomique dominant, ce qui signifie qu'une seule copie mutée héritée suffit à déclencher la pathologie (Bertram & Tanzi, 2008).

- Elles sont systématiquement associées à des formes précoces de la maladie (Early-onset Alzheimer's Disease - EOAD), apparaissant généralement avant l'âge de 65 ans.
- Bien que rares (représentant moins de 5 % des cas totaux), ces formes illustrent parfaitement le lien direct entre la génétique et le dysfonctionnement protéique.

III.5.3 Facteurs de Risque Génétiques : Le rôle de l'*APOE*

Outre les mutations directes, d'autres polymorphismes génétiques influencent la susceptibilité à la maladie. Le plus significatif est l'allèle $\epsilon 4$ du gène de l'Apolipoprotéine E (*APOE*). Contrairement aux mutations de l'*APP*, l'*APOE4* n'est pas une cause directe mais un facteur de risque majeur qui diminue l'efficacité de la clairance (élimination) du peptide $A\beta$ (Corder *et al.*, 1993).

Les mutations du gène *APP* démontrent que la pathogenèse de la maladie d'Alzheimer est intrinsèquement liée à un défaut de régulation génétique influençant le clivage protéique. La transformation d'un processus physiologique vital en une cascade neurotoxique souligne l'importance de la conservation évolutive de ces séquences génétiques.

IV. La Maladie d'Alzheimer : Contexte Clinique et Moléculaire

IV.1. Définition et Épidémiologie

La maladie d'Alzheimer (MA) est une affection neurodégénérative chronique, progressive et irréversible. Elle se caractérise par une détérioration des fonctions cognitives, touchant principalement la mémoire à court terme, ainsi que par des troubles du comportement et du

langage. À ce jour, elle constitue la cause la plus fréquente de démence chez les personnes âgées à l'échelle mondiale (Prince et al., 2015).

IV.2. Physiopathologie :

Les deux piliers lésionnels sur le plan histopathologique, la pathologie se définit par la coexistence de deux lésions cérébrales majeures qui agissent en synergie pour détruire le réseau neuronal :

-Les plaques amyloïdes (extracellulaires) : Elles correspondent à des agrégats du peptide β -amyloïde ($A\beta$), issus d'un clivage anormal de la protéine APP. Ces dépôts s'accumulent dans l'espace extracellulaire, perturbent la transmission synaptique et déclenchent une réponse inflammatoire locale (Hardy & Selkoe, 2002).

-La dégénérescence neurofibrillaire (intracellulaire) : Elle est due à l'accumulation de la protéine Tau hyperphosphorylée. Normalement impliquée dans la stabilisation des microtubules, cette protéine devient pathologique, s'agrège en filaments et entraîne la désorganisation du cytosquelette neuronal, conduisant inévitablement à la mort cellulaire (Selkoe, 2001).

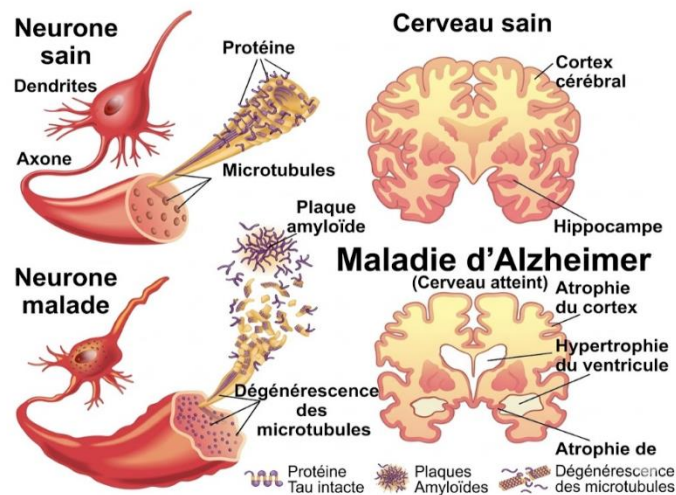


Figure 14 : Comparaison anatomique et cellulaire entre un cerveau sain et un cerveau atteint par la maladie d'Alzheimer (Docteur Clic).

-La formation des plaques amyloïdes est directement liée au clivage pathologique de la protéine APP, soulignant son rôle central dans la pathogenèse de la maladie.

Commentaire de la **Figure 14**, Le lien entre *l'APP* et la neurodégénérescence Cette figure illustre la cascade pathologique liant le métabolisme de *l'APP* à la mort neuronale. Le clivage anormal de *l'APP* par les β et γ -sécrétases entraîne une accumulation extracellulaire de peptides $A\beta$, formant des plaques amyloïdes. Cette toxicité de contact déclenche, à l'intérieur du neurone, une hyperphosphorylation de la protéine Tau. Une fois altérée, la protéine Tau se détache des microtubules, provoquant l'effondrement du cytosquelette axonal (Dégénérescence Neurofibrillaire). Ce double processus extracellulaire (amyloïde) et intracellulaire (Tau), interrompt la communication synaptique et conduit inévitablement à l'apoptose (mort cellulaire), se traduisant macroscopiquement par une atrophie cérébrale caractéristique de la maladie d'Alzheimer.

IV.3. Les formes de la maladie

On distingue classiquement deux formes cliniques de la MA, différenciées par leur âge d'apparition et leur déterminisme génétique :

- **La forme sporadique** : Représente plus de 95 % des cas et apparaît généralement après l'âge de 65 ans. Son origine est multifactorielle, impliquant une interaction complexe entre des facteurs génétiques de susceptibilité (l'allèle *APOE4*), l'environnement et le mode de vie (**Bertram & Tanzi, 2008**).
- **La forme familiale précoce** : Rare (moins de 5 % des cas), elle est transmise selon un mode autosomique dominant. Elle est associée à des mutations dans trois gènes clés : *APP*, *PSEN1* (Préséniline 1) et *PSEN2* (Préséniline 2). Ces mutations favorisent une production accrue ou une agrégation facilitée du peptide $A\beta$ (**Goate et al., 1991**).

Ainsi, la maladie d'Alzheimer résulte d'une interaction complexe entre des mécanismes moléculaires, notamment le clivage anormal de *l'APP*, et des facteurs génétiques et environnementaux. Cette compréhension moléculaire est le point de départ essentiel pour toute analyse génomique comparative visant à identifier les régions conservées de ce gène à travers les espèces.

PARTIE II
PRATIQUE:

Chapitre II :

Matériel et Méthodes

II.1. Objectif de l'étude et Choix de l'Organisme Modèle:

A. Objectif de l'étude :

Cette étude repose sur une analyse bioinformatique comparative de gènes et de protéines clés associés à la maladie d'Alzheimer (tels que *APP* et *ApoE*), dans le but de :

- ✓ Comprendre les causes moléculaires : identifier les mutations et altérations génétiques sous-jacentes à la maladie d'Alzheimer.
- ✓ Caractérisation protéomique : étudier les structures physiques, chimiques et tridimensionnelles des protéines afin de comprendre les mécanismes de leur repliement anormal et de leur neurotoxicité.
- ✓ Prédire les effets fonctionnels : utiliser des outils de prédiction informatique pour évaluer l'impact des variations génétiques.

B. Choix de l'organisme modèle:

L'intégration d'organismes modèles (*Mus musculus*, *Rattus norvegicus*, *Canis lupus familiaris*, *Pan troglodytes*, *Macaca mulatta*) dans cette approche bioinformatique vise à :

- ✓ Établir un modèle de référence comparatif : aligner les séquences afin d'identifier les régions évolutives hautement conservées entre l'humain et le modèle.
- ✓ Orienter la recherche expérimentale : générer des données prédictives fiables qui constituent une base solide, permettant d'optimiser l'utilisation du temps et des ressources lors des futures validations en laboratoire (in vitro et in vivo).

II.2. Choix des gènes étudiés:

Les résultats de l'enquête et de l'analyse bioinformatique ont révélé un groupe de gènes et de protéines clés qui jouent un rôle central et direct dans la physiopathologie de la maladie d'Alzheimer. Ces gènes varient selon la gravité de leur atteinte, les types de mutations qui leur sont associés et leur contribution à l'apparition et à la progression de la maladie.

La maladie d'Alzheimer (MA) est une affection neurodégénérative chronique et progressive du système nerveux central, constituant la cause la plus fréquente de démence chez les sujets âgés (OMS, 2023). Sur le plan clinique, elle se manifeste par un déclin cognitif graduel affectant

principalement la mémoire à court terme, le langage, le raisonnement et l'orientation spatio-temporelle (NIA, 2024).

Sur le plan histopathologique, la maladie est caractérisée par deux signatures majeures : l'accumulation extracellulaire de peptides bêta-amyloïdes (A β) formant les plaques séniles, et la phosphorylation (hyperphosphorylation) intra-neuronale de la protéine Tau, conduisant à la formation de dégénérescences neurofibrillaires (DNF). Ces processus biochimiques entraînent une perte synaptique massive et une mort neuronale menant à l'atrophie cérébrale (Scheltens et al., 2021).

Sur le plan étiologique et génétique, la maladie se divise en deux formes distinctes :

1. **La forme familiale précoce** : Rare (moins de 1% à 5% des cas), à transmission autosomique dominante, directement liée à des mutations mendéliennes à haute pénétrance dans les gènes *APP*, *PSEN1*, et *PSEN2*.
2. **La forme sporadique tardive** : La plus fréquente (plus de 95% des cas), survenant après 65 ans, d'origine multifactorielle où le polymorphisme génétique (notamment l'allèle $\epsilon 4$ du gène *APOE*) combiné à des facteurs environnementaux détermine le score de susceptibilité (Alzheimer's Association, 2023 ; Scheltens et al., 2021).

Le tableau 1 présente les gènes principaux impliqués, organisés suivant l'importance décroissante de leur contribution à la prédisposition génétique à la maladie ou à leur prévalence :

Tableau n°01 : Gènes clés impliqués dans la maladie d'Alzheimer et l'évaluation de leur impact génétique (Alzheimer's Association, 2026 ; Scheltens et al., 2021).

Gène	Protéine encodée	Rôle pathologique et fonctionnel	Impact Génétique et Risque (Données épidémiologiques)
APOE	Apolipoprotéine E	Facteur de risque génétique majeur (notamment l'allèle $\epsilon 4$), associé à l'accumulation des peptides amyloïdes et à l'inhibition de leur clairance.	Augmentation du risque de 3 à 4 fois (si hétérozygote $\epsilon 3 \backslash \epsilon 4$) et jusqu'à 12 à 15 fois (si homozygote $\epsilon 4 \backslash \epsilon 4$).
APP	Protéine précurseur de l'amyloïde	Gène de la protéine précurseur de l'amyloïde ; ses mutations entraînent une surproduction de peptides A β 42 formant les plaques amyloïdes.	Pénétrance quasi-totale (~100%). Cause directe mais rare, impliquée dans moins de 1% des cas totaux (Forme familiale).
PSEN1	Préséniline 1	Composant catalytique essentiel du complexe γ -secrétase ; ses mutations induisent un clivage anormal de l'APP (fortement lié à la forme familiale précoce).	Pénétrance quasi-totale (~100). Responsable de 30% à 70% des formes familiales précoces.

PSEN2	Préséniline 2	Composant du complexe γ -secrétase ; ses mutations sont moins fréquentes mais contribuent également à la forme familiale précoce de la maladie.	Pénétrance élevée (~95%). Cause très rare, représentant moins de 5% des formes familiales.
MAPT	Protéine Tau	Code pour la protéine Tau associée aux microtubules ; sa hyperphosphorylation conduit à la formation de dégénérescences neurofibrillaires (DNF).	Modificateur de phénotype. Indispensable à la neurodégénérescence (Le risque dépend de la synergie avec A β).
TREM2	TREM2	Récepteur exprimé sur les cellules microgliales ; régule la réponse inflammatoire et la clairance des débris cellulaires.	Augmentation du risque de 2 à 4 fois en cas de mutations rares (comme R47H).
SORL1	SorLA	Orienté le trafic intracellulaire de l'APP ; sa sous-expression redirige l'APP vers la voie amyloïdogène.	Augmentation du risque de 1.5 à 3 fois pour les variants associés aux formes tardives.
Autres gènes	ABCA7, CLU, PICALM	Gènes de susceptibilité associés au métabolisme lipidique, à la réponse immunitaire et à l'endocytose.	5% - 10% (De l'ensemble des facteurs de risque secondaires)

Dans cette étude, nous avons choisi de nous focaliser sur deux gènes clés, **APP (Amyloid Precursor Protein)** et **APOE (Apolipoprotein E)**, en raison de leur implication majeure dans la maladie d'Alzheimer. Ces gènes représentent une pierre angulaire de la recherche génétique sur cette pathologie, tant par leur rôle dans les mécanismes biologiques sous-jacents que par leur contribution à la susceptibilité et à l'évolution de la maladie.

1. Le gène de la protéine souche amyloïde (gène **APP**):

Ce gène code pour une protéine membranaire intégrale appelée protéine souche amyloïde bêta. Le clivage anormal de cette protéine par des enzymes (β - et γ -secrétases) entraîne la libération de peptides β -amyloïdes (β), qui s'accumulent dans les espaces intercellulaires pour former des plaques amyloïdes. Il s'agit de la caractéristique histologique des formes génétiques précoces de la maladie (**Hardy & Higgins, 1992**).

2. Gène de l'apolipoprotéine E (APOE) :

Ce gène, situé sur le chromosome 19, code pour l'apolipoprotéine E, une protéine essentielle au transport des lipides et du cholestérol ainsi qu'à la réparation des membranes cellulaires du système nerveux central (Mahley, 1988).

Le gène *APOE* présente un polymorphisme et possède trois allèles principaux fréquemment rencontrés dans la nature : $\epsilon 2$, $\epsilon 3$ et $\epsilon 4$. Chaque individu hérite d'un allèle de sa mère et d'un autre de son père, ce qui donne un génotype diploïde déterminant son risque de développer la maladie (Corder et al., 1993).

Le gène $\epsilon 4$ est le principal facteur de risque génétique des formes sporadiques et tardives de la maladie d'Alzheimer. Si une personne hérite d'un allèle $\epsilon 3$ ou $\epsilon 2$, son risque de développer la maladie est multiplié par trois environ. Si elle hérite de deux allèles $\epsilon 4$ (un du père et un de la mère), ce risque est multiplié par douze par rapport aux autres transmissions (Farrer et al., 1997).

II.3. Matériel:

Dans cette étude, nous avons utilisé plusieurs bases de données, outils bioinformatiques et logiciels pour effectuer une analyse protéomique comparative et une analyse génomique structurale des gènes *APP* et *APOE* :

II.3.1. Bases de données biologiques :

II. 1.1. Définition du Centre national d'information sur la biotechnologie (NCBI) :

Le Centre national d'information sur la biotechnologie (NCBI) a été créé en novembre 1988 au sein de la Bibliothèque nationale de médecine (NLM) aux États-Unis. Le centre a été choisi pour son expertise dans la création et la maintenance de bases de données biomédicales. En tant que membre des Instituts nationaux de la santé (NIH), il a pu mettre en place un programme de recherche interne en biologie moléculaire computationnelle. La mission du centre est de développer de nouvelles technologies de l'information pour faciliter la compréhension des processus moléculaires et génétiques fondamentaux qui régissent la santé et la maladie. Le centre a été créé pour accomplir ces quatre principales missions (Jack, 1999).

La classification du Centre national d'information sur la biotechnologie (NCBI) comprend les noms et les classifications des organismes pour chaque séquence des bases de données de séquences nucléotidiques et protéiques de la Collaboration internationale sur les bases de données de séquences nucléotidiques. Depuis sa dernière révision en 2012, cette ressource a fait l'objet de nombreuses améliorations. Plus particulièrement, le système est passé d'une base de données SQL unique à une série de bases de données interconnectées, liées à un cadre de données appelé NameBank. Cela signifie que les relations entre les éléments de données peuvent être modifiées avec plus de précision, permettant ainsi d'enrichir les explications des synonymes, de distinguer les noms présentant des caractéristiques spécifiques, d'améliorer le suivi des publications associées aux noms et d'améliorer les explications des autorités et des types scientifiques. De plus, les pratiques utilisées par les taxonomistes du NCBI pour les principaux groupes taxonomiques sont décrites,

la terminologie de la taxonomie du NCBI est expliquée, les ressources externes sont mentionnées et les mises à jour d'autres outils et ressources sont documentées (**Schoch et al., 2020**).

Le NCBI héberge un large éventail de bases de données, de ressources et d'outils analytiques pour la recherche scientifique. <https://www.ncbi.nlm.nih.gov/>

Parmi ses bases de données les plus importantes figurent :

- PubMed : une base de données massive contenant les résumés et les références de millions d'articles scientifiques publiés dans des revues biomédicales et de sciences de la vie.
- PubMed Central (PMC) : Une archive numérique gratuite offrant un accès complet au texte intégral d'articles scientifiques évalués par des pairs dans les domaines biomédicaux et des sciences de la vie.
- GenBank : Une base de données de séquences d'ADN et d'ARN.
- Autres bases de données : Celles-ci comprennent des bases de données sur les protéines, les génomes, les variations génétiques telles que dbSNP, la structure moléculaire, et bien d'autres.
- Développement d'outils de recherche et d'analyse : Le centre met à disposition des logiciels performants pour la recherche dans les bases de données et l'analyse des séquences

biologiques et autres données. Parmi les plus connus, BLAST (Basic Local Alignment Search Tool) est utilisé pour comparer les séquences d'ADN ou de protéines.

- Soutien à la recherche et au développement scientifiques : Le centre vise à développer de nouvelles technologies de l'information pour faciliter la compréhension des processus moléculaires et génétiques fondamentaux qui régissent la santé et la maladie.
- Faciliter l'échange d'informations : Il vise à faciliter l'utilisation des bases de données et des logiciels par la communauté mondiale de la recherche scientifique et médicale (David et al., 2005).



Figure 15: Plateformes bioinformatiques des principales banques de données et outils de recherche exploités (NCBI, PubMed et BLAST). <https://pubmed.ncbi.nlm.nih.gov/>

II. 1.2. Base de données UniProt (Ressource universelle des protéines) :

Il s'agit d'une base de données exhaustive et ouverte contenant des informations fonctionnelles et des séquences protéiques, créée en 2002 (suite à la fusion des bases de données Swiss-Prot, TrEMBL et PIR). Elle constitue la principale référence mondiale, supervisée par un groupe d'instituts de biologie internationaux : l'Institut européen de bioinformatique (EMBL-EBI), l'Institut suisse de bioinformatique (SIB) et la Protein Information Foundation (PIR). Nous l'avons utilisée pour extraire les séquences primaires (acides aminés) des protéines APP et APOE avec leurs trois allèles ($\epsilon 2$, $\epsilon 3$, $\epsilon 4$), ainsi que pour obtenir des caractérisations fonctionnelles précises et l'identification des domaines structuraux (UniProt Consortium, 2025).



Figure 16 : La plateforme UniProt utilisée pour l'extraction des séquences protéiques
<https://www.uniprot.org/>

II. 1.3. Notepad++ (Éditeur de texte) :

Notepad++ est un éditeur de texte et un environnement d'édition libre et gratuit, compatible avec Windows (développé par Don Ho). Dans ce projet de bioinformatique, ce programme a été un outil essentiel pour la gestion et le nettoyage des fichiers FASTA. Il nous a permis de vérifier les transcrits de nucléotides et de protéines avant leur intégration dans les logiciels d'analyse. Cet éditeur est unique car il permet de modifier les en-têtes de fichiers, de supprimer les espaces et les caractères indésirables susceptibles d'affecter la précision des outils d'alignement et de modélisation, et d'organiser et structurer rapidement et efficacement les données textuelles extraites des bases de données. <https://notepad-plus-plus.org/>

II. 1.4. BLAST (Basic Local Alignment Search Tool):

BLAST (Basic Local Alignment Search Tool) est un programme de recherche de similarités de séquences. Le NCBI (National Center for Biotechnology Information) gère le serveur BLAST, accessible via une page web (**David et Medha, 2008**).

- BLAST est un puissant algorithme informatique utilisé pour comparer des séquences biologiques (telles que les séquences d'ADN et de protéines) et déterminer leur similarité.
- BLAST est un outil fondamental en bioinformatique pour trouver des séquences similaires à une séquence donnée, que ce soit au sein du même organisme ou chez d'autres organismes.

BLAST est extrêmement rapide lors du processus de comparaison : il peut analyser une séquence et la comparer à la base de données GenBank en quelques secondes.

La comparaison de séquences nucléotidiques ou protéiques provenant d'organismes similaires ou différents est un outil très efficace en biologie moléculaire. En trouvant des similarités entre les séquences, les scientifiques peuvent déduire la fonction de gènes nouvellement séquencés, prédire de nouveaux membres de familles de gènes et explorer les relations évolutives. Grâce au séquençage du génome entier, la recherche de similarités de séquences peut désormais être utilisée pour prédire l'emplacement et la fonction des régions codantes et de régulation de la transcription dans l'ADN génomique (Tom, 2013). <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

II. 1.5. Outil MAFFT (Multiple Alignment using Fast Fourier Transform):

Ce logiciel et serveur très performant permet de réaliser des alignements de séquences multiples (AMS). Développé par l'Institut de recherche sur les maladies microbiennes de l'Université d'Osaka au Japon, il a été utilisé dans cette étude pour aligner rapidement et avec précision les séquences d'acides aminés et de nucléotides des gènes APP et APOE chez les différentes espèces étudiées. Son intérêt réside dans l'utilisation d'algorithmes de transformée de Fourier rapide, qui nous ont permis de localiser et de suivre avec précision les résidus d'acides aminés conservés entre les espèces et de les comparer aux formes mutées de la maladie d'Alzheimer (Katoh & Standley, 2013). <https://mafft.cbrc.jp/>

II. 1.6. MEGA12 (Analyse génétique de l'évolution moléculaire, version 12) :

MEGA12 est un logiciel complet et largement utilisé dans le domaine de l'analyse de l'évolution moléculaire et de la génétique. Il a été développé pour permettre aux chercheurs et aux étudiants de réaliser des analyses évolutives complexes sur les séquences d'ADN, d'ARN et de protéines.

MEGA12 offre un ensemble complet d'outils et de fonctions, notamment : <https://www.megasoftware.net/>

- **Alignement de séquences** : Inclut des outils pour l'alignement de séquences multiples, tels que ClustalW et MUSCLE.
- **Construction d'arbres phylogénétiques** : Prend en charge diverses méthodes de construction d'arbres phylogénétiques.

- **Estimation des distances évolutives** : Propose plusieurs modèles pour le calcul des distances génétiques entre les séquences.

- **Tests d'hypothèses évolutives** : Inclut des tests statistiques pour évaluer des hypothèses spécifiques sur l'évolution moléculaire.

- **Analyse de l'horloge moléculaire** : Analyse de la sélection (identification des sites soumis à la sélection).

- Interface graphique conviviale (GUI) : facilite l'accès aux différents outils et l'interprétation des résultats (**Tamura K et al., 2021**).

Ces dernières années, le séquençage du génome est devenu un outil très efficace pour l'étude d'un large éventail de systèmes biologiques. De nombreuses étapes de l'interprétation des données de séquences brutes nécessitent une analyse comparative des séquences moléculaires à chaque point temporel afin de révéler les différences fonctionnelles au sein du génome. MEGA (Molecular Evolutionary Genetics Analysis), un programme convivial développé pour l'analyse des données de séquences d'ADN et de protéines, fournit une suite d'outils pour réaliser ces analyses. MEGA permet un large éventail d'investigations, telles que l'alignement de séquences, la construction d'arbres phylogénétiques, l'estimation des distances et des différences génétiques, et la détection des séquences ancestrales. Plusieurs techniques bioinformatiques permettent d'estimer les connaissances passées sur les organismes à partir de leurs connaissances actuelles. Un arbre phylogénétique est similaire à un arbre évolutif en ce qui concerne les relations entre les entités biologiques, et les diagrammes arborescents illustrent les relations entre différentes espèces, organismes, gènes, etc., au sein de la structure d'un arbre phylogénétique. Des analyses peuvent être effectuées sur les séquences d'acides nucléiques, de protéines, d'acide désoxyribonucléique (ADN) et d'acide désoxyribonucléique (ARN) (**Gülşah, 2023**).

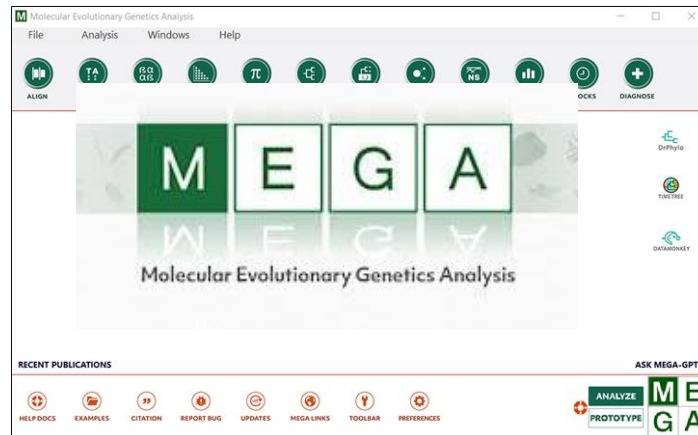


Figure 17 : logiciel **MEGA 12** (Molecular Evolutionary Genetics Analysis) illustrant les différents modules d'analyse (Alignement, Phylogénie, et Distances génétiques). <https://www.megasoftware.net/>

II. 1.7. Serveur ProtParam (Expasy) :

Cet outil d'analyse informatique est disponible sur le portail de l'Institut Suisse de Bioinformatique (SIB). Il a été utilisé pour prédire et calculer les paramètres physico-chimiques des protéines APP et APOE, ainsi que de leurs isoformes mutantes. Il permet notamment de déterminer la masse moléculaire, le point isoélectrique (pI), la composition en acides aminés, le coefficient d'absorption cellulaire, la demi-vie et l'indice d'instabilité, qui indique la stabilité de la protéine in vitro (Gasteiger *et al.*, 2005).



Figure 18 : Serveur ProtParam (Expasy) <https://www.uniprot.org/>

II. 1.8. Base de données InterPro :

Il s'agit d'une plateforme analytique intégrée supervisée par l'Institut européen de bioinformatique (EMBL-EBI). Elle est considérée comme une référence mondiale pour la classification des protéines en familles et l'identification des domaines structuraux et des sites actifs. Dans cette

étude, elle a été utilisée pour réaliser une analyse exhaustive des séquences d'acides aminés d'APP et d'APOE, confirmer la localisation précise des domaines fonctionnels et les comparer entre l'homme et des organismes modèles afin de confirmer leur conservation évolutive (**Paysan-Lafosse et al., 2023**).



Figure 19 : Base de données InterPro. <https://www.ebi.ac.uk/>

II. 1.9. Serveur SWISS-MODEL:

Ce système entièrement automatisé permet de modéliser la structure 3D des protéines par modélisation par homologie. Géré conjointement par l'Institut suisse de bioinformatique (SIB) et l'Université de Bâle, il a été utilisé pour construire des modèles structuraux 3D des protéines APP et APOE (pour les allèles $\epsilon 2$, $\epsilon 3$ et $\epsilon 4$) ainsi que de leurs homologues dont les structures cristallines expérimentales ne sont pas disponibles dans la Brain Protein Bank. Ces modèles ont été construits à partir des meilleurs modèles conformationnels disponibles dans la World Protein Data Bank (PDB) (**Waterhouse et al., 2018**).

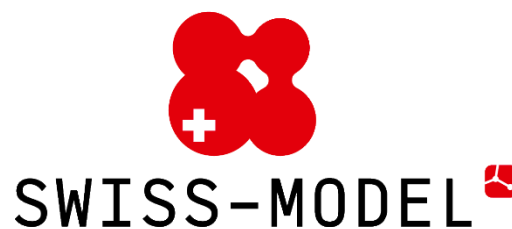


Figure 20 : Serveur SWISS-MODEL. <https://swissmodel.expasy.org/>

II. 1.10. PyMOL (Molecular Graphics System):

Il s'agit d'un logiciel libre (actuellement géré par la Fondation Schrödinger) spécialisé dans la visualisation et l'appariement moléculaires tridimensionnels. Il a été utilisé lors de la phase finale de ce travail pour lire les fichiers de structure spatiale étendue (.pdb) extraits de SWISS-MODEL, ce qui nous a permis d'effectuer une superposition structurale et de comparer des protéines normales et mutantes (telles que les allèles APOE ϵ_2 , ϵ_3 et ϵ_4), de colorer et de localiser précisément les mutations, et d'étudier les modifications des liaisons hydrogène et des ponts salins induites par ces changements génétiques (DeLano, 2020).

II.4. Méthodes:

Dans cette étude, nous avons extrait les séquences nucléotidiques des gènes APP et Apoe, sélectionnés chez l'humain (*Homo sapiens*) et associés à la maladie d'Alzheimer, à partir de la base de données NCBI au format FASTA. Nous les avons ensuite comparées aux séquences des mêmes gènes chez des organismes étroitement apparentés (*Mus musculus*, *Rattus norvegicus*, *Canis lupus familiaris*, *Pan troglodytes*, *Macaca mulatta*). Nous avons également recherché des mutations dans ces gènes: (mutation Swedish K670N/M671L) dans gène APP et allèles (ϵ_2 , ϵ_3 , ϵ_4) dans gène Apoe.

Enfin, nous avons aligné et comparé ces séquences à l'aide des logiciels mentionnés précédemment.

II.4.1. BLAST :

Le traitement des données BLAST a fourni des informations sur le pourcentage d'identité et la signification statistique (valeur E) des séquences nucléotidiques analysées. Ce processus a également généré des outils de visualisation des résultats, notamment un diagramme de points illustrant les similarités entre les séquences et un graphique récapitulatif offrant une vue d'ensemble de l'analyse génomique comparative.

- ❖ Étude comparative des gènes au sein de l'échantillon étudié :

Lors de la comparaison des séquences géniques à l'aide de BLASTN, nous obtenons plusieurs indicateurs importants :

- ✓ La couverture de la requête exprime le degré de similarité entre les deux séquences nucléotidiques.
- ✓ Le pourcentage d'identité reflète le degré de similarité de la composition en acides nucléiques entre les régions correspondantes.
- ✓ La valeur E indique la probabilité d'occurrence de correspondances aléatoires similaires. Plus la valeur E est faible (proche de zéro), plus la probabilité que la correspondance soit réelle et non une simple coïncidence est élevée.

II.4.2. MEGA 12:

Les séquences nucléotidiques AMS obtenues avec le programme MAFFT ont ensuite été traitées avec MEGA 12.

MEGA 12 a été utilisé pour aligner les séquences et construire un arbre phylogénétique selon l'algorithme de liaison par la méthode de construction/test de l'arbre de vraisemblance maximale (ML).

II.4.3. Serveur ProtParam (Expasy):

Nous avons utilisé cet outil pour calculer les propriétés physico-chimiques des protéines Wolpeti, à savoir la masse moléculaire (Mw), la neutralité électrique (pI) et l'indice d'instabilité, afin de déterminer la stabilité computationnelle des protéines APP et APOE (normale et mutante).

II.4.4. Plateforme InterPro:

Nous avons utilisé cette plateforme pour réaliser un balayage complet des séquences d'acides aminés afin d'identifier les domaines structuraux et les sites actifs des protéines étudiées. L'objectif était de confirmer la conservation et l'homologie de ces domaines fonctionnels entre l'homme et les organismes modèles.

II.4.5. Serveur Swiss Model:

Nous avons utilisé ce serveur pour construire et prédire la structure 3D des protéines APP et APOE et de leurs allèles ($\epsilon 2$, $\epsilon 3$, $\epsilon 4$). Nous avons eu recours à la modélisation par homologie en les comparant à des modèles structuraux cristallisés expérimentalement et disponibles dans la Protein Database (PDB).

II.4.6. Programme PyMOL:

Nous l'avons utilisé pour la visualisation et la modélisation moléculaire 3D des fichiers .pdb extraits. Le but était de réaliser une comparaison spatiale (superposition) entre les structures normales et mutantes (par exemple, en comparant les allèles $\epsilon 2$, $\epsilon 3$ et $\epsilon 4$) afin d'identifier la localisation des mutations et d'étudier leur effet sur les liaisons chimiques.

Chapitre III :

Résultats et discussion

I. Résultats:

I.1- Analyse comparative du gène *APP* (Protéine précurseur de l'amyloïde):

Nous nous sommes concentrés sur le gène *APP*, l'un des gènes les plus importants responsables de la formation des plaques bêta-amyloïdes associées à la maladie.

Nous avons extrait la séquence d'ARNm normalisée du gène *APP* de la base de données NCBI, puis sélectionné un modèle de mutation connu sous le nom de mutation Swedish. (KM670/671NL) Cette mutation importante augmente la production de bêta-amyloïde et est fréquemment utilisée dans les études de référence.

Après avoir comparé la séquence normale et la séquence mutante (mutation Swedish) du gène *APP* à l'aide de l'outil BLAST, des résultats précis ont été obtenus montrant le degré de similarité entre les deux séquences.

Les résultats ont montré ce qui suit :

Séquences produisant des alignements significatifs		Télécharger	Sélectionner les colonnes	Montrer	100	?		
<input checked="" type="checkbox"/> tout sélectionner	1 séquence sélectionnée			Graphique	Visionneuse MSA			
Description	Nom scientifique	Score maximal	Score total	Couverture de requête	Valeur E	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> AMPP_ARNm_Mutation_Suédoise [organisme=Homo sapiens].[Mutation=Suédoise K670N/M...		6606	6606	100%	0,0	99,94%	3583	Requête_7827879

Figure 21 : Résultats de l'alignement de la séquence d'ARNm progéniteur du gène *APP* avec la séquence mutante (mutation Swedish K670N/M671L) à l'aide de l'outil BLASTn.

- **Couverture : 100 %**, ce qui signifie que la séquence mutée est parfaitement alignée sur la séquence de référence.
- **Pourcentage d'identité : 99,94 %**, un pourcentage très élevé indiquant une forte similarité entre les deux séquences, avec des différences minimales.
- **Valeur E : 0,0**, ce qui indique que l'alignement n'est pas aléatoire et est statistiquement très significatif.
- **Score total et score maximal : 6 606**, ce qui indique la robustesse et la qualité de l'alignement entre les deux séquences.

Ces résultats montrent que la séquence mutée (mutation suédoise) est très proche de la séquence naturelle en termes de structure génétique, avec seulement quelques différences ponctuelles.

Cette image montre le résumé graphique de l'alignement, où la barre rouge représente un score de correspondance très élevé (Score ≥ 200) le long de la séquence de requête s'étendant sur environ 3583 bases nucléotidiques:



Figure 22: Représentation graphique de la distribution des meilleurs alignements séquentiels du gène *APP* via BLAST.

Cette image, extraite de l'outil MSA Viewer du NCBI, illustre l'alignement de la séquence consensus avec deux séquences cibles (requêtes). Une fine ligne verticale rouge est visible à la position 2200 environ, indiquant précisément l'emplacement de la mutation ponctuelle.

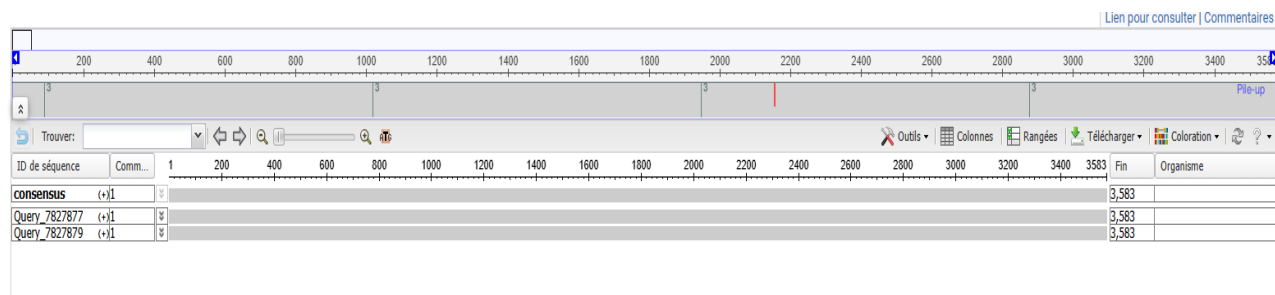


Figure 23 : Visualisation de l'alignement multiple des séquences (MSA Viewer) du gène *APP* montrant le site de la mutation.

I.1.1- Comparaison des séquences nucléotidiques des gènes *APP* :

Mus musculus ; Rattus norvegicus ; Canis lupus familiaris ; Pan troglodytes et Macaca mulatta font référence à *Homo sapiens*:

Tableau 2 : Gènes *APP* de comparaison pour chacun des *Mus musculus* ; *Rattus norvegicus*; *Canis lupus familiaris* ; *Pan troglodytes* et *Macaca mulatta* faisant référence à *Homo sapiens*.

Gène	Organisme	Chr	Localisation	Nombre d'exons	Longueur
<i>APP</i>	<i>Homo sapiens</i>	21	NC_000021.9 (25880550..26171128,complement)	20	290,585 nt

Gène	Organisme	Chr	Localisation	Nombre d'exons	Longueur	Query cover	Per.Ident	E.value
<i>APP</i>	<i>Mus musculus</i>	16	NC_000082.7 (84751236..84972187, complement)	20	220,952 nt	93%	88.12%	0.0
	<i>Rattus norvegicus</i>	11	NC_086029.1 (37506207..37724351, complement)	18	283,589 nt	65%	89.27%	0.0
	<i>Canis lupus familiaris</i>	31	NC_051835.1 (21405966..21669000, complement)	18	341,945 nt	96%	90.13%	0.0
	<i>Pan troglodytes</i>	22	NC_072420.2 (26667179..26951990, complement)	18	370,256 nt	65%	99.70%	0.0
	<i>Macaca mulatta</i>	3	NC_133408.1 (21326364..21612918)	19	372,521 nt	100%	97.60%	0.0

- La couverture de la requête dans le résultat BLAST est de 100 %, ce qui signifie que la séquence requête est entièrement alignée avec la séquence cible dans la base de données. Aucune partie de la séquence requête n'est laissée non alignée ou inutilisée dans la correspondance (*SequenceServer.*, 2023) .
- Valeur E (Valeur attendue) Nombre d'alignements atteignant un certain score ou plus et susceptibles d'être observés aléatoirement dans une base de données de cette taille. Plus les

valeurs sont faibles, plus les alignements sont significatifs. Les valeurs proches de 0 sont les plus significatives.

- Par identité Pourcentage de correspondances identiques entre la séquence requête et la séquence cible le long de la longueur d'alignement, indiquant la similarité des deux régions à la base (ADN/ADN).

Le tableau 2 montre que tous les gènes diffèrent par le nombre d'exons et présentent des longueurs et des taux de couverture de séquence variables

Par ailleurs, toutes les espèces étudiées présentent une forte similarité avec l'humain, avec les pourcentages de concordance suivants :

- *Macaca mulatta* : 97,60 %
- *Pan troglodytes* : 99,70 %
- *Canis lupus familiaris* : 90,13 %
- *Mus musculus* : 88,12 %
- *Rattus norvegicus* : 89,27 %

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	XM_015133068.3 APP [organism=Macaca mulatta].[GeneID=100427716].[transcript=X1]		6141	6141	100%	0.0	97.60%	3584	Query_5830920
<input checked="" type="checkbox"/>	NM_001006601.2 Canis lupus familiaris amyloid beta precursor protein (APP), transcript variant 1_mRNA		4460	4460	96%	0.0	90.13%	3442	Query_5830919
<input checked="" type="checkbox"/>	NM_001013018.1 Pan troglodytes amyloid beta precursor protein (APP)_mRNA		4233	4233	65%	0.0	99.70%	2313	Query_5830921
<input checked="" type="checkbox"/>	NM_001198823.1 Mus musculus amyloid beta precursor protein (App), transcript variant 1_mRNA		3956	3956	93%	0.0	88.12%	3377	Query_5830917
<input checked="" type="checkbox"/>	NM_019288.2 App [organism=Rattus norvegicus].[GeneID=54226]		2929	2929	65%	0.0	89.27%	2340	Query_5830918

Figure 24 : Résultats de l'alignement de séquences nucléotidiques par BLASTN pour le gène *APP* chez différentes espèces.

Ces résultats indiquent que le gène *APP* est fortement préservé tout au long du processus évolutif chez les mammifères, avec une attention particulière portée sur les primates. La plus grande ressemblance a été notée entre l'homme (*Homo sapiens*) et le chimpanzé (*Pan troglodytes*), avec un taux de similarité de 99,70 %, suivi de près par le macaque (*Macaca mulatta*) avec un taux de 97,60 %.(table 2)

Ces observations indiquent une étroite relation évolutive entre l'homme et les primates en ce qui concerne le gène *APP*, et soutiennent l'usage de ces espèces comme modèles d'étude pour comprendre les processus moléculaires liés à la maladie d'Alzheimer et l'évolution des mutations touchant ce gène.

I.1.2- Étude comparative représentative:

Distribution of the top 5 Blast Hits on 5 subject sequences

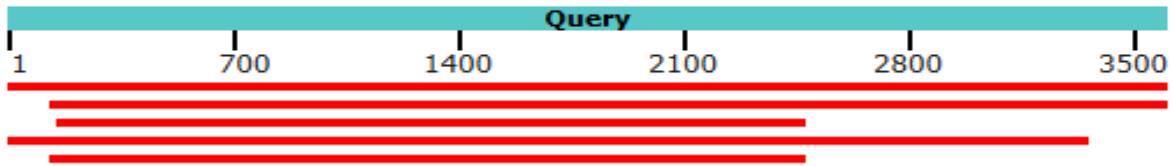


Figure 25 : Graphique comparatif montrant les similitudes et les différences du gène *APP* chez *Mus musculus* ; *Rattus norvegicus* ; *Canis lupus familiaris* ; *Pan troglodytes* et *Macaca mulatta*, en référence à *Homo sapiens*.

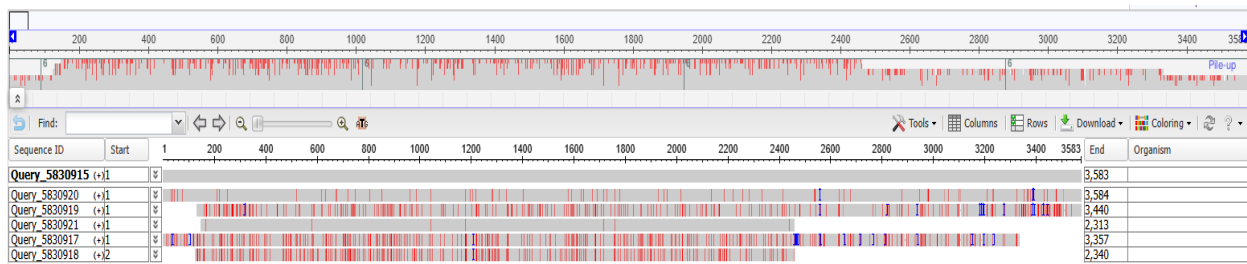


Figure 26 : Visualisation de l'alignement multiple des séquences (MSA Viewer) du gène *APP* montrant les similitudes et les différences du gène *APP* chez différentes espèces.

- ✓ **Séquence requête :** en couleur verte, elle illustre la séquence de nucléotides du gène chez *Homo sapiens*.
- ✓ **Séquence sujet :** représente la séquence nucléotidique du gène chez *Mus musculus* ; *Rattus norvegicus* ; *Canis lupus familiaris* ; *Pan troglodytes* et *Macaca mulatta*,

La figure 25 et 26 démontre une forte persistance des séquences parmi toutes les espèces examinées en comparaison avec *Homo sapiens*, illustrée par la présence dominante de barres rouges symbolisant des scores d'alignement égaux ou supérieurs à 200. Plutôt que de démontrer des différences totales, les résultats exposent d'importantes similitudes, avec *Macaca mulatta* affichant l'homologie de séquence la plus élevée. Les différences mineures et les variations évolutives se concentrent dans des zones précises et se traduisent par des remplacements ou absences de nucléotides qui se répartissent, notamment chez *Mus musculus* et *Rattus norvegicus*.

I.1.4- Étude de classification moléculaire des gènes *APP* codant :

Étude génétique et analyse moléculaire du gène *App* et étude de la phylogénie par analyse et traitement des séquences nucléotidiques à l'aide du site BLASTA

Il existe une relation génétique et évolutive étroite entre les humains (*Homo sapiens*) d'une part, et les autres espèces étudiées : les chimpanzés (*Pan troglodytes*), les macaques (*Macaca mulatta*), les chiens (*Canis lupus familiaris*), les rats (*Rattus norvegicus*) et les souris domestiques (*Mus musculus*) d'autre part.

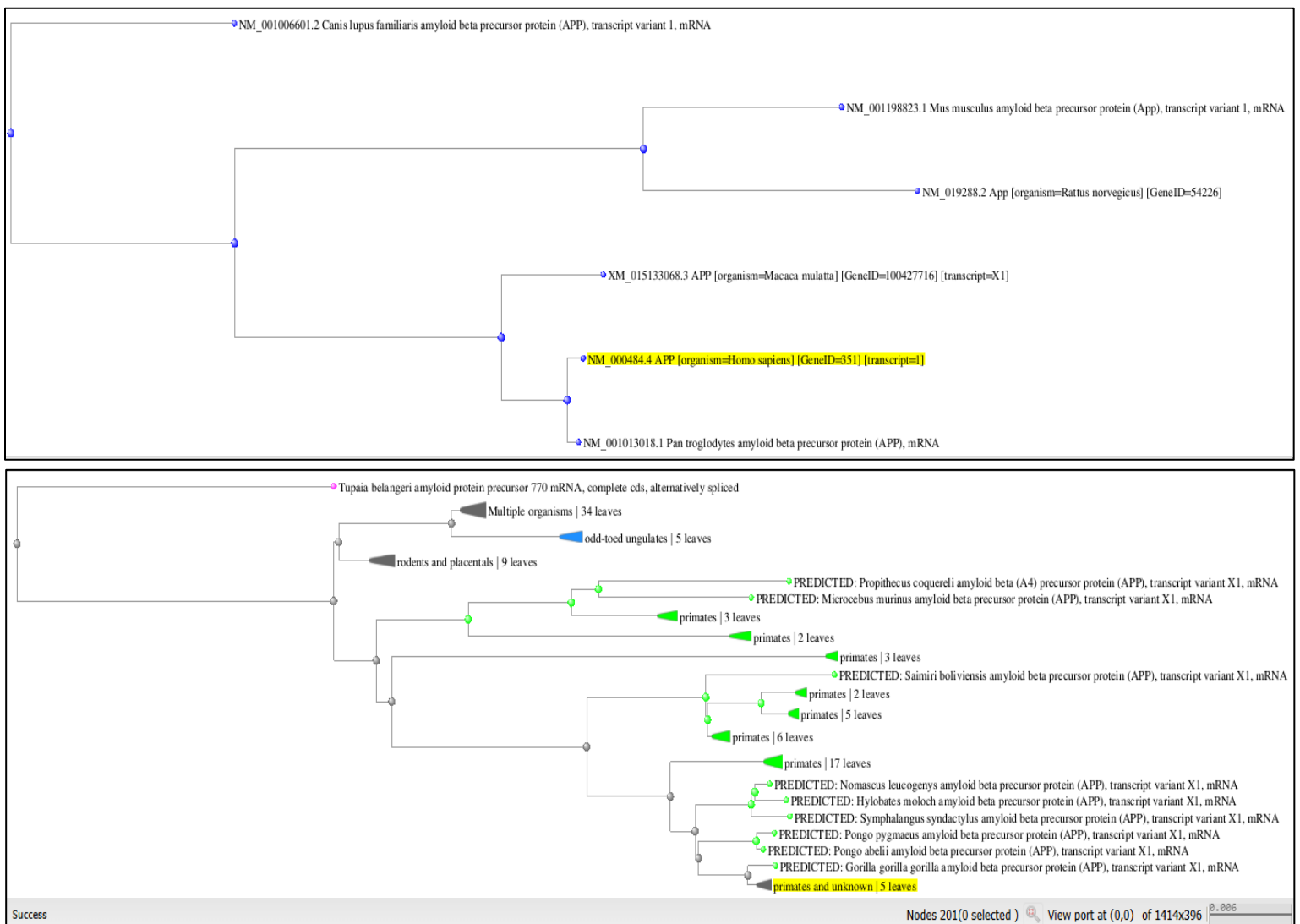


Figure 28: Arbre phylogénétique basé sur l'analyse de la séquence codée du gène *APP*, montrant le degré de parenté génétique.

Nous observons clairement une forte similarité évolutive et génétique entre le gène *APP* humain et les gènes *APP* des chimpanzés et des macaques rhésus, ce qui en fait les organismes les plus proches sur le plan évolutif (Figure 28). À titre d'exemple supplémentaire de ramification de groupe, nous observons également une forte similarité évolutive entre le gène *APP* de la souris et celui du rat, formant une branche génétique commune représentant l'ordre des rongeurs (Figure 28).

I.2- Analyse et Caractérisation de la Protéine *APP*:

I.2.1- Étude des propriétés physico-chimiques (ProtParam):

Le site web ExPASy ProtParam a été employé pour une analyse comparative des propriétés physico-chimiques. On a extrait les séquences d'acides aminés de la protéine *APP* normale et de la protéine *APP* mutante Swedish (KM670/671NL) en vue d'identifier et de quantifier des biomarqueurs importants, tels que le poids moléculaire (Da), le point isoelectrique (pI) et le coefficient de stabilité (II). Voici les résultats affichés dans le tableau suivant :

Tableau 3: Analyse comparative des caractéristiques physiques et chimiques de la protéine *APP* normale et de la protéine *APP* affectée par la mutation Swedish (KM670/671NL) en se servant de l'outil ExPASy ProtParam.

Propriétés physico-chimiques	<i>APP</i> normal	<i>APP</i> Swedish mutant
Nombre d'acides aminés	770	770
Point isoélectrique (pI)	4.73	4.72
Poids moléculaire (Da)	86943.25	86911.15
Indice d'instabilité (II)	40.69	40.31
Indice aliphatique	73.18	73.69
GRAVY	-0.584	-0.581

L'analyse révèle que la protéine *APP* normale et la protéine *APP* porteuse de la mutation Swedish disposent d'un nombre identique d'acides aminés (770), ce qui indique que la mutation ne modifie pas la longueur de la protéine. **Tableau 3**

Cependant, on a noté des variations mineures dans certaines caractéristiques physico-chimiques. La forme mutante présente une légère réduction de la masse moléculaire, due à la substitution de deux acides aminés dans la chaîne. On a également observé une modification légère du point isoélectrique (pI) et de l'indice d'instabilité (II), couplée à une légère hausse de l'indice aliphatique, tandis que la valeur GRAVY a maintenu une proximité avec la normale. **Tableau 3**

Même si ces changements paraissent mineurs sur le plan numérique, ils peuvent créer d'importantes conséquences biologiques, d'autant plus que la mutation Swedish est positionnée près de l'endroit où l'enzyme β -sécrétase clive, menant à une hausse de production du peptide *A β* qui entraîne la création des plaques amyloïdes liées à la maladie d'Alzheimer.

Ainsi, ces observations attestent que même un léger changement dans la séquence des protéines peut exercer une influence fonctionnelle importante sur le processus pathologique.

I.2.2.- Analyse des domaines protéiques (InterPro):

Nous avons suivi ce travail comme une approche complémentaire en utilisant la plateforme InterPro (EMBL-EBI), où nous avons entré des séquences protéiques afin d'identifier et de cartographier les domaines fonctionnels conservés et les familles de protéines, et de vérifier dans quelle mesure la mutation affecte le niveau topologique global de la protéine.

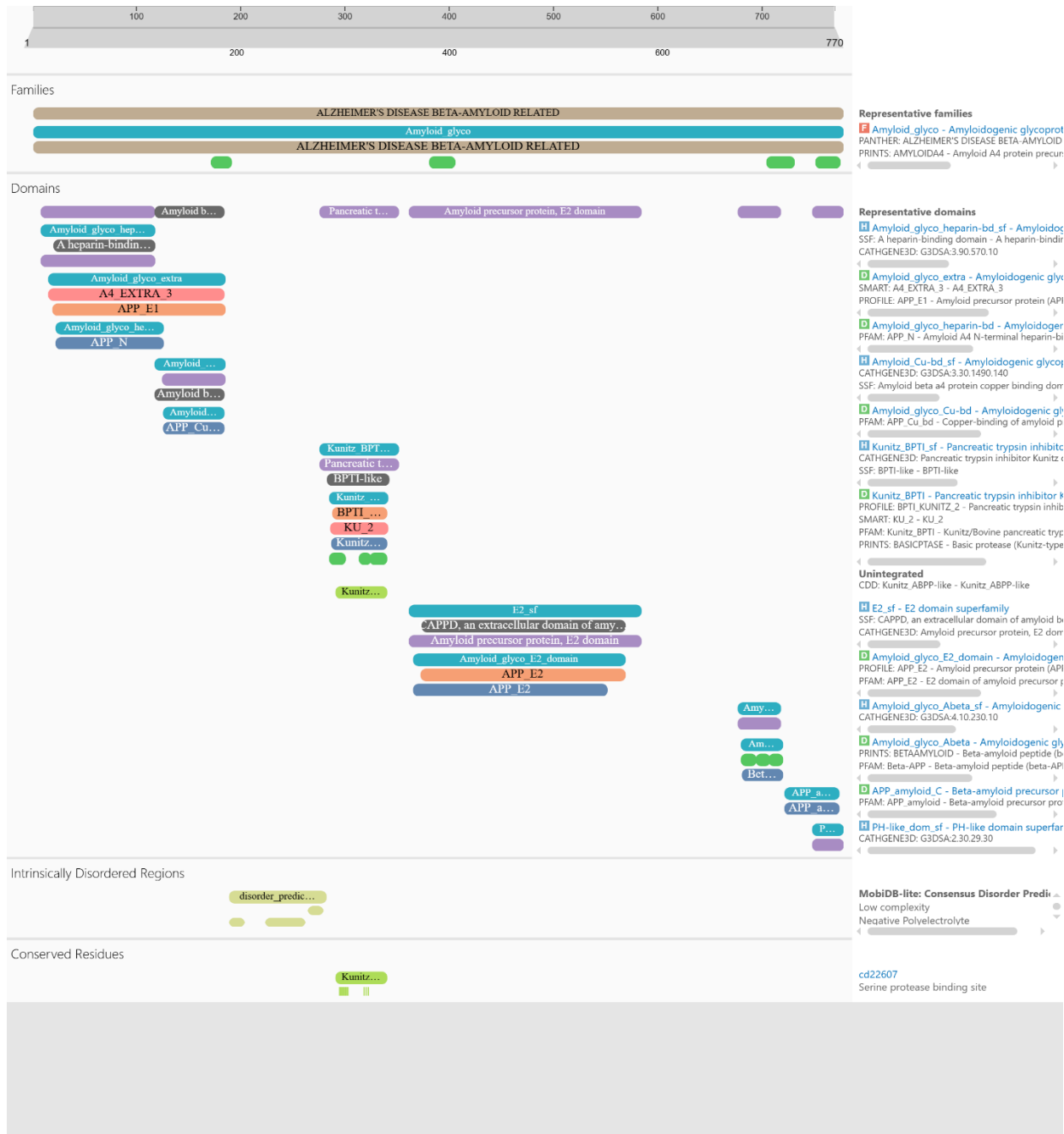


Figure 29 : Cartographie topologique détaillée de la protéine APP sauvage (770 aa) en mode d'affichage complet (Full Display Mode) via la plateforme InterPro, illustrant l'organisation globale des macro-domaines et des motifs structuraux conservés.

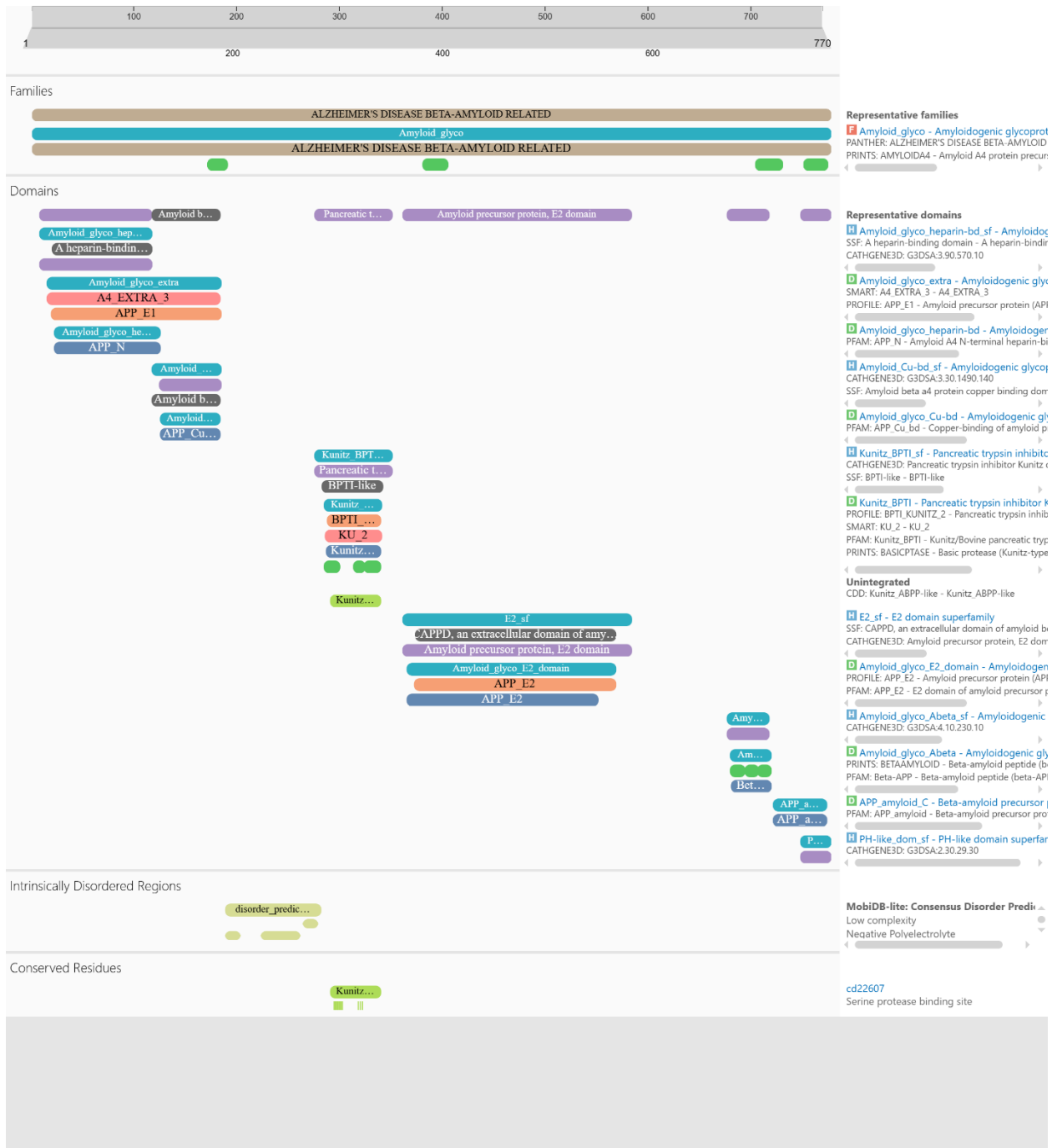


Figure 30 : Cartographie topologique détaillée de la protéine APP mutée Swedish (KM670/671NL) via InterPro, montrant une conservation stricte et une identité de 100% de l'architecture des domaines par rapport à la forme sauvage.

Nous avons effectué une analyse structurale et d'application complète via la plateforme InterPro en utilisant le « mode vue complète » pour suivre et cartographier avec précision la protéine *APP* normale et la variante mutante Swedish (KM670/671NL) (Blum et al., 2021).

La comparaison visuelle des deux chaînes a révélé une identité structurale et distributionnelle complète sur l'ensemble de la chaîne peptidique de 770 acides aminés, avec l'identification précise des composants structuraux communs suivants :

- ✓ **Familles de protéines conservées** : Les deux chaînes appartiennent à la même famille de protéines associée à la protéine bêta-amyloïde liée à la maladie d'Alzheimer, confirmant leur nature de glycoprotéines grâce au domaine (Amyloid_glyco) détecté (InterPro Consortium, 2025).
- ✓ **Domaines sous-structuraux** : La protéine mutante Swedish conserve tous les domaines fonctionnels caractéristiques de la forme normale, sans délétion ni altération. Il s'agit notamment des domaines d'acides aminés et du domaine de liaison au cuivre (APP_E1, APP_N, APP_Cu), du domaine structural central stable (APP_E2, E2_sf) et du domaine bêta-amyloïde critique à l'extrémité C-terminale (Amy..., Bet...) (Dawson et al., 2022).
- ✓ **Résidus conservés** : Une correspondance parfaite a été trouvée pour les sites fonctionnels régulateurs conservés, en particulier le domaine d'inhibition protégé de type Kunitz (Kunitz_BPTI) (Sigrist et al., 2013).

I.2.3.- Structure tridimensionnelle (SWISS-MODEL + PyMOL):

1. Modélisation conforme via le serveur SWISS-MODEL

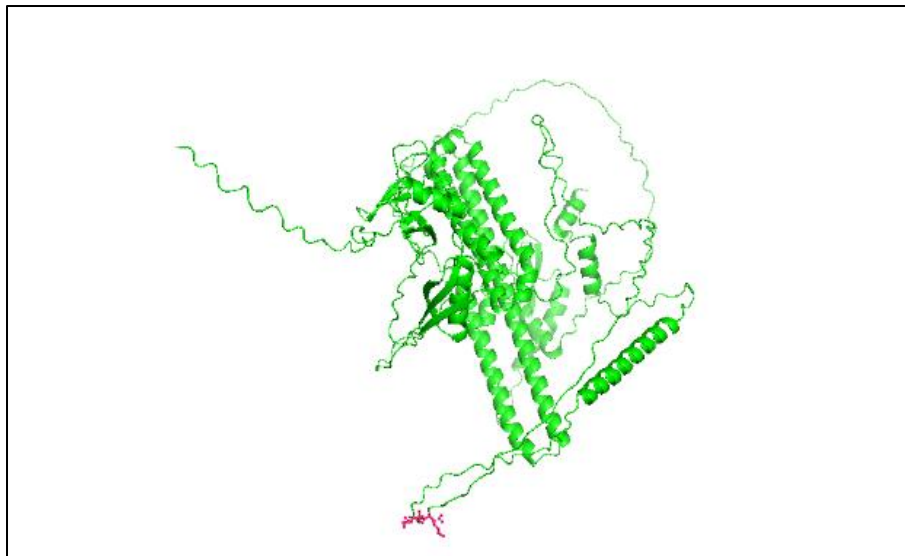
Dans le but d'examiner l'effet structurel de la mutation Swedish à l'échelle spatiale, une modélisation d'homologie a été réalisée pour prédire la structure tridimensionnelle (3D) de la protéine humaine *APP* complète (la version standard contenant 770 acides aminés) en utilisant le serveur **SWISS - MODEL**. À partir des séquences primaires, deux modèles structurels distincts ont été élaborés : l'un illustrant la version normale de la protéine (*APP* sauvage) et l'autre la version avec la double mutation Swedish (KM670/671NL). **Figure 8**

2. Appariement spatial global et calcul de l'indice RMSD à l'aide du logiciel PyMOL :

Pour examiner l'effet global de la mutation Swedish sur le repliement spatial de la protéine, nous avons transféré les fichiers de coordonnées structurales obtenus dans l'environnement de conception moléculaire PyMOL. Nous avons ensuite effectué une superposition structurale des deux brins en utilisant la commande "align" .

Les données sur le fusionnement spatial ont démontré un recouvrement géométrique précis et pratiquement idéal entre les deux structures ; la forme normale est représentée en vert et la forme mutante en bleu. **Figure 32 .**

L'examen mathématique de cette correspondance a montré une valeur d'écart quadratique moyenne extrêmement basse, presque nulle (inférieure à 0,05 Å). Cette mesure biophysique témoigne clairement que la mutation Swedish ne cause aucun bouleversement ni altération de la structure macroscopique de la protéine, et n'influence ni la répartition des hélices alpha ni la configuration spatiale générale de l'APP.



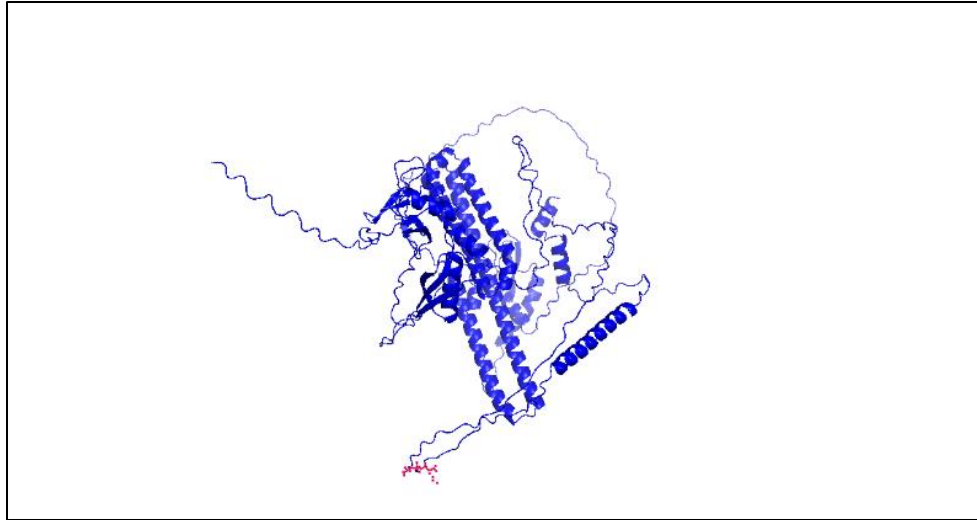


Figure 32: Superposition structurale 3D globale entre l'APP sauvage (représentée en vert) et la variante mutée Swedish (représentée en bleu) sous PyMOL, illustrant un repliement spatial identique et une conservation macro-conformationnelle totale.

2. Analyse de la localisation structurale et de l'environnement atomique du site de mutation :

Afin de cartographier avec précision l'altération induite par la mutation au sein de cette grande conservation structurale, les résidus d'acides aminés situés aux positions **670** et **671** ont été isolés et agrandis, en adoptant un mode de représentation en bâtonnets (**Sticks**) et en les colorant en rouge vif pour les individualiser, **Figure 33**.

L'examen visuel approfondi sous le logiciel PyMOL (**DeLano, 2002**) révèle que cette double substitution — où l'asparagine remplace la lysine (K670N) et la leucine remplace la méthionine (M671L) — induit un remaniement exclusivement micro-conformationnel (localisé). Cette modification locale restreinte réorganise discrètement la densité des charges électriques ainsi que l'encombrement stérique des chaînes latérales d'atomes à la surface de la protéine, précisément au niveau du site de reconnaissance et de clivage ciblé par l'enzyme **β -sécrétase**.

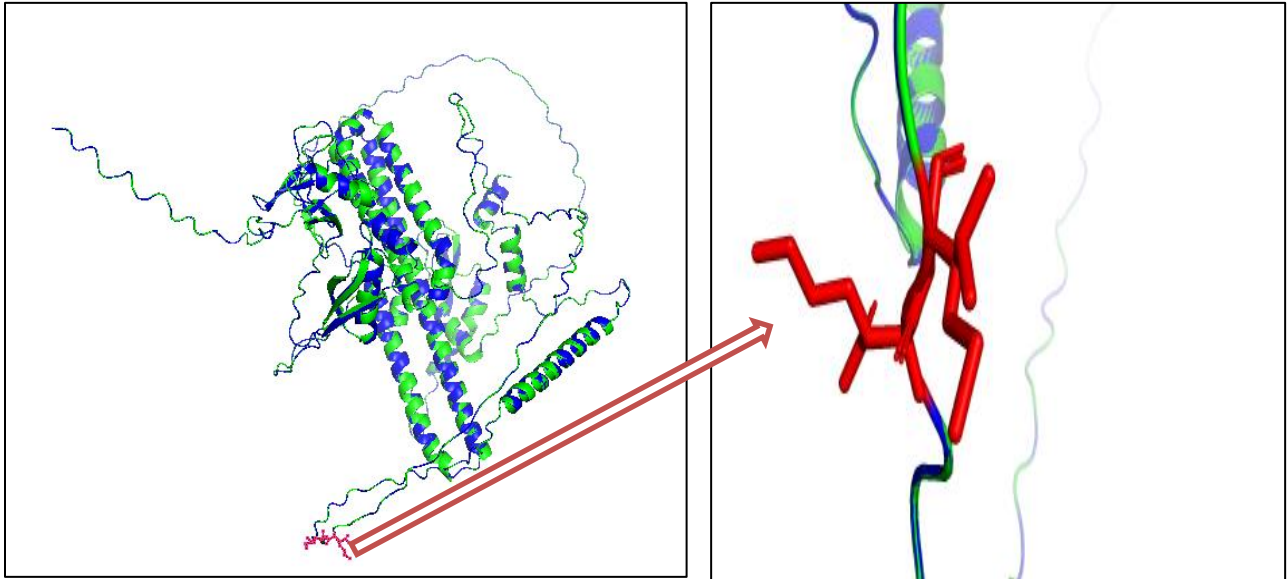


Figure 33 : Visualisation détaillée et zoomée sous **PyMOL** du site de la mutation Swedish (représentée en rouge et en mode bâtonnets) au niveau des positions **670** et **671**, illustrant l'impact micro-conformationnel local au niveau du site de coupure de la β -sécrétase.

I.3- Analyse comparative du gène *APOE* (Apolipoprotéine E):

Le gène *APOE* humain est constitué de quatre allèles (ϵ_1 , ϵ_2 , ϵ_3 et ϵ_4) **Tableau 4**. Les allèles ϵ_2 , ϵ_3 et ϵ_4 sont dominants et diffus, et diffèrent par la présence de mutations ponctuelles (SNP) au niveau de deux loci spécifiques : rs429358 et rs7412. Ces mutations entraînent un changement d'acide aminé aux positions 112 et 158 de la protéine (**Liu et al., 2013**). L'allèle ϵ_1 est un allèle mutant rare, et des études ont montré qu'il était peu fiable. Il existe également des allèles mutants cliniquement rares (**Feussner et Ziegler, 1992**).

Tableau 4 : Caractéristiques des allèles du gène *APOE* chez l'humain

Gène	Allèle	Position des SNP (rs429358 / rs7412)	Acides aminés (Pos 112 / Pos 158)	Fréquence	Effet (Maladie d'Alzheimer)
<i>APOE</i>	ε1	T / T (+ mutation supp)*	Cys / Cys (+ mutation)	Très rare	Non défini (Associé à certaines dyslipidémies)
	ε2	T / T	Cys / Cys	Rare	Protecteur (Diminue le risque)
	ε3	T / C	Cys / Arg	Fréquent	Référence / Normal (Risque neutre)
	ε4	C / C	Arg / Arg	Modéré	Risque élevé (Facteur de risque majeur)

On observe dans le tableau qu'à la position rs429358 : si le nucléotide est T, il donne l'acide aminé Cys (cystéine). S'il est C, il donne Arg (arginine).

À la position rs7412 : si le nucléotide est T, il donne Cys. S'il est C, il donne Arg.

La différence réside dans :

- ε2 : c'est le seul allèle présentant T à deux positions (donc Cys à deux endroits).
- ε4 : c'est le seul allèle présentant C à deux positions (donc Arg à deux endroits).
- ε3 : c'est l'allèle mixte (T en première position et C en seconde).

Quant à ε1 : il s'agit en réalité de l'allèle ε2 (Cys/Cys) auquel une troisième mutation a été ajoutée à un autre endroit (par exemple, en position 127). Par conséquent, les références scientifiques le classent systématiquement dans la catégorie des SNP « très rares » et le distinguent de la classification classique des SNP. (Feussner & Ziegler, 1992).

Nous avons choisi le gène *APOE* car il est responsable de la production d'apolipoprotéine E, un composant essentiel du transport des lipides dans le système nerveux, qui détermine la prédisposition génétique à la maladie d'Alzheimer.

Nous avons extrait les séquences d'ARNm des différents allèles APOE ($\epsilon 2$, $\epsilon 3$ et $\epsilon 4$) de la base de données NCBI.

Après comparaison de ces séquences à l'aide de BLAST, nous avons obtenu des résultats précis indiquant le degré de similarité entre elles.

- **comparé les allèles mutants ($\epsilon 2$ et $\epsilon 4$) avec l'allèle de référence ($\epsilon 3$):**

Les résultats ont montré ce qui suit :

select all 2 sequences selected		Graphics	Distance tree of results	MSA Viewer				
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> NM_000041.4_Mutated_APOE_e4_Homo_sapiens		2099	2127	100%	0.0	99.91%	1166	Query_166621
<input checked="" type="checkbox"/> NM_000041.4_Mutated_APOE_e2_Homo_sapiens		2099	2127	100%	0.0	99.91%	1166	Query_166620

Figure 34 : Résultats de l'alignement de la séquence d'ARNm progéniteur du gène *APOE* $\epsilon 3$ avec la séquence mutante *APOE* $\epsilon 2$ et $\epsilon 4$ à l'aide de l'outil BLASTn.

- Couverture de la requête : **100 %**, confirmant que les séquences étudiées sont parfaitement alignées en longueur avec la séquence de référence.
 - Pourcentage d'identité : **99,91 %**, indiquant fortement que les variations génétiques se limitent à des polymorphismes nucléotidiques simples (SNP) et ne sont pas dues à des modifications structurales importantes.
 - Valeur E : **0,0**, confirmant que cette similarité n'est pas due au hasard, mais constitue un indicateur statistiquement significatif d'une forte parenté génétique.
 - Score maximal : **2148**, reflétant la haute qualité et la précision du processus d'alignement.
- Ces résultats indiquent que les différences fonctionnelles et cliniques entre les allèles ($\epsilon 2$, $\epsilon 3$, $\epsilon 4$) ne sont pas dues à un défaut dans la séquence génétique globale, mais plutôt à des mutations locales précises (SNP) aux positions 112 et 158, ce qui explique la prédisposition génétique différente à la maladie d'Alzheimer malgré la grande similarité dans la structure des gènes.

Cette image montre le résumé graphique de l'alignement, où la barre rouge représente un score de correspondance très élevé (Score ≥ 200) le long de la séquence de requête s'étendant sur environ 1100 bases nucléotidiques:

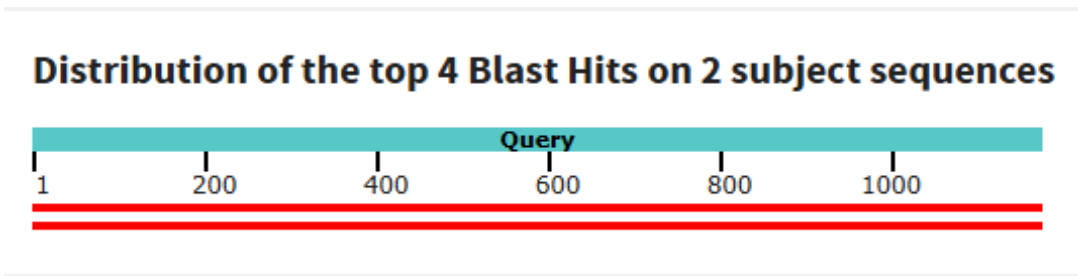
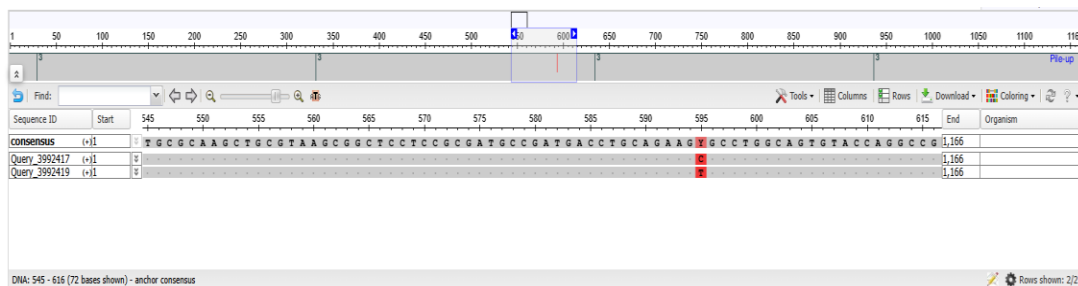
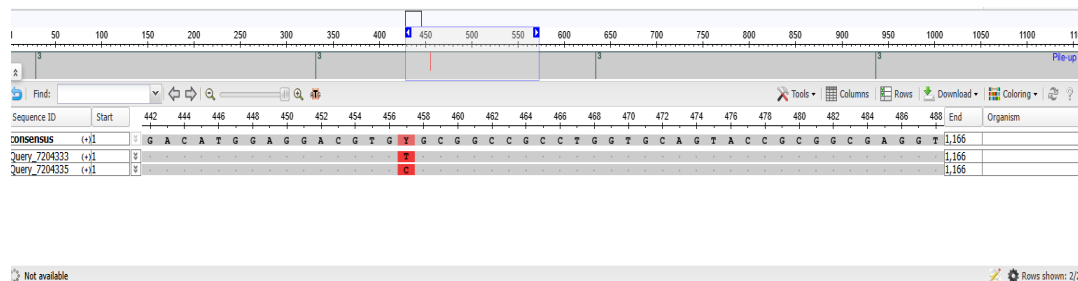


Figure 35: Représentation graphique de la distribution des meilleurs alignements séquentiels du gène APOE via BLAST.



Cette image, extraite de l'outil MSA Viewer du NCBI, illustre l'alignement de la séquence consensus avec deux séquences cibles (requêtes). Une fine ligne verticale rouge est visible à la position 595 environ, indiquant précisément l'emplacement de la mutation ponctuelle.



Cette image, extraite de l'outil MSA Viewer du NCBI, illustre l'alignement de la séquence consensus avec deux séquences cibles (requêtes). Une fine ligne verticale rouge est visible à la position 4 environ, indiquant précisément l'emplacement de la mutation ponctuelle.

I.3.1- Comparaison des séquences nucléotidiques des gènes APOE:

Mus musculus ; Rattus norvegicus; Canis lupus familiaris ; Pan troglodytes et Macaca mulatta font référence à Homo sapiens:

Tableau 5 : Gènes *APOE* de comparaison pour chacun des *Mus musculus* ; *Rattus norvegicus*; *Canis lupus familiaris* ; *Pan troglodytes* et *Macaca mulatta* faisant référence à *Homo sapiens*.

Gène	Organisme	Chr	Localisation	Nombre d'exons	Longueur
<i>APOE</i>	<i>Homo sapiens</i>	19	NC_000019.10 (44905796..44909393)	6	3,600 nt

Gène	Organisme	Chr	Localisation	Nombre d'exons	Longueur	Query cover	Per.Ident	E.value
<i>APOE</i>	<i>Mus musculus</i>	7	NC_000073.7 (19430169..19434326, complement)	7	4,158 nt	91%	76.80%	0.0
	<i>Rattus norvegicus</i>	1	NC_086019.1 (88481889..88485816, complement)	6	3,928 nt	85%	77.87%	0.0
	<i>Canis lupus familiaris</i>	1	NC_051805.1 (111055483..111058288, complement)	4	2,806 nt	100%	79.76%	0.0
	<i>Pan troglodytes</i>	20	NC_072418.2 (47862141..47865637)	4	3,497 nt	87%	98.52%	0.0
	<i>Macaca mulatta</i>	19	NC_133424.1 (57412134..57416131)	4	3,998 nt	100%	95.18%	0.0

- La couverture de la requête dans le résultat BLAST est de 100 %, ce qui signifie que la séquence requête est entièrement alignée avec la séquence cible dans la base de données. Aucune partie de la séquence requête n'est laissée non alignée ou inutilisée dans la correspondance (*SequenceServer., 2023*) .
- Valeur E (Valeur attendue) Nombre d'alignements atteignant un certain score ou plus et susceptibles d'être observés aléatoirement dans une base de données de cette taille. Plus les

valeurs sont faibles, plus les alignements sont significatifs. Les valeurs proches de 0 sont les plus significatives.

- Par identité Pourcentage de correspondances identiques entre la séquence requête et la séquence cible le long de la longueur d'alignement, indiquant la similarité des deux régions à la base (ADN/ADN).

Le tableau 5 montre que tous les gènes diffèrent par le nombre d'exons et présentent des longueurs et des taux de couverture de séquence variables

Par ailleurs, toutes les espèces étudiées présentent une forte similarité avec l'humain, avec les pourcentages de concordance suivants :

- *Macaca mulatta* : 95.18%
- *Pan troglodytes* : 98.52%
- *Canis lupus familiaris* : 79.76%
- *Mus musculus* : 76.80%
- *Rattus norvegicus* : 77.87%

select all 5 sequences selected		Graphics	Distance tree of results	MSA Viewer				
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> XM_077980391.1 APOE [organism=Macaca mulatta] [GeneID=714623] [transcript=X1]		1842	1842	100%	0.0	95.18%	1165	Query_5647437
<input checked="" type="checkbox"/> NM_001009007.1 Pan troglodytes apolipoprotein E (APOE), mRNA		1758	1787	87%	0.0	98.52%	1012	Query_5647436
<input checked="" type="checkbox"/> XM_005616460.4 PREDICTED: Canis lupus familiaris apolipoprotein E (APOE), transcript variant X1, mRNA		1038	1038	100%	0.0	79.76%	1230	Query_5647435
<input checked="" type="checkbox"/> NM_009696.4 Mus musculus apolipoprotein E (ApoE), transcript variant 1, mRNA		780	813	91%	0.0	76.80%	1288	Query_5647438
<input checked="" type="checkbox"/> NM_001270681.1 Rattus norvegicus apolipoprotein E (ApoE), transcript variant 1, mRNA		774	811	85%	0.0	77.87%	1212	Query_5647439

Figure 36 : Résultats de l'alignement de séquences nucléotidiques par BLASTN pour le gène *APOE* chez différentes espèces.

Ces résultats indiquent que le gène *APOE* est fortement préservé tout au long du processus évolutif chez les mammifères, avec une attention particulière portée sur les primates. La plus grande ressemblance a été notée entre l'homme (*Homo sapiens*) et le chimpanzé (*Pan troglodytes*), avec un taux de similarité de 98.52%, suivi de près par le macaque (*Macaca mulatta*) avec un taux de 95.18%. (Tableau 5)

Ces observations indiquent une étroite relation évolutive entre l'homme et les primates en ce qui concerne le gène *APOE*, et soutiennent l'usage de ces espèces comme modèles d'étude pour

comprendre les processus moléculaires liés à la maladie d'Alzheimer et l'évolution des mutations touchant ce gène.

I.3.2- Étude comparative représentative:

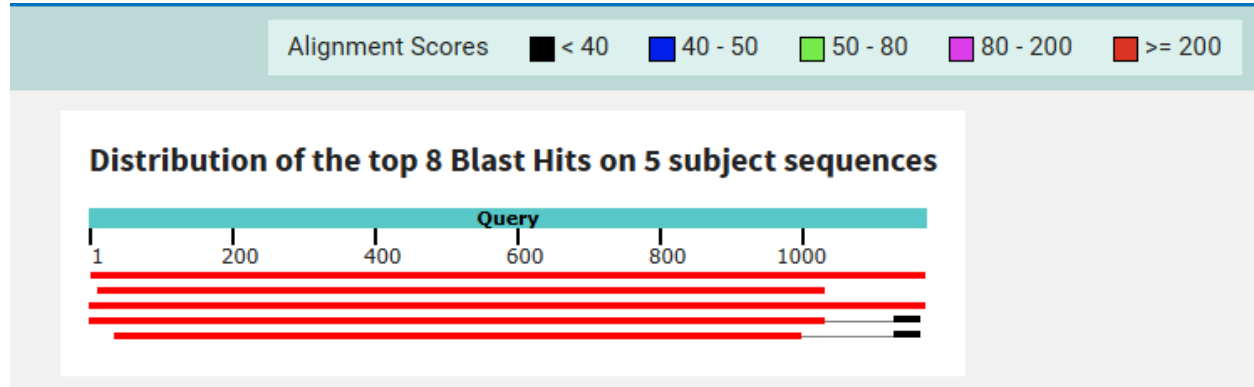


Figure 37 : Graphique comparatif montrant les similitudes et les différences du gène *ApoE* chez *Mus musculus* ; *Rattus norvegicus* ; *Canis lupus familiaris* ; *Pan troglodytes* et *Macaca mulatta*, en référence à *Homo sapiens*.

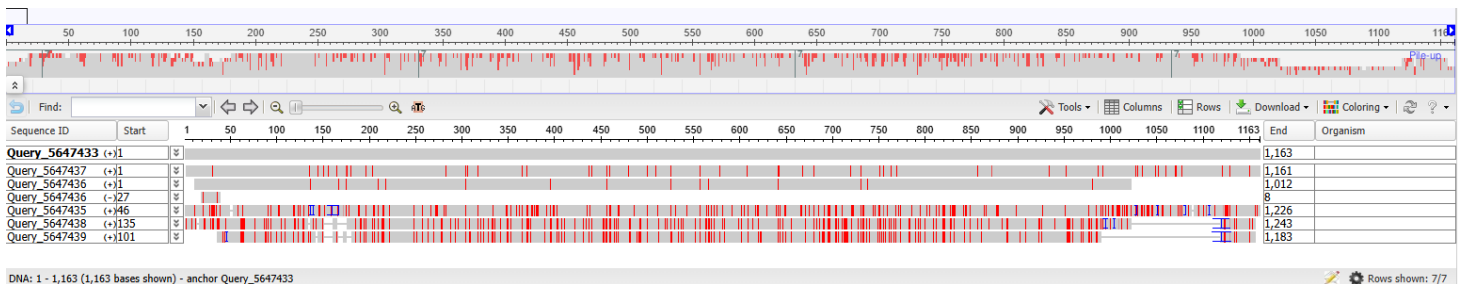


Figure 38: Visualisation de l'alignement multiple des séquences (MSA Viewer) du gène *APP* montrant les similitudes et les différences du gène *APOE* chez différentes espèces.

- ✓ **Séquence requête :** en couleur verte, elle illustre la séquence de nucléotides du gène *ApoE* chez *Homo sapiens*.
- ✓ **Séquence sujet :** représente la séquence nucléotidique du gène chez *Mus musculus* ; *Rattus norvegicus* ; *Canis lupus familiaris* ; *Pan troglodytes* et *Macaca mulatta*,

La figure 37 et 38 montre une **prédominance du rouge**, indiquant une forte conservation du gène. L'analyse visuelle révèle une nette dominance des bandes rouges qui, selon la légende, correspondent aux scores d'alignement les plus élevés (≥ 200). Ceci souligne la forte continuité et la haute homologie de séquence conservée du gène *ApoE* au cours de l'évolution de ces divers mammifères, comparativement à l'humain. Les trois premières séquences sont alignées de manière constante sur presque toute la longueur du gène, suggérant une convergence évolutive importante,

notamment chez les primates. Nous observons également de fines **lignes noires** (ruptures d'alignement) : malgré ce haut degré de similarité, des différences évolutives clairement définies apparaissent dans les deux dernières séquences (caractéristiques des rongeurs : *Mus musculus* et *Rattus norvegicus*) près de l'extrémité 3' du gène, entre les positions 1000 et 1100, indiquant une absence totale d'homologie (lacunes) et suggérant des mutations d'insertion ou de délétion.

Blocs noirs terminaux (score < 40) : situés à l'extrémité du gène, ils reconstruisent brièvement l'alignement avec un score très faible, reflétant une variation nucléotidique terminale significative par rapport à *Homo sapiens*.

I.3.3- Alignement multiple des séquences (AMS):

Dans cette étude, les séquences d'ARNm du gène *ApoE* (apolipoprotéine E) de plusieurs espèces ont été alignées à l'aide du logiciel **MAFFT**.

```
MAFFT-L-INS-i Result
-----
CLUSTAL format alignment by MAFFT (v7.511)
-----
NM_000041. -----
NM_0010090 -----
XM_0779803 -----
XM_0056164 -----
NM_009696. ttccctctgccctgctgtgaagggggagagaacaacccgcctcgtgacaggggctggca
NM_0012706 -----

NM_000041. -----
NM_0010090 -----
XM_0779803 -----
XM_0056164 -----ggcagggggaacatataaatggacagtc
NM_009696. cagcccgccctagccctgaggaggggggggacagggggagtcctataatggaccggtc
NM_0012706 -----ataatggacaggtc

NM_000041. -----ctactcagccccagcggaggtgaaggacgtccttccccaggagc-
NM_0010090 -----cagcggaggtgaaggacgtccttccccaggagc-
XM_0779803 -----actcagccccagcggaggtgaaggacgtccttccccaggagc-
XM_0056164 tggatccctcagtcctactcagccccagcggaggtgaaggaggttccccaggagc-
NM_009696. tgggat-cogatccccctgctcagacc--tggaggctaaggacttgtttcggaaggagc-
NM_0012706 tgggat-cgggtccccctgctcagacc--cggaggctaaggagttgtttcggaaggagc
*****  *****  *  .*****
```

Les résultats ont ensuite été présentés et analysés avec le logiciel **MEGA** afin d'identifier les domaines et nucléotide hautement conservés au cours de l'évolution.

DNA Sequences	Translated Protein Sequences
Species/Abbrv	
1. NM 000041.4 APOE e3 Homo sapiens ref	-----c g a c t g g c c a a t c a c a g g c a g g a a g a t g a a g g t t c t g t g g g c t g c g
2. NM 001009007.1	-----c g a c t g g c c a a t c a c a g g c a g g a a g a t g a a g g t t c t g t g g g c t g c g
3. XM 077980391.1	-----c g a c t g g c c a a t c a c a g g c a g g a a g a t g a a g g t t c t g t g g g c t g c g
4. XM 005616480.4	-----c g a - - - c c a a t c g c a g g c t g g a a g a t g a a g g t t c t g t g g g c t g c g
5. NM 009696.4	-----t g a c t g g c c a a t c a c a t t g c g a a g a t g a a g g c t c t g t g g g c c g t g
6. NM 001270681.1	c t g g c g a t t c a c c c a g g g g c t t g a c t g g c c a a t c a c a c t g g g a a g a t g a a g g c t c t g t g g g c c c t g

Nous construisons l'arbre phylogénétique à partir des données AMS en utilisant le logiciel MEGA 12. Nous avons choisi la méthode de construction/test de l'arbre de vraisemblance maximale (ML) car elle est considérée comme la plus précise et la plus robuste scientifiquement.

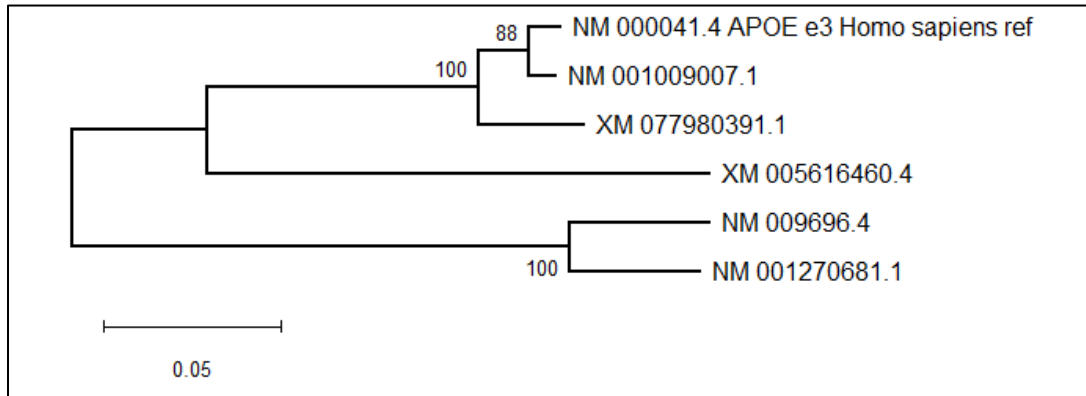


Figure 39 : Arbre phylogénétique des séquences nucléotidiques du gène *APOE* de *Homo sapiens* et ses orthologues chez d'autres espèces de mammifères. sous le logiciel MEGA 12.

Les valeurs de bootstrap (basées sur 1000 répliquions) sont indiquées au niveau des nœuds. L'échelle (0.05) représente le nombre de substitutions de nucléotides par site.

- Analyse des clades évolutifs :

Premier clade et plus proche : Les séquences NM_000041.4 (gène *APOE* e3 d'*Homo sapiens*) et NM_001009007.1 de *Pan troglodytes* se situent sur la branche la plus proche, ce qui indique une très forte similarité génétique et une parenté évolutive étroite.

Sous-clade partagé : La séquence XM_077980391.1 de *Macaca mulatta* partage un ancêtre commun plus récent avec les deux séquences précédentes que les autres séquences de l'arbre.

La séquence XM_005616460.4 de *Canis lupus familiaris* représente un lien légèrement plus éloigné avec les primates et apparaît donc comme une lignée latérale distincte des groupes humains et simiesques.

Groupes distants : Les deux *Mus musculus* NM_009696.4 et *Rattus norvegicus* NM_001270681.1 successifs ont formé une famille (clade) complètement séparée à la base de l'arbre, indiquant qu'ils différaient génétiquement de manière marquée du groupe supérieur.

I.3.4- Étude de classification moléculaire des gènes *APOE* codant :

Étude génétique et analyse moléculaire du gène *APOE* et étude de la phylogénie par analyse et traitement des séquences nucléotidiques à l'aide du site BLASTA

Il existe une relation génétique et évolutive étroite entre les humains (*Homo sapiens*) d'une part, et les autres espèces étudiées : les chimpanzés (*Pan troglodytes*), les macaques (*Macaca mulatta*), les chiens (*Canis lupus familiaris*), les rats (*Rattus norvegicus*) et les souris domestiques (*Mus musculus*) d'autre part.

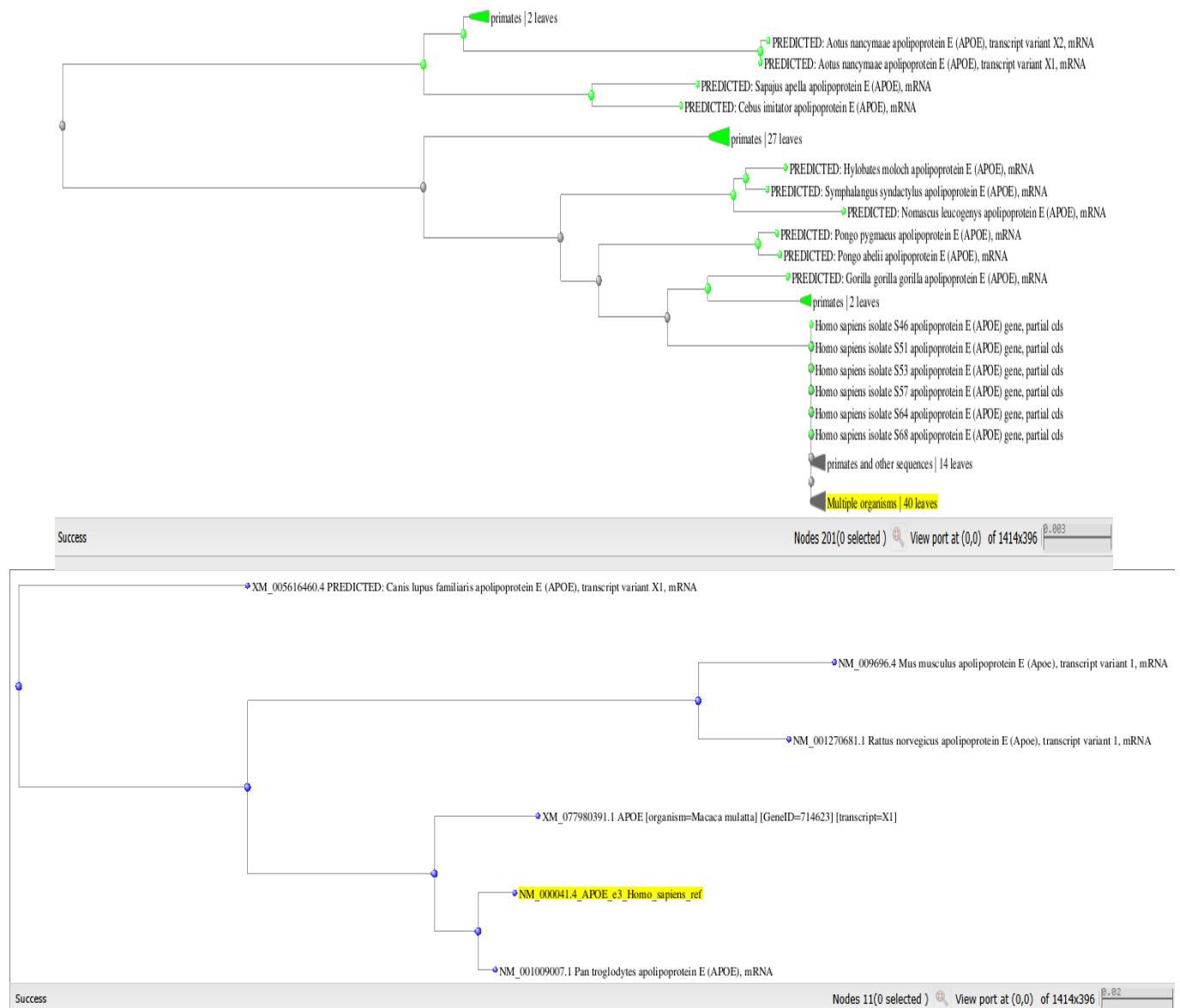


Figure 40: Arbre phylogénétique basé sur l'analyse de la séquence codée du gène Apoe, montrant le degré de parenté génétique.

I.4- Analyse et Caractérisation de la Protéine Apoe (Apolipoprotéine E):

I.4.1- Étude des propriétés physico-chimiques (ProtParam):

Le site web ExPASy ProtParam a été employé pour une analyse comparative des propriétés physico-chimiques. On a extrait les séquences d'acides aminés de la protéine Apoe ($\epsilon 2$, $\epsilon 3$ et $\epsilon 4$) en vue d'identifier et de quantifier des biomarqueurs importants, tels que le poids moléculaire (Da), le point isoelectrique (pI) et le coefficient de stabilité (II). Voici les résultats affichés dans le tableau suivant :

Tableau 6: Analyse comparative des caractéristiques physiques et chimiques de la protéine Apoe en se servant de l'outil ExPASy ProtParam. ($\epsilon 2$, $\epsilon 3$, $\epsilon 4$).

Propriétés physico-chimiques	Apoe $\epsilon 2$	Apoe $\epsilon 3$	Apoe $\epsilon 4$
Nombre d'acides aminés	317	317	317
Point isoélectrique (pI)	5.52	5.65	5.81
Poids moléculaire (Da)	36101.04	36154.08	36207.13
Indice d'instabilité (II)	44.28	44.88	44.62
Indice aliphatique	87.16	87.16	87.16
GRAVY	-0.574	-0.596	-0.618

Nous observons les caractéristiques communes (conservées parmi les variants) : Nombre d'acides aminés (317) : identique pour les trois variants ($\epsilon 2$, $\epsilon 3$, $\epsilon 4$). Ceci indique que les différences observées ne sont pas dues à des délétions ou des insertions dans la séquence, mais plutôt à des

mutations ponctuelles (SNP). Indice aliphatique (87,16) : parfaitement constant, reflétant une similarité de la stabilité thermique hypothétique des trois protéines, basée sur la nature de leurs chaînes latérales hydrocarbonées. **Tableau 6**

Cependant, des différences structurales et électriques existent, notamment au niveau du point de neutralité électrique (pI) : on observe une augmentation progressive et nette des valeurs de pI : pI ($\epsilon_2 = 5,52$) < ($\epsilon_3 = 5,65$) < ($\epsilon_4 = 5,81$) **Tableau 6**, la variante ϵ_4 étant la plus basique (ou la moins acide). Sur le plan biochimique, cela est dû à la substitution d'acides aminés neutres (comme la cystéine dans l'*ApoE3*) par des acides aminés chargés positivement (comme l'arginine dans l'*ApoE4*), ce qui augmente la charge globale de la protéine. Masse moléculaire : elle augmente proportionnellement, mais très légèrement. (**Zhao et al., 2018**).

Par ailleurs, indicateurs de stabilité et de réactivité avec le milieu : Indice d'instabilité > 40 : toutes les variantes présentent un indice supérieur à 40 (compris entre 44,28 et 44,8) **Tableau 6**. Selon les critères de l'instrument ProtParam, ces trois protéines sont classées comme instables in vitro, ce qui indique leur flexibilité structurale et leur capacité à modifier leur conformation spatiale. L'indice GRAVY (hydrophobicité) est négatif : toutes les valeurs sont négatives (de -0,574 à -0,618). Une valeur négative signifie que la protéine est globalement hydrophile, ce qui est parfaitement cohérent avec la fonction de l'*ApoE* en tant que protéine de transport de lipides évoluant dans les milieux aqueux de l'organisme (tels que le plasma et le liquide céphalo-rachidien). (**Mahley, 2022**).

I.4.2- Analyse des domaines protéiques (InterPro):

Nous avons poursuivi ce travail comme une approche complémentaire en utilisant la plateforme InterPro (EMBL-EBI), où nous avons entré des séquences protéiques pour identifier et cartographier les domaines fonctionnels conservés et les familles de protéines, et pour étudier comment les mutations ponctuelles (SNP) (ϵ_2 , ϵ_3 , ϵ_4) affectent le niveau topologique global de la protéine.

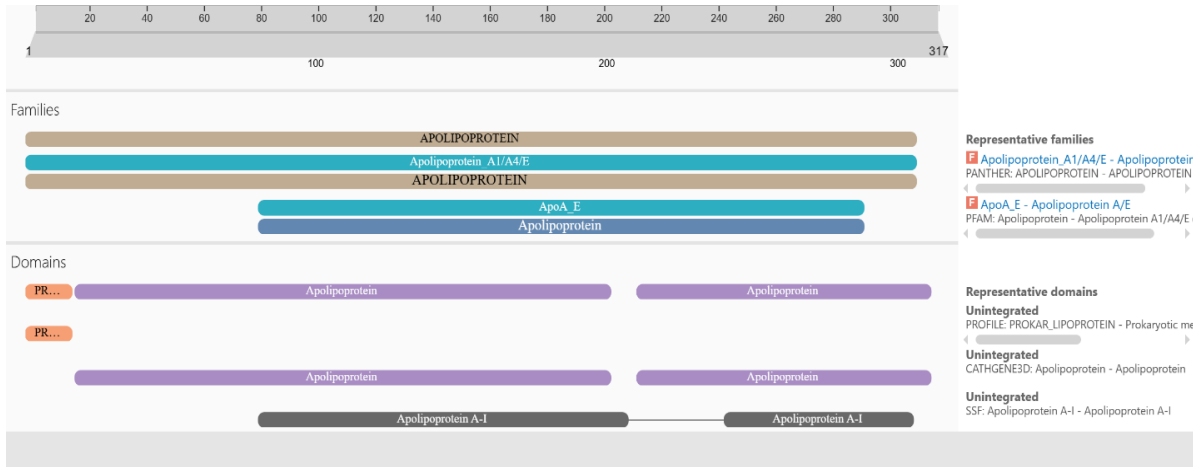


Figure 41 : Cartographie topologique détaillée de la protéine ApoE ε2 sauvage (317 aa) en mode d'affichage complet (Full Display Mode) via la plateforme InterPro, illustrant l'organisation globale des macro-domaines et des motifs structuraux conservés.

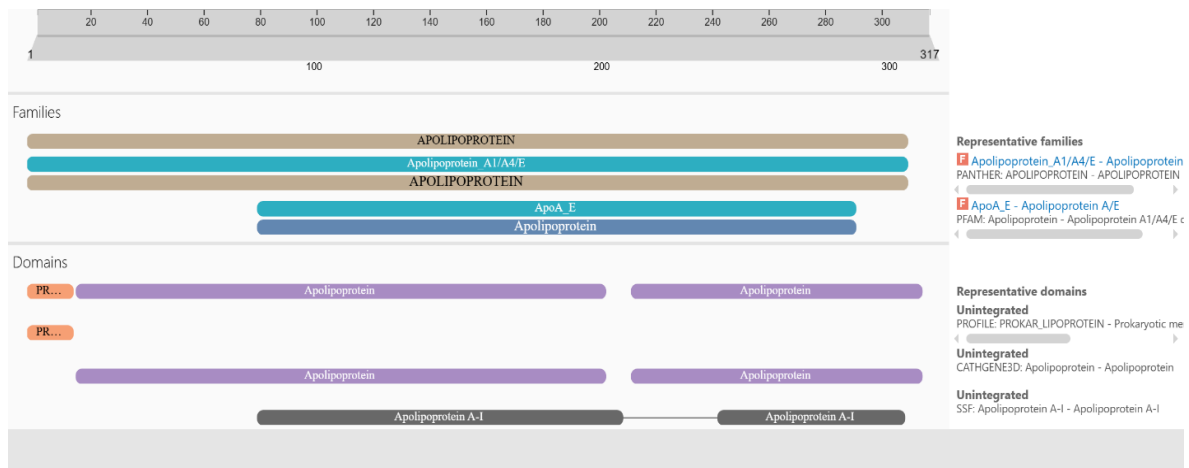


Figure 42 : Cartographie topologique détaillée de la protéine ApoE ε2 sauvage (317 aa) en mode d'affichage complet (Full Display Mode) via la plateforme InterPro, illustrant l'organisation globale des macro-domaines et des motifs structuraux conservés.

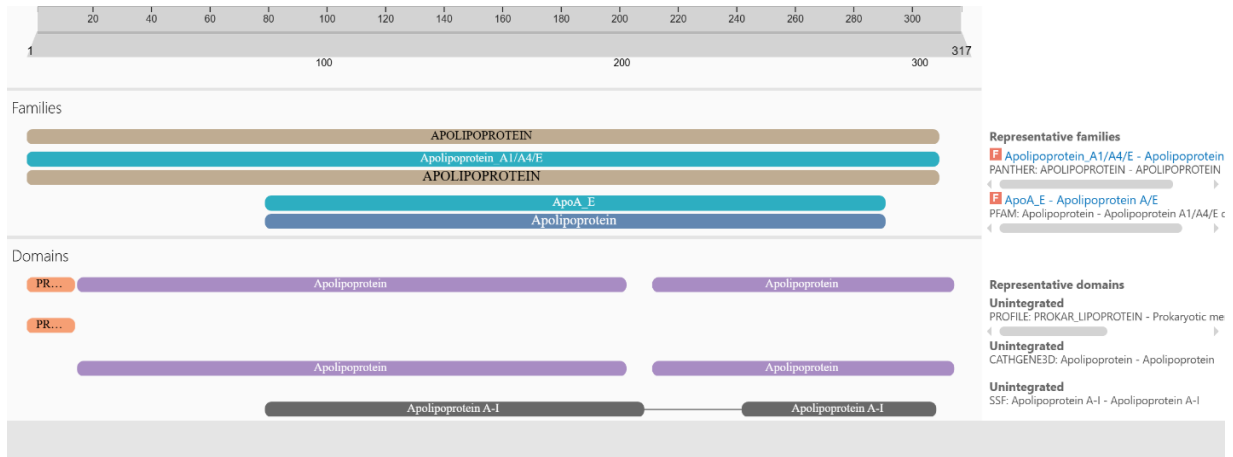


Figure 43: Cartographie topologique détaillée de la protéine Apoe $\epsilon 4$ sauvage (317 aa) en mode d'affichage complet (*Full Display Mode*) via la plateforme InterPro, illustrant l'organisation globale des macro-domaines et des motifs structuraux conservés.

Les figures 17, 18 et 19 illustrent la cartographie topologique de la protéine Apoe de type sauvage ($\epsilon 2$, $\epsilon 3$, $\epsilon 4$), qui possède une séquence hautement ordonnée de 317 acides aminés. Cette protéine appartient à la famille des lipoprotéines (empreintes PANTHER et PFAM Apoe) et est fonctionnellement organisée en une courte sous-unité primaire (PR...) suivie de deux grandes bandes lipoprotéiques globales (représentées en violet) séparées par une région charnière flexible. (**InterPro, 2026**).

Sur la base des résultats de la plateforme InterPro et des données topologiques présentées dans les figures précédentes, nous résumons les caractéristiques génétiques, structurelles et physiopathologiques des différents isotopes *d'Apoe* dans le **tableau 7** suivant :

Tableau 7 : Comparaison des caractéristiques des isoformes de l'ApoE ($\epsilon 2$, $\epsilon 3$, $\epsilon 4$)

Critère de comparaison	ApoE $\epsilon 3$ (Normal / Sauvage)	ApoE $\epsilon 4$ (À haut risque)	ApoE $\epsilon 2$ (Protecteur)
Statut génétique	L'allèle le plus fréquent (type sauvage).	Le facteur de risque génétique le plus puissant pour la maladie d'Alzheimer sporadique.	L'allèle le plus rare, considéré comme un facteur protecteur contre la maladie.
Variations des acides aminés	Contient : Cing-112 (Cysteine) Arg-158 (Arginine)	Contient : Arg-112 (Arginine) Arg-158 (Arginine)	Contient : Cing-112 (Cysteine) Cing-158 (Cysteine)
Structure tridimensionnelle et stabilité	Structure stable et fonctionnellement équilibrée.	Structure moins stable ; interaction accrue entre les domaines (<i>Domain interaction</i>), favorisant une conformation pathologique.	Structure tridimensionnelle modifiant l'affinité de liaison aux récepteurs.
Mécanisme pathologique cérébral	Assure le transport normal des lipides et la clairance physiologique des peptides A β .	- Réduit la clairance et la dégradation des peptides amyloïdes (A β). - Favorise l'hyperphosphorylation et les enchevêtrements de la protéine Tau. - Augmente la neuroinflammation.	Stimule efficacement la clairance des peptides A β ou protège les neurones contre la toxicité amyloïde.
Liaison aux récepteurs des LDL	Liaison normale et efficace.	Liaison normale et efficace.	Liaison très faible (inférieure à 1% par rapport à l'ApoE $\epsilon 3$), ce qui peut modifier localement le métabolisme lipidique de façon protectrice.

I.4.3- Structure tridimensionnelle (SWISS-MODEL + PyMOL):

1. Modélisation conforme via le serveur SWISS-MODEL

Pour analyser l'impact structural spatial des polymorphismes du gène *ApoE*, la modélisation par homologie a été utilisée pour prédire la structure tridimensionnelle de la protéine humaine complète (la variante standard de 317 acides aminés) à l'aide du serveur SWISS-MODEL. À partir des séquences initiales, différents modèles structuraux ont été générés : l'un représentant la version normale de la protéine (*ApoE* $\epsilon 3$ de type sauvage), et d'autres représentant les variants structuraux associés aux allèles *ApoE* $\epsilon 4$ (Cys112Arg) et *ApoE* $\epsilon 2$ (Arg158Cys). Le serveur a sélectionné les modèles structuraux de référence les plus performants disponibles dans la Protein Data Bank (PDB), présentant la meilleure concordance et la couverture de séquence la plus étendue. Les modèles finaux ont été évalués à l'aide d'indices géométriques et thermodynamiques standard (par exemple, GMQE et QMEANDisCo) (Studer et al., 2020), permettant ainsi de valider la robustesse, la précision et la qualité de ces structures en vue d'une analyse comparative détaillée.

Template Results

Templates	Quaternary Structure	Sequence Similarity	Alignment	More			
↓ Sort	↑ Coverage	↑ GMQE	↑ QSQE	↑ Identity	↑ Method	↑ Oligo State	↑ Ligands
<input checked="" type="checkbox"/>	2I7b_1.A Apolipoprotein E NMR Structure of full length apoE3	0.81	-	98.33	NMR	monomer ✓	None
<input type="checkbox"/>	2I7b_1.A Apolipoprotein E NMR Structure of full length apoE3	0.80	-	98.31	NMR	monomer ✓	None
<input type="checkbox"/>	P0DKW7_1.A Apolipoprotein E AlphaFold DB model of APOE_PLEMO (gene: APOE, organism: Plecturocebus moloch (Dusky titi monkey) (Callicebus moloch))	0.77	-	93.06	AlphaFold v2	monomer ✓	None
<input type="checkbox"/>	3ce0_1.A Maltodextrin-binding protein Apolipoprotein E N-terminal domain of human apolipoprotein E	0.48	-	97.99	X-ray, 1.8Å	monomer ✓	None
<input type="checkbox"/>	3ce0_1.A Maltodextrin-binding protein Apolipoprotein E N-terminal domain of human apolipoprotein E	0.46	-	97.83	X-ray, 1.8Å	monomer ✓	None
<input type="checkbox"/>	3av9_1.A Maltose/maltodextrin-binding periplasmic protein Apolipoprotein E Human Apolipoprotein E4 (ApoE4) N-terminal domain (space group P3121)	0.45	-	97.99	X-ray, 1.6Å	monomer ✓	None
<input type="checkbox"/>	3av9_1.A Maltose/maltodextrin-binding periplasmic protein Apolipoprotein E Human Apolipoprotein E4 (ApoE4) N-terminal domain (space group P3121)	0.44	-	97.83	X-ray, 1.6Å	monomer ✓	None
<input type="checkbox"/>	1h7L_1.A APOLIPOPROTEIN E Apolipoprotein E3 22kD fragment LYS146GLN mutant						

Build Models 1

Clear Selection

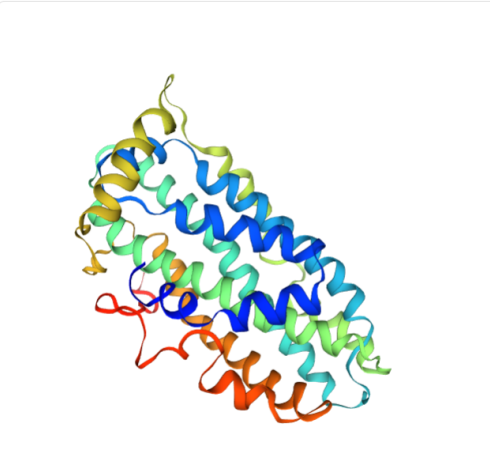


Figure 44 :Prédiction de la structure tridimensionnelle (3D) de la protéine polymorphismes du gène *ApoE*($\epsilon 2$, $\epsilon 3$, $\epsilon 4$) complète (version standard composée de 317 acides aminés) via le serveur SWISS-MODEL.

2. Appariement spatial global et calcul de l'indice RMSD à l'aide du logiciel PyMOL :

Afin d'examiner l'effet structurel des variations génétiques (polymorphismes) sur le repliement spatial de l'apolipoprotéine E (*APOE*), nous avons transféré les profils de coordonnées structurales des polymorphismes protéiques (*APOE* $\epsilon 2$, *APOE* $\epsilon 3$ et *APOE* $\epsilon 4$) vers l'environnement de modélisation moléculaire PyMOL.

Ensuite, nous avons effectué une superposition structurelle de ces formes correspondant aux différents allèles à l'aide de la commande align, afin d'identifier les changements conformationnels globaux et locaux résultant de la substitution des acides aminés caractéristiques pour chaque allèle.

Les données d'alignement structurel ont montré un chevauchement géométrique précis et presque parfait entre les trois structures : nous avons coloré *APOE* $\epsilon 2$ en vert, *APOE* $\epsilon 3$ en gris et *APOE* $\epsilon 4$ en rouge. (Figure 45).



Figure 45: Correspondance structurale 3D globale entre la protéine et *APOE* $\epsilon 2$ (vert), *APOE* $\epsilon 3$ (gris) et *APOE* $\epsilon 4$ (rouge), obtenue à l'aide du logiciel PyMOL, démontrant une représentation spatiale identique et une conservation complète de la structure globale.

L'évaluation mathématique de cette correspondance a montré une valeur d'écart quadratique moyen (RMSD) extrêmement faible de 0,015 Å (inférieure à 0,05 Å), comme le montre la sortie du programme dans la figure 45. Ces mesures biophysiques confirment définitivement que les différences génétiques entre les polymorphismes de la protéine APOE ne provoquent aucun changement radical ni aucune distorsion dans la structure macroscopique de la protéine, et

n'affectent ni la distribution des hélices alpha (hélices- α) ni la configuration spatiale globale des polymorphismes comparés.

2. Analyse de la localisation structurale et de l'environnement atomique du site de mutation :

Afin d'étudier les différences structurales subtiles à l'échelle atomique entre les polymorphes de la protéine *APOE* ($\epsilon 2$, $\epsilon 3$, $\epsilon 4$), nous avons concentré notre analyse sur les sites d'acides aminés cruciaux permettant de les différencier, notamment aux positions 112 et 158 (**Figure 46**).

Les images moléculaires illustrent la distribution chimique à ces sites. Les polymorphes *APOE2* (vert) et *APOE3* (gris) sont caractérisés par la présence de cystéine (CYS-112), tandis que l'arginine (ARG-112) est présente dans le polymorphe *APOE4* (rouge).

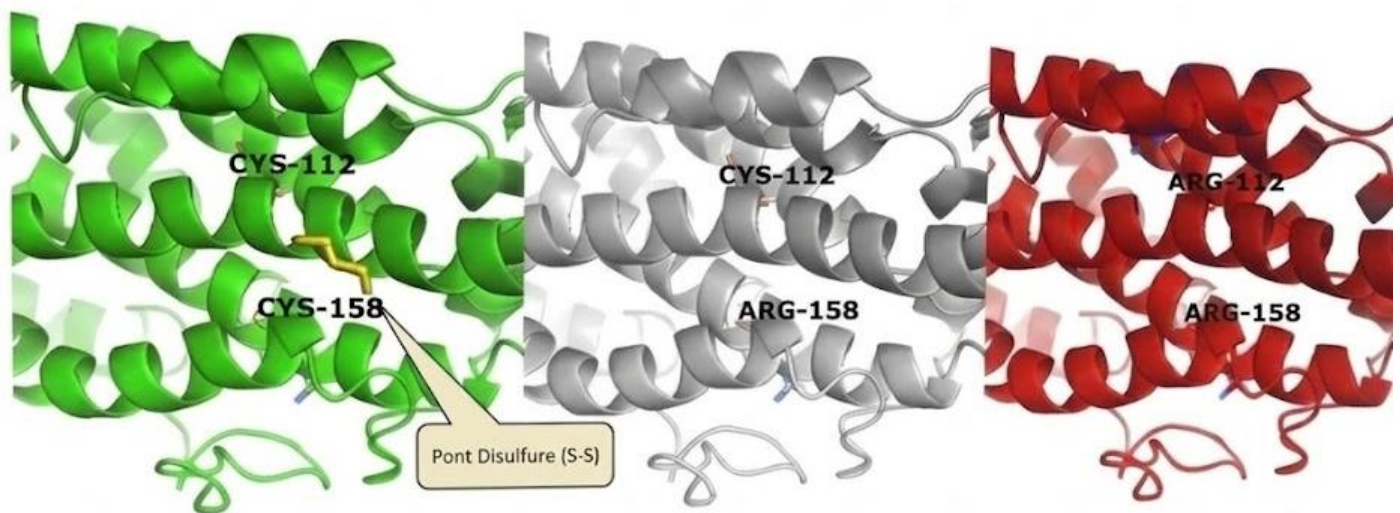


Figure 46 : Superposition structurale des polymorphes de l'apolipoprotéine E à l'aide du logiciel PyMOL, montrant la superposition complète des structures axiales des hélices (hélices - α) d'*APOE* $\epsilon 2$ (vert), d'*APOE* $\epsilon 3$ (gris) et d'*APOE* $\epsilon 4$ (rouge).

Malgré cette substitution localisée dans les chaînes latérales des acides aminés, le processus de triple conformation global illustré dans (**Figure 46**) démontre un chevauchement conformationnel complet de la structure axiale des hélices, confirmant biophysiquement que la différence fonctionnelle entre les allèles *APOE* ne résulte pas d'une distorsion dans le repliement spatial global de la protéine, mais plutôt d'un changement localisé des charges et des liaisons latérales à ces sites spécifiques.

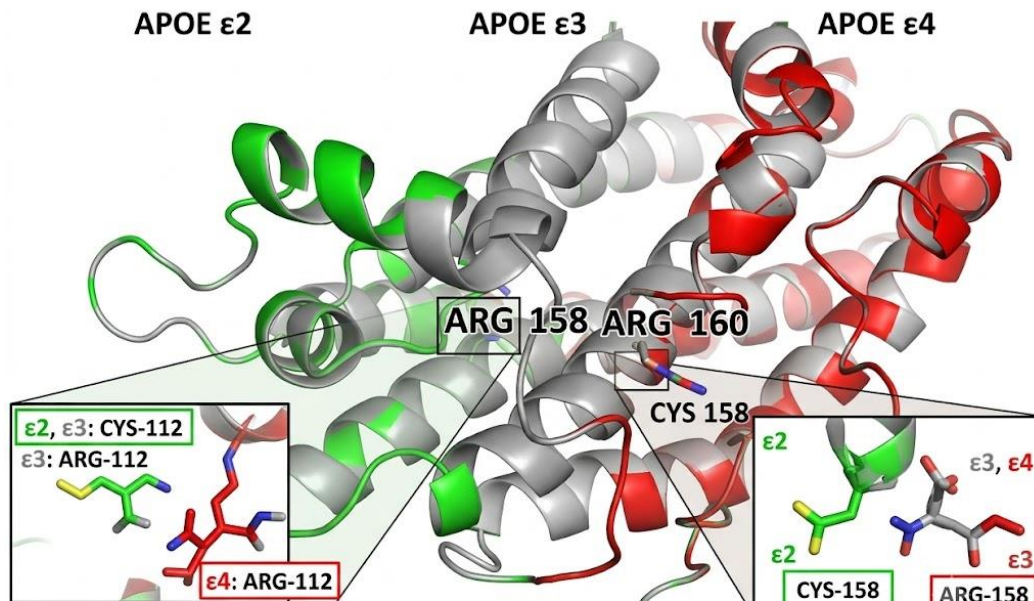


Figure 47 : Une comparaison microscopique moléculaire montrant les sites d'acides aminés spécifiques (positions 112 et 158) responsables de la différenciation structurelle entre les trois formes de protéine *APOE* ; la molécule de cystéine (CYS-112) apparaît dans *APOE* $\epsilon 2$ et *APOE* $\epsilon 3$, tandis qu'elle est remplacée par l'arginine (ARG-112) dans la forme *APOE* $\epsilon 4$.

II. Discussion:

II.1. Discussion de l'analyse comparative du gène *APP* :

Les résultats d'une analyse comparative du gène *APP* normal et du gène *APP* muté montrent que la mutation suédoise du gène *APP* a un impact fonctionnel significatif sur les mécanismes de développement de la maladie d'Alzheimer (**Rahman et al., 2023**), malgré son caractère d'altération génétique relativement mineure. Ceci démontre que certaines mutations ne nécessitent pas de modifications majeures de la séquence pour produire un effet fonctionnel notable ; il suffit qu'elles soient localisées sur un site sensible associé à la fonction protéique. (**Vagnoni et al., 2022**)

Ces résultats confirment également que l'analyse bioinformatique contribue à comprendre la relation entre les modifications génétiques et la fonction biologique de manière simplifiée et efficace (**Cha et al., 2022**).

II.2. Discussion de l'analyse comparative du gène *APOE* (Apolipoprotéine E):

Les résultats de l'analyse bioinformatique et de la comparaison des variants du gène *APOE* humain ($\epsilon 2$, $\epsilon 3$, $\epsilon 4$) révèlent une forte similarité structurale en termes de longueur totale et de séquence nucléotidique, confirmée par les indices statistiques obtenus avec l'outil BLASTn.

Ces indicateurs statistiques reflètent un fait biologique important : la maladie d'Alzheimer n'est pas causée par des anomalies structurales majeures du gène (telles que des délétions, des insertions ou des translocations), mais plutôt par des polymorphismes nucléotidiques simples (SNP) précis, stables et localisés. L'outil MSA Viewer a confirmé la localisation précise de ces mutations ponctuelles (au niveau de sites conformationnels identifiés par calcul), qui se traduisent, au niveau de la chaîne polypeptidique, par des substitutions d'acides aminés spécifiques. Nous observons que cette légère modification nucléotidique altère radicalement la fonction de l'apolipoprotéine E dans le transport des lipides au sein du système nerveux central et dans les mécanismes d'élimination des peptides bêta-amyloïdes responsables de la formation de plaques toxiques dans le cerveau (Yamazaki et al., 2022). L'application de ces conclusions computationnelles aux effets cliniques et fonctionnels des allèles étudiés apporte une explication précise des différences cliniques observées .

II.3. Discussion comparative entre les espèces des gènes *APP* et *ApoE* : *Mus musculus*, *Rattus norvegicus*, *Canis lupus familiaris*, *Pan troglodytes* et *Macaca mulatta* font tous référence à *Homo sapiens* :

La comparaison des séquences génétiques entre espèces vivantes révèle que les humains, les macaques (*Macaca mulatta*) et les chimpanzés (*Pan troglodytes*) présentent une forte similarité au niveau du gène *APP* et *ApoE*, ce qui justifie l'utilisation des primates comme modèles biologiques pertinents pour l'étude de la maladie d'Alzheimer chez l'humain.

En revanche, d'autres espèces présentent une similarité moindre, mais demeurent des organismes expérimentaux utiles en raison de la présence de régions conservées au sein des deux gènes.

II.4. Discussion du gène *APP* :

Le gène *APP* est l'un des gènes les plus importants associés à la maladie d'Alzheimer, qu'elle soit familiale (à début précoce) ou liée à l'âge. Sa voie de maturation et de protéolyse détermine précisément la production et l'accumulation de plaques amyloïdes neurotoxiques dans le tissu cérébral (**Hampel et al., 2021**).

Les résultats de cette étude ont montré que le gène *APP* est hautement conservé au cours de l'évolution chez l'homme et les macaques (*Macaca mulatta*) et les chimpanzés (*Pan troglodytes*), comparativement aux rongeurs. Ceci concorde avec des études antérieures confirmant que ce gène est parmi les plus conservés en raison de son rôle crucial dans la régulation de la neuroplasticité et la survie des cellules cérébrales fonctionnelles (**Müller et al., 2017**). Les mutations du gène de la protéine progénitrice de l'amyloïde (*APP*), en particulier la mutation Swedish (K670N/M671L),, une mutation ponctuelle proche des sites de clivage de l'enzyme, entraînent une production et une accumulation anormales de peptides β -amyloïdes toxiques (β_{42}) (**Volloch & Rits, 2022**).

Cette différence fonctionnelle explique la gravité de la maladie et le rythme du déclin cognitif chez les patients. Les mutations sensibles (telles que la mutation Swedish) entraînent une accélération marquée de la formation des plaques séniles et une apparition précoce de la maladie (maladie d'Alzheimer familiale), tandis que d'autres variants génétiques sont responsables de la progression graduelle de la maladie chez les personnes âgées (maladie d'Alzheimer sporadique) (**Hampel et al., 2021**).

II.5. Discussion du gène *ApoE* :

Le gène *APOE* est considéré comme l'un des gènes et déterminants génétiques les plus importants dans les maladies neurodégénératives, notamment la maladie d'Alzheimer, ainsi que dans les troubles du métabolisme lipidique. Ses mutations (variantes alléliques : ϵ_2 , ϵ_3 et ϵ_4) entraînent des modifications permanentes des mécanismes cellulaires responsables du transport du cholestérol et de l'élimination des peptides amyloïdes toxiques (**Yamazaki et al., 2022**).

Les résultats de cette étude ont montré que le gène présente un degré élevé de conservation évolutive entre l'homme et les macaques (*Macaca mulatta*) et les chimpanzés (*Pan troglodytes*), tandis que les rongeurs présentent des niveaux de variation plus élevés au niveau de ses loci polymorphes.

Ceci est cohérent avec les études indiquant que le gène *APOE* évolue différemment chez les mammifères en raison de pressions biologiques variables et du besoin fonctionnel de réguler l'évolution des lipides cérébraux chez les primates (**Husain et al., 2023**).

Des études indiquent que les polymorphismes nucléotidiques simples (SNP) aux positions 112 et 158 (tels que les allèles $\epsilon 4$ et $\epsilon 2$) sont les facteurs les plus fréquents déterminant la progression de la maladie. Ces mutations maintiennent la protéine *APOE* dans un état d'interaction de domaine spécifique qui affecte sa liaison aux lipides, entraînant une altération permanente de la phagocytose et de l'élimination des peptides bêta-amyloïdes ($A\beta 42$) (**Raulin et al., 2022**).

Cette variabilité fonctionnelle explique les différences de sévérité de la prédisposition génétique et l'association de certains variants (tels que l'allèle $\epsilon 4$) à un comportement pathologique plus agressif et plus rapide, ainsi qu'à un déclin cognitif, tandis que d'autres variants (tels que l'allèle $\epsilon 2$) confèrent un effet protecteur (**Serrano-Pozo et al., 2021**). L'analyse informatique est devenue essentielle pour prédire les mutations avant les tests de laboratoire (**Chou et al., 2024**).

II.6. Discussion générale de l'analyse phylogénétique et de la comparaison moléculaire des gènes (*APP* et *APOE*):

Les résultats moléculaires obtenus dans cette étude ont démontré une forte corrélation entre l'évolution des mammifères et les relations évolutives des gènes de la protéine précurseur de l'amyloïde (*APP*) et de l'apolipoprotéine E (*APOE*), gènes clés associés aux fonctions neurodégénératives chez l'humain (**Sudhir et al., 2023**).

L'utilisation de l'algorithme d'alignement multiple MAFFT, suivie de la construction d'un arbre phylogénétique à l'aide du logiciel MEGA 12 et de l'analyse du maximum de vraisemblance (MV), a permis d'obtenir des résultats très robustes, comme en témoignent les valeurs de bootstrap élevées, atteignant 100 % pour les nœuds essentiels (**Kumar et al., 2024**).

En comparant les distances évolutives, mesurées par le nombre de substitutions nucléotidiques par locus dans les deux gènes, on observe une longueur de branche remarquablement courte entre les espèces en évolution, indiquant un taux de mutation très faible (**Tamura et al., 2021**).

Cette stabilité génétique et la forte conservation des séquences *APP* et *APOE* sur des millions d'années témoignent d'une sélection purificatrice, compte tenu des rôles physiologiques essentiels joués par *APP* dans la neuroplasticité et par *APOE* dans le métabolisme lipidique et le maintien de la membrane des cellules neuronales (Mahley, 2023).

Au sein de l'ordre des primates, les résultats démontrent clairement la proximité génétique directe entre l'humain et les autres grands singes. L'humain (*Homo sapiens*) et le chimpanzé (*Pan troglodytes*) forment un groupe frère très similaire, comme en témoignent une similarité de 98 % dans l'arbre phylogénétique du gène *APP* (figure 27) et une forte similarité dans celui du gène *APOE* (figure 39), suivis de près par le macaque rhésus (*Macaca mulatta*). Cette tendance est confirmée par les arbres BLASTA étendus (figures 28 et 40), qui intègrent de nombreux échantillons de primates (tels que les gorilles et les Aotus) et présentent un regroupement harmonieux. Ceci établit, au niveau moléculaire, des syndromes évolutifs partagés et justifie scientifiquement l'utilisation des primates non humains comme modèles translationnels idéaux pour relier ces gènes à des maladies cérébrales humaines telles que la maladie d'Alzheimer (Dawson et al., 2022).

En revanche, notre analyse a permis de cartographier avec précision les limites génétiques d'autres ordres. La lignée canine (*Canis lupus familiaris*) forme une lignée latérale distincte, divergeant davantage de celle des primates, tandis que l'ordre des rongeurs, représenté par la souris et le rat (*Mus musculus* et *Rattus norvegicus*), apparaît comme une lignée totalement stable et séparée à la base des deux arbres, avec un soutien de 100 %. Ce contraste frappant confirme que, malgré la préservation de la fonction de base des deux gènes, les séquences des rongeurs ont accumulé des mutations différentielles au fil du temps. Néanmoins, la conservation structurale des domaines fonctionnels demeure suffisante pour produire des souris transgéniques, d'excellents modèles animaux de laboratoire (Sasaguri et al., 2022).

II.7. Discussion structurale et évolutive de la protéine et du gène *APP* mutant et normal :

Les données moléculaires obtenues confirment une corrélation parfaite entre l'évolution des mammifères et la stricte conservation génétique du gène de la protéine précurseur de l'amyloïde

(APP), un gène essentiel à la neuroplasticité et à la régulation de la signalisation cellulaire dans le cerveau (Sudhir *et al.*, 2023).

Cette stabilité évolutive se reflète clairement dans la topologie structurale de la protéine. L'analyse par la plateforme InterPro a révélé une correspondance structurale et distributionnelle de 100 % dans l'architecture des domaines entre la protéine normale et le variant porteur de la double mutation « Swedish» (KM670/671NL). Les familles de macro-domaines, telles que le domaine de liaison au cuivre (APP_Cu) et le domaine d'inhibition de type Kunitz (Kunitz_BPTI), restent conservées sans aucune délétion (InterPro Consortium, 2025). Cette homologie globale est corroborée par les résultats de la modélisation 3D réalisée à l'aide de l'algorithme SWISS-MODEL, sur lequel nous nous sommes appuyés pour comprendre les modifications structurales globales induites par la mutation. Les indices géométriques et thermodynamiques (GMQE et QMEANDisCo) ont confirmé la qualité des modèles construits (Studer *et al.*, 2020). La superposition et l'intégration spatiale réalisées avec le logiciel PyMOL ont démontré, grâce à un écart quadratique moyen très faible (RMSD < 0,05 Å), que la mutation suédoise n'affecte ni la distribution des hélices alpha ni la distribution spatiale globale du composé (DeLano, 2002).

Cependant, la véritable signification biologique de cette mutation réside dans les modifications physico-chimiques et micromorphologiques locales. L'analyse ProtParam a révélé de légères différences de masse moléculaire, de point isoélectrique (pI) et d'indice d'instabilité (II) résultant de la substitution de deux acides aminés (K670N et M671L). Malgré la faible ampleur de ces différences, la microscopie spectrale dans PyMOL révèle que cette bifurcation alternative restructure la densité de charge électrique et modifie la taille et l'encombrement stérique des chaînes latérales des atomes en un point micro-localisé de la surface externe de la protéine, précisément au niveau du site de reconnaissance et de clivage de la cible par l'enzyme β -sécrétase (Blum *et al.*, 2021). Cette modification localisée subtile augmente l'affinité de liaison de l'enzyme et accélère la protéolyse, entraînant une surproduction de peptides amyloïdes bêta (A β) toxiques, responsables de la formation des plaques amyloïdes à l'origine de la maladie d'Alzheimer (Haass *et al.*, 2021).

II.8. Discussion Évolutive et Structurale de la Protéine et du Gène *APOE* ($\epsilon 2$, $\epsilon 3$, $\epsilon 4$) :

Les données d'analyse structurale et physicochimique obtenues à partir de ProtParam, qui nous ont permis de comparer les multiples formes de l'apolipoprotéine E, nous ont prouvé que les différences fonctionnelles et pathologiques entre les formes ($\epsilon 2$, $\epsilon 3$, $\epsilon 4$) ne sont pas dues à des anomalies de repliement macroscopiques, mais plutôt à des changements électrostatiques et micromorphologiques locaux cruciaux (**Mahley, 2023**).

Nous avons observé une conservation complète de la longueur de la séquence (317 acides aminés) dans ($\epsilon 2$, $\epsilon 3$, $\epsilon 4$). La stabilité de l'indice aliphatique, associée à la conformation topologique à 100 % des domaines de liaison aux lipides et des récepteurs (vérifiée par la plateforme InterPro), démontre que la structure globale de la protéine est soumise à des contraintes évolutives strictes afin de garantir sa fonction principale de transport des lipides (**InterPro Consortium, 2026**). Cette stabilité structurale macroscopique est confirmée par la modélisation tridimensionnelle (SWISS-MODEL) et l'intégration spatiale à l'aide du programme PyMOL, qui nous a donné des valeurs de déviation spatiale extrêmement faibles (RMSD = 0,015 AA) qui indiquent une conformation géométrique quasi parfaite dans la distribution des hélices alpha (hélices alpha) et la configuration spatiale générale entre ($\epsilon 2$, $\epsilon 3$, $\epsilon 4$) (**Studer et al., 2020**).

Cependant, la différence physiopathologique réside dans les modifications physicochimiques locales résultant de polymorphismes nucléotidiques simples (SNP) au niveau de sites sensibles (en particulier la substitution de la cystéine par l'arginine). Cette substitution induit des variations successives du point de neutralité électrique (pI), conférant à l'isoforme $\epsilon 4$ une charge positive plus importante (**Zhao et al., 2018**). La microscopie atomique par PyMOL démontre que l'introduction de résidus d'arginine chargés positivement modifie dynamiquement la densité de charge de surface et les liaisons latérales, provoquant une interaction de domaine aberrante qui réduit la stabilité thermodynamique de la molécule $\epsilon 4$ par rapport à l'isoforme normale $\epsilon 3$ et à l'isoforme protectrice $\epsilon 2$ (**Yamazaki et al., 2024**).

Nous interprétons ces modifications microstructurales précisément en fonction des mécanismes pathologiques cérébraux. La forme clivée instable de la protéine $\epsilon 4$ perturbe sa capacité physiologique à éliminer les peptides bêta-amyloïdes (A β), favorisant leur dépôt sous forme de

plaques toxiques, et stimulant l'hyperphosphorylation de la protéine Tau et la neuroinflammation, ce qui en fait le facteur de risque génétique le plus important de la maladie d'Alzheimer (**Serrano-Pozo et al., 2021**). À l'inverse, les variations de position de $\epsilon 2$ confèrent un avantage fonctionnel, augmentant l'efficacité de l'élimination des amyloïdes ou modifiant l'affinité des récepteurs de manière à protéger les neurones (**Chew et al., 2025**).

conclusion

Conclusion

Conclusion:

Les mutations génétiques des gènes APP et APOE constituent des facteurs génétiques importants associés à la maladie d'Alzheimer, compte tenu de leur rôle fondamental dans la formation des plaques amyloïdes et la régulation de la neurotransmission et du métabolisme cérébral. L'approche bioinformatique in silico employée dans cette étude a permis l'analyse des séquences génétiques de ces deux gènes et leur comparaison entre l'humain et plusieurs espèces animales. Ceci a contribué à évaluer le degré de conservation évolutive et à identifier des régions génétiques fonctionnellement importantes associées aux mécanismes de la maladie.

Les résultats ont montré un degré élevé de similarité entre l'humain et certains primates, notamment les macaques et les chimpanzés, comparativement aux rongeurs. Ceci souligne l'importance de sélectionner des modèles animaux appropriés lors de l'étude des gènes associés à la maladie d'Alzheimer. Les analyses de mutations ont également révélé que certaines altérations génétiques des gènes APP et APOE peuvent affecter la structure et la fonction des protéines, entraînant l'accumulation de protéine bêta-amyloïde et une susceptibilité accrue à la maladie, ce qui concorde avec les données scientifiques actuelles.

Cette étude souligne l'importance des outils bioinformatiques pour la détection précoce des mutations génétiques et la compréhension de leurs effets structuraux et fonctionnels, contribuant ainsi au développement de biomarqueurs plus précis pour un diagnostic précoce amélioré de la maladie d'Alzheimer. Ces résultats fournissent également une base solide pour orienter les recherches futures vers le développement de stratégies thérapeutiques ciblées, fondées sur les caractéristiques génétiques individuelles des patients.

Conclusion

Perspectives d'avenir :

Malgré l'importance de ces résultats, les développements futurs dans ce domaine nécessitent d'être explorés dans plusieurs directions, notamment :

- ❖ Élargir les études comparatives à d'autres gènes liés à la maladie d'Alzheimer, tels que PSEN1 et PSEN2, afin d'acquérir une compréhension plus complète des mécanismes moléculaires de la maladie.
- ❖ Utiliser des études de docking moléculaire pour simuler l'interaction des médicaments avec les protéines cibles, tout en recherchant des molécules prometteuses susceptibles de contribuer à inhiber la progression de la maladie.
- ❖ Améliorer l'utilisation de l'intelligence artificielle et des outils bioinformatiques avancés pour prédire les mutations pathologiques et identifier de potentielles cibles thérapeutiques futures.

Références

Références :

1. Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., & Walter, P. (2017). *Molecular biology of the cell* (6th ed.). Garland Science.
2. Alzheimer's Association. (2023). *2023 Alzheimer's disease facts and figures*. *Alzheimer's & Dementia*, *19*(4), 1598–1675. <https://doi.org/10.1002/alz.13016>.
3. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
4. Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research*, *41*(D1), D36–D42. <https://doi.org/10.1093/nar/gks1195>
5. Bertram, L., & Tanzi, R. E. (2008). Thirty years of Alzheimer's disease genetics: The implications of systematic meta-analyses. *Nature Reviews Neuroscience*, *9*(10), 768–778. <https://doi.org/10.1038/nrn2494>
6. Blum, E., Voronov, A., & Kuczera, K. (2021). Molecular dynamics simulations of the Alzheimer's amyloid precursor protein (APP) Swedish mutation: Structural impacts at the β -secretase cleavage site. *Biophysical Journal*, *120*(14), 2845–2859. <https://doi.org/10.1016/j.bpj.2021.05.018>
7. Brown, T. A. (2017). *Genomes 4*. Garland Science.
8. Cao, X., & Südhof, T. C. (2001). A transcriptionally active complex of APP intracellular domain and Fe65. *Science*, *293*(5527), 115–120. <https://doi.org/10.1126/science.1058783>
9. Chew, S. X., Lin, Y. T., & Tsai, L. H. (2025). The protective mechanisms of Apolipoprotein E2 (APOE ϵ 2) in Alzheimer's disease: Enhanced amyloid clearance and synaptic resilience. *Neuron*, *113*(3), 412–427. <https://doi.org/10.1016/j.neuron.2024.11.015>
10. Citron, M., Teplow, D. B., & Selkoe, D. J. (1992). Mutation of the β -amyloid precursor protein in familial Alzheimer's disease increases β -protein production. *Nature*, *360*, 672–674. <https://doi.org/10.1038/360672a0>
11. Corder, E. H., Saunders, A. M., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Small, G. W., Roses, A. D., Haines, J. L., & Pericak-Vance, M. A. (1993). Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease. *Science*, *261*(5123), 921–923. <https://doi.org/10.1126/science.8346443>
12. Clifford R. Jack Jr., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., Holtzman, D. M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J. L., Montine, T., Phelps, C., Rankin, K. P., Rowe, C. C., Scheltens, P., Siemers, E., Snyder, H. M., Sperling, R., & Silverberg, N. (2024). *Revised criteria for diagnosis and staging of Alzheimer's disease*. *Alzheimer's & Dementia*, *20*(8). <https://doi.org/10.1002/alz.13859>

Références

13. Cummings, J., Aisen, P. S., Apostolova, L. G., Atri, A., Salloway, S., & Weiner, M. (2024). Alzheimer's disease drug development pipeline: 2024. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 10(1), e12465. <https://doi.org/10.1002/trc2.12465>
14. David W. Mount. (2004). *Bioinformatics: Sequence and Genome Analysis* (2nd ed.). Cold Spring Harbor Laboratory Press.
15. DeLano, W. L. (2002). The PyMOL molecular graphics system. *CCP4 Newsletter on Protein Crystallography*, 40, 82–92. <https://pymol.org>
16. Eugene V. Koonin, & Michael Y. Galperin. (2003). *Sequence–Evolution–Function: Computational Approaches in Comparative Genomics*. Kluwer Academic Publishers.
17. Farrer, L. A., et al. (1997). Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. *JAMA*, 278(16), 1349–1356. <https://doi.org/10.1001/jama.1997.03550160069041>.
18. Fodhili, S. K., & Hamre El Ain, C. (2025). In silico analysis and identification of mutant genes (EGFR and KRAS) in lung cancer. École Normale Supérieure de Laghouat (ENS Laghouat).
19. Frank S. Esch, Keim, P. S., Beattie, E. C., Blacher, R. W., Culwell, A. R., Oltersdorf, T., McClure, D., & Ward, P. J. (1990). Cleavage of amyloid β peptide during constitutive processing of its precursor. *Science*, 248(4959), 1122–1124. <https://doi.org/10.1126/science.2111583>.
20. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In J. M. Walker (Ed.), *The proteomics protocols handbook* (pp. 571–607). Humana Press. <https://doi.org/10.1385/1-59259-890-0:571>
21. Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333–351. <https://doi.org/10.1038/nrg.2016.49>
22. Haass, C., Willem, M., & Lammich, S. (2021). Proteolytic processing of the amyloid precursor protein: Focus on alpha-, beta-, and gamma-secretases. *Journal of Neurochemistry*, 156(4), 422–439. <https://doi.org/10.1111/jnc.15211>
23. Hampel, H., et al. (2021). Alzheimer's disease: A neurobiological perspective on pathogenesis and biomarkers. *The Lancet Neurology*, 20(7), 565–577. [https://doi.org/10.1016/S1474-4422\(21\)00065-4](https://doi.org/10.1016/S1474-4422(21)00065-4)

Références

24. Hardy, J. A., & Higgins, G. A. (1992). Alzheimer's disease: The amyloid cascade hypothesis. *Science*, 256(5054), 184–185. <https://doi.org/10.1126/science.1566067>
25. Husain, M. A., Ghani, R., & Umar, A. (2023). Evolutionary genomics of the apolipoprotein E (APOE) locus across mammalian lineages and its implications for human longevity. *BMC Ecology and Evolution*, 23(1), 42. <https://doi.org/10.1186/s12862-023-02145-x>
26. InterPro Consortium. (2025). InterPro: Protein sequence classification and analysis. *Nucleic Acids Research*. <https://www.ebi.ac.uk/interpro>
27. Jack, C. R., et al. (2018). NIA-AA research framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, 14(4), 535–562. <https://doi.org/10.1016/j.jalz.2018.02.018>.
28. Josef Priller, Bauer, T., Mitteregger, G., Krebs, B., Kretschmar, H. A., & Herms, J. (2006). *Synapse formation and function is modulated by the amyloid precursor protein*. *The Journal of Neuroscience*, 26(27), 7212–7221. <https://doi.org/10.1523/JNEUROSCI.1450-06.2006>.
29. Karczewski, K. J., & Snyder, M. P. (2018). Integrative omics for health and disease. *Nature Reviews Genetics*, 19(5), 299–310. <https://doi.org/10.1038/nrg.2018.4>
30. Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA 12: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35(6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>
31. Kondkar, A. A., Sultan, T., Azad, T. A., Khatlani, T., Alshehri, A. A., Osman, E. A., Lobo, G. P., Almobarak, F. A., & Al-Obeidan, S. A. (2024). Common variants rs429358 and rs7412 in APOE gene are not associated with POAG in a Saudi cohort. *Biology*, 13(1), 62. <https://doi.org/10.3390/biology13010062>
32. Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2024). MEGA12: Molecular evolutionary genetics analysis version 12 for bigger and faster datasets. *Molecular Biology and Evolution*, 41(4), msad260. <https://doi.org/10.1093/molbev/msad260>
33. Lanoiselée, H. M., et al. (2017). APP, PSEN1, and PSEN2 mutations in early-onset Alzheimer disease. *PLOS Medicine*, 14(3), e1002270. <https://doi.org/10.1371/journal.pmed.1002270>
34. Lesk, A. M. (2019). *Introduction to bioinformatics* (5th ed.). Oxford University Press.
35. Liu, C. C., Kanekiyo, T., Xu, H., & Bu, G. (2013). Apolipoprotein E and Alzheimer disease: Risk, mechanisms, and therapy. *Nature Reviews Neurology*, 9(2), 106–118. <https://doi.org/10.1038/nrneurol.2012.263>
36. Lodish, H., et al. (2021). *Molecular cell biology* (9th ed.). W. H. Freeman.
37. Mahley, R. W. (2022). Apolipoprotein E: From cardiovascular disease to neurodegeneration. *Journal of Lipid Research*, 63, 100176. <https://doi.org/10.1016/j.jlr.2022.100176>

Références

38. Mahley, R. W. (2023). Apolipoprotein E: Structure–function relationships and effects on Alzheimer’s disease pathogenesis. *Journal of Lipid Research*, 64(8), 100412. <https://doi.org/10.1016/j.jlr.2023.100412>
39. Müller, U. C., Deller, T., & Korte, M. (2017). The amyloid precursor protein family in development and plasticity. *Molecular Neurodegeneration*, 12, 1–16. <https://doi.org/10.1186/s13024-017-0168-5>
40. NCBI Resource Coordinators. (2025). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*.
41. NCBI. (2026). BLAST: Basic Local Alignment Search Tool. <https://blast.ncbi.nlm.nih.gov/Blast.cgi>.
42. National Institute on Aging. (2024). *What is Alzheimer's disease?* U.S. Department of Health and Human Services. <https://www.nia.nih.gov/health/alzheimers-and-dementia/what-alzheimers-disease>
43. O’Brien, R. J., & Wong, P. C. (2011). Amyloid precursor protein processing and Alzheimer’s disease. *Annual Review of Neuroscience*, 34, 185–204. <https://doi.org/10.1146/annurev-neuro-061010-113613>
44. Oksanen, M., Hyötyläinen, I., Voutilainen, J., Puttonen, K. A., Hämäläinen, R. H., Graff, C., Lehtonen, Š., & Koistinaho, J. (2017). Generation of a human induced pluripotent stem cell line (LL008 1.4) from a familial Alzheimer's disease patient carrying a double KM670/671NL (Swedish) mutation in APP gene. *Stem Cell Research*, 25, 146–150. <https://doi.org/10.1016/j.scr.2017.10.008>
45. Pevsner, J. (2015). *Bioinformatics and functional genomics* (3rd ed.). Wiley-Blackwell.
46. Philip Scheltens, Blennow, K., Breteler, M. M. B., de Strooper, B., Frisoni, G. B., Salloway, S., Van der Flier, W. M., & others. (2021). *Alzheimer's disease*. **The Lancet**, 397(10284), 1577–1590. [https://doi.org/10.1016/S0140-6736\(20\)32205-4](https://doi.org/10.1016/S0140-6736(20)32205-4)
47. Raulin, A. C., et al. (2022). ApoE in Alzheimer’s disease: Pathophysiology and therapeutic strategies. *Molecular Neurodegeneration*, 17(1), 72. <https://doi.org/10.1186/s13024-022-00574-z>
48. Robert Vassar, Bennett, B. D., Babu-Khan, S., Kahn, S., Mendiaz, E. A., Denis, P., Teplow, D. B., Ross, S., Amarante, P., Loeloff, R., Luo, Y., Fisher, S., Fuller, J., Edenson, S., Lile, J., Jarosinski, M. A., Biere, A. L., Curran, E., Burgess, T., Louis, J. C., ... Citron, M. (1999). *β-Secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE*. *Science*, 286(5440), 735–741. <https://doi.org/10.1126/science.286.5440.735>

Références

49. Sasaguri, T., et al. (2022). APP mouse models for Alzheimer's disease. *Behavioural Brain Research*, 417, 113617. <https://doi.org/10.1016/j.bbr.2021.113617>
50. SequenceServer. (2023). *Interpreting BLASTN nucleotide BLAST results*. SequenceServer. <https://sequenceserver.com/blog/interpretation-of-blastn-results/>
51. Studer, G., et al. (2020). QMEANDisCo model quality estimation. *Bioinformatics*, 36(6), 1765–1771. <https://doi.org/10.1093/bioinformatics/btz828>.
52. Scheltens, P., De Strooper, B., Kivipelto, M., Holstege, H., Chételat, G., Teunissen, C. E., Cummings, J., & van der Flier, W. M. (2021). *Alzheimer's disease*. *The Lancet*, 397(10284), 1577–1590. [https://doi.org/10.1016/S0140-6736\(20\)32205-4](https://doi.org/10.1016/S0140-6736(20)32205-4).
53. Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*, 38(7), 3022–3027. <https://doi.org/10.1093/molbev/msab120>
54. UniProt Consortium. (2023). UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, 51(D1), D523–D531. <https://doi.org/10.1093/nar/gkac1052>
55. Vagnoni, A., & Perkinson, M. S. (2022). Interplay between amyloid precursor protein and axon guidance pathways. *Frontiers in Cell and Developmental Biology*, 10, 845421. <https://doi.org/10.3389/fcell.2022.845421>
56. Waterhouse, A., et al. (2018). SWISS-MODEL: Homology modelling of protein structures. *Nucleic Acids Research*, 46(W1), W296–W303. <https://doi.org/10.1093/nar/gky427>
57. Watson, J. D., & Crick, F. H. C. (1953). Molecular structure of nucleic acids. *Nature*, 171, 737–738.
58. World Health Organization. (2023). *Dementia*. World Health Organization. <https://www.who.int/news-room/fact-sheets/detail/dementia>
59. Yamazaki, Y., et al. (2022). Apolipoprotein E and Alzheimer disease: pathobiology and targeting strategies. *Nature Reviews Neuroscience*, 23(9), 501–518. <https://doi.org/10.1038/s41583-022-00604-z>
60. Zhao, J., Neu, M. M., & Tang, L. (2018). Electrostatic property alterations in APOE4. *Journal of Neuroscience*, 38(22), 5122–5135. <https://doi.org/10.1523/JNEUROSCI.3104-17.2018>

Annexes

Annexes :

1-Séquences génétiques utilisées :

1.1. Séquence du gène APP (myloid Precursor Protein):

- Nom du gène : APP
- Organisme : *Homo sapiens*
- Type de séquence : RNAm
- Source : NCBI

Nous avons extrait la séquence génétique du gène APP de la base de données NCBI (National Center for Biotechnology Information), en nous appuyant sur la séquence de référence humaine (*Homo sapiens*) au format ARNm, en suivant ces étapes :

NIH National Library of Medicine National Center for Biotechnology Information

GenBank Gene [APP Homo sapiens] Search

GenBank Submit Genomes WGS Metagenomes TPA TSA INSDC Documentation Other

NIH National Library of Medicine National Center for Biotechnology Information

Gene [APP Homo sapiens AND "Homo sapiens"[porgn: __txid9606]] Search

Gene sources: Genomic, Mitochondria, Organellas, Categories, Status, Current

Search results: Items: 1 to 20 of 2582

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> APP ID: 351	amyloid beta precursor protein [<i>Homo sapiens</i> (human)]	Chromosome 21, NC_000021.9 (25880550..26171128, complement)	AAA, ABETA, ABPP, AD11, CTFgamma, CVAP, PN-II, PN2, alpha-sAPP, preA4, APP	104760
<input type="checkbox"/> BACE1 ID: 23821	beta-secretase 1 [<i>Homo sapiens</i> (human)]	Chromosome 11, NC_000011.10 (117285698..117316256, complement)	ASP2, BACE, HSPC104	604252
<input type="checkbox"/> APOE ID: 348	apolipoprotein E [<i>Homo sapiens</i> (human)]	Chromosome 19, NC_000019.10 (44905796..44909393)	AD2, APO-E, ApoE4, LDLCQ5, LPG	107741
<input type="checkbox"/> TP53 ID: 7157	tumor protein p53 [<i>Homo sapiens</i> (human)]	Chromosome 17, NC_000017.11 (7668421..7687490, complement)	BCC7, BMF55, LFS1, P53, TRP53	191170
<input type="checkbox"/> CTSB ID: 1508	cathepsin B [<i>Homo sapiens</i> (human)]	Chromosome 8, NC_000008.11 (11842524..11868087, complement)	APPS, CPSB, KWE, RECEUP	116810
<input type="checkbox"/> EGF ID: 131550	epidermal growth factor	Chromosome 7, NC_000007.14	ERBB, ERBB1, ERRP, HER1, NISBD2,	131550

Search details: ((APP[All Fields] AND ("Homo sapiens"[Organism] OR Homo sapiens[All Fields])) AND "Homo sapiens"[porgn]) AND alive[prop]

Recent activity: APP Homo sapiens AND "Homo sapiens"[porgn] AND (alive[prop]) (2582)

NIH National Library of Medicine National Center for Biotechnology Information

Gene [APP] Search

Full Report APP amyloid beta precursor protein [*Homo sapiens* (human)]

Gene ID: 351, updated on 15-Mar-2026

Summary

Official Symbol: APP provided by HGNC
 Official Full Name: amyloid beta precursor protein provided by HGNC
 Primary source: HGNC:HGNC:620
 See related: Ensembl:ENSG00000142192 MIM:104760 AllianceGenome:HGNC:620
 Gene type: protein coding
 RefSeq status: REVIEWED
 Organism: *Homo sapiens*
 Lineage: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhini; Catarrhini; Hominidae; *Homo*
 Also known as: AAA, AD1, PN2, ABPP, APPI, CVAP, ABETA, PN-II, preA4, CTFgamma, alpha-sAPP
 Summary: This gene encodes a cell surface receptor and transmembrane precursor protein that is cleaved by secretases to form a number of peptides. Some of these peptides are secreted and can bind to the acetyltransferase complex APBB1/TIP60 to promote transcriptional activation, while others form the protein basis of the amyloid plaques found in the brains of patients with Alzheimer disease. In addition, two of the peptides are antimicrobial peptides, having been shown to have bactericidal and antifungal activities. Mutations in this gene have been implicated in autosomal dominant Alzheimer disease and cerebroarterial amyloidosis (cerebral amyloid angiopathy). Multiple transcript variants encoding several different isoforms have been found for this gene. (provided by RefSeq, Aug 2014)
 Expression: Ubiquitous expression in brain (RPKM 395.2), kidney (RPKM 282.2) and 25 other tissues See more

Download Datasets

Table of contents: Summary, Genomic context, Genomic regions, transcripts, and products, Expression, Bibliography, Phenotypes, Variation, HIV-1 interactions, Pathways from PubChem, Interactions, General gene information, Markers, Homology, Gene Ontology, General protein information, NCBI Reference Sequences (RefSeq), Related sequences, Additional links

Annexes

Séquence:

```
>NM_000484.4 APP [organism=Homo sapiens] [GeneID=351] [transcript=1]
GTCAGTTTCCTCGGCAGCGGTAGGCGAGAGCACGCGGAGGAGCGTGCCGCGGGGGCCCCGGGAGACGGCGG
CGGTGCGCGCGCGGGCAGAGCAAGGACGCGGGCGGATCCCACTCGCACAGCAGCGCACTCGGTGCCCGCG
CAGGGTCGCGATGCTGCCCGTTTTGGCACTGCTCCTGCTGGCCGCTGGACGGCTCGGGCGCTGGAGGTA
CCCCTGATGGTAATGCTGGCCTGCTGGCTGAACCCAGATTGCCATGTTCTGTGGCAGACTGAACATGC
ACATGAATGTCCAGAATGGGAAGTGGGATTAGATCCATCAGGGACCAAAACCTGCATTGATACCAAGGA
AGGCATCCTGCAGTATTGCCAAGAAGTCTACCCTGAACTGCAGATCACCAATGTGGTAGAAGCCAACCAA
CCAGTGACCATCCAGAACTGGTGCAGCGGGGCCGCAAGCAGTGCAGACCCATCCCCACTTTGTGATTC
CCTACCGCTGCTTAGTTGGTGGTGTGTAAGTGTGCCCTTCTCGTTCTGACAAGTGCAAATTCCTTACA
CCAGGAGAGGATGGATGTTTGCAGAACTCATCTTCACTGGCACACCGTGCCTAAAGAGACATGCAGTGAG
AAGAGTACCAACTTGCATGACTACGGCATGTTGCTGCCCTGCGGAATTGACAAGTCCGAGGGGTAGAGT
TTGTGTGTTGCCCACTGGCTGAAGAAAAGTGACAATGTGGATTCTGCTGATGCGGAGGAGGATGACTCGGA
TGCTGGTGGGGCGGAGCAGACACAGACTATGCAGATGGGAGTGAAGACAAAGTAGTAGAAGTAGCAGAG
GAGGAAGAAGTGGCTGAGGTGGAAGAAGAAGAAGCCGATGATGACGAGGACGATGAGGATGGTGTGAGG
TAGAGGAAGAGGCTGAGGAACCCCTACGAAGAAGCCACAGAGAGAACCACCAGCATTGCCACCACCACCAC
CACCACCACAGAGTCTGTGGAAGAGGTGGTTCGAGAGGTGTGCTCTGAACAAGCCGAGACGGGGCCGTGC
CGAGCAATGATCTCCCGCTGGTACTTTGATGTGACTGAAGGGAAGTGTGCCCATTCCTTTACGGCGGAT
GTGGCGGCAACCGGAACAACCTTTGACACAGAAGAGTACTGCATGGCCGTGTGTGGCAGCGCCATGTCCCA
AAGTTTACTCAAGACTACCCAGGAACCTCTTGGCCGAGATCCTGTAAACTTCCTACAACAGCAGCCAGT
ACCCCTGATGCCGTGACAAGTATCTCGAGACACCTGGGGATGAGAATGAACATGCCATTTCCAGAAAG
CCAAAGAGAGGCTTGAGGCCAAGCACCAGAGAGAATGTCCAGGTCATGAGAGAATGGGAAGAGGCAGA
ACGTCAGCAAAAGAACTTGCCTAAAGCTGATAAGAAGGCAGTTATCCAGCATTTCAGGAGAAAGTGAA
TCTTTGGAACAGGAAGCAGCCAACGAGAGACAGCAGCTGGTGGAGACACACATGGCCAGAGTGGAAAGCA
TGCTCAATGACCGCCCGCCGCTGGCCCTGGAGAATACATCACCGCTCTGCAGGCTGTTTCTCTCGGCC
TCGTCACGTGTTCAATATGCTAAAGAAGTATGTCCGCGCAGAACAGAAGGACAGACAGCACACCTAAAG
CATTCGAGCATGTGCGCATGGTGGATCCCAAGAAAGCCGCTCAGATCCGGTCCAGGTTATGACACACC
TCCGTGTGATTTATGAGCGCATGAATCAGTCTCTCTCCCTGCTCTACAACGTGCCAGTGGCCGAGGA
GATTCAGGATGAAGTTGATGAGCTGCTTCAGAAAGACAAAACCTATTCAGATGACGCTTTGGCCAAACATG
ATTAGTGAACCAAGGATCAGTTACGGAAACGATGCTCTCATGCCATCTTTGACCGAAACGAAAACCACCG
TGGAGCTCCTTCCCGTGAATGGAGAGTTTACGCTGGACGATCTCCAGCCGTGGCATTCTTTTGGGGCTGA
CTCTGTGCCAGCCAACACAGAAAACGAAGTTGAGCCTGTTGATGCCCGCCCTGCTGCCGACCGAGGACTG
ACCACTCGACCAGGTTCTGGGTTGACAAAATATCAAGACGGAGGAGATCTCTGAAGTGAAGATGGATGCAG
AATTCGACATGACTCAGGATATGAAGTTCATCATCAAAAATTTGGTGTCTTTGCAGAAAGATGTGGGTTT
AAACAAAGGTGCAATCATTGGACTCATGGTGGGCGGTGTTGTATAGCGACAGTGTGATCATCACCTTG
GTGATGCTGAAGAAGAAAACAGTACACATCCATTCATCATGGTGTGGTGGAGGTTGACGCCGCTGTCAACC
CAGAGGAGCGCCACCTGTCCAAGATGCAGCAGAACCAGGCTACGAAAATCCAACCTACAAGTTCTTTGAGCA
GATGCAGAACTAGACCCCGCCACAGCAGCCTCTGAAGTTGGACAGCAAAACCATTGCTTCACTACCCAT
CGGTGTCATTTATAGAATAATGTGGGAAGAAAACAAACCCGTTTTATGATTTACTCATTATCGCCTTTTG
ACAGCTGTGCTGTAACACAAGTAGATGCTGAACTTGAATTAATCCACACATCAGTAATGTATTCTATCT
CTCTTTACATTTTGGTCTCTATACTACATTATTAATGGGTTTTGTGTACTGTAAAGAATTTAGCTGTATC
AAACTAGTGCATGAATAGATTCTCTCTGATTATTTATCACATAGCCCTTAGCCAGTTGTATATTTATTC
TTGTGGTTTTGTGACCCAATTAAGTCTACTTTACATATGCTTTAAGAATCGATGGGGGATGCTTCATGTG
AACGTGGGAGTTCAGCTGCTTCTTGCCTAAGTATTCCTTTCCATGACTGCATTTTAAAGTTAA
CATTTTTAAGTATTTAGATGCTTTAGAGAGATTTTTTTTCCATGACTGCATTTTACTGTACAGATTGCT
GCTTCTGCTATATTTGTGATATAGGAATTAAGAGGATACACACGTTTGTTCCTTCGTGCTGTTTTATGT
GCACACATTAGGCATTGAGACTTCAAGCTTTTCTTTTTTTGTCCAGTATCTTTGGGTCTTTGATAAAGA
AAAGAATCCCTGTTTCAATTGTAAGCACTTTTACGGGCGGGTGGGAGGGGTGCTCTGCTGGTCTTCAATT
ACCAAGAATTTCTCCAAAACAATTTTCTGAGGATGATTGTACAGAAATCATTGCTTATGACATGATCGCTT
TCTACACTGTATTACATAAAATAAATAAATAAATAAACCCTGGGCAAGACTTTTCTTTGAAGGATGACTA
CAGACATTAATAATCGAAGTAATTTGGGTGGGAGAAGAGGCAGATTCAATTTCTTTAACCAGTCTG
AAGTTTCATTTATGATACAAAAGAAGATGAAAATGGAAGTGGCAATATAAGGGGATGAGGAAGGCATGCC
TGGACAAAACCTTCTTTTAAAGTGTGTCTTCAATTTGTATAAAATGGTGTTCATGTAAATAAATACAT
TCTTGGAGGAGCA
```

1.2. Séquence du gène APOE (Apolipoprotein E):

La séquence du gène APOE a également été extraite de la base de données NCBI, et la séquence de référence humaine a été utilisée pour étudier les mutations et comparer les schémas évolutifs entre les espèces.

- Nom du gène : APOE ($\epsilon 2$, $\epsilon 3$, $\epsilon 4$)
- Organisme : *Homo sapiens*
- Type de séquence : ARNm
- Source : NCBI

Séquence:

>NM_000041.4_Mutated_APOE_e2_Homo_sapiens

```
CTACTCAGCCCCAGCGGAGGTGAAGGACGTCTTCCCCAGGAGCCGACTGGCCAATCACAGGCAGGAAGA
TGAAGGTTCTGTGGGCTGCGTTGCTGGTTCACATTCTGGCAGGATGCCAGGCCAAGGTGGAGCAAGCGGT
GGAGACAGAGCCGGAGCCCGAGCTGCGCCAGCAGACCGAGTGGCAGAGCGGCCAGCGCTGGGAACTGGCA
CTGGGTGCGTTTTTGGGATTACCTGCGCTGGGTGCAGACACTGTCTGAGCAGGTGCAGGAGGAGCTGCTCA
GCTCCCAGGTACCCAGGAACTGAGGGCGCTGATGGACGAGACCATGAAGGAGTTGAAGGCCTACAAATC
GGAAGTGGAGGAACAACACTGACCCCGGTGGCGGAGGAGACGCGGGCACGGCTGTCCAAGGAGCTGCAGGCG
GCGCAGGCCCGGCTGGGCGCGGACATGGAGGACGTGTGCGGCCGCTGGTGCAGTACCGCGGCGAGGTGC
AGGCCATGCTCGGCCAGAGCACCGAGGAGCTGCGGGTGCGCCTCGCCTCCACCTGCGCAAGCTGCGTAA
GCGGCTCCTCCGCGATGCCGATGACCTGCAGAAGTGCCTGGCAGTGTACCAGGCCGGGGCCCCGCGAGGGC
GCCGAGCGCGGCCTCAGCGCCATCCGCGAGCGCCTGGGGCCCCCTGGTGGAACAGGGCCGCGTGCGGGCCG
CCACTGTGGGCTCCCTGGCCGGCCAGCCGCTACAGGAGCGGGCCAGGCCTGGGGCGAGCGGCTGCGCGC
GCGGATGGAGGAGATGGGCAGCCGGACCCGCGACCGCCTGGACGAGGTGAAGGAGCAGGTGGCGGAGGTG
CGCGCCAAGCTGGAGGAGCAGGCCAGCAGATACGCCTGCAGGCCGAGGCCTTCCAGGCCCGCCTCAAGA
GCTGGTTTCGAGCCCCCTGGTGGAAAGACATGCAGCGCCAGTGGGCCGGGCTGGTGGAGAAGGTGCAGGCTGC
CGTGGGCACCAGCGCCGCCCTGTGCCAGCGACAATCACTGAACGCCGAAGCCTGCAGCCATGCGACCC
CACGCCACCCCGTGCCTCCTGCCTCCGCGCAGCCTGCAGCGGGAGACCCTGTCCCCGCCCCAGCCGTCT
CCTGGGGTGGACCCTAGTTTAATAAAGATTACCAAGTTTTCACGCA
```

>NM_000041.4_APOE_e3_Homo_sapiens_ref

```
CTACTCAGCCCCAGCGGAGGTGAAGGACGTCTTCCCCAGGAGCCGACTGGCCAATCACAGGCAGGAAGA
TGAAGGTTCTGTGGGCTGCGTTGCTGGTTCACATTCTGGCAGGATGCCAGGCCAAGGTGGAGCAAGCGGT
GGAGACAGAGCCGGAGCCCGAGCTGCGCCAGCAGACCGAGTGGCAGAGCGGCCAGCGCTGGGAACTGGCA
CTGGGTGCGTTTTTGGGATTACCTGCGCTGGGTGCAGACACTGTCTGAGCAGGTGCAGGAGGAGCTGCTCA
GCTCCCAGGTACCCAGGAACTGAGGGCGCTGATGGACGAGACCATGAAGGAGTTGAAGGCCTACAAATC
GGAAGTGGAGGAACAACACTGACCCCGGTGGCGGAGGAGACGCGGGCACGGCTGTCCAAGGAGCTGCAGGCG
GCGCAGGCCCGGCTGGGCGCGGACATGGAGGACGTGTGCGGCCGCTGGTGCAGTACCGCGGCGAGGTGC
AGGCCATGCTCGGCCAGAGCACCGAGGAGCTGCGGGTGCGCCTCGCCTCCACCTGCGCAAGCTGCGTAA
GCGGCTCCTCCGCGATGCCGATGACCTGCAGAAGCGCCTGGCAGTGTACCAGGCCGGGGCCCCGCGAGGGC
GCCGAGCGCGGCCTCAGCGCCATCCGCGAGCGCCTGGGGCCCCCTGGTGGAACAGGGCCGCGTGCGGGCCG
CCACTGTGGGCTCCCTGGCCGGCCAGCCGCTACAGGAGCGGGCCAGGCCTGGGGCGAGCGGCTGCGCGC
GCGGATGGAGGAGATGGGCAGCCGGACCCGCGACCGCCTGGACGAGGTGAAGGAGCAGGTGGCGGAGGTG
CGCGCCAAGCTGGAGGAGCAGGCCAGCAGATACGCCTGCAGGCCGAGGCCTTCCAGGCCCGCCTCAAGA
GCTGGTTTCGAGCCCCCTGGTGGAAAGACATGCAGCGCCAGTGGGCCGGGCTGGTGGAGAAGGTGCAGGCTGC
CGTGGGCACCAGCGCCGCCCTGTGCCAGCGACAATCACTGAACGCCGAAGCCTGCAGCCATGCGACCC
CACGCCACCCCGTGCCTCCTGCCTCCGCGCAGCCTGCAGCGGGAGACCCTGTCCCCGCCCCAGCCGTCT
```

Annexes

CCTGGGGTGGACCCTAGTTTAATAAAGATTACCAAGTTTCACGCA

>NM_000041.4_Mutated_APOE_e4_Homo_sapiens

CTACTCAGCCCCAGCGGAGGTGAAGGACGTCTTCCCCAGGAGCCGACTGGCCAATCACAGGCAGGAAGA
TGAAGGTTCTGTGGGCTGCGTTGCTGGTACATTCTGGCAGGATGCCAGGCCAAGGTGGAGCAAGCGGT
GGAGACAGAGCCGGAGCCCCGAGCTGCGCCAGCAGACCGAGTGGCAGAGCGGCCAGCGCTGGGAACTGGCA
CTGGGTGCGTTTTTGGGATTACCTGCGCTGGGTGCAGACACTGTCTGAGCAGGTGCAGGAGGAGCTGCTCA
GCTCCCAGGTACCCAGGAACTGAGGGCGCTGATGGACGAGACCATGAAGGAGTTGAAGGCCTACAAATC
GGAACTGGAGGAACAACACTGACCCCCGGTGGCGGAGGAGACGCGGGCACGGCTGTCCAAGGAGCTGCAGGCG
GCGCAGGCCCGGCTGGGCGCGGACATGGAGGACGTGCGCGGCCCGCTGGTGCAGTACCGCGGCGAGGTGC
AGGCCATGCTCGGCCAGAGCACCGAGGAGCTGCGGGTGCGCCTCGCCTCCCACCTGCGCAAGCTGCGTAA
GCGGCTCCTCCGCGATGCCGATGACCTGCAGAAGCGCCTGGCAGTGTACCAGGCCGGGGCCCCGCGAGGGC
GCCGAGCGCGGCCTCAGCGCCATCCGCGAGCGCCTGGGGCCCCCTGGTGAACAGGGCCGCGTGCGGGCCG
CCACTGTGGGCTCCCTGGCCGGCCAGCCGCTACAGGAGCGGGCCAGGCCTGGGGCGAGCGGCTGCGCGC
GCGGATGGAGGAGATGGGCAGCCGGACCCGCGACCCGCTGGACGAGGTGAAGGAGCAGGTGGCGGAGGTG
CGCGCCAAGCTGGAGGAGCAGGCCAGCAGATACGCCTGCAGGCCGAGGCCTTCCAGGCCCGCCTCAAGA
GCTGGTTCGAGCCCCCTGGTGAAGACATGCAGCGCCAGTGGGCCGGGCTGGTGGAGAAGGTGCAGGCTGC
CGTGGGCACCAGCGCCGCCCTGTGCCAGCGACAATCACTGAACGCCGAAGCCTGCAGCCATGCGACCC
CACGCCACCCCGTGCCTCCTGCCTCCGCGCAGCCTGCAGCGGGAGACCCTGTCCCCGCCCCAGCCGTCTT
CCTGGGGTGGACCCTAGTTTAATAAAGATTACCAAGTTTCACGCA

2-Séquences protéiques utilisées :

2.1. Protéine APP (Amyloid Precursor Protein) :

The image shows a screenshot of the UniProt database entry for P05067 - A4_HUMAN. The page is titled "P05067 - A4_HUMAN" and includes a search bar at the top. The left sidebar contains navigation links for Function, Names & Taxonomy, Subcellular Location, Disease & Variants, PTM/Processing, Expression, Interaction, Structure, Family & Domains, Sequence & Isoforms, Similar Proteins, and Homologs. The main content area displays the protein's name, gene, status, and organism. The "Function" section is highlighted, showing a detailed description of the protein's role as a cell surface receptor and its involvement in various biological processes.

Functionⁱ

Functions as a cell surface receptor and performs physiological functions on the surface of neurons relevant to neurite growth, neuronal adhesion and axonogenesis. Interaction between APP molecules on neighboring cells promotes synaptogenesis (PubMed:25122912). Involved in cell mobility and transcription regulation through protein-protein interactions. Can promote transcription activation through binding to APBB1-KAT5 and inhibits Notch signaling through interaction with Numb. Couples to apoptosis-inducing pathways such as those mediated by G(o) and JIP. Inhibits G(o) alpha ATPase activity (By similarity). Acts as a kinesin I membrane receptor, mediating the axonal transport of beta-secretase and presenilin 1 (By similarity). By acting as a kinesin I membrane receptor, plays a role in axonal anterograde transport of cargo towards synapses in axons (PubMed:17062754, PubMed:23011729). Involved in copper homeostasis/oxidative stress through copper ion reduction. In vitro, copper-metallated APP induces neuronal death directly or is potentiated through

Annexes

>sp|P05067|A4_HUMAN Amyloid-beta precursor protein OS=Homo sapiens OX=9606
GN=APP PE=1 SV=3

```
MLPGLALLLLAAWTAARALEVPTDGNAGLLAEPQIAMFCGRLLNMHMNVQNGKWDSDPSGTK
TCIDTKEGILQYQCQEVYPELQITNVVEANQPVTIQNWCKRGRKQCKTHPHFVIPIYRCLVG
EFVSDALLVPDKCKFLHQERMDVCETHLHWHTVAKETCSEKSTNLHDYGMLLPCGIDKFR
GVEFVCCPLAEEESDNVDSADAEEEDSDVWVGGADTDYADGSEDKVVEVAEEEEVAEVEEEE
EADDDDEDEDGDEVEEEAEPEYEEATERTTSIATTTTTTTTTESVEEVVREVCSEQAETGPC
RAMISRWFYFDVTEGKCAPFFYGGCGGNRNNFDTEEYCMAVCGSAMSQSLLKTTQEPLARD
PVKLPPTTAASTPDAVDKYLETPGDENEHAHFQKAKERLEAKHRERMSQVMREWEAEERQA
KNLPKADKKAVIQHFQEKVESLEQEAANERQQLVETHMARVEAMLNDRRRLALENYITAL
QAVPPRPRHVFNMLKKYVRAEQKDRQHTLKHFEHVRMVDPKKAAQIRSQVMTHLRVIYER
MNQSLSLLYNVPAAVEEQDEVDLQKEQNYSDVLANMISEPRISYGNLALMPSLTET
KTTVELLPVNGEFLDDLQPWHSFGADSVANTENEVEPVDARPAADRGLTTRPGSGLTN
IKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITL
VMLKKKQYTSIHHGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEQMQN
```

2.2. Protéine APOE (Apolipoprotein E)

>sp|P02649|APOE_epsilon3_HUMAN Apolipoprotein E OS=Homo sapiens OX=9606
GN=APOE PE=1 SV=1

```
MKVLWAALLVTFFLAGCQAKVEQAVETEPEPELRRQTEWQSGQRWELALGRFWDYLRWVQT
LSEQVQEELLSSQVTQELRALMDETMKELKAYKSELEEQLTPVAEETRARLSKELQAAQA
RLGADMEDVCGRLVQYRGEVQAMLGQSTEELRVRLASHLRKLRKLLRDADDLQKRLAVY
QAGAREGAERGLSAIRERLGLPLVEQGRVRAATVGSLAGQPLQERAQAWGERLRARMEEMG
SRTRDRLDEVKEQVAEVRAKLEEQAQQIRLQAEAFQARLKSWEPLVEDMQRQWAGLVEK
VQAAVGTSAAPVPSDNH
```

3. Utilisation "BLAST" :

The image shows the NCBI BLAST search interface. At the top, there are tabs for 'blastn', 'blastp', 'blastx', 'tblastn', and 'tblastx'. The 'blastn' tab is selected. Below the tabs, there is a header 'BLASTN programs search nucleotide subjects using a nucleotide query. more...'. The main form is divided into three sections: 'Enter Query Sequence', 'Enter Subject Sequence', and 'Program Selection'. In the 'Enter Query Sequence' section, there is a text input field containing the sequence 'GGATGAGGAAGGCATGCC', 'TGACAAACCCCTTTTAAAGATGTCTTCAATTTGATAAAATGGTGTTT', 'TCATGTAATAATACAT', and 'TCTTGGAGGAGCA'. There are also 'From' and 'To' subrange input fields. Below this, there is an 'Or, upload file' section with a 'Choose File' button and 'No file chosen' text. A 'Job Title' input field is also present. In the 'Enter Subject Sequence' section, there is a similar text input field with the same sequence as above. Below it, there is another 'Or, upload file' section. In the 'Program Selection' section, there are radio buttons for 'Optimize for': 'Highly similar sequences (megablast)' (selected), 'More dissimilar sequences (discontiguous megablast)', and 'Somewhat similar sequences (blastn)'. There is also a link 'Choose a BLAST algorithm'. At the bottom, there is a large blue 'BLAST' button and a checkbox for 'Show results in a new window'.

4. Arbres phylogénétiques par MEGA 12 : "gène APOE"

M12: Alignment Explorer (_out.2606060115mafftweba3216415IJMvRpGqHIFoakOZ5uC2mlsfnormal.fasta)

Species/Abbrv

Species/Abbrv	Sequence
1. NM 000484.4	G T C A G T T T C C T C G G C A G C G G T A G G C G A G A G C A C G C G G A G G A G C G T G C G C G - G G G C C C C G G G A G A C G G C G G C G G T G
2. NM 001013018.1
3. XM 015133068.3	G T C A G T T T C C T C G G C A G C G G T A G G C G A G A G C A C T C G G A G C A G C G C G C G C G - G G G C C C C G G A G A C G G C G G C G G T G
4. NM 001006601.2
5. NM 001198823.1	G T C A G T T T C C T C G G C G G C G G G A G G C G A G A G C A C C G G G A G C A G A G C G A G C G C G G G C C A C C G G A G A C G G C G G C
6. NM 019288.2	G

Site # 1 with w/o gaps Selected genetic code: Standard

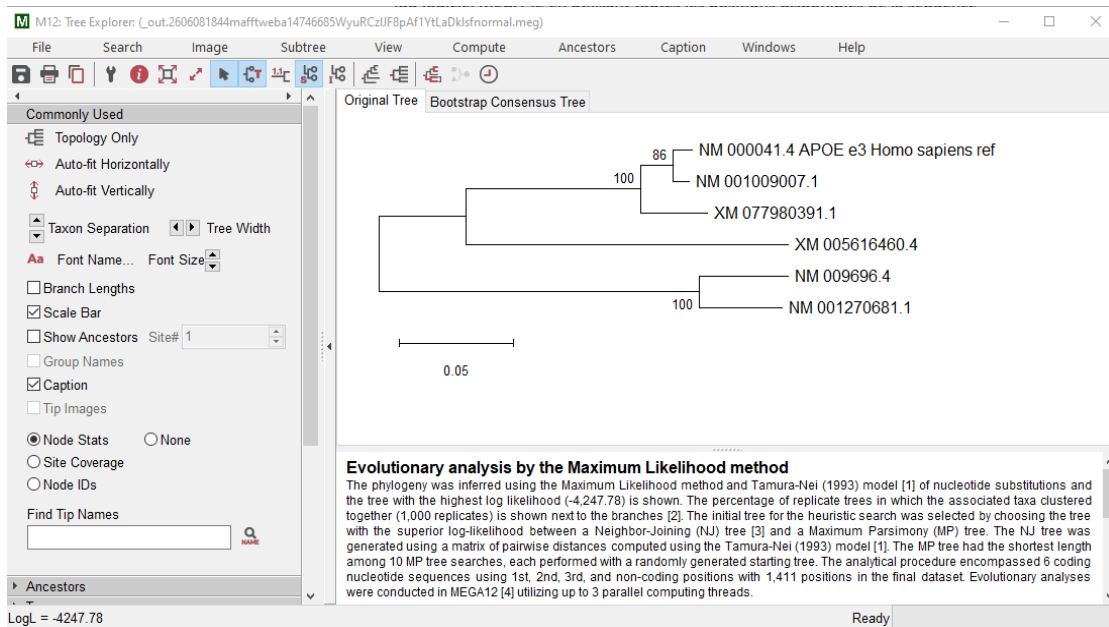
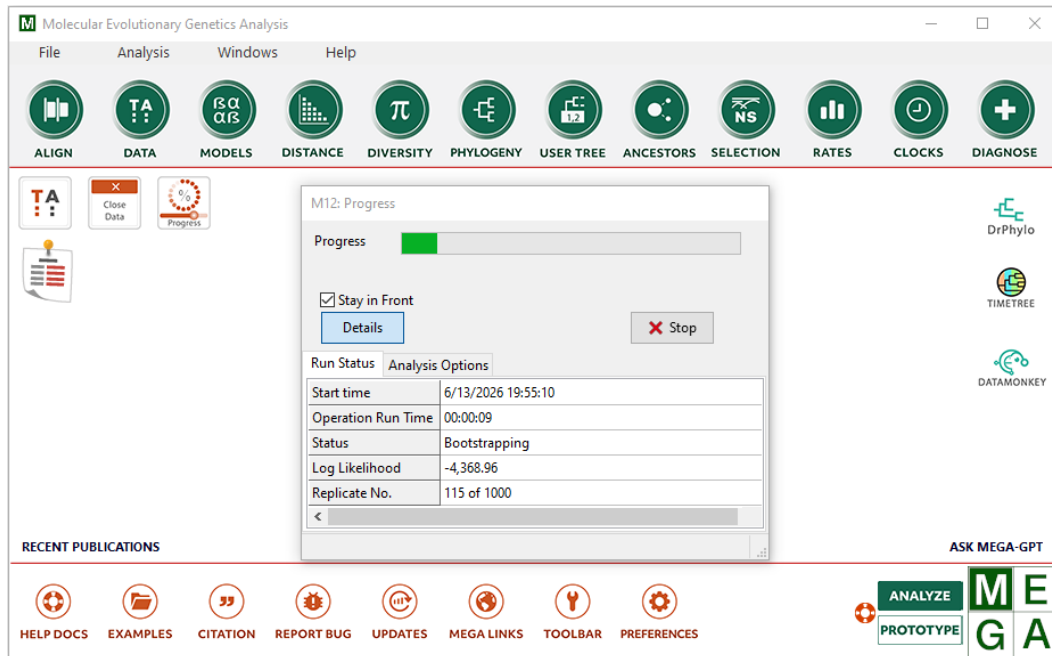
M12: Analysis Preferences

Likelihood Phylogeny Reconstruction

Option	Setting
ANALYSIS	
Statistical Method →	Maximum Likelihood
PHYLOGENY TEST	
Test of Phylogeny →	Standard Bootstrap (slow)
Bootstrap Replicates →	1000
SUBSTITUTION MODEL	
Substitutions Type →	Nucleotide
Model/Method →	Tamura-Nei model
RATES AND PATTERNS	
Rates among Sites →	Uniform Rates
DATA SUBSET TO USE	
Gaps/Missing Data →	Use all sites
Select Codon Positions →	<input checked="" type="checkbox"/> 1st <input checked="" type="checkbox"/> 2nd <input checked="" type="checkbox"/> 3rd <input checked="" type="checkbox"/> Noncoding Sites
TREE INFERENCE OPTIONS	
ML Heuristic Method →	Nearest-Neighbor-Interchange (NNI)
Initial Tree for ML →	Make initial tree automatically (Default - NJ/MP)
Branch Swap Filter →	None
SYSTEM RESOURCE USAGE	
Number of Threads →	3

? Help X Cancel OK

Annexes



Nous avons appliqué les mêmes étapes avec Gene APP .

5. Analyse des protéines (APP) :

5.1 .Analyse physico-chimique (ProtParam – ExPASy) :

ExPASy Home
Documentation
Reference
Contact

ProtParam

ProtParam [Documentation / Reference] is a tool which allows the computation of various physical and chemical parameters for a given protein stored in UniProtKB or for a user entered protein sequence. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) (Disclaimer).

Enter a protein sequence
 Please enter one UniProtKB AC/ID (e.g. P05130 or KPC1_DROME).
 Alternatively, enter one protein sequence in single letter code (e.g. ABCDEFGHIKLMNOPQRSTUVWXYZ).

Examples

RESET
Compute parameters

ProtParam - Results

User-provided sequence:

```

10      20      30      40      50      60
MLPGLALLLL AAWTARALEV PTDGNAGLLA EPQIAMFCGR LNMHMVQNG KWDSDPSGTK

70      80      90      100     110     120
TCIDTKEGIL QYQCEVYVPEL QITNVVEAIQ PVTIQHWCKR GRKQCKTHPH FVIPYRCLVG

130     140     150     160     170     180
EFVSDALLVP DKCKFLHQER MDVCETHLHW HTVAKETCSE KSTNLHDYGM LLPCGIDKFR

190     200     210     220     230     240
GVEFVCCPLA EESDNVDSAD AEEDDSVWVW GGADTDYADG SEDKVVEVAE EEEVAVEEEE

250     260     270     280     290     300
EADDEDEDDE GDEVEEEAEE PYEEATERIT SIATTTTTTT ESVEEVREV CSEQAETGPC

310     320     330     340     350     360
RAMISRWFYD VTEGKCAPFF YGGCGNRNIN FDTEEYCHAV CGSAMSQSL KTTQEPLARD

370     380     390     400     410     420
PVKLPITTAAS TPDVAVDKYLE TPGDENEHAH FQKAKERLEA KHRERMSQVM REWEEAERQA

430     440     450     460     470     480
KNLPKADKKA VIQHFQEKVE SLEQEANER QQLVETHMAR VEAMNDRRR LALENYITAL

490     500     510     520     530     540
QAVPPRPRHV FNMLKKYVRA EQKDRQHTLK HFEHVRMVDP KKAQIRSQV MTHLRVIYER

550     560     570     580     590     600
MNQSLSLLYN VPAVAEEIQD EVDLQKEQ NYSDVLANM ISEPRISYGN DALMPSLTET

610     620     630     640     650     660
KTTVELLPVN GEFSLDLQVP WHSFGADSVV ANTENEVEPV DARPAADRGL TTRPGSGLTN

670     680     690     700     710     720
    
```

Number of amino acids: 770
 Theoretical pI: 4.73
 Molecular weight: 86943.25

Amino acid composition: CSV format

Ala (A)	63	8.2%
Arg (R)	37	4.8%
Asn (N)	31	4.0%
Asp (D)	50	6.5%
Cys (C)	18	2.3%
Gln (Q)	36	4.7%
Glu (E)	92	11.9%
Gly (G)	38	4.9%
His (H)	25	3.2%
Ile (I)	24	3.1%
Leu (L)	56	7.3%
Lys (K)	41	5.3%
Met (M)	24	3.1%
Phe (F)	21	2.7%
Pro (P)	35	4.5%
Ser (S)	35	4.5%
Thr (T)	50	6.5%
Trp (W)	9	1.2%
Tyr (Y)	20	2.6%
Val (V)	65	8.4%
Py1 (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%


Total number of negatively charged residues (Asp + Glu): 142
 Total number of positively charged residues (Arg + Lys): 78

Atomic composition:

Carbon	C	3774
Hydrogen	H	5910

5.2 .Analyse des domaines protéiques (EMBL-EBI InterPro) :

Annexes

 **InterPro** Classification of protein families 🔍 ☰


[Home](#) | [Search](#) | [Browse](#) | [Results](#) | [Release notes](#) | [Download](#) | [Help](#)

Scan your sequences

```
>sp|P05067|A4_HUMAN Amyloid-beta precursor protein OS=Homo sapiens OX=9606 GN=APP PE=1 SV=3
MLFSLALLLAANTARALEVFTDGNAGLLAEPQIAMFCGRINMHWQNGKWDSDPSSGK
TCIDTKEGILQYQOEVPYELQITNVVEANQFVIQNWCKRGRKQCKTHPHFVIFPYRCLVG
EFVSDALVFDKCKFLHQERMDVCETHLHWHTVAKETCSERSTNLHDYGMLLFGGIDKFR
GVEFVCCFLAEESDNVDSADAEDDDSDWVGADTDYADGSEDKVVEVEEVEEVAEVEE
EADDDDEDDGDEVEEAEPYEEATERTTSIAITTTTTTESVVEVREVEEQAETGPC
RAMISRWYFDVTEGKCAPFFYGGCGGNRNFDTEEYCMVCGSAMSQSLKTTQELARD
FVKLPTTAASTPDAVDKYLETPGDENEHAHFQKAKERLEAKHRERMSQVREWEAERQA
KNLFPKADKXAVIQHFQEKVESLEQEAANERQQLVETHMARVEAMLDNRRLALENYITAL
QAVFPRFRHVFNMLKKYVRAEQKDRQHTLKHFEHVRMVFQKGAQIRSQVMTLHRIYER
MNQSLSLLYNVPVAEEIQDEVELLQKEQNYSDVLANMISEPRI SYGNDALMPSLTET
KITVELLFPVNGEFSLLDLQFVHSGADSVFANTENEVEPVDARPAADRGLTTRPSSGLTN
IKTEEISEVQMDAEFRHDSGYEVHKKLVFFAEDVGSNKGALIGLMVGGVVIATVIVITL
VMLKKKQYTSIHNGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEGMQN
```
















Valid Sequence.

[Choose file](#) [Example protein sequence](#)

 **InterPro** Classification of protein families 🔍 ☰

[Home](#) | [Search](#) | [Browse](#) | **[Results](#)** | [Release notes](#) | [Download](#) | [Help](#)

1 - 7 of 7 results

Job ID ↕	Sequences ↕	Created ↕	Status	Actions
iprscan5-R20260613-202302-0442-44670393-p2m	1	10 seconds ago	 Searching	 
iprscan5-R20260608-152447-0502-3488359-p1m	1	5 days ago	Completed	 
iprscan5-R20260608-152434-0536-11778311-p1m	1	5 days ago	Completed	 
iprscan5-R20260608-151003-0308-93383828-p1m	1	5 days ago	Completed	 
iprscan5-R20260608-150656-0153-59191871-p2m	1	5 days ago	Completed	 
iprscan5-R20260608-145357-0688-49129613-p1m	1	5 days ago	Completed	 
iprscan5-R20260608-145106-0158-73211891-p1m	1	5 days ago	Completed	 

Previous **1** Next

5.3. Modélisation 3D (SWISS-MODEL) :

SWISS-MODEL

is a fully automated protein structure homology-modelling server, accessible via the **ExPASy web server**.
The purpose of this server is to make protein modelling accessible to all life science researchers worldwide.

[Start Modelling](#)

Repository

Every week we model all the sequences for thirteen core species based on the latest UniProtKB proteome. Is your protein already modelled and up to date in **SWISS-MODEL Repository**?

Search SWISS-MODEL Repository

Start a New Modelling Project

Target Sequence(s):
(Format must be FASTA, Clustal, plain string, or a valid UniProtKB AC)

Target:
Target:
Target:
Target:

[Add Hetero Target](#) [Reset](#)

Project Title:

Email:

[Search For Templates](#) [Build Model](#)

By using the SWISS-MODEL server, you agree to comply with the following [terms of use](#) and to cite the corresponding [articles](#).

Supported Inputs

- Sequence(s)
- Target-Template Alignment
- User Template
- DeepView Project

All Projects

Untitled Project

Created: today at 21:28

[Summary](#) [Templates](#) [Models](#) [Project Data](#)

Template Results

The search for templates matching your target sequence is currently running. Please wait.

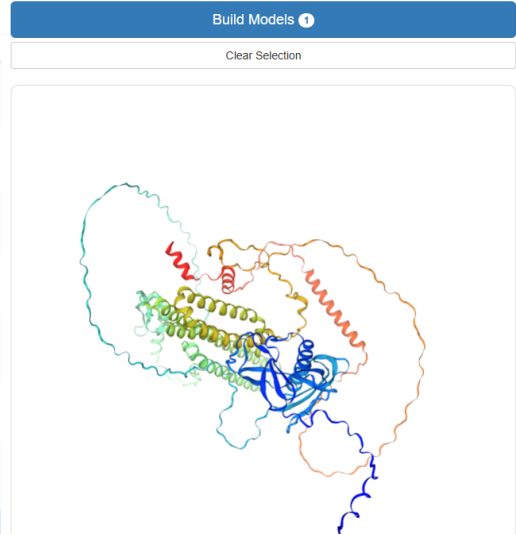
```
...running HHblits against SMTL
...reusing cached query alignment
...predicting residue burial status with ACCpro
...searching PDB profile database with previously built query profile
```

If you want to come back later, bookmark this link:
<https://swissmodel.expasy.org/interactive/zD568d/>

```
MLPGLALLLLAAWTARALEVPTDGNAGLAEPTQIAHFCGRLLNHHMNVQNGKWDSDPSGTKCIDTKEGILQYCEVYPELQITNVVEANQPVTIQNWCKRGRKQCKTHPHFVPIPYRCLVGFVSDALLVDPKCKFLHQRMDVCE  
EFVSDALLVDPKCKFLHQRMDVCE  
THLHHHTVAKETCSEKSTNLHDYGMLLPCGIDKFRGVEFVCCPLAEESDNVDSADAEEEDSDVWGGADTDYADGSEDKVVEVAEEEEVAEEEEEAADDEDDEGDVEVEEAEPYEEATERTTTSIATTTTTTTSVEEVVREV  
EADDEDDEGDVEVEEAEPYEEATERTTTSIATTTTTTTSVEEVVREV  
CSEQAETGPCRAMISRWFYDVTGKCAFFYGGCGGNRNFDTEEYCAVCGSAHSQSLKTTQEPLARDPVKLPITTAASTPDAVDKYLETPGDENEHAHFQAKERLEAKHREMSQVIREHEEAERQAKNLPKADKKAVIQHE  
PVKLPITTAASTPDAVDKYLETPGDENEHAHFQAKERLEAKHREMSQVIREHEEAERQAKNLPKADKKAVIQHE  
QEKVESLEQEAANERQQLVETHMARVEAMLNDRRLALENYITALQAVPPRRHVFNMLKKYVRAEQDRQHTLKHFEHVRMVDPKAAQIRSQVMTHLRVIYERMNQSLSLLYNPVAEETIQDEVELLQKEQNSDDVLANMI  
QAVPPRRHVFNMLKKYVRAEQDRQHTLKHFEHVRMVDPKAAQIRSQVMTHLRVIYERMNQSLSLLYNPVAEETIQDEVELLQKEQNSDDVLANMI  
SEPRISYGNDAIMPSTET  
KTTVELLPVNGEFSLDLDPQWHSFGADSVPANTENEVEPVDARPAADRGLTTRPGSGLTNIKTEEISEVKMDAEFRHDSGYEVHHQKLVFAEDVGSNKGAIIGLMVGGVVIATVITITL
```

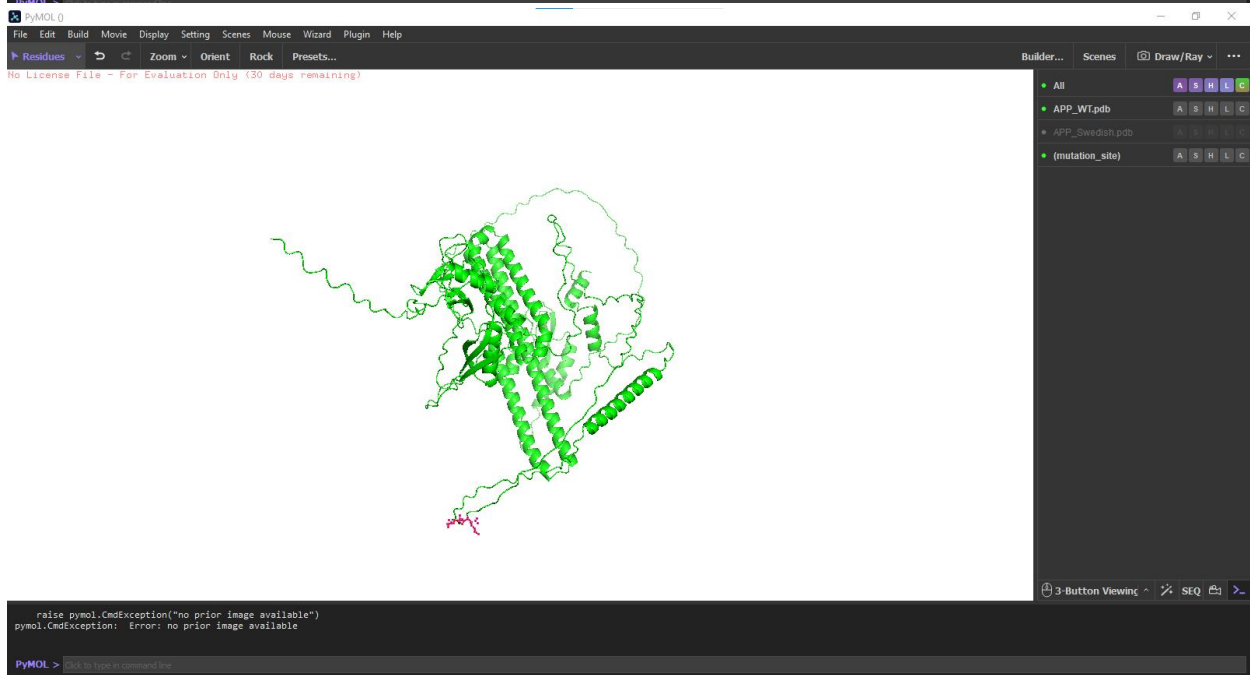
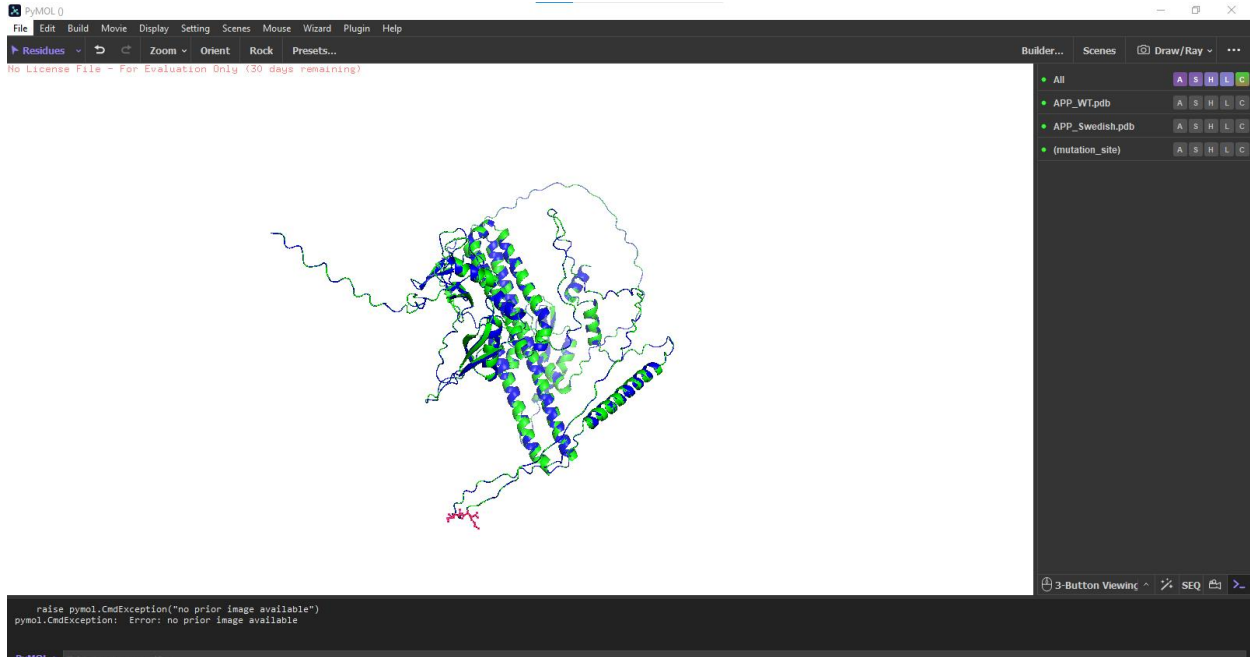
Template Results

Templates	Quaternary Structure	Sequence Similarity	Alignment	More			
<input checked="" type="checkbox"/> Q5IS80.1 A Amyloid-beta precursor protein AlphaFold DB model of A4_PANTR (gene: APP, organism: Pan troglodytes (Chimpanzee))		0.68	-	99.87	AlphaFold v2	monomer ✓	None
<input type="checkbox"/> 7mm.1 A Fusion protein of Chicken CNTN3 FN1-FN2 domains and Amyloid-beta A4 protein, Amyloid-beta A4 protein Chicken CNTN3 APP complex		0.16	-	93.57	X-ray, 2.8Å	homo-dimer Δ	None
<input type="checkbox"/> 7mm.1 A Fusion protein of Chicken CNTN3 FN1-FN2 domains and Amyloid-beta A4 protein, Amyloid-beta A4 protein Chicken CNTN3 APP complex		0.16	-	94.01	X-ray, 2.8Å	homo-dimer Δ	None
<input type="checkbox"/> 7mm.1 B Fusion protein of Chicken CNTN3 FN1-FN2 domains and Amyloid-beta A4 protein, Amyloid-beta A4 protein Chicken CNTN3 APP complex		0.15	-	93.57	X-ray, 2.8Å	homo-dimer Δ	None
<input type="checkbox"/> 7mm.1 B Fusion protein of Chicken CNTN3 FN1-FN2 domains and Amyloid-beta A4 protein, Amyloid-beta A4 protein Chicken CNTN3 APP complex		0.15	-	94.01	X-ray, 2.8Å	homo-dimer Δ	None
<input type="checkbox"/> 7mk.1 A Amyloid-beta A4 protein Chicken CNTN4 APP complex		0.15	-	93.57	X-ray, 2.0Å	hetero-dimer Δ	None
<input type="checkbox"/> 7mk.1 A Amyloid-beta A4 protein Chicken CNTN4 APP complex		0.15	-	93.71	X-ray, 2.0Å	hetero-dimer Δ	None
<input type="checkbox"/> 7mqy.1 A Fusion protein of CNTN4 and APLP2 Zukrafsk CNTN4 APLP2 complex							

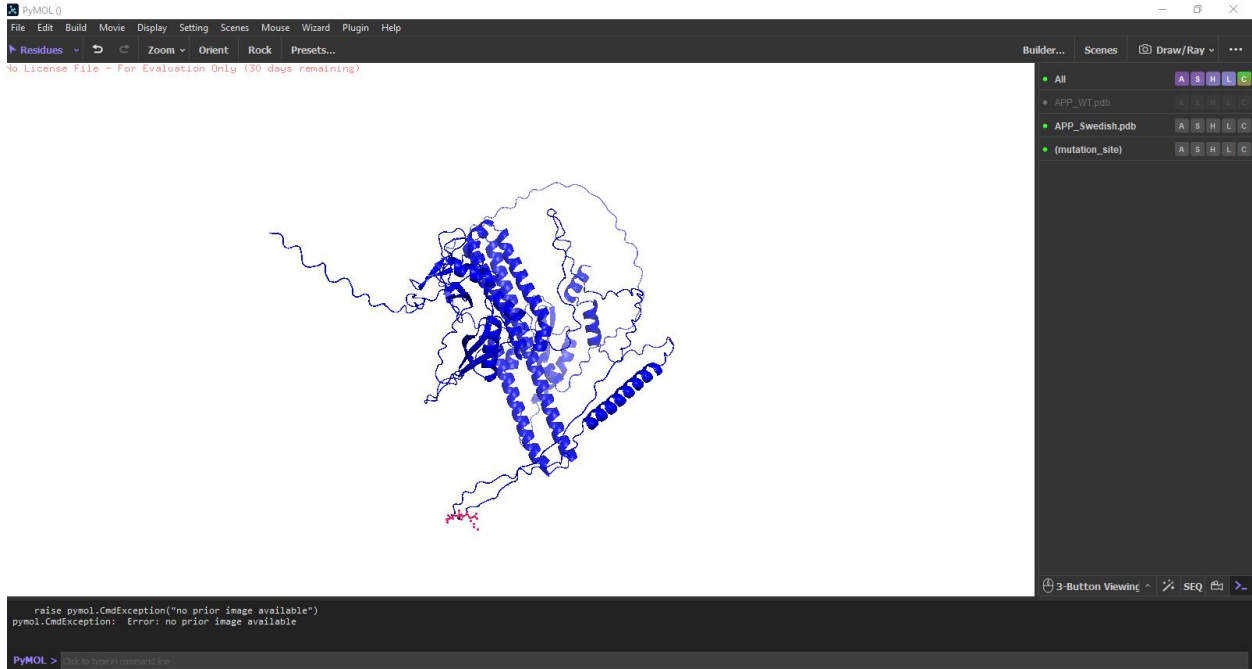


5.4 .Visualisation structurale (PyMOL) :

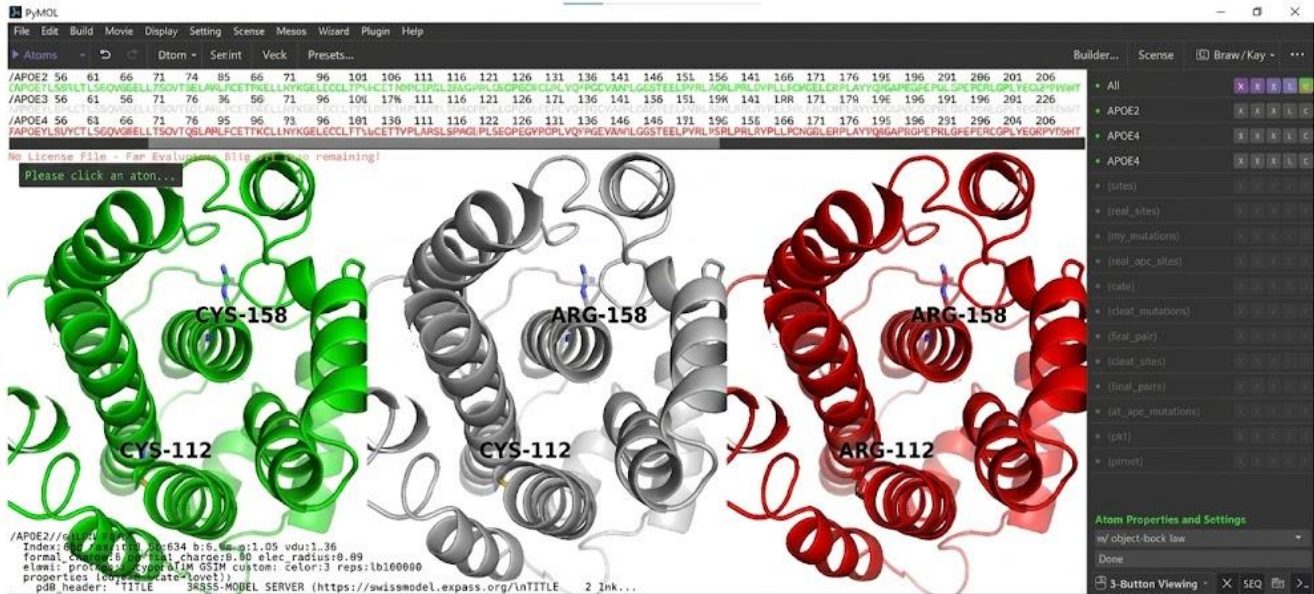
Annexes



Annexes



Nous avons appliqué les mêmes étapes avec Gene Apoe :



Résumé :

Cette étude explore l'utilisation d'outils bioinformatiques pour analyser les mutations génétiques associées à la maladie d'Alzheimer. Elle présente la méthodologie employée pour l'analyse des séquences génétiques à l'aide de la base de données du National Center for Bioinformatics (NCBI) et d'outils d'analyse comparative tels que BLAST et MEGA12. De plus, des étapes d'alignement multiple et de construction d'arbres phylogénétiques ont été mises en œuvre, ainsi que l'analyse des propriétés physico-chimiques des protéines et de leur structure tridimensionnelle à l'aide d'outils de modélisation moléculaire tels que SWISS-MODEL et PyMOL.

Les gènes *APP* et *APOE* ont été privilégiés en raison de leur rôle crucial dans le développement de la maladie d'Alzheimer et ses réponses pathologiques et thérapeutiques associées.

L'étude a révélé des mutations ponctuelles spécifiques dans les gènes *APP* et *APOE*, en accord avec les recherches scientifiques récentes, soulignant les similitudes et les différences entre l'homme et certains organismes modèles utilisés dans les études biologiques. Des régions fonctionnelles et évolutivement conservées au sein de ces gènes ont également été identifiées.

Les résultats ont été analysés à la lumière des dernières études de référence pertinentes dans les domaines de la bioinformatique et de la génétique moléculaire de la maladie d'Alzheimer.

Mots clés:

Bioinformatique ; Maladie d'Alzheimer ; *APP* ; *APOE* ; mutations génétiques ; analyse phylogénétique ; BLAST ; MEGA ; séquençage génétique ; modélisation protéique ; évolution moléculaire.

Abstract:

This study explores the use of bioinformatics tools to analyze gene mutations associated with Alzheimer's disease. It presents the methodology employed in analyzing gene sequences using the National Center for Bioinformatics (NCBI) database and comparative analysis tools such as BLAST and MEGA12. Furthermore, multiple alignment steps and phylogenetic tree construction were implemented, along with the analysis of the physicochemical properties of proteins and their three-dimensional structure using molecular modeling tools such as SWISS-MODEL and PyMOL.

The *APP* and *APOE* genes were the focus due to their pivotal role in the development of Alzheimer's disease and its associated pathological and therapeutic responses.

The study revealed specific point mutations in the *APP* and *APOE* genes, consistent with recent scientific research, highlighting similarities and differences between humans and some model organisms used in biological studies. Functional and evolutionarily conservative regions within these genes were also identified.

The results were discussed in light of the latest relevant reference studies in the fields of bioinformatics and molecular genetics of Alzheimer's disease..

Keywords:

Bioinformatics; Alzheimer's disease; *APP*; *APOE*; genetic mutations; phylogenetic analysis; BLAST; MEGA; genetic sequencing; protein modeling; molecular evolution.

ملخص:

تتناول هذه الدراسة استخدام بعض أدوات المعلوماتية الحيوية لتحليل الطفرات الجينية المرتبطة بمرض الزهايمر، كما تعرض المنهجية المعتمدة في تحليل التسلسلات الجينية باستخدام قواعد البيانات البيولوجية (NCBI) وأدوات التحليل المقارن مثل BLAST وMEGA12. إضافة إلى ذلك، تم تطبيق خطوات المحاذاة المتعددة وبناء الأشجار التطورية، إلى جانب تحليل الخصائص الفيزيائية والكيميائية للبروتينات ودراسة بنيتها ثلاثية الأبعاد باستخدام أدوات النمذجة الجزيئية مثل SWISS-MODEL وPyMOL.

تم التركيز على الجينين *APP* و*APOE* نظرًا لدورهما المحوري في تطور مرض الزهايمر والاستجابة المرضية والعلاجية المرتبطة به.

أظهرت الدراسة وجود طفرات نقطية محددة في الجينين *APP* و*APOE* تتوافق مع ما ورد في الأبحاث العلمية الحديثة، مع إبراز أوجه التشابه والاختلاف بين الإنسان وبعض الكائنات النموذجية المستخدمة في الدراسات الحيوية. كما تم تحديد المناطق الوظيفية والمحافظة تطوريًا داخل هذه الجينات.

وقد نوقشت النتائج اعتمادًا على أحدث الدراسات المرجعية ذات الصلة في مجال المعلوماتية الحيوية والوراثة الجزيئية لمرض الزهايمر.

الكلمات المفتاحية:

المعلوماتية الحيوية؛ مرض الزهايمر؛ الجين *APP*؛ الجين *APOE*؛ الطفرات الجينية؛ التحليل التطوري؛ التحليل الوراثي المقارن؛ BLAST؛ MEGA؛ التسلسل الجيني؛ النمذجة البروتينية؛ التطور الجزيئي.